

XXII Workshop de Investigadores En Ciencias De La Computación
Junio 2020- El Calafate - Santa Cruz – Argentina

LIBRO DE ACTAS

Universidad Nacional De La Patagonia Austral

Red de universidades con carreras de informática (RedUNCI)





XXII Workshop de Investigadores en Ciencias de la Computación: WICC 2020 /
Rodolfo Bertone... [et al.] ; compilado por Marta Lasso.- 1a ed.- Río Gallegos :
Universidad Nacional de la Patagonia Austral, 2020.
Libro digital, PDF

Archivo Digital: descarga y online
ISBN 978-987-3714-82-5

1. Computación. 2. Tecnología Informática. I. Bertone, Rodolfo. II. Lasso, Marta, comp.
CDD 004.071



Universidad Nacional de la Patagonia Austral (UNPA)

Rector:
Ing. Hugo Santos Rojas
Vice Rectora
Prof. Roxana Puebla

Autoridades Red de Universidades con Carreras de Informática (Red UNCI)

Coordinador Titular
Pesado Patricia (UNLP)
Coordinador Alterno
Estayno Marcelo (UNLZ)

Coordinadores de Área WICC 2020

Agentes y Sistemas Inteligentes
Marcelo Falappa (UNS)
Marcelo Errecalde (UNSL)
Daniel Pandolfi (UNPA)

Arquitectura, Redes y Sistemas Operativos
Luis Marrone (UNLP)
Daniel Arias Figueroa (UNSa)
Orlando Micolini (UNC)

Computación Gráfica, Imágenes y Visualización
Martın Larrea (UNS)
Marıa J. Abasolo (UNLP-UNCPBA)
Roberto Guerrero (UNSL)

Ingeniería de Software
Pablo Fillottrani (UNSur)
Pablo Thomas (UNLP)
Fernanda Carmona (UNdeC)

Innovación en Sistemas de Software
Mónica Tugnarelli (UNER)
Guillermo Feierherd (UNTDF)
Marisa Panizzi (UK)
Gladys Dapozo (UNNE)

Procesamiento Distribuido y Paralelo
Marcelo Naiouf (UNLP)
Verónica Gil Costa (UNSL)
Javier Balladini (UNCOMA)

Tecnología Informática aplicada en Educación
Zulma Cataldi (UBA-UTN)
Alejandra Zangara(UNLP)
Gustavo Gil (UNSa)

Procesamiento de señales y Sistemas de Tiempo Real
Oscar Bria (INVAP)
Fernando Tinetti (UNLP)
Nelson Rodríguez (UNSJ)

Bases de Datos y Minería de Datos
Laura Lanzarini (UNLP)
Claudia Deco (UNR)
Norma Herrera (UNSL)

Innovación en Educación Informática
Claudia Russo (UNNOBA)
Elena Durán (UNSE)
Lucía Malbernat (UCAECE)

Seguridad Informática
Paula Venosa (UNLP)
Jorge Eterovic (UNLaM)
Javier Echaiz (UNS)
Antonio Castro Lechtaller (IESE) J

Jurado de Tesis de Doctorado
Marcela Printista (UNSL)
Laura de Giusti (UNLP)
Silvia Castro (UNS)
Alejandra Cechich (UNCOMA)
Horacio Kuna (UNaM)

Comité Académico

UBA – Cs. Exactas	Garbervetsky, Diego
UBA – Ingeniería	Echeverria, Adriana
UN La Plata	Pesado, Patricia
UN Sur	Rueda, Sonia
UN San Luis	Piccoli, Fabiana
UNCPBA	Aciti, Claudio
UN Comahue	Grosso, Guillermo
UN La Matanza	Spositto, Osvaldo
UN La Pampa	Alfonso, Hugo
UN Lomas de Zamora	Estayno, Marcelo
UNTierra del Fuego	Feierherd, Guillermo
UN Salta	Gil, Gustavo
UN Patagonia Austral	Lasso, Marta
UN SanJuan	Rodriguez, Nelson
UADER	Noriega, Jorge
UN Patagonia SJB	Buckle, Carlos
UN Entre Ríos	Tugnarelli, Mónica
UN Nordeste	Dapozo, Gladys
UN Rosario	Zanarini, Dante
UN Misiones	Kuna, Horacio
UNNOBA	Russo, Claudia
UN Chilecito	Carmona, Fernanda
UN Lanús	Azcurra, Diego
UN Santiago del Estero	Duran, Elena
Esc. Sup. Ejército	Arroyo Arzubi, Alejandro
UN Litoral	Loyarte, Horacio
UN Rio IV	Arroyo, Marcelo
UN Córdoba	Fridlender, Daniel
UN Jujuy	Herrera Cognetta, Analía
UN Rio Negro	Vivas, Luis
UN Villa María	Prato, Laura
UN Lujan	Panessi, Wálter
UN Catamarca	Poliche Maria Valeria
UN La Rioja	Martinez, Marcelo
UNTres de Febrero	Oliveros, Alejandro
UN Tucumán	Luccioni, Griselda María
UNAJ	Morales, Martín
UN Chaco Austral	Zachman Patricia
UN del Oeste	Foti, Antonio
UN de Cuyo	García Garino, Carlos
UN de Mar del Plata	Ríos, Carlos
UN de Quilmes	
UN Hurlingham	Medrano Gustavo
U Morón	Padovani Hugo
UAI	De Vincenzi, Marcelo
U Belgrano	Guerci, Alberto
U kennedy	Panizzi, Marisa
U Adventista del Plata	Bournissen Juan
UCAECE	Finocchieto, Jorge
UPalermo	Alvarez Adriana
UCA Rosario	Grieco, Sebastián
U Salvador	Zanitti, Marcelo
U Aconcagua	Giménez, Rosa
U Gastón Dachary	Ruidías, Hector Javier
UADE	
UCEMA	Guglianone, Ariadna
U Austral	Cosentino, Juan Pablo
U Atlántida Argentina	Rathmann, Liliana
UCA La Plata	Bertone, Rodolfo
ITBA	Mon, Alicia
U Champagnat	Pinciroli, Fernando

Comité Organizador UNPA – UACO

Ma. Eugenia de San Pedro
Eugenia María Márquez
Marta Lasso
Natalia Serón
Sergio Orozco
Miriam Diaz vivar
Daniel Gonzalez

Coordinadores Locales

Altamirano Walter
Bilbao Martín
Casas Sandra
Cruz Diana
de San Pedro Eugenia
Diaz Vivar Miriam
Gaetan Gabriela
Gimenez Javier
González Leonardo
Hallar Karim
Laguia Daniel
Lasso Marta
Márquez Eugenia M
Mercado Viviana
Molina Daniel
Orozco Sergio
Pandolfi Daniel
Pires Analía
Porzuelo Gabriela
Ramos Luis
Rasjido José
Rivadeneira Gabriela
Saldaño Viviana
Serón Natalia Alicia
Sierpes Luis
Soto Héctor
Valdez Jorge
Varas Valeria Susana
Vilanova Gabriela
Villagra Andrea
Villagra Silvia

ID	ACTAS	PAG.
	AGENTES Y SISTEMAS INTELIGENTES	1
12938	Agente conversacional inteligente como herramienta de ayuda al proceso de atención al aspirante de las carreras (FTyCA)(UNCa)	84
12773	Agentes Inteligentes para Recuperación de Información y Analítica Visual en Big Data	39
12748	Análisis de Sentimientos en Twitter: Desarrollo de Recursos en el Españolrioplatense de Argentina	26
12883	Analíticas para Sistemas de Atención con Grandes Volúmenes de Eventos	65
12806	Aplicaciones Inteligentes sobre Internet de las Cosas y Grandes Volúmenes de Datos: Un Enfoque Riguroso	49
12877	Aprendizaje Automático Profundo y Visión por Computadora. Aplicaciones en el Reconocimiento de Lengua de Señas e Imágenes Astronómicas	60
12729	Cambio de Creencias Múltiple No-Priorizada	12
12895	Ciudades Inteligentes, Eficientes y Sostenibles	70
12941	Evolución Diferencial para Problemas Restringidos Dinámicos	90
12845	Inteligencia Artificial aplicada a IoT	54
12708	Inteligencia Artificial aplicada al Poder Judicial	7
12741	Inteligencia Artificial y Computación Cuántica en Finanzas	17
12897	Inteligencia Colectiva para una Recuperación Sostenible de Residuos Tecnológicos	75
12754	Metaheurísticas aplicadas al monitoreo en plantas químicas	31
12937	Minería de texto y Deep Learning aplicados a determinar la pertenencia de las consultas realizadas a un metabuscador a cada área temática dentro de las ciencias de la computación	80
12770	Procesamiento Digital de Imágenes Citogenéticas para su clasificación según la técnica ensayo cometa para la detección de daños en Eladn	35
12796	Supervisión y Control de Procesos	44
12701	Técnicas de Aprendizaje Automatizado Aplicadas a Problemas de Visión por Computadora	2
12745	Utilización de Sistemas Inteligentes para optimizar el diseño de redes de distribución de agua en General Pico - La Pampa	21
	ARQUITECTURA, REDES Y SISTEMAS OPERATIVOS	94
12778	Análisis de componente de hardware para la utilización con Frameworks de IoT	120
12764	Análisis, simulación y estudio experimental del comportamiento de métricas de QoS y QoE de streamings de video multicast IPTV- Caso de estudio en la red de la UTN-Mendoza	110
12794	Determinación de la eficiencia en el procesamiento sobre Arquitecturas Multiprocesador y Estrategias de Tolerancia a Fallos en HPC	135
12830	Entorno de integración continua para validación de sistemas embebidos de tiempo real	140
12737	Explorando las redes definidas por software (SDN)	100
12782	Implementación de un Servidor DNS Seguro basado en Pi-Hole utilizando un entorno virtualizado	130
12779	Implementación del protocolo AODV en el simulador de redes Shawn	125
12746	Modelo de Comunicación basado en Internet de las Cosas para la Gestión de Catástrofes	105
12916	Monitoreo a Distancia de Centros de Transformación Eléctrica	165
12716	Programa de Investigación, Desarrollo y Enseñanza en Compatibilidad Electromagnética e Integridad de señal FCEIA/UNR	95

12777	Propuesta de métricas para comparación de Framewoks IoT	115
12870	Red de IoT: Estimación de Consumo de Energía	145
12875	Sistema de Entrenamiento para Navegación a Vela Ligera de Alto Rendimiento basados en Dispositivos Multisensoriales de bajo costo(categ olímpicas, veleros de 1 o 2 tripulantes)	150
12917	Smart Micro Grid de Campus Universitario	169
12801-12903	Tecnologías de la información y las comunicaciones mediante IoT aplicadas a soluciones en el medio productivo y medioambiental	155
12904	Wireless Wine: Estimación de rendimiento y ubicación de sensores para la predicción de heladas en los viñedos	160
	BASE DE DATOS Y MINERIA DE DATOS	174
12879	Almacenamiento y procesado automático de imágenes satelitales para proyectos de monitoreo de bosque nativo a escala regional y local	241
12930	Análisis cienciométrico de la producción en investigación científica y tecnológica	284
12774	Análisis de relaciones intra-institucionales e interdisciplinares de una universidad a partir de la producción registrada en Microsoft Academic: el caso de la Universidad Nacional de La Plata	221
12926	Aplicación de Técnicas Descriptivas de Minería de Textos sobre Contenido Digital Realizando Análisis Inteligente	279
12767	Avances en el proyecto de Análisis y elaboración de datos para el desarrollo de un sistema de indicadores de ayuda social	216
12908	Bases de Datos Espaciales y Espacio Temporales	261
12712	Bioingeniería Informática Aplicada a la predicción de enfermedades cardiológicas y su implementación en el Hospital Delicia Concepción Masvernat de la Ciudad de Concordia, Provincia de Entre Ríos	185
12829	Búsqueda y Recopilación de Información sobre Legislación referida a Residuos Informáticos	231
12755	Búsquedas Selectivas sobre Flujos de Documentos	200
12780	Ciencia de datos aplicada al mejoramiento genético de la raza Aberdeen Angus	226
12900	Cluster para Aprendizaje y Práctica de BigData y Servicios de Learning Analytics	252
12907	Contribuciones a las Bases de Datos Métricas	256
12914	Estrategias de Desambiguación de perfiles y similitud temática para un Metabuscador de las Ciencias de la Computación	265
12763	Guía de recomendaciones para el tratamiento del Big Data como evidencia digital	205
12925	Herramientas Informáticas para el Estudio de la Biodiversidad utilizando Datos Abiertos Enlazados	275
12899	Modelos para Aprendizaje Automático en Tiempo Real sobre Entornos de Big Data	246
12765	Modelos, Algoritmos y Aplicaciones en Búsquedas a Gran Escala	210
12753	Plataforma de Datos Abiertos Enlazados para la Gestión y Visualización de Datos Primarios de Ciencias del Mar	195
12751	Predicción de la enfermedad de Parkinson utilizando redes neuronales convolucionales	190
12703	Predicción del resultado de oseointegración en implantes dentales mediante múltiples clasificadores	180
12921	Recuperación de Información en Grandes Volúmenes de Datos	270
12836	Sistemas Inteligentes. Aplicaciones en Minería de Datos y Big Data	236

12700	Tratamiento de Secuencias de ADN y Clustering de Pacientes con Cáncer Colorrectal	175
12951	WEB-GIS Orientado al monitoreo y evaluar el control de la población de perros callejeros en Patagonia	289
	COMPUTACIÓN GRAFICA, IMÁGENES Y VISUALIZACIÓN	294
12944	Aplicaciones de la Televisión Interactiva y aplicaciones móviles para el mejoramiento de los pueblos latinoamericanos	322
12945	Aplicaciones de Realidad Extendida y Aplicaciones Móviles	327
12728	Avances en Línea de Investigación Doctoral: Integración Escalable de Realidad Aumentada Basada en Images y Rostros	298
12811	Evaluación de Visualizaciones eficientes en Ciencia de Datos	302
12826	Generación Gráfica de partículas artificiales de arena para su utilización en Redes Neuronales	307
12888	Realidad Virtual: Maximizando Presencia, Inmersión y Usabilidad	318
12844	Realidades Alternativas como Soporte para el Desarrollo Sostenible	313
12705	Verificación y Validación de representaciones visuales y sus interacciones	295
	INGENIERÍA DE SOFTWARE	334
12805	Adecuación de la forma de la construcción de glosarios al estilo del discurso	435
12812	Análisis de calidad y recuperación de información en foros de discusión	440
12706	Aplicación de Gestión de conocimiento a la fase de prueba de la Ingeniería de Software	350
12852	Aplicaciones Móviles 3D y Realidad Virtual	490
12771	Aporte de los Sistemas de Inteligencia de negocios a los Sistemas de Información Organizacionales para la toma de decisiones	395
12934	Aportes Ágiles en Etapas Iniciales del desarrollo de Software a través de implementaciones de lenguajes específicos de dominio	557
12848	Arquitectura Tecnológica para la implementación de Gestión del Conocimiento en pequeñas y medianas fábricas de software	473
12854	Aspectos de Ingeniería de Software, Bases de Datos Relacionales y Bases de Datos No Relacionales para el Desarrollo de Sistemas de Software en Escenarios Híbridos	494
12797	Automatización de la Medición de Software para Flotas Dinámicas mediante un Modelo de Calidad Mixto para la Movilidad en Smart Cities	415
12784	Construcción de grafos de glosarios guiada por el estilo del discurso	405
12843	Construcción de Interfaces Gráficas de Usuario aplicando técnicas de la Ingeniería del Software y utilizando Metodologías Centradas en el Usuario	464
12815	Creación de Herramientas de Software para la construcción de Sistemas de IOT Escalables y Seguros	448
12726	Desagregación de información en los escenarios futuros	361
12881	Desarrollo de Recursos Humanos para la Gobernanza de Ciudades Inteligentes Sostenibles. Proyecto CAP4CITY	509
12694	Desarrollo de requisitos en Aspect-Oriented Process for a Smooth Transition	335
12747	Desarrollo de Videojuegos Innovadores	386
12905	Diseño de Métodos, Procedimientos y Herramientas. Aportes al desarrollo Regional	543
12892	Diseño UX: Una Guía para Mejorar la Experiencia de los Usuarios haciendo un uso Eficiente de los Recursos Disponibles	523
12849	Enfoques y Tendencias en el Desarrollo de Aplicaciones para Dispositivos Móviles.	478

12736	Equipos de trabajo 4.0: Nuevas configuraciones	376
12889	Evaluación de la reusabilidad de los conjuntos de datos abiertos de un portal de Infraestructura de Datos Espaciales. Caso de Estudio: IDERA	513
12798	Evaluación de Seguridad de Aplicaciones Web	420
12838	Evaluación de técnicas para la validación de requerimientos en entornos de trabajo para el desarrollo de software	459
12738	Evaluación del Impacto de las emociones en la calidad de Software desde el punto de vista del usuario	381
12847	Experiencia de Usuario en APPs Móviles basado en emociones.	469
12898	Fiabilidad en la Calidad del Software: Modelos, Métodos y Estrategias	533
12802	Herramienta De Búsqueda en Repositorios Académicos basada en Web Semántica y Sistemas NoSQL	425
12894	Heurística de Evaluación de la Experiencia de Usuario en Sistemas e-Learning	528
12695	Implantación de Sistemas de Software: Nuevas Tendencias	340
12890	Ingeniería de Software al Servicio de la Informática Forense y la Evidencia Digital	518
12936	Ingeniería de software para sistemas embebidos, sistemas críticos e Internet de las Cosas	567
12880	Ingeniería de Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube	504
12752	Internet del futuro: aplicaciones de IoT en la Patagonia Austral	390
12902	La Experiencia de usuario en el contexto de los Multi-Dispositivos	538
12940	Mejora de procesos, Calidad de Software y Gobernanza Digital	574
12803	Metodología para el Desarrollo de Aplicaciones con RA en Educación (Meduc_AR)	430
12935	Métodos Emergentes en calidad del Software: Pruebas continuas y modelado de procesos de negocios	562
12783	Modelo Ontológico como garantía de completitud funcional en la elicitación de requerimiento de software	400
12915	Optimización global de unidades de negocio interrelacionadas de PyMEs de la región aplicando modelos de redes colaborativas	552
12787	Plataforma para la definición y ejecución de actividades orientadas a la recolección y análisis de datos, con intervención humana	410
12727	Reducción de la subjetividad en los procesos de requisitos	366
12851	Reuso, Composición y Refactorización de Servicios Heterogéneos	485
13003	Reuso orientado a dominios: hacia un proceso de desarrollo basado en reusabilidad a nivel de subdominos	579
12707	Revisión Sistemática para el diseño de una metodología de requerimientos referenciales en dominios complejos	355
12906	Técnicas de Gestión del Conocimiento en la Industria del Software Argentina aplicada a la Mejora de Requisitos	548
12819	Técnicas, Estrategias y Herramientas de Comprensión de Programas para facilitar el entendimiento de Sistemas Multiparadigmas	453
12859	Toma de decisiones multicriterio en problemas de la Ingeniería de Software utilizando Computación Blanda	499
12732	Tratamiento de Entrevistas en la Ingeniería de Requisitos	371
12697	Validación de la Técnica de Análisis del dominio (JODA) al proceso UML Components	345

	(DSBC)	
12814	Validación y Verificación con Alloy de Ontologías en OWL que unifican modelos de datos heterogéneos	445
	INNOVACIÓN EN EDUCACIÓN INFORMÁTICA	584
12776	Enseñanza y Aprendizaje del Pensamiento Computacional y la Programación en los distintos Niveles Educativos	595
12924	Estrategias de enseñanza basadas en Aprendizaje Activo para favorecer la accesibilidad académica en la enseñanza de la programación	615
12775	Formación de docentes y alumnos de grado como Investigadores Científicos Iniciales en las áreas de Informática y Ciencias de la Computación	590
12873	Plataforma para la Programación Tangible	605
12922	Recursos Educativos Desenchufados para la Enseñanza de las Ciencias de la Computación en la Escuela Secundaria	610
12720	SCRUM como metodología de enseñanza y aprendizaje de la Programación	585
12791	Tecnicatura Superior en Programación (TSP) Modalidad Semipresencial	600
12929	Tecnologías del aprendizaje y el conocimiento aplicado a la enseñanza de infraestructuras IT	620
12956	Tecnologías Emergentes para la Educación	625
	INNOVACIÓN EN SISTEMAS DE SOFTWARE	630
12713	Aplicaciones Web Progresivas Enfocadas en el Uso y Optimización de Cache	640
12846	Avances en Métodos y Técnicas para la construcción de Aplicaciones basadas en Computación Ubicuas	678
12809	Computación Afectiva aplicada a la valoración emocional en contextos gastronómicos.	664
12789	Desarrollo, Evaluación y Puesta en Marcha de Herramientas para la Toma de Decisiones del Sistema Productivo Ganadero de la Pcia. de Santa Cruz	660
12871	Detección de productos Software para la Industria 4.0	693
12714	Diseño y Desarrollo de aplicaciones móviles basadas en GEOFENCING	645
12731	Enfoque para el diseño de aplicaciones móviles usando técnicas de Design Thinking dirigido a usuarios con dificultades/desafíos en habilidades sociales (Fase I).	649
12865	Herramienta Web para Interoperabilidad Conceptual entre UML, EER y ORM 2	688
12696	Impacto de las TICs en la sustentabilidad del uso del agua potable en la República Argentina	631
12823	Integración de una BCI para rehabilitación de problemas cognitivos en pacientes neurológicos infantiles implementando VRPN como protocolo de comunicaciones	669
12928	Inteligencia y tecnologías aplicadas al deporte de alto rendimiento	698
12756	Interfaces no convencionales aplicadas a la captura de datos en procesos productivos industriales aplicados a las PyMEs	655
12704	Tecnologías Vestibles aplicadas al cuidado de la salud construcción de un prototipo de monitoreo	636
12824	Un marco de trabajo para el desarrollo de software de dominio específico en el contexto de Gobierno Digital	673
12860	Uso de la tecnología Blockchain Federal (BFA) para dejar pistas de auditoría y trazabilidad a sentencias y acordadas de la Suprema Corte de la provincia de Mendoza	683
	PROCESAMIENTO DE SEÑALES Y SISTEMAS DE TIEMPO REAL	703

12943	Detección de patologías en señales biomédicas mediante técnicas de machine learning	732
12833	Identificación y compensación del sesgo en sensores inerciales MEMS de muy bajo costo mediante el uso de algoritmos de aprendizaje automático	718
12733	Modelo mejorado para la estimación de puntos de borde en imágenes Sar Polarimetricas	704
12800	Modelos Matemáticos y Métodos Computacionales en Ingeniería	714
12769	Monitoreo de Catástrofes Naturales a partir de la Obtención y Procesamiento de Imágenes Satelitales	709
12840	Redes de Sensores, Vehículos móviles y Simulación en Sistemas de Tiempo Real	724
12872	Síntesis de sonido en dispositivos móviles	728
	PROCESAMIENTO DISTRIBUIDO Y PARALELO	737
12825	Computación Serverless para tratamiento de datos provenientes de dispositivos de IoT	753
12855	Fundamentos, Algoritmos y Evaluación de rendimiento en diferentes plataformas de HPC	763
12744	Hacia la Optimización de un Sistema de Recuperación de Información	738
12856	Hardware, Software, Modelos, Métricas y Tendencias en Arquitecturas Multiprocesador	769
12750	Métodos y Algoritmos para Procesamiento Distribuido de Grafos Masivos y Evolutivos	743
12932	Modelo lattice-Boltzmann para flujo multifásico con transferencia de calor en GPU	797
12901	Procesamiento de datos masivos en tiempo real y consumo energético de sistemas paralelos	792
12874	Rendimiento de Cloud Computing para HPC en IaaS privados y públicos	787
12790	Sintonización de Aplicaciones científico/ingenieriles: un proceso de medición-mejora para incrementar la eficiencia	748
12832	Software y aplicaciones en Computación de Altas Prestaciones para el contexto de la UNdeC	758
12858	Solución de Grandes Problemas aplicando HPC Multi-tecnología	776
12868	Técnicas de modelado y simulación en entornos HPC	782
	SEGURIDAD INFORMÁTICA	802
12781	Anáisis de la seguridad del protocolo de comunicaciones CAN	818
12896	Análisis de Frameworks de Nube: Microsoft Azure y Amazon Web Services, mediante versiones privadas de prueba en entornos educativos	852
12698	Aplicación de redes neuronales profundas para la detección automática de Nombres de Dominio generados de manera algorítmica.	803
12835	Avances en Aspectos de Seguridad Aplicados a Sistemas de Voto Electrónico	838
12807	Avances en Reconocimiento de Patrones de Tecleo para la Identificación de Personas en Ambientes Web	828
12768	Blockchain para aseguramiento de Evidencia Digital en entornos Forensic Readiness	813
12867	Criptología Simétrica Encriptado-Autenticado	848
12766	Detección de vulnerabilidades en especificaciones de contratos inteligentes de la plataforma Ethereum	808
12821	Implementación de un nodo minero institucional en la red Ethereum Blockchain Federal Argentina	833
12939	Métodos y Herramientas para el análisis forense de dispositivos móviles	857
12863	Participación y despliegue de CTFs como herramienta para fortalecer la formación en ciberseguridad	843

12788	Seguridad en Internet de las Cosas usando soluciones Blockchain	823
	TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN	862
12795	Análisis desde la perspectiva interacción Hombre-Máquina de los procesos de inclusión en el contexto universitario	907
12882	Análisis y detección de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico - Fase 2	977
12793	Analítica de aprendizaje aplicada al contexto de la enseñanza superior mediante la definición de variables y métricas para la valoración de rendimiento académico	902
12772	Aplicación de rúbrica C.O.d.A para evaluación de calidad objetos de aprendizajes basados en realidad aumentada	883
12876	Avances y Uso de las Tecnologías Informáticas en la Educación	972
12869	Computación Ubicua aplicada al aprendizaje: Implementación ene l curso de ingreso universitario	967
12810	Construcción de una Herramienta de Capacitación a Distancia y Recursos Educativos Abiertos bajo un Enfoque de Diseño Universal orientados a Personas con Discapacidad Visual	917
12785	Desarrollo de Herramientas IoT para la Enseñanza de Sistemas de Control	888
12820	Diseño de un sistema de medición de desempeño para Moodle en Educación Superior	932
12804	Dispositivos móviles para la educación en colegios de nivel secundario	912
12721	El proceso de desarrollo de serious games Modelos, herramientas y analíticas de aprendizaje	878
12831	Enseñanza de TICs, mediante el desarrollo de videojuegos utilizando metodologías STEAM	941
12886	Entornos y herramientas digitales para el aprendizaje y la colaboración	987
12853	ESDEU: Sistema de Gestión tutorial caso de prueba en comisiones de ingresantes 2020 de UTN La Plata	950
12816	Evaluación Cuantitativa y Sistemática de software educativo libre	927
12841	Evento de Transferencia DeCoDev, Desarrollo Colaborativo de un Videojuego	945
12693	Incorporación de la Tecnología Móvil en el Proceso Educativo	863
12919	Juegos Serios Móviles. Diseño, Desarrollo e Integración en escenarios educativos	1004
12828	Los Estilos de Aprendizaje en la Construcción de Tutores Inteligentes. Una Revisión Bibliográfica	937
12933	Materiales educativos digitales: Macro narrativas y Micro narrativas para la construcción de conocimientos en la Universidad	1014
12709	Metodología de implementación de un Chatbot como tutor virtual en el ámbito educativo	873
12885	Minería de Datos educacional para determinar perfiles de alumnos recursantes de Carreras de Ingeniería	982
12813	Modelo de Sistema de Gestión de calidad para la Cátedra de Programación Numérica	922
12893	Plataforma Web de entrenamiento en el uso de pulsadores virtuales para personas con alteraciones en el desarrollo del lenguaje	999
12864	Por un Diseño inclusivo. Caso de uso en un MOOC de Accesibilidad Web	960
12957	Proyectos de Innovación y Transferencia de Tecnologías de la región del Noroeste de Buenos Aires (PRITT NOBA)	1026
12702	Puesta en valor de un simulador de entrenamiento mediante la incorporación de experiencia inmersiva y analíticas de aprendizaje	868

12931	Realidad Aumentada y Realidad Virtual: experiencias en diferentes ámbitos de aplicación	1009
12946	REFORTICCA: Recursos para el Empoderamiento de FORmadores en TIC, Ciencias y Ambiente	1019
12792	Simulación y equipo real Mikrotik en la enseñanza de redes de computadoras en el nivel universitario (resultados parciales)	897
12891	Tecnologías emergentes y modelos de interacción avanzados para contextos educativos Tesis de Maestría: Diseño e implementación de juguetes interactivos para actividades educativas basadas en interacción tangible	992 998
12786	Uso de las TICs para la Construcción de Espacios Institucionales - Generación Automática de Sitios Web Mediante Frameworks	892
12857	Visualización de datos en un Tablero de Comando aplicado a plataformas de educación a distancia en el nivel Superior	955
	TESIS DOCTORALES	1031
12884	Bocetado de interacciones enactivas	1072
12822	Empoderamiento de la Conciencia Situacional en operaciones militares usando Realidad Aumentada	1042
12862	Generación automática inteligente de resúmenes de textos con técnicas de Soft Computing	1062
12927	Medidas de Invarianza y Equivarianza a Transformaciones en Redes Neuronales Convolucionales	1082
12837	Modelos de Madurez para la Mejora de Calidad de los Datos de los Indicadores de Desarrollo Sostenible	1052
12759	Robustez de las métricas de clasificación de cadencia de tecleo frente a variaciones emocionales	1032
12948	SEDAR: Detección y Recuperación Automática de Fallos Transitorios en Sistemas de Cómputo de Altas Prestaciones	1092

Agentes y Sistemas Inteligentes

Técnicas de Inteligencia Artificial aplicadas a Problemas de Visión por Computadora

Javier Izetta Riera, Nilda Pérez Otero, Abigail Verazay, Fabiola Paz, Virginia Battezzati, Ronaldo Chuca, Facundo Llampa, Facundo Contreras, Osvaldo Cuellar, Cristian Ponce, Abel Díaz, Fabián Arjona
GIDIA / Facultad de Ingeniería / Universidad Nacional de Jujuy
Ítalo Palanca 10, +54 (388) 4221587
{javierizetta, nilperez, abigailrn, fabyppaz, virginiavir, ronaldochuca, a.facundollampa, facucontreras21, o.e.cuellar94, cristian.marcelo.ponce, abelguillermodiaz, nestorfabianar }@gmail.com

Resumen

El Aprendizaje Automatizado (AA) es uno de los campos científicos y técnicos de más rápido crecimiento en la actualidad. Está situado en el núcleo de la inteligencia artificial y ciencia de datos. Sus técnicas han surgido como las más elegidas para dar solución a problemas en distintas disciplinas. Una de ellas es la visión por computadora cuya finalidad es crear modelos del mundo real a partir de imágenes. Entre las tareas más importantes de esta disciplina están la detección de puntos claves faciales, el reconocimiento de emociones humanas y la detección de plagas en cultivos. El presente proyecto tiene como objetivo el desarrollo de nuevos algoritmos basados en AA que permitan realizar estas tareas en forma automática.

Palabras clave: Aprendizaje Automatizado, Aprendizaje Profundo, Visión por Computadora.

Contexto

Las líneas de investigación aquí presentadas se encuentran insertas en el proyecto *Técnicas de Inteligencia Artificial Aplicadas a Visión por Computadora*, ejecutado a partir de 2019 por el Grupo de Investigación y Desarrollo en Informática Aplicada (GIDIA) de la Facultad de Ingeniería de la Universidad Nacional de Jujuy.

El proyecto se encuentra acreditado por la Secretaría de Ciencia y Técnica y Estudios Regionales de la Universidad Nacional de Jujuy, enmarcado en la convocatoria "Desafíos de Investigación".

Introducción

El Aprendizaje Automatizado (AA) o *Machine Learning* es la parte central de la nueva revolución tecnológica basada en el uso inteligente de la información. Su objetivo principal es la construcción de algoritmos que automáticamente mejoren su eficiencia en la solución de un problema a través de la experiencia acumulada. Se puede definir AA como un proceso de inducción de conocimiento, ya que busca crear modelos que sean capaces de generalizar comportamientos o tomar decisiones sobre nuevas situaciones, a partir de una información no estructurada e implícita proporcionada en forma de ejemplos [1]. Dentro de la inteligencia artificial (IA), el AA ha surgido como el campo más elegido para dar solución a

problemas de visión por computadora, reconocimiento de voz, lenguaje natural, control de robots, y otras aplicaciones [2]. En particular, la visión por computadora es una disciplina científica que surge debido a la necesidad de desarrollar métodos que permitan adquirir, procesar, analizar y comprender las imágenes en formato digital con el fin de obtener información simbólica o numérica que pueda ser tratada en una computadora [3]. Está inspirada en la capacidad humana de comprender el mundo que lo rodea por medio de su sentido de visión.

Detección de puntos clave faciales

El problema del reconocimiento de rostros es una tarea que el ser humano puede llevar a cabo de forma sencilla. Sin embargo, no es posible brindar una explicación simple del procedimiento con el cual es llevada a cabo y mucho menos llevarlo de forma directa a un programa computacional. Para dar solución a esta problemática se han creado diversas aplicaciones basadas en AA que permiten, a partir de una imagen, la extracción de características particulares del rostro que ayuden a la identificación de personas [4], [5]. Una de las tareas clave, para la extracción de características de rostros, es la detección de puntos claves o *keypoint*. Estos puntos son aquellos que están presentes en las estructuras sobresalientes o protrusiones del rostro y pueden ser utilizados para desarrollar modelos. La detección de estos puntos, es el proceso inicial que sirve para desarrollar tareas subsecuentes como el reconocimiento automático de personas, la detección de signos faciales dismórficos para el diagnóstico médico, el seguimiento de rostros en imágenes y videos de forma automatizada para la implementación de políticas de seguridad, entre otros. En la literatura existen varias propuestas que aplican métodos de AA para la detección de puntos clave en imágenes [6], [7] que evidencian un creciente interés en el desarrollo de algoritmos basados en la detección de estos puntos que brinden soluciones a una amplia gama de problemas de visión artificial. Por todo esto, la primera línea de investigación del presente proyecto es el desarrollo de algoritmos que detecten de forma automática puntos clave faciales en imágenes digitales de rostros humanos.

Reconocimiento de emociones humanas

Además de la identificación de las personas, existen muchas situaciones en las que se necesita extraer otro tipo de información. Por ejemplo, el estudio del comportamiento humano analiza cómo reacciona una persona ante ciertas situaciones que ocurren en su entorno. Esta reacción, no es siempre de la misma manera en todas las personas ya que dependen de factores tales como edad biológica, estado de salud o estado emocional, entre otros. Con respecto a este último, Paul Ekman propuso seis tipos de emociones panculturales: enojo, disgusto, miedo, felicidad, tristeza y sorpresa [8]. Éstas constituyen las bases de diferentes investigaciones y recientemente se agregó la emoción neutral [9]. En los últimos años, los métodos de Aprendizaje Profundo (AP) o *Deep Learning* han sido utilizados con éxito para la detección de emociones en expresiones faciales. Estos métodos utilizan modelos computacionales que están compuestos de múltiples capas de procesamiento para aprender representaciones de datos con múltiples niveles de abstracción [10]. Entre las técnicas más utilizadas del AP, las Redes Neuronales Convolucionales o *Convolutional Neural Networks* (CNN) [11] han sido aplicadas en forma efectiva para la detección automática de expresiones faciales en imágenes digitales

[12], [13], [14]. Los resultados obtenidos han sido muy buenos, sin embargo, no hay un consenso en cuanto a qué arquitectura usar, lo que implica la necesidad de buscar nuevas arquitecturas adecuadas para este problema. Por ello, la segunda línea de investigación del presente proyecto es el desarrollo de arquitecturas basadas en CNN que clasifique de forma automática emociones en imágenes digitales de rostros humanos.

Detección de plagas o enfermedades en cultivos

Por último, la tercera línea de investigación de este proyecto está enfocada a los cultivos andinos. Son considerados como tales por su resistencia y adaptación a sequías, heladas y suelos con características salinas [15]. En especial la papa, que posee excelentes características nutricionales para su consumo ya que presenta altos porcentajes de materia seca, almidón, proteínas y minerales, así como cualidades culinarias destacables [16]. Sin embargo, cuando este cultivo se ve afectado por plagas y enfermedades, provoca pérdidas en el rendimiento y calidad de los productos antes y después de la cosecha. Por lo tanto, tomar decisiones oportunas es vital para mantener y conservar sus propiedades como así también, una producción más eficiente [17]. En consecuencia y considerando la importancia de conservar el valor alimenticio y cultural del cultivo de papa andina, es que esta línea de investigación se centra en el desarrollo de un algoritmo para la identificación de plagas o enfermedades mediante el empleo de técnicas de procesamiento de imágenes.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación antes mencionadas consisten en analizar técnicas de AA y AP para aplicarlas al desarrollo de algoritmos robustos para resolver automáticamente los problemas de detección de puntos clave faciales, el reconocimiento de emociones humanas y la detección de plagas en cultivos.

Resultados y Objetivos

El proyecto, que se desarrolla en el período 2019-2020, tiene como objetivo crear algoritmos que implementen técnicas de IA, una combinación de las ya existentes u otras técnicas para resolver los problemas anteriormente descriptos. Los resultados que se lograron hasta el presente son los siguientes:

En la primera de línea de investigación se desarrollaron nuevos algoritmos basados en *Random Forest* (RF) [18] y *Support Vector Machine* (SVM) [19] que permitieron la detección automática de puntos clave faciales en imágenes de rostros humanos. Para el desarrollo de los algoritmos se utilizó el conjunto de imágenes digitales obtenidas de *Kaggle* [20]. Este conjunto de datos fue utilizado por diferentes trabajos de investigación, lo que permitió comparar nuestros resultados con los obtenidos en trabajos anteriores. El mismo contiene 7049 ejemplos de imágenes de rostros humanos con 15 puntos clave. Las experimentaciones realizadas en [21] permitieron demostrar que nuestra propuesta basada en RF logró mejorar significativamente los resultados de trabajos anteriores [22], [23], es decir, mejoró la precisión de las posiciones de todos los puntos clave, disminuyendo la tasa de error a 2 pixeles.

En la segunda línea se desarrollaron dos nuevas arquitecturas basadas en CNN para

la clasificación automática de emociones en imágenes digitales de rostros humanos. Para el desarrollo de las arquitecturas se utilizó el conjunto de imágenes desarrollado por *Jeffrey Cohn y Takeo Kanade* [24], tomado como referencia por la mayoría de los trabajos científicos sobre reconocimiento de emociones. Este conjunto se caracteriza por contener imágenes de rostros humanos asociadas a una emoción. Para mejorar el entrenamiento se aplicaron técnicas de procesamiento de imágenes, tales como ecualización de histograma, realce por modificación de contraste, rotaciones, redimensionamiento, traslaciones, entre otras. Las experimentaciones realizadas permitieron alcanzar una exactitud alrededor del 99%. Este año se prevé profundizar en las experimentaciones y realizar un estudio comparativo de nuestros resultados con los obtenidos por trabajos anteriores.

En la tercera línea se desarrolló un algoritmo para la detección de enfermedades o plagas en fotos de hojas de plantas de cultivos de papas andinas. Para ello se aplicaron técnicas bien conocidas de procesamiento de imágenes digitales. Este proceso inicia con la captura de la imagen y luego mediante un análisis del color de la imagen capturada se determina la existencia o no de alguna enfermedad o plaga. En el caso que exista alguna anomalía en la hoja, se cuantifica y muestra el límite del área dañada. Los experimentos de este trabajo, se realizaron con varias muestras de hojas de plantas de cultivos dañadas y sanas extraídas de la web. El algoritmo propuesto logró detectar las plantas de cultivo de papas que presentan alguna enfermedad o plaga y determinar una aproximación aceptable del porcentaje del área dañada [25].

Formación de Recursos Humanos

El equipo de trabajo está integrado por docentes-investigadores, egresados y alumnos de la Universidad Nacional de Jujuy, cuenta con cinco docentes investigadores (un doctor en informática y cuatro ingenieros en informática), dos egresados (ingenieros en informática) y cinco alumnos de la carrera de ingeniería en informática. En el marco de este proyecto se finalizaron tres tesinas de grado. Se prevé realización de dos tesinas de grado y una tesis de maestría.

Referencias

[1] Mitchell, T., Buchanan, B., DeJong, G., Dietterich, T., Rosenbloom, P., & Waibel, A. (1990). Machine learning. Annual review of computer science, 4(1), 417-433.

[2] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

[3] Gonzalez, R. C., & Richard, E. (2002). Woods, digital image processing. ed: Prentice Hall Press, ISBN 0-201-18075-8.

[4] Muñoz-Salinas, R., Aguirre, E., & García-Silvente, M. (2007). People detection and tracking using stereo vision and color. Image and Vision Computing, 25(6), 995-1007.

[5] Han, C., Shan, S., Kan, M., Wu, S., & Chen, X. (2018). Face recognition with contrastive convolution. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 118-134).

[6] Gall, J., & Lempitsky, V. (2013). Class-specific hough forests for object detection. In Decision forests for computer vision and medical image analysis (pp. 143-157). Springer London.

[7] Dantone, M., Gall, J., Fanelli, G., & Van Gool, L. (2012, June). Real-time

facial feature detection using conditional regression forests. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 2578-2585). IEEE.

[8] Pellejero, N. F., Grinblat, G. L., & Uzal, L. C. (2017). Análisis semántico en rostros utilizando redes neuronales profundas,

[9] Raghuvanshi, A., & Choksi, V. (2016). Facial Expression Recognition with Convolutional Neural Networks. CS231n Course Projects.

[10] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436.

[11] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[12] Li, J., & Lam, E. Y. (2015, September). Facial expression recognition using deep neural networks. In Imaging Systems and Techniques (IST), 2015 IEEE International Conference on (pp. 1-6). IEEE.

[13] Lopes, A. T., de Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. Pattern Recognition, 61, 610-628.

[14] Jain, R., Rangachar, K., and Schunk, B. (1995). Computer Vision, McGraw-Hill, New York.

[15] Alvarez, S., Cáceres, M., Calle, M., Córdoba, P., Giménez, L., López, A., Méndez, D., Quispe, J., Rodríguez, J., Singh, G. (2015). Conociendo los cultivos andinos de mi provincia. Calidad y educación alimentaria. Comisión Nacional de Bibliotecas Populares (CONABIP). ISBN 978-987-3926-00-6. Universidad Nacional de Jujuy.

[16] INTA informa (2012). Las papas nativas son de Interés Nacional. Recuperado en junio 2019 de intainforma.inta.gov.ar/?p=14123.

[17] Pérez, W., & Forbes, G. Guía de identificación de plagas que afectan a la papa en la zona andina. International Potato Center.

[18] Breiman, L. (2001). Random Forests Machine Learning. 45: 5–32. View Article PubMed/NCBI, Google Scholar.

[19] Cortes, C. & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3):273– 297.

[20] Kaggle. Facial Keypoint Detection Competition. URL: https://www.kaggle.com/c/facialkeypointsdetection. (2016, September).

[21] N. Fabián Arjona, Abel G. Díaz, C. Javier Izetta & N. Revollo Sarmiento. (2019). Detección de puntos clave en rostros humanos. VIII IPCTIIC. Cordoba, Argentina.

[22] Longpre, S., & Sohmshetty, A. (2016). Facial Keypoint Detection.

[23] Wang, Y., & Song, Y. Facial Keypoints Detection.

[24] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotionspecified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on (pp. 94-101). IEEE.

[25] Fabiola Patricia Paz, C. Javier Izetta. (2019). Detección de enfermedades o plagas en cultivos de papas andinas del norte argentino utilizando procesamiento de imágenes digitales. VIII IPCTIIC. Cordoba, Argentina.

Inteligencia Artificial aplicada al Poder Judicial

Osvaldo Spositto, Viviana Ledesma, Gastón Procopio, Julio Bossero

Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza
spositto@unlam.edu.ar, vledesma@unlam.edu.ar,
gprocopio@ unlam.edu.ar, jbossero@unlam.edu.ar.

RESUMEN

El uso de la tecnología en el ámbito judicial ha tenido un enorme desarrollo en la actualidad. Las actividades informatizadas son numerosas y variadas: desde la planificación de la agenda de jueces y magistrados, hasta la redacción automatizada de textos jurídicos, pasando por la gestión de las causas, incluso su utilización como apoyo en la toma de decisiones relacionadas a las sentencias. El presente proyecto tiene como objetivo principal la generación de conocimiento especializado en el análisis, diseño y construcción de una herramienta informática que ayude a la sistematización y optimización de varios de los procesos judiciales que actualmente se realizan en forma manual o semiautomática en un juzgado perteneciente al Poder Judicial de la Provincia de Buenos Aires. A partir de las bases de información generadas por este sistema a futuro se pretende aplicar otras técnicas asociadas a la Inteligencia Artificial en búsqueda de la optimización de los procesos.

Palabras clave: Nuevas Tecnologías de la Información y la Comunicación, Informática Jurídica, Justicia Digital, Inteligencia Artificial, Sistemas Expertos.

CONTEXTO

La línea de investigación aquí presentada está enmarcada en el proyecto de investigación "*Diseño e Implementación de un Sistema Experto como Apoyo al Proceso de Despacho de Trámites de un* Organismo Judicial", dentro del Programa de Incentivos para Docentes Investigadores de la Secretaría de Políticas Universitarias (PROINCE) 2020-2021. Este proyecto es financiado por la Universidad Nacional de La Matanza (UNLaM), se lleva a cabo por investigadores de dos departamentos, el Departamento de Ingeniería e Investigaciones Tecnológicas y el Departamento de Derecho y Ciencias Sociales. A su vez el mismo se realiza con la estrecha colaboración del Departamento de Desarrollo Informático dependiente de la Subsecretaría de Tecnología Informática del Poder Judicial de la Provincia de Buenos Aires y del Juzgado de Ejecución Nº2 del Departamento Judicial Morón.

El mismo se desarrolla en el Polo Tecnológico dependiente del Departamento de Ingeniería e Investigaciones Tecnológicas de la UNLaM.

1. INTRODUCCIÓN

Hoy en día, son indiscutibles los beneficios que la sociedad del conocimiento ha recibido de las Nuevas Tecnologías de la Información y la Comunicación. Esto queda demostrado

en las diferentes ramas científicas de la sociedad, donde ahora sería impensable una aplicación y desarrollo apropiado de las mismas sin la ayuda de los avances informáticos. Tal es el caso del uso de estas tecnologías en el Poder Judicial, lo que se conoce como informática jurídica, un instrumento del derecho, que puede dividirse en tres áreas de la siguiente manera [1, 2]:

1. *Informática Documentaria*, trata de crear un corpus jurídico documentario, relativo a diversas fuentes del derecho a los efectos de análisis y recuperación de información en función a criterios propios acordes a esa información y su relevancia jurídica.
2. *Informática de Control y Gestión*, se utiliza para seguimiento de los tramites y procesos con la finalidad de mantener la información actualizada y a su vez tener un mayor control de las actividades.
3. *Informática Decisoria*, se conforma por bases de conocimiento jurídico. Sus ámbitos de aplicación se relacionan con cinco subáreas: soporte para la decisión, para la redacción, ayuda en la previsión, en la investigación y en la educación.

Uno de los sectores más dinámicos y en constante evolución de la informática jurídica decisional es el que se refiere a la aplicación al derecho de la Inteligencia Artificial, ya que la finalidad, en este caso, consiste en la creación de *Sistemas Expertos* [3] capaces de simular el razonamiento jurídico (decisional). En este contexto, un sector con importantes innovaciones es el de los sistemas expertos legales. Para tal fin es imprescindible una intensa colaboración entre juristas e informáticos, dado que el objetivo que se persigue es reproducir de manera automática las actividades del jurista, pretendiendo brindar una o más alternativas de solución a los problemas y no simplemente aportar documentación referida a dichos problemas.

En este contexto surge el presente trabajo, que se encuadra como un Proyecto de Desarrollo Tecnológico y Social, dado que tiene por objeto cubrir una necesidad social y tecnológica real de los tribunales de la provincia de Buenos Aires, en particular, consiste en asistir a la tarea de despacho de trámites del Juzgado de Ejecución Nº 2 del Departamento Judicial Morón.

En [3] encontramos la siguiente definición "…*el del despacho judicial, en el que se sitúan las verdaderas unidades de producción del sistema de justicia, donde los aspectos de gestión cobran mayor importancia y donde ha sido más difícil introducir mejoras significativas al sistema tradicional de organización y funcionamiento...*".

A partir de la constatación de esta centralidad, existen varios proyectos que se ocupan de la tecnología jurídica, en [4] encontramos la siguiente definición: "*...El término se refiere a la adopción de tecnologías y programas informáticos innovadores para racionalizar y mejorar los servicios jurídicos...*".

Es decir, el desempeño de la función judicial por parte del juez exige la realización de numerosos actos materiales y una actividad ordenatorio-judicial que es desarrollada por un personal que se halla bajo las órdenes directas del mismo o de una persona que responde directamente al juez en relación con esa actividad. Esa entidad, denominada *despacho judicial*, se refiere al conjunto de personas que participan subordinadamente en la administración de justicia mediante el desempeño de funciones materializadas en la instrumentación del proceso.

Con lo anterior presente, la investigación aborda el análisis, diseño y construcción

de una herramienta informática que de soporte a las decisiones y ayude a la sistematización y optimización de varios de los procesos judiciales que actualmente se realizan en forma manual o semiautomática en dicho juzgado. De esta manera se busca estandarizar el proceso de despacho de trámites, y a la vez agilizar y reducir los tiempos de carga, minimizando posibles errores, tanto durante la toma de decisiones, como en el ingreso de datos.

Se espera que el sistema desarrollado a partir de este proyecto sea implementado para su utilización tanto por del Juzgado de Ejecución Nº2 del Departamento Judicial Morón como también por la Secretaría de Planificación de la Suprema Corte de Buenos Aires.

En una etapa posterior el sistema Experticia se integrará con el Sistema Informático preexistente denominado AUGUSTA[1]. Se trata de un sistema para la asistencia integral en la Gestión de las Causas de los organismos jurisdiccionales de las diferentes instancias y fueros. Es un Sistema de Gestión Integral en el cual se registran datos de los Casos a partir de la Demanda y luego se registran todos los pasos procesales, las partes o personas intervinientes, documentación anexa y toda aquella información que contribuya a la gestión del mismo. Asiste en el despacho del organismo con una biblioteca de modelos propios al organismo o genéricos.

Por lo tanto, con este proyecto se pretende establecer un punto de partida para que, en un futuro, a partir de las bases de información generadas durante la utilización de este sistema, sea posible incursionar en otras áreas de la *Inteligencia Artificial* ampliando la optimización del proceso sin perder la calidad esperada.

[1] Disponible en: http://www.scba.gov.ar/subinformacion/augusta.asp.

2. LÍNEAS de INVESTIGACIÓN y DESARROLLO

El presente trabajo tiene como eje central el análisis de distintas aplicaciones de la Inteligencia Artificial al ámbito del poder judicial.

Entre las líneas de investigación a considerar en este proyecto se pueden mencionar:

a) Diferentes técnicas aplicadas a Sistemas Expertos.
b) Algoritmos utilizados en la Minería de Datos y en Machine Learning para el descubrimiento de patrones.
c) Aplicación de la Indexación Semántica Latente para la recuperación de información.

Este proyecto tiene como objetivo general establecer los criterios para coordinar el proceso de ingeniería de software para el desarrollo de una aplicación para el Juzgado de Ejecución Nro. 2 del Departamento Judicial Morón que facilite la toma de decisiones, incremente la calidad de los servicios jurisdiccionales, agilicen las comunicaciones, mejore la relación interjurisdiccional y reduzca el tiempo de tramitación de las causas. De este se desprenden los siguientes objetivos específicos:

a) Analizar, diseñar e implementar una infraestructura tecnológica y humana que garantice la viabilidad, continuidad y sustentabilidad del modelo propuesto.
b) Desarrollar procesos para brindar entrenamiento y capacitación a jueces, funcionarios, empleados judiciales y usuarios.
c) Actualizar la infraestructura de comunicaciones, la de datos, la de seguridad informática y la de

equipamiento necesaria para el funcionamiento de la aplicación y su integración con el Sistema Augusta.

d) Estudiar diferentes tendencias respecto al desarrollo de sistemas expertos y de algoritmos de minería de datos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Este grupo de investigación viene trabajando, en proyectos previos asociados a las líneas de investigación mencionadas en el apartado anterior, lo que dio lugar a las siguientes publicaciones:

1. "*Predicción del riesgo de abandono universitario utilizando métodos supervisados*" En colaboración con Edwards, Diego y Pérez, Silvia (UNLaM). Trabajo presentado en el Workshop de la V Jornadas Nacionales y I Latinoamericanas de Ingreso y Permanencia en Carreras Científico – Tecnológicas. Facultad Regional Bahía Blanca. Universidad Tecnológica Nacional. Bahía Blanca. Mayo de 2016. IPECyT 2016.
2. "*Modelos de minería de datos para el diagnóstico de enfermedad de Parkinson mediante el análisis de voz*". En colaboración con el Ing. Osvaldo Spositto, Ing. Gabriel Blanco, Mg. Mónica Giuliano y el Ing. Luis Fernández (UNLaM). Trabajo presentado en el Workshop del V Congreso Nacional de Ingeniería en Informática/Sistemas de Información. Publicación en línea - ISSN. CONAIISI 2017. Santa Fe. Argentina.
3. "*Comparación de Algoritmos de Aprendizaje Supervisado para la obtención de perfiles de alumnos desertores*". En colaboración con el Ing. Osvaldo Spositto (UNLaM). Trabajo presentado en el Workshop del IV Congreso Nacional de Ingeniería en Informática/Sistemas de Información. Publicación en línea - ISSN 2347-0372. CONAIISI 2016. Salta. Argentina.
4. "*Aceleración en la Recuperación de Información utilizando Algoritmos de Minería de Datos de R*". En colaboración con el Ing. Osvaldo Spositto, Lic. Julio Bossero, J., Ing. Hugo Ryckeboer. (UNLaM). Trabajo presentado en el Congreso Argentino de Ciencias de la Computación - CACIC 2018. Universidad Nacional del Centro, Tandil.

Este proyecto, en particular, como se mencionó anteriormente es un paso inicial que servirá para posteriormente aplicar distintas técnicas de Inteligencia Artificial a los sistemas del poder judicial. El sistema a implementar un sistema, se denominará "Experticia" y pretende ser desarrollado mediante una estructura de Árbol Binario.

A futuro se buscará integrar dicho sistema con el sistema ya existente AUGUSTA, incorporando la funcionalidad de árbol de decisión, considerando la posibilidad de aplicar mecanismos basados en algoritmos de machine learning para:

- Aplicación de ajustes automáticos en los árboles de decisión basados en las acciones de los usuarios.
- Clasificación de documentos electrónicos y su correspondiente asociación automática a los circuitos de despacho establecido.
- Sugerencias asistidas para la selección de un trámite de despacho en función del estado de la causa.

4. FORMACIÓN DE RECURSOS HUMANOS

La presente línea de investigación la lleva adelante un equipo de 12 integrantes provenientes de dos

departamentos de la UNLaM, el Departamento de Ingeniería e Investigaciones Tecnológicas y el Departamento de Derecho y Ciencia Política.

- 8 docentes (2 de posgrado y 6 de grado).
- 2 (dos) alumnos de grado.
- 2 (dos) asesores especialistas externos. (1 perteneciente al Poder Judicial de la Provincia de Buenos Aires y un Secretario de Juzgado).

En esta línea de trabajo actualmente un integrante está desarrollando su tesis de doctoral, otros dos han finalizado recientemente sus Maestrías en Informática en la UNLaM.

Se prevé la capacitación y formación de recursos humanos, a través de cursos de actualización y posgrado en el área de estudio. Se espera la consolidación de los integrantes del equipo, en especial de los alumnos, como investigadores.

La transferencia de conocimiento y resultados serán compartidos en distintas cátedras, Análisis de Sistemas, Diseño de Sistemas y Base de Datos, entre otras. Se analizará la posibilidad de brindar charlas informativas del desarrollo e implementación del proyecto a distintos sectores interesados y cátedras afines a la investigación.

Por otra parte, dado que este desarrollo será implementado para el Poder Judicial de la Provincia de Buenos Aires se planifica brindar entrenamiento y capacitación a jueces, funcionarios, empleados judiciales y otros usuarios.

5. BIBLIOGRAFÍA

[1]. Tellez J. (1996) Derecho Informático, 2ª. Ed. McGraw Hill. México.

[2]. Anzalone, A. (2019) ¿Robotización Judicial? Breves Reflexiones Críticas. Journal of Ethics and Legal Technologies – Volume 1(1). Universita degli Studi di Padova. Italia.

[3]. Guibourg, R., Alende, J. & Campanella, E. (1996) Manual de Informática Jurídica. Informática Jurídica Decisoria. Tomo ASTREA pág. 151. Disponible en: http://www.saij.gob.ar/doctrina/daca960114-guibourg-manual_informatica_juridica_informatica.htm Consultado el 15/10/2019.

[4]. Vianco, J. (2005) Herramientas para el Diseño de Despachos Judiciales. III Seminario de Gestión Judicial. Centro de Estudio de Justicia de las Américas (CEJA). Disponible en: http://biblioteca.cejamericas.org/handle/2015/1124?show=full. Consultado el 25/02/2020.

[5]. Moisés Barrio, A. (2020). La tecnología jurídica no acabará con los abogados, solo facilitará su trabajo. Disponible en: https://theconversation.com/la- tecnologia-juridica-no-acabara-con-los-abogados-solo-facilitara-su-trabajo-129269. Consultado el 25/02/2020.

Cambio de Creencias Múltiple No-Priorizada

Néstor Jorge Valdez† **Diego Emanuel Peralta**†

† Departamento de Ciencias de la Computación, Fac. de Ciencias Exactas y Naturales
Universidad Nacional de Catamarca (UNCa)
Av. Belgrano 300 - San Fernando del Valle de Catamarca
Tel.: (03834)420900
e-mails: njvaldez@exactas.unca.edu.ar
diegodep2003@hotmail.com

Resumen

Esta investigación se desarrolla en el campo de *cambio de creencias*. La teoría de cambio de creencias estudia como un agente racional cambia el estado de sus creencias cuando ella es expuesta a *nueva información*. El modelo AGM (Alchourrón, Gärdenfors and Makinson) es considerado como el marco de trabajo estándar en cambio de creencias. No obstante, en los últimos años este modelo a sido criticado, en el sentido de su falta de realismo, y aplicabilidad en la realidad. Por ello, es que el modelo AGM fue objeto de diversas modificaciones y extensiones a fin de que su uso resulte computacionalmente factible. Entre esas modificaciones y extensiones al modelo AGM se puede mencionar las siguientes: generalizar sus operaciones a cambios múltiples de sentencias de entradas; cambios de creencias no-priorizada (el agente acepta o rechaza la nueva información) y cambio de creencias parcial (el agente a veces acepta parte de la nueva información). Esta línea de investigación tiene como objetivo principal el estudio y desarrollo de nuevos tipos de operadores de revisión múltiple no-priorizada. Así también, establecer sus diferencias con otros métodos alternativos que caracterizan funciones de revisión de creencias múltiple no-priorizada, como por ejemplo: Mult Semi-Revision (Fuhrmann 1997), Merge Partial Meet y Kernel (Falappa et al. 2012) y Making Up One's Mind (Zhang and Hansson 2015). Por último, se espera que los resultados obtenidos brinden una nueva perspectiva para desarrollar herramientas tecnológicas que funcionen para un enfoque cuyos modelos representen cambios en un mundo dinámico.

Palabras Claves Cambio de Creencias No-Priorizada, Operadores de Revisión Múltiple No-Priorizada , Dinámica de Conocimiento.

1. Contexto

Esta línea de investigación se realiza dentro del ámbito del Laboratorio de Investigación del Departamento de Ciencias de la Computación de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Catamarca. Está asociado con el proyecto de investigación Anual: "*Cambio de Creencias No-Priorizada en Sistemas Argumentativos Aplicado para Programación Lógica Rebatible*". Financiado por el programa de desarrollo científico y tecnológico de la Secretaría de Cien-

cia y Tecnología: Consejo de Investigación, de la Universidad Nacional de Catamarca. *Periodo* :01/01/2019 al 31/12/2020.

2. Introducción

La lógica de la teoría de cambio de creencias se enfrenta con el problema de modelar el conocimiento dinámico, es decir, como el estado epistémico de un agente es modificado en un escenario dinámico. El modelo AGM (Alchourrón, Gärdenfors and Makinson) [1] es considerado como el marco de trabajo estándar en cambio de creencias. Este modelo, establece que en una revisión de creencias se incorpora nueva información y se descarta cierta información existente para acomodar la recién ingresada. En los últimos años este modelo ha sido criticado, en el sentido de su falta de realismo, y aplicabilidad en la realidad. Por ello, es que el modelo AGM fue objeto de diversas modificaciones y extensiones a fin de que su uso resulte computacionalmente factible.

Algunos investigadores argumentan que la información nueva *no siempre debería tener prioridad sobre la información existente* por lo que se han propuesto algunos métodos de revisión de creencias *no-priorizada* en los que no se acepta necesariamente la nueva información. Por ejemplo, la *semi-revisión* introducida por Hansson [6, 7], ella difiere de la revisión de creencia clásica en dos aspectos: primero, la información original está representada como una base de creencias en lugar de un conjunto de creencias, y segundo, la nueva información no es siempre aceptada. Fuhrmann [4] propone una forma diferente de revisión no-priorizada, él define un *operador de mezcla* (merge operator) en el cual dos bases de conocimiento pueden combinarse en una. El operador de mezcla abre la posibilidad de que la nueva información sea parcialmente o totalmente ignorada si la vieja información es más fuerte. Este tipo de operador permite dos caminos *no posibles* en los operadores de revisión AGM clásica: aceptar parcialmente la nueva información o rechazarla totalmente. La operación de mezcla une la vieja información con la nueva, sin darle precedencia o prioridad a ninguna de ellas y elimina posibles contradicciones.

Los operadores de revisión no-priorizada vistos hasta ahora tienen una característica distintiva: o aceptan plenamente la sentencia a incorporar o la rechazan totalmente. Esto es, o son muy osados o muy cautos en su actitud epistémica. No aceptan una solución intermedia al problema de aceptación de una sentencia. Por tal motivo, Fermé y Hansson [3] proponen un operador de revisión que puede aceptar parte de la información que constituye la entrada epistémica. Por este motivo, es que los autores denominaron a este operador *revisión selectiva* (selective revision). El operador de revisión con *límite de credibilidad* es otro tipo de operador no-priorizada [8]. El mismo utiliza una construcción alternativa del operador de *screened revision* definido por Makinson [5]. Si debemos revisar un conjunto de creencias K con respecto a una sentencia α debemos recurrir a un conjunto C, denominado conjunto de creencias creíbles. Este conjunto, determina que sentencias podrían ser aceptadas en un proceso de revisión.

En resumen, podríamos mencionar como modificaciones y extensiones al modelo AGM a las siguientes:

- Generalizar sus operaciones a cambios múltiples de sentencias de entradas.
- Cambios de creencias no-priorizada (el agente acepta o rechaza la nueva información).
- Cambio de creencias parcial (el agente a veces acepta parte de la nueva información).

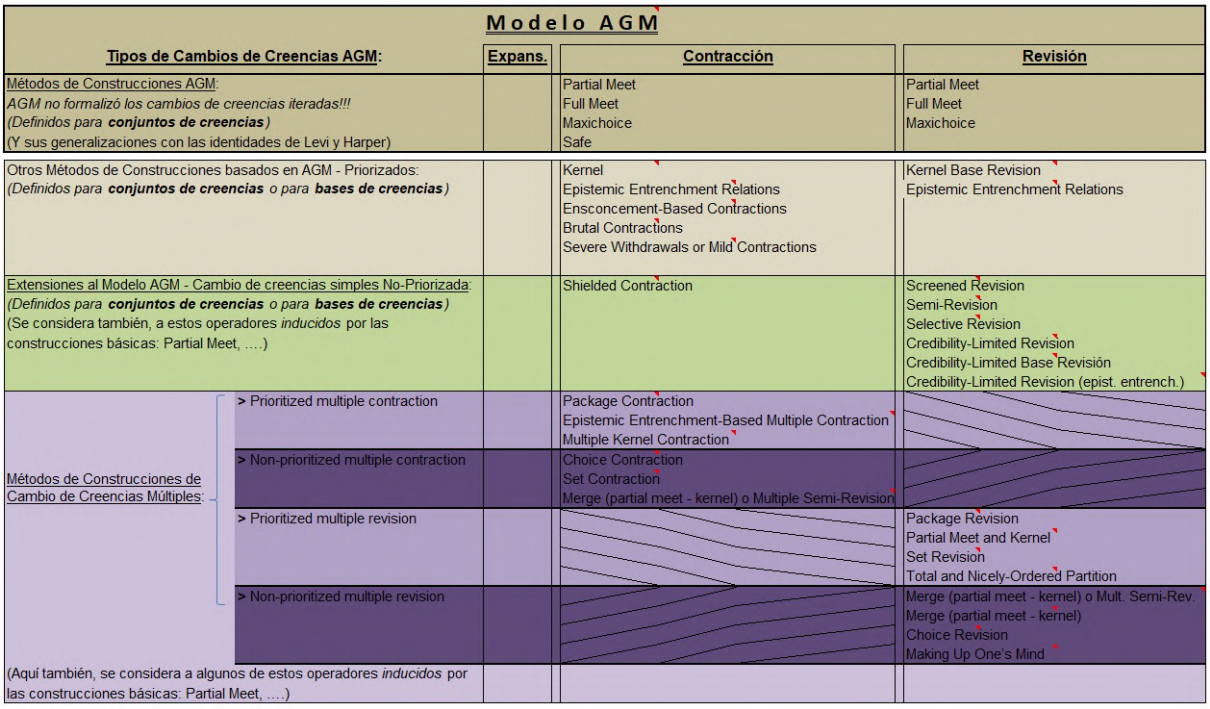


Figura 1: *Modelo AGM - modificaciones y extensiones.*

En la figura 1 se puede observar el estado del arte, con sus modificaciones y extensiones que resultaron en el desarrollo de modelos formales de operadores para los distintos tipos de cambio de creencias con sus teoremas de representación y caracterizaciones axiomáticas.

3. Línea de Investigación y Desarrollo

Esta línea de investigación toma como punto de partida nuestros aportes en la temática de operadores de contracción y revisión múltiple priorizada de la teoría de cambio de creencias bajo cláusulas Horn que ha sido ampliamente investigado por el autor principal de este artículo y cuyos resultados fueron partes de su tesis de Magister en Ciencias de la Computación en la UNS - Argentina [9, 10, 11, 13, 12].

El objetivo principal del proyecto es el estudio de operadores de cambio de creencias múltiple no-priorizada. Esto significa, proponer nuevas construcciones (con sus caracterizaciones axiomáticas) de funciones de revisión múltiple de tipo no-priorizada.

En este contexto, consideramos dos tipos de cambios: el primero, los operadores de cambio no-priorizada en donde todas las nuevas creencias no siempre deben ser aceptadas (como por ejemplo, los operadores con *credibilidad limitada*), y el segundo, a los operadores de *mezcla* (merging) que permite que creencias antiguas y nuevas jueguen roles simétricos dentro de un proceso de cambio. Así también, establecer sus diferencias con otros métodos alternativos que caracterizan funciones de cambio de creencias múltiple no-priorizada, como por ejemplo: Mult Semi-Revision [4], Merge Partial

Meet y Kernel [2] y Making Up One's Mind [14].

Estos objetivos nos permitirán proponer nuevas líneas de investigación dentro de la temática de este proyecto de investigación a los docentes-investigadores en ciencias de la computación que cursan carreras de posgrado en nuestro ámbito educativo.

Por último, se espera que los resultados obtenidos brinden una nueva perspectiva para desarrollar herramientas tecnológicas que funcionen para un enfoque cuyos modelos representen cambios en un mundo dinámico.

4. Resultados y Objetivos

Diversas contribuciones relacionado a la temática de esta investigación, fueron presentados en el ámbito de las ciencias básicas, provocando un impacto directo en el desarrollo de áreas tecnológicas de vanguardia en ciencias de la computación, tales como la aplicaciones de minería de datos, aprendizaje automático, framework machine learning (para transacciones comerciales), mecanismos de toma de decisión automática, etc.

En esta línea de investigación se espera obtener los siguientes resultados:

- Desarrollo de modelos formales (con sus caracterizaciones axiomáticas) para las operaciones de revisión múltiple no-priorizada.

- Determinar sus diferencias con otros métodos alternativos que caracterizan funciones de revisión de creencias múltiple no-priorizada, como por ejemplo: Mult Semi-Revision (Fuhrmann 1997), Merge Partial Meet y Kernel (Falappa et al. 2012) y Making Up One's Mind (Zhang and Hansson 2015).

- Establecer nuevas líneas de investigación dentro de la temática de este proyecto de investigación a los docentes-investigadores en ciencias de la computación que cursan carreras de posgrado en nuestro ámbito educativo.

Respecto a los objetivos de esta investigación, se espera obtener contribuciones en el área de las ciencias básicas y en el ámbito de aplicaciones tecnológicas.

Por último, se integrará programas de capacitación e intercambio para los integrantes del proyecto y otras áreas afines. Estas actividades serán de carácter:

- técnica: estudio y manejo de herramientas informáticas tanto para automatizar, procesar información (e.g. DeLP) y demás programas existentes para publicar resultados de investigación (e.g. LaTex, Beamer, etc) ; y

- metodológica: estrategias de investigación científica, mecanismos de publicación de artículos científicos, planificación de tutorías en ambientes presenciales y virtuales.

5. Formación de Recursos Humanos

Es intención primordial que los integrantes de este proyecto que no están en la actualidad categorizados por el programa de incentivos de investigación, adquieran los antecedentes científicos suficientes en esta línea de investigación lo que les posibilitaría poder ingresar al sistema del programa de incentivos en el futuro. Así también, los docentes-investigadores que integran el proyecto, realicen cursos de posgrados relacionados con la temática de esta investigación, con el objetivo de consolidar su formación

en investigación y el cursado de sus carreras de posgrado.

Por último, dentro de la temática de esta línea de investigación, se espera que los resultados científicos a obtener del proyecto contribuyan al proceso de elaboración de la tesis de doctorado de Néstor Jorge Valdez (uno de los autores de este artículo).

Referencias

[1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50, 1985.

[2] Falappa, Kern-Isberner, and Simari. Prioritized and non-prioritized multiple change on belief bases. *Journal of Philosophical Logic, ISSN 0022-3611, New York: Springer*, 41:77–113, 2012.

[3] Eduardo L. Fermé and Sven Ove Hansson. Selective revision. *Studia Logica*, 63:331–342, 1998.

[4] Fuhrmann. An essay on contraction. studies in logic, language and information. *CSLI Publications, Stanford*, 1997.

[5] David Makinson. Screened revision. *In Theoria of Hansson S. 1997*, pages–, 1997.

[6] Hansson S. Semi-revision. *Journal of Applied Non-Classical Logic*, pages 151–175, 1997.

[7] Hansson S. A survey of non-prioritized belief revision. *Erkenntnis*, 50:413–427, 1999.

[8] John Cantwell Sven Ove Hansson, Eduardo Fermé and Marcelo Falappa. Credibility limited revision. pages–, 1998.

[9] Valdez and Falappa. Dinámica de conocimiento: Contracción múltiple en lenguajes horn. *XIX Congreso Argentino de Ciencias de la Computación, XIV Workshop Agentes y Sistemas Inteligentes (WASI), CACiC'2013*, 2013.

[10] Valdez and Falappa. Dinámica de conocimiento: Contracciones horn a partir de ordenamientos epistémicos. *42JAIIO Jornadas Argentinas de Informáticas, ASAI 2013, 42 JAIIO'2013*, pages 206–209, 2013.

[11] Valdez and Falappa. Implementación para bases de creencias horn de operadores de contracciń múltiple. *XX Congreso Argentino de Ciencias de la Computación, XV Workshop Agentes y Sistemas Inteligentes (WASI), CACiC'2014*, 2014.

[12] Valdez and Falappa. Multiple revision on horn belief bases. *XXII Congreso Argentino de Ciencias de la Computación, XVII Workshop Agentes y Sistemas Inteligentes (WASI), CACiC'2016*, 2016.

[13] Valdez, Lara, Pedraza, and Teseira. Dinámica de conocimiento: Cambio de creencias múltiples. temática de investigación de inteligencia artificial en las ciencias de la computación. *JUCEN'15, Facultad de Ciencias Exactas y Naturales, UNCa. Catamarca 9 y 10 de diciembre de 2015 - Argentina*, pages 30–36, 2015.

[14] Li Zhang and Sven Ove Hansson. How to make up one's mind. *Logic Journal of the IGPL*, 23(4):705–717, 2015.

Inteligencia Artificial y Computación Cuántica en Finanzas.

Juan Pablo Braña, Alejandra M.J. Litterio y Alejandro Fernández
Centro de Altos Estudios en Tecnología Informática- Facultad de Tecnología Informática - U.A.I
Av. Montes de Oca 745 - (C1270AAH) Ciudad Autónoma de Buenos Aires, República Argentina
{juan.brana, alejandra.litterio, alejandroa.fernandez}@uai.edu.ar

RESUMEN

El proyecto de investigación en curso, que aquí se presenta, propone estudiar las aplicaciones de la Computación Cuántica (QC) y la Inteligencia Artificial (IA) utilizando los simuladores de la Plataforma Cloud de IBM-Q en el campo de las Finanzas y el Trading Algorítmico para los Mercados Financieros. Con el fin de realizar avances y mejoras en las aproximaciones actuales se diseñará un modelo híbrido de ambas tecnologías que permita codificar algoritmos de Machine Learning en computadoras cuánticas, siempre manteniendo el criterio comparativo entre aproximaciones tradicionales y cuánticas, para (1) la administración de carteras de inversión y el análisis de riesgo; y (2) el procesamiento y análisis de datos alternativos, como por ejemplo redes sociales, earnings conference calls o noticias financieras.

Palabras clave: Aprendizaje Automático, Automático, Computación Cuántica, Procesamiento de Lenguaje Natural, Trading Algorítmico, Aprendizaje Profundo

CONTEXTO

El presente proyecto, cuyo inicio es Marzo 2020, se desarrolla en el Centro de Altos Estudios en Tecnología Informática (CAETI) dependiente de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI). Se enmarca dentro de la línea de Algoritmos y Software y continúa las investigaciones iniciadas en el Proyecto "Modelo de Sentiment Analysis para la clasificación de noticias en tiempo real". Es financiado y evaluado por la Secretaría de Investigación de la Universidad. Cuenta con la participación de docentes y alumnos de la Maestría en Tecnología Informática y de la Diplomatura en Análisis de Datos para Negocios, Finanzas e Investigación de Mercados.

1. INTRODUCCION

La conjunción de la Computación Cuántica (QC) y la Inteligencia Artificial (IA) ha abierto un nuevo campo de investigación que resulta prometedor tanto en el ámbito académico como en la industria [1], [2], [3], [11], [12].

Si bien gran parte de los trabajos al día de hoy se centran en estudiar potenciales aplicaciones de la QC y se estima que su máxima capacidad se hará visible en los próximos cinco años, ya se han realizado destacados avances en casos reales, en especial en el área de Seguridad y en el tema que nos compete: Finanzas Cuantitativas, donde se han llevado a cabo desarrollos vinculados con "Pricing de Derivados" y Optimización de

Portfolios (Simulaciones de Monte Carlo) [4], [5], [6], [7], [8], [9], [10], [13].

En relación al híbrido propuesto en el presente trabajo entre IA y QC ya se han logrado codificar circuitos cuánticos de algoritmos populares de Machine Learning, como Vecinos Más Cercanos (KNN) y Redes Neuronales Artificiales (ANN), aunque como mencionábamos este campo se encuentra actualmente en pleno desarrollo e investigación académica.

Por otro lado, el vínculo entre Procesamiento del Lenguaje Natural (NLP) y QC es uno de los campos interdisciplinarios que más se está expandiendo, incluso, varios autores ya han denominado a esta disciplina como "Quantum NLP". Una de las aplicaciones más destacadas se presenta en "Detección y Prevención de Fraudes", como así también en el procesamiento de datos alternativos [14], [15].

Finalmente, es importante mencionar, que la empresa IBM ha lanzado lo que se denomina IBM *Quantum Experience*, un servicio en la nube y que brinda acceso libre a los simuladores de computadoras cuánticas, lo cual sumado al framework open source Qiskit, el cual permite programar computadoras cuánticas, brindan un entorno de trabajo eficiente e ideal para llevar proyectos de investigación [9].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El presente trabajo se enmarca en la investigación de aplicación de técnicas y algoritmos de minería de datos y pretende dar un enfoque y soporte académico a toda la comunidad sea que esté relacionada a la industria financiera o dentro del marco académico. En este estudio, en particular sentamos las bases para codificar algoritmos de Machine Learning en computadoras cuánticas para la administración de carteras de inversión y su relación con trading algorítmico.

3. RESULTADOS OBTENIDOS/ESPERADOS

Pregunta Problema:

La computación cuántica podría tener el potencial de acelerar y optimizar los algoritmos de Machine Learning en general y aquellos relacionados con gestión de carteras de inversión y análisis de riesgo, en particular.

Objetivo General:

Estudiar las posibles aplicaciones en el campo de las Finanzas y el Trading Algorítmico de la Computación Cuántica en conjunto con la Inteligencia Artificial utilizando los simuladores de la Plataforma Cloud de IBM-Q.

Objetivos Específicos:

a. Determinar los casos donde la aplicación de Computación Cuántica (QC) en la actualidad sea eficiente y posible en el campo de las Finanzas.

b. Determinar los casos donde la aplicación de la Computación Cuántica (QC) no resulte posible o eficiente en el campo de las Finanzas.

c. Estudiar y desarrollar algoritmos para Finanzas donde se utilizan circuitos cuánticos en conjunto con algoritmos de Machine Learning.

d. Realizar un benchmark entre algoritmos de Machine Learning utilizando circuitos cuánticos y sin utilizarlos.

e. Disponibilizar una interfaz escrita en Python y Qiskit donde diferentes usuarios puedan realizar simulaciones dentro del entorno Cloud de IBM.

f. Avanzar en la relación entre NLP y Computación Cuántica (QC) utilizando los lexicones desarrollados en las etapas anteriores a este proyecto.

Metodología de Trabajo

El enfoque teórico-metodológico del presente trabajo se basa en aplicar herramientas de Procesamiento de lenguaje Natural y Machine Learning desarrolladas en los lenguajes de programación R y Python y descripto en los siguientes puntos:

1. Algoritmos de Machine Learning y Text Mining:

Algoritmos de Aprendizaje Supervisado y Deep Learning: Reinforcement Learning (Q-Learning), Redes Neuronales, LSTM, KNN.

2. Plataformas de Análisis de Datos:

-R (librerías: twitteR, RTextTools: A Supervised Learning Package for Text Classification, OpenNLP: It supports the most common NLP tasks, such as tokenization, sentence segmentation, part-of-speech tagging, TM: A framework for text mining applications within R.)

- BERT
- Tensorflow
- Plataforma Cloud Watson IBM Q Experience
- Python (Librería Qiskit y Jupyter Notebook para desarrollar el código en los simuladores cuánticos de IBM)

Resultados Esperados

En una primera instancia esperamos poder llevar a cabo tareas de comparación (benchmark) entre algoritmos codificados en los lenguajes Python y R en ordenadores tradicionales y los mismos desarrollados en computadoras cuánticas. Debido a que el entorno propuesto por IBM es un entorno de simulación y no posee la potencia de trabajar en entornos productivos, no se esperan obtener resultados realmente significativos a nivel performance de aquellos desarrollados con QC, aunque sí nos permitirá detectar posibles aplicaciones, estrategias y algoritmos donde los mismos puedan brindar resultados superadores a los actuales.

Por último, en un estadio más avanzado, nos proponemos profundizar en aplicaciones específicas relacionadas con nuestro dominio específico, siempre manteniendo el criterio comparativo entre aproximaciones tradicionales y cuánticas en los siguientes dos puntos: (a) creación de carteras de inversión, tanto para mercados argentinos como del exterior; (b) el procesamiento y análisis de datos alternativos, como por ejemplo redes sociales, earnings conference calls o noticias financieras.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo del proyecto, multidisciplinario, se compone principalmente de docentes de la Diplomatura en Análisis de Datos para Negocios, Finanzas e Investigación de Mercado, y la Maestría en Tecnología Informática así como expertos del área de Lingüística y Finanzas. Por su parte, señalamos que el proyecto contará con la participación de alumnos avanzados de la Maestría en Tecnología Informática, quienes llevan a cabo su pasantía de investigación al tiempo que identifican temas en los que puedan desarrollar su tesis. Además se cuenta con la colaboración de alumnos de la mencionada Diplomatura.

5. BIBLIOGRAFÍA

[1] Alcazar, J., Leyton-Ortega, V. and A. Perdomo-Ortiz (2020). Classical versus Quantum Models in Machine Learning: Insights from a Finance Application en *arXiv:1908.10778v2 [quant-ph] 8 Jan 2020.*

[2] Dasgupta, S., Banerjee, A. (2019). Quantum Annealing Algorithm for Expected Shortfall based

Dynamic Asset Allocation en *arXiv:1909.12904v1 [q-fin.RM] 27 Sep 2019.*

[3] Egger, D., García Gutiérrez, R., Cahué Mestre, J. and S. Woerner (2019). Credit Risk Analysis using Quantum Computers en *arXiv:1907.03044v1 [quant-ph] 5 Jul 2019.*

[4] Hao, W., Lefevre, C., Tamturk, M. and S. Utev, (2019). Quantum option pricing and data analysis en *Quantitative Finance and Economics*, QFE, 3(3): 490–507. DOI:10.3934/QFE.2019.3.490

[5] Hodson, M., Ruck, B., Ongy, H., Garvin, D., and S. Dulman. (2019). Portfolio rebalancing experiments using the Quantum Alternating Operator Ansatz en *arXiv: 1911.05296v1 [quant-ph] 13 Nov 2019*.

[6] Kerenidis, I., Prakash, A. and D. Szilagyi. (2019). Quantum Algorithms for Portfolio Optimization en arXiv: 1908.08040v1 [math.OC] 22 Aug 2019.

[7] Mahajan, R.P. (2011). A Quantum Neural Network Approach for Portfolio Selection en *International Journal of Computer Applications*, Volume 29– No.4, September 2011.

[8] Marzec, M. (2016). "Portfolio Optimization: Applications in Quantum Computing", en Ionut Florescu, Maria C. Mariani, H. Eugene Stanley, Frederi G. Viens (eds.) Handbook of High-Frequency Trading and Modeling in Finance, John Wiley & Sons, Inc.

[9] Martin, A., Candelas, B., Rodríguez-Rozas, Martín-Guerrero, J. Chen, X., Lamata, L., Orus, R., Solano, E. and M. Sanz (2019). Towards Pricing Financial Derivatives with an IBM Quantum Computer en *arXiv: 1904.05803v1 [quant-ph] 11 Apr 2019*.

[10] Orrell, D. (2020). A Quantum Walk Model of Financial Options (January 1, 2020). Available at SSRN: https://ssrn.com/abstract=3512481 or http://dx.doi.org/10.2139/ssrn.3512481

[11] Orus, R., Mugel, S. and E. Lizaso (2019). Quantum computing for finance: overview and prospects en Reviews in Physics, Volume 4, November 2019, 100028. https://doi.org/10.1016/j.revip.2019.100028

[12] Rosenberg, G., Haghnegahdar, P., Goddard, P., Carr, P., Wu, K, and M. López de Prado. (2016). Solving the Optimal Trading Trajectory Problem Using a Quantum Annealer en IEEE Journal of Selected Topics in Signal Processing (JSTSP), Volume 10, Issue 6, 2016, and Proc. of the 8th Workshop on High Performance Computational Finance (WHPCF), p. 7, ACM, 2015

[13] Stamatopoulos, N., Egger, D.J., Sun, Y., Zoufal, C., Iten, R., Shen, N. and S. Woerner. (2020). Option Pricing using Quantum Computers en *arXiv: 1905.02666 [quant-ph], 17 Feb 2020*.

[14] Woerner, S., Egger, D.J. Quantum risk analysis. En *NPJ Quantum Inf,* 5, 15 (2019). https://doi.org/10.1038/s41534-019-0130-6.

[15] Zeng, W. and Coecke, B. (2016) Quantum Algorithms for Compositional Natural Language Processing en Kartsaklis, D., Lewis, M. and Rimell, L. (Eds) *2016 Workshop on Semantic Spaces at the Intersection of NLP, Physics and Cognitive Science* (SLPCS'16) EPTCS 221, 2016, pp. 67–75, doi:10.4204/EPTCS.221.8

Utilización de Sistemas Inteligentes para optimizar el diseño de redes de distribución de agua en General Pico - La Pampa

Hugo Alfonso, Carlos Bermúdez, Franco Morero, Gabriela Minetti, Carolina Salto
Laboratorio de Investigación en Sistemas Inteligentes (LISI)
Facultad de Ingeniería - Universidad Nacional de La Pampa
Calle 110 Esq. 9 (6360) General Pico - La Pampa - Rep. Argentina
Te. / Fax: (02302) 422780/422372, Int. 6302
e-mail: {alfonsoh, bermudezc, minettig, saltoc}@ing.unlpam.edu.ar

Resumen Esta línea de investigación aborda la resolución del problema del diseño óptimo de redes de distribución de agua multi-período, usando un algoritmo híbrido basado en la metaheurística *Simulated Annealing*. La propuesta algorítmica inicialmente es testeada satisfactoriamente con benchmarks presentes en el estado del arte. En consecuencia y dado que en la provincia de La Pampa el problema del acceso al agua es de tratamiento prioritario, los resultados obtenidos en esta línea de investigación son transferidos y aplicados a una variante de este problema para optimizar el diseño de redes de distribución de agua en un nuevo barrio de 505 hectáreas de la ciudad de General Pico.

Palabras claves: Sistemas Inteligentes, Optimización, Diseño de Red de Distribución de Agua

Contexto

Esta línea de investigación surge de una de las líneas de investigación del Proyecto "Técnicas inteligentes avanzadas y sistemas distribuidos aplicados a la resolución de problemas de decisión complejos" que se desarrolla en la Facultad de Ingeniería de la UNLPam, y puntualmente se enmarca en la convocatoria para Proyectos Orientados en Investigación Regional (POIRe) financiados por la Universidad Nacional de La Pampa (UNLPam). El objetivo de los POIRes es incentivar la investigación en áreas prioritarias para la resolución de problemas regionales con integración de diversos actores sociales. Nuestra propuesta para la convocatoria 2019 fue seleccionada para su financiamiento y se lleva adelante en el Laboratorio de Investigación de Sistemas Inteligentes (LISI) de la Facultad de Ingeniería de la mencionada universidad. Este POIRe es dirigido por la Dra. Salto. Cabe destacar que desde hace varios años, los integrantes de este laboratorio mantienen una importante vinculación con investigadores de la Universidad Nacional de San Luis (Argentina) y de la Universidad de Málaga (España), con quienes se realizan trabajos conjuntos.

En el LISI, desde su creación en 1998, nos hemos abocado al estudio de algoritmos cada vez más eficientes para la solución de problemas complejos, tanto de optimización como de diseño. En este dominio, el objetivo consiste en obtener algoritmos nuevos que den solución al problema y que necesiten un esfuerzo computacional más pequeño que los algoritmos existentes, así como caracterizar su comportamiento para las clases de problemas que demanda la comunidad científica e industrial en general. Actualmente, es común enfrentarnos a problemas de alta complejidad en el que intervienen varias variables con un conjunto de restricciones definidas sobre ellas, muchas veces contrapuestas, que deben ser consideradas para evaluar la factibilidad de la solución aportada. En este sentido, las líneas de investigación del LISI se encargan de proponer, adaptar y analizar distintas metaheurísticas con el fin de resolver eficaz y eficientemente diferentes problemas.

Una de las líneas es la que investiga la optimización del diseño de redes de distribución de agua, un campo de investigación muy activo desde

hace algunas décadas, con resultados prometedores. Estas redes están compuestas por reservorios y tuberías que tratan de brindarles a los usuarios un flujo constante de agua con una determinada presión. El problema de optimizar el diseño de estas redes de agua consiste en encontrar el diámetro óptimo de cada tubería, seleccionándolo de un conjunto limitado de tamaños disponibles comercialmente, con el objetivo de reducir el costo global brindando la prestación adecuada a la demanda.

Dado que en nuestra provincia, La Pampa, y la región el problema del acceso al agua es de tratamiento prioritario, en este POIRe se propone abordar una variante del mismo, al transferir los conocimientos adquiridos para desarrollar un sistema inteligente que optimice el diseño de redes de distribución de agua para CORPICO, la prestataria de este servicio en la ciudad de General Pico.

Introducción

El problema de optimización del diseño de redes de distribución de agua (Water Distribution Network Design Optimization - WDND) es un problema de gran interés en la comunidad científica. El objetivo del Problema WDND es minimizar el costo total de inversion (Total Investment Cost - TIC) de una red de distribución de agua. La red puede ser modelada con un grafo conexo el cual está conformado por un conjunto de nodos $N = \{n_1, n_2, ...\}$, un conjunto de tuberias $P = \{p_1, p_2, ...\}$, un conjunto de subredes o loops internos $L = \{l_1, l_2, ...\}$, y un conjunto de tipos de tuberías disponibles en el mercado $T = \{t_1, t_2, ...\}$. La función a minimizar TIC es obtenida por la fórmula mostrada en el Ecuación 1.

$$\min TIC = \sum_{p \in P} \sum_{t \in T} L_p \;\; IC_t \;\; x_{p,t} \tag{1}$$

donde IC_t es el costo de un tubo p del tipo t, L_p es la longitud del tubo, y $x_{p,t}$ en una variable binaria que indica si el tubo p es del tipo t o no. La función objetivo está limitada por: leyes físicas de conservación de masa y energía, demanda de presión mínima en cada nodo, y la máxima velocidad en la cañería, en cada momento $\tau \in \mathcal{T}$. Estas leyes son explicadas en los siguientes parágrafos.

Ley de conservación de la masa: Debe ser satisfecha por cada nodo N en cada período de tiempo τ. Esta ley establece que el volumen de agua que fluye hacia un nodo en un momento dado debe ser igual al flujo que traslada (ver Ecuación 2).

$$\sum_{n_1 \in N/n} Q_{(n_1,n),\tau} - \sum_{n_2 \in N/n} Q_{(n,n_2),\tau} = WD_{n,\tau} - WS_{n,\tau} \quad \forall\, n \in N \quad \forall\, \tau \in \mathcal{T} \tag{2}$$

donde $Q_{(n_1,n),\tau}$ es el flujo desde el nodo n_1 al nodo n en el tiempo τ, $WS_{n,\tau}$ es el agua externa provista y $WD_{n,\tau}$ es la demanda de agua externa.

Ley de conservación de energía: Establece que la suma de la presion del flujo en un circuito cerrado en un instante de tiempo τ es cero. Este flujo puede ser aproximado usando la ecuación de Hazen-Williams con los parámetros usados por el software EPANET 2.0 [1] (ver Ecuación 3).

$$\sum_{p \in l} \left[\frac{10.6668 y_{p,\tau} Q_{p,\tau}^{1.852} L_p}{\sum_{t \in T} (x_{p,t} C_t^{1.852} D_t^{4.871}} \right] = 0 \quad \forall l \in L \quad \forall \tau \in \mathcal{T} \tag{3}$$

donde $y_{p,\tau}$ es el signo de $Q_{p,\tau}$ que indica cambios en la dirección del flujo de agua en relación a la dirección del flujo definido, $Q_{p,\tau}$ es la cantidad de agua que fluye a traves de la cañeria p in time τ, L_p es la longitud de la cañeria, C_t es el coeficiente de rugosidad Hazen-Williams según el tipo de caño t, y D_t es el diámetro del tipo de caño t.

Presión mínima requerida en cada nodo: para cada nodo n en cada período de tiempo τ, se debe satisfacer la siguiente condición (ver Ecuación 4).

$$H_{n,\tau}^{min} \leq H_{n,\tau} \quad \forall n \in N \quad \forall \tau \in \mathcal{T} \tag{4}$$

siendo H^{min} la mínima presión en el nodo y $H_{n,\tau}$ la presión actual del nodo.

Máxima velocidad del agua: La velocidad del agua $v_{p,\tau}$ no puede exceder la máxima velocidad estipulada $v_{p,\tau}^{max}$ (ver Ecuación 5).

$$v_{p,\tau} \leq v_{p,\tau}^{max} \quad \forall p \in P \quad \forall \tau \in \mathcal{T} \tag{5}$$

Para resolver este problema de alta complejidad se requiere un método eficaz que sea confiable y fácil de usar, el cual no solo considere los costos

de capital y operativos junto con el rendimiento y la confiabilidad hidráulica, sino también la gestión competente de la energía.

Este problema, aún para sus versiones más simples, es muy difícl de resolver y ha sido clasificado como NP-duro [2]. Muchas versiones del mismo se han focalizado solo en considerar algunas de sus características. Existen variantes que consideran una demanda constante o diferentes demandas en determinadas franjas horarias, las que se identifican como WDND de período simple o multi-período. Las mismas han sido resueltas en varios trabajos aplicando técnicas inteligentes, tales como *Simulated Annealing* [3], [4], [5], *Tabu Search* [6], Algoritmos Genéticos [7], [8], [9], Optimización basada en Colonias de Hormigas [10], [11], *Scatter Search* [12] y Evolución Diferencial [13], entre otras.

Trabajos más recientes abordan variantes que contemplan la variabilidad de los flujos y consideran múltiples objetivos, muchas veces contrapuestos [14], [8]. Una de las más recientes utiliza una metaheurística que optimiza el diseño, al atender la demanda de agua variable durante las 24 horas y contemplar los límites mínimos y máximos de circulación del agua a través de los ductos [15].

En consecuencia, las metaheurísticas brindan, una vez más, una alternativa de solución eficiente. Por este motivo, en esta línea de investigación se analiza y diseña un algoritmo metaheurístico híbrido basado en *Simulated Annealing* para resolver el problema WDND en cuestión.

Dada las características propias de este tipo de redes, una vez que la metaheurística arma una determinada solución, es necesario evaluar su factibilidad y costo. Para ello se utiliza el simulador EPANET 2.0 [16], el cual es de dominio público y desarrollado por la Agencia de Protección Ambiental de Estados Unidos. Este simulador, además, resuelve todas las ecuaciones hidráulicas de forma externa y se puede configurar para resolver distintas variantes del problema.

Línea de Investigación y Desarrollo

Inicialmente, el problema de optimización WDND multi-período es abordado al hibridar una metaheurística basada en trayectoria en [17], [18], [19] y testearla con benchmarks estándares propuestos por la comunidad científica (ver [15]). Esta propuesta, denominada HSA, consiste en un algoritmo de *Simulated Annealing* adaptado e hibridado para resolver el problema planteado.

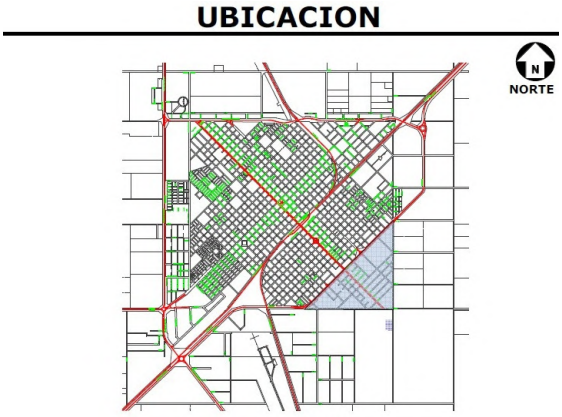


Fig. 1. Mapa tendido redes de agua en General Pico.

En particular, para resolver el caso presentado por CORPICO resulta necesario elaborar las instancias a considerar en función de la topología del área fijada por la entidad adoptante CORPICO. El área para la que se diseñará la red de distribución de agua abarca la zona de General Pico denominada "Barrio Quintas Sur", como puede verse en la región sombreada de la Fig. 1. Inicialmente se requiere realizar el diseño sobre un espacio de 165 hectáreas (triángulo delimitado por Ruta Provincial N°1, Diagonal 201 y Diagonal 500, mostrado en la Fig. 2), previendo para los próximos 10 años una extensión de 340 hectáreas colindantes (espacio comprendido desde Diagonal 500 a Calle H. Gandini y desde Diagonal 201 a Calle 233). Al ser una red independiente, debe contemplar la ubicación de un tanque y cisterna en la intersección de Diagonales 201 y 501, desde donde se realizará la distribución a los distintos usuarios.

Resultados obtenidos y esperados

En cuanto al estudio inicial, HSA ha resuelto el problema WDND multi-período en instancias de variada complejidad que contemplan desde 73 a 285 nodos ([17], [18], [19]). Se ha verificado empírica y estadísticamente la eficiencia de este

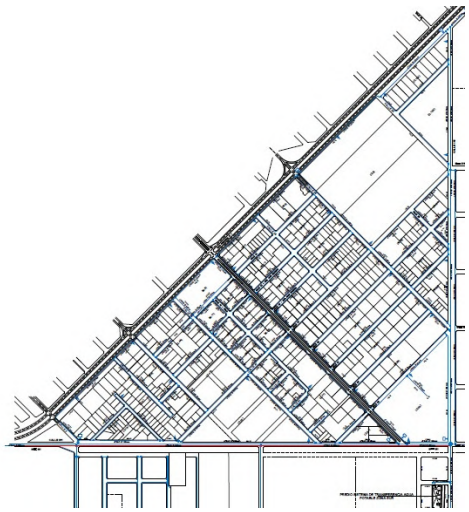


Fig. 2. Mapa Zona 2 de red de agua a diseñar.

algoritmo, así como también su competitividad en comparación a los presentados en la literatura.

En cuanto al trabajo con la institución adoptante, que debe gestionar dicho financiamiento a nivel provincial y nacional, contar con un instrumento que justifique la óptima selección de materiales se estima facilitará dicha gestión. En consecuencia, para lograr un diseño de la red de distribución de agua en una zona de la ciudad de General Pico, que minimice los costos de la inversión a enfrentar en su construcción, que atienda las demandas propias de los actuales habitantes de la zona y que, también, considere su futuro crecimiento, ha sido necesario armar la instancia del problema con los datos reales proporcionados por CORPICO.

Esta nueva instancia consta de 222 nodos que, además de la distribución domiciliaria considera nodos hidrantes, nodos exclusas y válvulas de limpieza. La red total implica unos 35.000 metros de cañerías de diferentes diámetros para atender la demanda de flujo de cada nodo. Los diámetros de cañerías considerados van entre los 50 y 630 mm, cuyos precios oscilan entre 2,85 y 273,28 dólares el metro lineal. El caudal demandado por cada nodo se ha determinado en función de la cantidad de habitantes en las parcelas y el consumo promedio de cada individuo en las diferentes franjas horarias, motivo por el cual se implentará la búsqueda de la solución en la versión multi-períodos. Actualmente se están llevando a cabo los ajustes algorítmicos de HSA, que nos permitirá evaluar la calidad de las soluciones y además evaluar los algoritmos utilizados en instancias reales.

Formación de Recursos Humanos

Cada año se incorporan al proyecto alumnos avanzados en la carrera Ingeniería en Sistemas, quienes trabajan en temas relacionados a la resolución de problemas de optimización usando técnicas inteligentes, con el objeto de guiarlos en el desarrollo de sus tesinas de grado y, también, de formar futuros investigadores científicos. Por otra parte, los docentes-investigadores que integran el proyecto realizan diversos cursos de posgrado relacionados con la temática del proyecto, con el objetivo de sumar los créditos necesarios para cursar carreras de posgrado.

REFERENCES

[1] L. A. Rossman, *The EPANET Programmer's Toolkit for Analysis of Water Distribution Systems*, 1999.

[2] D. F. Yates, A. B. Templeman, and T. B. Boffey, "The computational complexity of the problem of determining least capital cost designs for water supply networks," *Engineering Optimization*, vol. 7, no. 2, pp. 143–155, 1984.

[3] G. Loganathan, J. Greene, and T. Ahn, "Design heuristic for globally minimum cost water-distribution systems," *Journal of Water Resources Planning and Management*, vol. 121, no. 2, pp. 182–192, 1995.

[4] M. Cunha and J. Sousa, "Water distribution network design optimization: Simulated annealing approach," *Journal of Water Resources Planning and Management*, vol. 125, p. 215.221, 07 1999.

[5] M. d. C. Cunha and J. Sousa, "Hydraulic infrastructures design using simulated annealing," *Journal of Infrastructure Systems*, vol. 7, no. 1, pp. 32–39, 2001.

[6] M. da Conceicao Cunha and L. Ribeiro, "Tabu search algorithms for water network optimization," *European Journal of Operational Research*, vol. 157, no. 3, pp. 746–758, 2004.

[7] G. C. Dandy, A. R. Simpson, and L. J. Murphy, "An improved genetic algorithm for pipe network optimization," *Water Resources Research*, vol. 32, no. 2, pp. 449–458, 1996.

[8] I. Gupta, A. Gupta, and P. Khanna, "Genetic algorithm for optimization of water distribution systems," *Environmental Modelling & Software*, vol. 14, no. 5, pp. 437–446, 1999.

[9] W. Bi, G. C. Dandy, and H. R. Maier, "Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge," *Environmental Modelling & Software*, vol. 69, pp. 370–381, 2015.

[10] H. R. Maier, A. R. Simpson, A. C. Zecchin, W. K. Foong, K. Y. Phang, H. Y. Seah, and C. L. Tan, "Ant colony optimization for design of water distribution systems," *Journal of Water Resources Planning and Management*, vol. 129, no. 3, pp. 200–209, 2003.

[11] A. C. Zecchin, A. R. Simpson, H. R. Maier, and J. B. Nixon, "Parametric study for an ant algorithm applied to water distribution system optimization," *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 2, pp. 175–191, 2005.

[12] M.-D. Lin, Y.-H. Liu, G.-F. Liu, and C.-W. Chu, "Scatter search heuristic for least-cost design of water distribution networks," *Engineering Optimization*, vol. 39, no. 7, pp. 857–876, 2007.

[13] A. Vasan and S. P. Simonovic, "Optimization of water distribution network design using differential evolution," *Journal of Water Resources Planning and Management*, vol. 136, no. 2, pp. 279–287, 2010.

[14] R. Farmani, G. A. Walters, and D. A. Savic, "Trade-off between total cost and reliability for anytown water distribution network," *Journal of Water Resources Planning and Management*, vol. 131, no. 3, pp. 161–171, 2005.

[15] A. De Corte and K. Sörensen, "An iterated local search algorithm for water distribution network design optimization," *Network*, vol. 67, no. 3, pp. 187–198, May 2016.

[16] L. Rossman, *EPANET2: Users Manual*, 2000.

[17] C. A. Bermúdez, G. F. Minetti, and C. Salto, "SA to optimize the multi-period water distribution network design," in *XXIX Congreso Argentino de Ciencias de la Computación (CACIC 2018).*, 2018, pp. 12–21.

[18] C. Bermúdez, C. Salto, and G. Minetti, *Communications in Computer and Information Science*, 2019, ch. Solving the Multi-Period Water Distribution Network Design Problem with a Hybrid Simulated Anealling, pp. 3–16.

[19] ——, "Designing a multi-period water distribution network with a hybrid simulated annealing," in *XLVIII JAIIO: XX Simposio Argentino de Inteligencia Artificial (ASAI 2019)*. Universidad Nacional de Salta, 2019, pp. 39–52.

ANÁLISIS DE SENTIMIENTOS EN TWITTER: DESARROLLO DE RECURSOS EN EL ESPAÑOL RIOPLATENSE DE ARGENTINA

Rojo, V.[1,2], Pollo-Cattaneo, Ma. F[1,2]; Britos, P.[3]

[1] Programa de Maestría en Ingeniería en Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina.

[2] Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS). Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina.

[3] Universidad Nacional de Río Negro. Laboratorio de Informática Aplicada. Río Negro, Argentina.

{vmrojo, flo.pollo}@gmail.com, pbritos@unrn.edu.ar

RESUMEN

Twitter se ha posicionado como una de las redes sociales más importantes para el intercambio y concentración de opiniones, convirtiéndose en un ambiente idóneo para el procesamiento automático de estos textos subjetivos a través del análisis de sentimientos. El desafío de la clasificación de tweets se trata de afrontar con ayuda de recursos especializados, tales como corpora, léxicos y herramientas de análisis, los cuales suelen tener un impacto significativo en los resultados de los clasificadores. Este, junto con futuros trabajos, busca colaborar a equiparar las condiciones del análisis de sentimientos en Twitter en español a sus homólogos en otros idiomas por medio del desarrollo de un nuevo recurso léxico y corpus enfocados en el lenguaje informal de Argentina.

Palabras clave: Análisis de Sentimientos en Twitter, Minería de Opiniones, Procesamiento del Lenguaje Natural, Léxicos de Sentimientos

CONTEXTO

En el marco de las actividades conjuntas que realizan el Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS) perteneciente a la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional, y el Grupo de Estudio en Ciencias de Datos (GECS) perteneciente al Laboratorio de Informática Aplicada de la Universidad Nacional de Río Negro, se comienza una nueva línea de trabajo que se articula dentro de los objetivos de ambos grupos en el campo de la Informática, vinculando la Inteligencia Artificial (que está asociada al Procesamiento del Lenguaje Natural y específicamente, al análisis de sentimientos) y la aplicación de sistemas de información (buscando la automatización en el procesamiento de textos subjetivos). En este contexto, se prevé generar nuevo conocimiento en el área de la Ingeniería de Software con la aplicación de tecnologías no convencionales provenientes del Aprendizaje Automático (o Machine Learning en inglés), por lo que sus actividades se desarrollan dentro del ámbito del PID con incentivos UTN UTI5103TC, y UNRN PI 40-C-542.

Por un lado, el Grupo GEMIS busca la sistematización de cuerpos de conocimientos y promoción sobre el campo de la Ingeniería en Sistemas de Información y la Ingeniería en Software, sus aplicaciones y abordajes metodológicos en todo tipo de escenarios (convencionales y no convencionales).

Por otro lado, el Grupo GECS se encuentra conformado por un equipo de docentes y alumnos dentro del ámbito de la Universidad Nacional de Río Negro. Este grupo busca la sistematización de cuerpos de conocimientos y promoción sobre el campo de la Ciencia de

Datos e Inteligencia Artificial, sus aplicaciones y abordajes metodológicos en todo tipo de escenarios (convencionales y no convencionales).

1. INTRODUCCIÓN

1.1. Análisis de sentimientos

Una de las redes sociales más populares para el intercambio de opiniones en torno a una gran variedad de temas es el servicio de *microblogging* Twitter[1]. Actualmente, según cifras oficiales, la plataforma cuenta con cerca de 335 millones de usuarios activos al mes, de los cuales 68 millones de ellos pertenecen a los Estados Unidos y los 267 millones restantes a la comunidad internacional [1]. Recolectar y analizar datos de la plataforma, a diferencia de métodos más tradicionales, es una alternativa que permite encuestar un amplio número de participantes con un menor número de recursos [2]. Es por este motivo que áreas relacionadas a las ciencias políticas, económicas, sociales y de investigación de mercado estudian la red social activamente con un interés especial en las estadísticas agregadas que surgen de los millones de mensajes (o tweets) producidos por sus usuarios todos los días [3]. El conjunto de las opiniones extraídas durante la ejecución de las técnicas sirve para delinear perfiles, conocer sentimientos e ideas de futuros consumidores o votantes, relevar expectativas y realizar predicciones. Esta información resulta de gran valor para empresas, gobiernos y demás organizaciones por sus posibles aplicaciones, las cuales, a menudo, buscan auxiliar en la toma de decisiones estratégicas que se alineen con los objetivos de la organización.

[1] http://twitter.com/

Una de las formas utilizadas para analizar el contenido de los datos de la red social es por medio del análisis de sentimientos, aunque los términos empleados en los trabajos que estudian y exploran el tratamiento computacional de las opiniones, sentimientos y subjetividad en textos incluyen expresiones como minería de opiniones (*opinion mining*), análisis de subjetividad (*subjectivity analysis*) y minería de reseñas (*review mining*), entre otras [4]. En este sentido, los enfoques de las soluciones actuales al problema del análisis de la información subjetiva abarcan una variedad de técnicas pertenecientes a distintas ciencias, campos y subcampos interdisciplinarios como la Lingüística Computacional, el Procesamiento del Lenguaje Natural (PLN), la Inteligencia Artificial (IA) y la ciencia de los datos, o *Data Science*.

El objetivo principal del análisis de sentimientos consiste en la clasificación automática de textos subjetivos en categorías de polaridad previamente determinadas. Normalmente, el contenido puede ser etiquetado como positivo, negativo o neutral [5], aunque estas agrupaciones pueden variar dependiendo de la granularidad que busque el estudio.

Las soluciones encontradas en la bibliografía perteneciente a dicho tema se pueden categorizar, en líneas generales, en tres grupos según las técnicas empleadas para afrontar el desafío del reconocimiento de polaridad en opiniones: enfoque basado en léxico, en aprendizaje automático y el abordaje híbrido. Si bien gran parte de los trabajos pueden ser representados por una de estas categorías, las clasificaciones nombradas no son exhaustivas y en la literatura es posible encontrar soluciones que no se adaptan por completo a ninguno de los tres paradigmas mencionados [6].

1.2. Recursos para el análisis de sentimientos en Twitter

En el análisis de sentimientos en Twitter se tratan de complementar las tareas de clasificación de tweets por medio del uso de distintos recursos.

Muchas de las soluciones hacen uso de léxicos, los cuales se conforman con listas de palabras o frases clasificadas en categorías que comúnmente denotan una polaridad y que han sido calificadas en una escala para describir la intensidad de la misma. Los métodos de elaboración de este tipo de recursos pueden alcanzar distintos grados de automatización, desde lo manual (cuando las etiquetas de polaridad son colocadas por humanos), a lo mecánico (al aplicar motores de traducción automática a léxicos en otros idiomas).

En aquellos trabajos que hacen uso de técnicas de aprendizaje automático, se vuelve indispensable el uso de conjuntos de datos compuestos de tweets recolectados de la plataforma, también llamados corpora, para el entrenamiento y validación de modelos estadísticos. Si la solución lo demanda, estos sets de datos son enriquecidos con anotaciones, normalmente de polaridad, por medio de una variedad de técnicas. Al igual que en el caso de los léxicos, esta pueden involucrar distintos grados de intervención humana.

Por último, existen otras herramientas que ofrecen un amplio espectro de funcionalidades como transformaciones de texto, implementaciones de modelos y algoritmos utilizados para el procesamiento del lenguaje natural y extracción de características comunes para el análisis de sentimientos.

Estos recursos son empleados en las soluciones para el diseño, implementación, entrenamiento y validación de características y clasificadores, por lo que tienen un impacto en los resultados obtenidos.

1.3.Relevancia del problema

Según estimaciones del año 2019, la cantidad de usuarios de Internet que se comunican en español es de alrededor de 344 millones, representando cerca de un 8% de la población total de usuarios, posicionándose en el tercer puesto de los lenguajes más utilizados [7].

La exploración del AST en lenguajes diferentes al inglés es una necesidad que ha sido reconocida por investigadores dentro de la comunidad hispanohablante, así como aquellos que no forman parte de ella [8]. Debido a que la mayoría de los trabajos en el dominio se enfocan en solucionar el problema considerando únicamente un idioma, los recursos creados varían en calidad y cantidad al pasar de un lenguaje a otro. Al contar con el mayor número de herramientas en inglés, la solución de algunos investigadores ha sido, en ocasiones, optar por traducciones automáticas de recursos léxicos para adaptarlos al análisis de tweets en español [9], [10]; sin embargo, como han señalado algunos autores, los resultados pueden variar dependiendo de la calidad de la traducción automática [9]. Es por esta razón que los investigadores con foco en el AST en español han puesto en evidencia en varios estudios [11], [12] la falta de recursos dedicados a este lenguaje.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Luego de la revisión de trabajos presentandos a la edición 2017 de la competencia organizada por la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN), el Taller de Análisis Semántico en la SEPLN (TASS), surge del trabajo una herramienta comparativa enfocada en los recursos utilizados para el análisis de sentimientos en Twitter en español [13]. Por medio de este análisis se respalda la necesidad de direccionar esfuerzo hacia la creación de nuevos recursos en español, y específicamente se pone en evidencia la falta de léxicos en los que se incluya lenguaje informal, común en las redes sociales, para el idioma.

3. RESULTADOS Y OBJETIVOS

La presente línea de trabajo posee los siguientes objetivos.

3.1. Objetivo general

El objetivo general es proponer un recurso léxico compuesto por palabras informales focalizado en el español rioplatense de Argentina para su aplicación en el análisis de sentimientos en español en Twitter.

3.2.Objetivos específicos

Asociados al presente objetivo general se definen los siguientes objetivos específicos:

- Relevar los desafíos y el estado del arte del análisis de sentimientos en Twitter en español y los recursos asociados.
- Identificar y documentar las características que formarán parte del recurso léxico.
- Analizar las características identificadas para determinar su aplicabilidad.
- Desarrollar un recurso léxico que haga uso de las características evaluadas.
- Desarrollar un conjunto de textos subjetivos anotado para el entrenamiento y validación de un clasificador utilizado para la validación.
- Aplicación y validación del léxico creado en un proceso de análisis de sentimientos en español.
- Reportar los resultados y conclusiones.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por 2 investigadores formados, 3 investigadores en formación, 5 alumnos avanzados de carreras de grado, 4 estudiantes avanzados de carreras de postgrado, 1 becario CIN. En su marco se desarrollan 2 Tesis de Maestría, 2 Trabajos Finales de Especialidad y 3 trabajos de Fin de Carrera de grado.De esta manera se espera generar un verdadero espacio integrado de investigación en carreras de grado y posgrado.

5. REFERENCIAS

1. Twitter Investor Relations. Twitter Q2 2018 Earnings Report. [Online].; 2018 [cited 2018 Octubre 03. Available from: https://twitter.com/i/moments/1022804623717875712.
2. Karami A, Bennett LS, He X. Mining Public Opinion about Economic Issues: Twitter and the U.S. Presidential Election. International Journal of Strategic Decision Sciences. 2018 Enero; 9(1): p. 18-28.
3. Rosenthal S, Farra N, Nakov P. SemEval-2017 Task 4: Sentiment Analysis in Twitter. In 11th International Workshop on Semantic Evaluation (SemEval-2017); 2017. p. 502-518.

4. Pang B, Lee L. Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval. 2008; 2(1-2): p. 1-135.
5. Escortell Pérez MA, Giménez Fayos M, Rosso P. El Impacto de las Emociones en el Analisis de la Polaridad en Textos con Lenguaje Figurado en Twitter. Procesamiento del Lenguaje Natural. 2017 Marzo;(58): p. 85-92.
6. Giachanou A, Crestani F. Like It or Not: A Survey of Twitter Sentiment Analysis Methods. ACM Computing Surveys (CSUR). 2016 Noviembre; 49(2).
7. Miniwatts Marketing Group. Top Ten Internet Languages in the World - Internet Statistics. [Online].; 2019 [cited 2020 02. Available from: https://www.internetworldstats.com/stats7.htm.
8. Nakov P. Semantic Sentiment Analysis of Twitter Data. In Encyclopedia on Social Network Analysis and Mining (ESNAM).; 2017.
9. Vilares D, Alonso MA, Gómez-Rodríguez C. Supervised Sentiment Analysis in Multilingual Environments. Information Processing & Management. 2017 Mayo; 53(3): p. 595-607.
10. Wehrmann J, Becker W, Cagnini HEL, Barros RC. A Character-Based Convolutional Neural Network for Language-Agnostic Twitter Sentiment Analysis. In Neural Networks (IJCNN), 2017 International Joint Conference; 2017; Anchorage, AK, USA: IEEE. p. 2384-2391.
11. Jiménez-Zafra SM, Martín-Valdivia MT, Martínez-Cámara E, Ureña-López LA. Studying the Scope of Negation for Spanish Sentiment Analysis on Twitter. IEEE Transactions on Affective Computing. 2017.
12. Sidorov G, Galicia Haro SN, Camacho Vázquez VA. Construcción de un corpus marcado con emociones para el análisis de sentimientos en Twitter en español. Revista Escritos BUAP. 2016; 1(1).
13. Rojo V, Britos P, Pollo-Cattaneo MF. Revisión de enfoques y comparación de recursos para el análisis de sentimientos en español en Twitter. In Desarrollo e Innovación en Ingeniería – Cuarta Edición.: Editorial Instituto Antioqueño de Investigación; 2019. p. 5-16.

Metaheurísticas aplicadas al monitoreo en plantas químicas

Carlos Bermudez[1], Gabriela Minetti[1], Carolina Salto[1]
José Luis Hernandez[2], Mercedes Carnero[2],
Laboratorio de Investigación en Sistemas Inteligentes (LISI)
[1]Facultad de Ingeniería - Universidad Nacional de La Pampa
e-mail: [1]{bermudezc, minettig, saltoc,@ing.unlpam.edu.ar}
[2]$FacultaddeIngeniera - UniversidadNacionalRoCuarto$
e-mail: [2]{jlh, mcarnero@ing.unrc.edu.ar}

Resumen Esta línea de investigación se enfoca en la resolución del problema de diseño de la red de sensores en las plantas de procesos, incluyendo la determinación de las variables de proceso que deben medirse para lograr el grado de conocimiento requerido sobre la planta. Este problema lo abordamos mediante el uso de la metaheurística *Simulated Annealing* hibridizada con estrategias especialmente diseñadas para el problema tales como Oscilación Estratégica. Además, se están incorporando métodos de simulación que permiten evaluar el diseño bajo un amplio abanico de situaciones.

Palabras claves: Metaheurísticas, Optimización, Diseño de red de sensores, Simulated Annealing.

Contexto

Esta línea de investigación se desarrolla en el marco de dos proyectos de investigación, acreditados en distintas universidades argentinas. Uno llevado a cabo en el Grupo de Optimización (GOp), de la Facultad de Ingeniería de la Universidad Nacional de Río Cuarto. El segundo se desarrolla en el Laboratorio de Investigación de Sistemas Inteligentes (LISI), de la Facultad de Ingeniería de la Universidad Nacional de La Pampa, acreditado por dicha facultad y dirigido por la Dra. Minetti.

El monitoreo en una planta química es absolutamente crucial y tiene considerables impactos en aspectos económicos, de seguridad o de control. Este monitoreo también debe garantizar el conocimiento sobre el estado del proceso en cualquier momento. La información es recopilada por sensores distribuidos en toda la planta, responsables de medir y transmitir los valores de magnitudes como temperatura, humedad, presión, etc. El conjunto de dispositivos utilizados en la medición se denomina red de sensores (*Sensor Network*, SN). El diseño de una SN se realiza sistemáticamente formulando un problema de optimización llamado problema de diseño de la red de sensores (*Sensor Network Design Problem*, SNDP), que es un problema de optimización discreta. En escenarios de trabajo reales, el número de variables involucradas es bastante grande y la formulación de SNDP puede ser más o menos compleja dependiendo de los criterios de desempeño y las restricciones que se le imponen.

SNDP fue formulado por Bagajewicz [1] y resuelto por Nguyen [2], usando un nuevo método de búsqueda de árbol. Este método explota ciertas propiedades de costo de los diferentes nodos en el árbol para podar eficientemente nodos no óptimos. Otros enfoques modelaron el problema combinando la programación entera y no lineal, resolviéndolo por medio de búsquedas de árbol que utilizan estrategias de inspección tanto en profundidad como en amplitud. La principal desventaja de todos estos métodos es la excesiva demanda de tiempo de cómputo, así como la poca flexilibilidad ante cambios en la formulación del problema. En consecuencia, para lograr una optimización eficiente es importante contar con una herramienta

que permita resolver el problema para diferentes formulaciones de diferente complejidad y tamaño, en períodos de tiempo razonables.

Los diferentes métodos metaheurísticos mitigan la desventaja mencionada anteriormente porque resuelven eficientemente casos de problemas mono o multiobjetivos. En este sentido, la literatura presenta algoritmos genéticos [3], inteligencia basada en cúmulo de partículas o enjambres [4], entre otras metaheurísticas basadas en la población. Además, en la literatura se encuentran metaheurísticas hibridadas que resuelven este problema. En una de las más recientes, Carnero et al. [5] proponen la técnica PBIL_SOTS, que combina el algoritmo de estimación de distribuciones con el algoritmo *Tabu Search* (TS) mejorado mediante el uso de una oscilación estratégica (SOTS).

En contraste con las metaheurísticas basadas en población ya mencionadas, el algoritmo *Simulated Annealing* (SA) [6] puede clasificarse en el grupo basado en trayectoria y ha demostrado ser un método eficiente para resolver muchos problemas de optimización combinatoria dura [7]. En esta línea de investigación, presentamos un algoritmo híbrido basado en *Simulated Annealing* (HSA) como metaheurística principal y Tabu Search integrada con una oscilación estratégica (SOTS), como metaheurística subordinada, para resolver el SNDP. Su adaptación a la resolución de un problema particular implica, entre otros aspectos, hacer un ajuste de sus parámetros que puedan obtener el mejor rendimiento de la técnica propuesta. Por lo tanto, el trabajo también se enfoca en el ajuste de los parámetros de control algorítmico para alcanzar un equilibrio entre la calidad de la solución y el consumo de tiempo. El algoritmo de optimización requiere utilizar en todos los casos un método de reconciliación de datos y verificación de cumplimiento de las restricciones impuestas. Dependiendo de la naturaleza de éstas últimas, el método de verificación puede ser analítico o bien llevado a cabo a través de técnicas de simulación. En particular, si se imponen restricciones sobre la confiabilidad asociada a un conjunto de variables claves, es posible evaluarlos mediante la implementación de métodos de Montercarlo para su evaluación.

Desarrollo

En esta sección se describe la línea de investigación ya mencionada, introduciendo una explicación más detallada del problema y del algoritmo propuesto para resolverlo.

Problema de diseño de la red de sensores (SNDP)

El SNDP se puede definir como el problema de encontrar una estructura de sensores que satisfaga un conjunto de restricciones a la vez que se optimiza un criterio de desempeño determinado. Una formulación particular en la que se ha trabajado es aquella donde se pretende diseñar una red de instrumentación de costo mínimo que satisfaga restricciones de precisión y estimabilidad sobre un conjunto de variables claves. Formalmente esto pude ser expresado tal como se muestra en la Ecuación (1), donde $\mathbf{q}$ es un vector n-dimensional de variables binarias tal que $q_i = 1$ si se mide la variable i y $q_i = 0$ en caso contrario, $\mathbf{c}^T$ es el vector de costo; $\hat{\sigma}_k$ es el desvío estandar de la variable k contenida in S_σ después de aplicar un procedimiento de conciliación de datos [8], y E_l representa el grado de estimabilidad de la variable l-ésima incluida en S_E. Además, S_σ y S_E son el conjunto de variables clave del proceso con requisitos de precisión y capacidad para estimarse, respectivamente.

$$\min \mathbf{c}^T\mathbf{q} \tag{1}$$

sujeto a:

$$\begin{aligned} \hat{\sigma}_k(\mathbf{q}) \leq \sigma_k^*(\mathbf{q}) \qquad & \forall\, k \in S_\sigma \\ E_l \geq 1 \qquad & \forall\, l \in S_E \\ \mathbf{q} \in \{0,1\}^n & \end{aligned}$$

En esta formulación, las mediciones están sujetas a errores aleatorios no correlacionados; solo hay un dispositivo de medición potencial para cada variable, y no hay restricciones para la localización de instrumentos. La viabilidad de las restricciones se puede verificar ejecutando una clasificación variable y un procedimiento de reconciliación de datos.

SA híbrido para resolver SNDP

SA funciona como metaheurísta principal con una búsqueda local ad hoc sobordinada, inspirada en la búsqueda tabú con la técnica de oscilación estratégica, SOTS, dando lugar al HSA (*Hybrid Simulated Annealing* [9]. Esta hibridación se aplica en dos niveles: en el primero, para generar una solución inicial, y en el segundo, para mejorar la solución durante el proceso de enfriado.

El esquema de perturbación de la solución actual se lleva a cabo para generar una solución candidata a partir de la actual. Ésta consiste en realizar una cierta cantidad de intercambios, que van desde variables medidas a otras no medidas y viceversa. Esta mutación se aplica sobre cada variable con una cierta probabilidad (llamada Pswap) [10]. Además, la temperatura se actualiza al aplicar el esquema de planificación geométrico.

La primera cuestión de diseño en HSA se relaciona con los componentes de búsqueda que se fijan en función del SNDP, es decir, definir la representación de una solución. En este caso, una solución al problema SNDP es un vector n-dimensional $\mathbf{q}= \{q_1, q_2, \ldots, q_n\}$ de variables binarias, como se especifica en la definición del problema. Para evaluar una solución se usa la ecuación (2) como función objetivo, donde $\sum_{i=1}^{n} c_i$ es el costo de la SN cuando se miden todas las variables (límite superior de la función objetivo del SNDP). Cuando no se cumplen las restricciones, la ecuación (3) calcula $S(\mathbf{q})$, donde rno y rnp representan el número de restricciones de observabilidad y precisión que no se cumplen para un vector solución.

$$H = \begin{cases} \sum_{i=1}^{n} c_i q_i & \text{si } \mathbf{q} \text{ es factible} \\ \sum_{i=1}^{n} c_i (1 + S(\mathbf{q})) & \text{si } \mathbf{q} \text{ no es factible} \end{cases} \tag{2}$$

$$S(\mathbf{q}) = \frac{rno}{S_E} + \frac{1}{rnp} \sum_{i=1}^{r} np \frac{\hat{\sigma}_i - \sigma_i^*}{\hat{\sigma}_i} \tag{3}$$

El segundo punto de diseño involucra los principales componentes del HSA que son variables durante el proceso de búsqueda: temperatura inicial, esquema de enfriamiento y longitud de la cadena de Markov. A fin de estudiar el impacto de diferentes temperaturas iniciales en el rendimiento de HSA, utilizamos valores iniciales, T_s, pertenecientes a $\{1, 900\}$. A la hora de analizar el efecto producido por el esquema de enfriamiento, proponemos cuatro versiones de HSA: $\boldsymbol{HSA}_{Prop}$ adopta una planificación proporcional [6], $\boldsymbol{HSA}_{Exp}$ usa el esquema exponencial [6], $\boldsymbol{HSA}_{Log}$ emplea la opción logarítmica [11] y, por último, $\boldsymbol{HSA}_{Rand}$ que aplica un esquema aleatorio [12]. Finalmente, también consideramos tres formas diferentes de calcular la longitud de la cadena de Markov (MCL). Al principio, MCL se calcula estáticamente y se mantiene constante durante 30 iteraciones (número comúnmente utilizado en la literatura), denominado MCL_s. Las otras dos formas implementan las técnicas adaptativas propuestas por Cardoso et al. [13] y Ali et al. [14], denominadas MCL_{a1} y MCL_{a2}, respectivamente. Al combinar todos los enfoques antes mencionados, se obtienen un total de 24 variantes de HSA para resolver el SNDP [15].

Resultados Obtenidos

Con el propósito de evaluar el rendimiento de HSA para resolver SNDP, se han considerado un conjunto de 5 casos de estudio que comprenden procesos de diferente complejidad y tamaño, cuya operación puede ser representada tanto por modelos lineales como no lineales [16], [17]. La complejidad del conjunto de restricciones impuestas a todos los casos de estudio se encuentra en [5].

La efectividad de las variantes de HSA se evalúa, empíricamente y con soporte estadístico, en términos de calidad de la solución, tiempo de ejecución y cantidad de evaluaciones [15]. A partir de esta evaluación se concluye que las variantes de HSA, que calculan estáticamente la longitud de la cadena de Markov (MCL), siempre obtienen la mejor solución con el mínimo esfuerzo para cuatro casos. En cambio, para el caso restante, el más complejo, se logra una compensación entre calidad y tiempo cuando HSA establece la temperatura inicial en un valor alto ($T_s = 900$), usa un MCL adaptativo (MCL_{a1}) y aplica el esquema de enfriamiento proporcional. Todo lo anterior permite

inferir que HSA es un algoritmo competitivo para resolver casos realistas de SNDP.

Formación de Recursos Humanos

Cada año se incorporan a los proyectos alumnos avanzados en la carrera Ingeniería en Sistemas, quienes trabajan en temas relacionados a la resolución de problemas de optimización usando técnicas inteligentes, con el objeto de guiarlos en el desarrollo de sus tesinas de grado y, también, de formar futuros investigadores científicos. Por otra parte, los docentes-investigadores que integran los proyectos realizan diversos cursos de posgrado relacionados con la temática del proyecto, con el objetivo de sumar los créditos necesarios para cursar carreras de posgrado.

REFERENCES

[1] M. Bagajewicz, "Design and retrofit of sensor networks in process plants," *AIChE Journal*, vol. 43, no. 9, pp. 2300–2306, 1997.

[2] D. Nguyen and M. Bagajewicz, "New efficient breadth-first/level traversal tree search method for the design and upgrade of sensor networks," *AIChE Journal*, vol. 57, no. 5, pp. 1302–1309, 2011. [Online]. Available: https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.12332

[3] C. Gerkens and G. Heyen, "Use of parallel computers in rational design of redundant sensor networks," *Computers and Chemical Engineering*, vol. 29, no. 6 SPEC. ISS., pp. 1379–1387, 2005.

[4] Y.-J. He and Z.-F. Ma, "Optimal design of linear sensor networks for process plants: A multi-objective ant colony optimization approach," *Chemometrics and Intelligent Laboratory Systems*, vol. 135, pp. 37–47, 2014.

[5] M. Carnero, J. L. Hernández, and M. Sánchez, "Optimal sensor location in chemical plants using the estimation of distribution algorithms," *Industrial & Engineering Chemistry Research*, vol. 57, no. 36, pp. 12 149–12 164, 2018.

[6] S. Kirkpatrick, C. G. Jr, and M. Vecchi, "Optimization by simulated annealing," *Science*, no. 220, pp. 671–680, 1983.

[7] E. Talbi, *Metaheuristics: From Design to Implementation*. Wiley Publishing, 2009.

[8] M. Bagajewicz and M. Sánchez, "Reallocation and upgrade of instrumentation in process plants," *Computers Chemical Engineering*, vol. 24, no. 8, pp. 1945 – 1959, 2000.

[9] J. Hernandez, C. Salto, G. Minetti, M. Carnero, and M. Sanchez, "Hybrid simulated annealing for optimal cost instrumentation in chemical plants," *Chemical Engineering Transactions*, vol. 74, pp. 709–714, May 2019.

[10] K.-L. Du and M. N. S. Swamy, *Search and Optimization by Metaheuristics: Techniques and Algorithms Inspired by Nature*, 1st ed. Birkhäuser Basel, 2016.

[11] B. Hajek, "Cooling schedules for optimal annealing," *Mathematics of Operations Research*, vol. 13, no. 2, pp. 311–329, 1988.

[12] C. Bermudez, C. Salto, and G. Minetti, "Solving the multi-period water distribution network design problem with a hybrid simulated anealling," in *Computer Science – CACIC 2018*, P. Pesado and C. Aciti, Eds. Cham: Springer International Publishing, 2019, pp. 3–16.

[13] M. F. Cardoso, R. L. Salcedo, and S. F. de Azevedo, "Nonequilibrium simulated annealing: A faster approach to combinatorial minimization," *Ind. Eng. Chem. Res*, vol. 33, pp. 1908–1918, 1994.

[14] M. Ali, A. Törn, and S. Viitanen, "A direct search variant of the simulated annealing algorithm for optimization involving continuous variables," *Computers & Operations Research*, vol. 29, no. 1, pp. 87 – 102, 2002.

[15] J. Hernandez, C. Salto, G. Minetti, M. Carnero, C. Bermudez, and M. Sanchez, "Optimal instrumentation: Adjustment and hybridization of a simulated annealing based technique," in *XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019)*, Oct. 2019, pp. –.

[16] M. Bhushan and R. Rengaswamy, "Design of sensor network based on the signed directed graph of the process for efficient fault diagnosis," *Industrial & Engineering Chemistry Research*, vol. 39, no. 4, pp. 999–1019, 2000. [Online]. Available: https://doi.org/10.1021/ie990383d

[17] H. Smith and N. Ichiyen, "Computer adjustment of metallurgical balances," *Computer Applications and Process Control. (CIM) Bulletin*, vol. 66, pp. 97–100, 1973.

PROCESAMIENTO DIGITAL DE IMÁGENES CITOGENÉTICAS PARA SU CLASIFICACIÓN SEGÚN LA TÉCNICA ENSAYO COMETA PARA LA DETECCIÓN DE DAÑOS EN ELADN.

Vera Laceiras María Silvia [1] , Pisarello María Inés [2] , Caffetti Jacqueline Diana [1,3,4] .

[1]Facultad de Exactas Químicas y Naturales. Universidad Nacional de Misiones. Posadas, Misiones.

[2] Facultad de Ciencias Exactas y Naturales y Agrimensura. Universidad Nacional del Nordeste. Corrientes. [3]Facultad de Ciencias Exactas, Físicas y Naturales. Universidad Nacional de Córdoba. Córdoba.

[4]Becaria Postdoctoral CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Instituto de Biología Subtropical (UNaM-IBS-CONICET).

vlhsilvia@gmail.com mainespisarello@exa.unne.edu.ar jacqui_caffetti@fceqyn.unam.edu.a

RESUMEN:

En este trabajo se estudian y analizan diferentes técnicas de clasificación, segmentación y extracción de contornos de imágenes digitales de origen biológico. En particular se centra la atención en los procesos que utilizan funciones con algoritmos neuro-fuzzy y de redes neuronales. Se presentan aquí algunos resultados logrados hasta el momento a través del desarrollo e implementación de un prototipo aplicado a imágenes obtenidas mediante microscopios, específicamente a aquellas parametrizadas según la técnica de ensayo cometa empleada por el Laboratorio de Citogenética General y Monitoreo Ambiental de la UNaM-IBS-CONICET para la detección de daños en el ADN con validación visual y numérica que avalan lo logrado.

Palabras claves: imágenes microscópicas, patrones de imágenes, algoritmos de segmentación, lógica difusa, extracción de contornos, redes neuronales, clasificación de imágenes.

CONTEXTO

La línea de investigación en que se enmarca el proyecto corresponde a la propuesta de tesis para acceder al título de Magíster en Tecnologías de la Información dictado conjuntamente por la Universidad Nacional de Misiones (UNaM) y la Universidad Nacional del Nordeste (UNNE). Las muestras pertenecen al Laboratorio de Citogenética General y Monitoreo Ambiental que se emplaza en la FCEQyN UNaM. El área de Monitoreo Ambiental analiza el daño genético en organismos acuáticos y su impacto potencial sobre la salud humana, como consecuencia de la exposición a contaminantes urbanos e industriales presentes en los ríos y arroyos de la Provincia de Misiones. Se desarrolla el procesamiento digital de imágenes citogenéticas en Ensayo Cometa usando para este estudio, imágenes de moléculas de ADN donde el nucleoide de ADN se presenta como una cabeza y/o cola de cometa. Se busca distinguir cabeza, cola y fondo de la molécula de ADN y medirlas para su clasificación, a fin de facilitar el trabajo del Laboratorio. La investigación se inicia con la posibilidad de aplicar técnicas de clasificación, segmentación, algoritmos neuro fuzzy, matemáticos y computacionales a las imágenes digitales producidas por microscopio y registradas a través de una cámara de fotografía.

INTRODUCCIÓN

Descripción de actividades del Laboratorio de Citogenética General y Monitoreo Ambiental, Instituto de Biología Subtropical; Facultad de Ciencias Exactas, Químicas y Naturales; Universidad Nacional de Misiones; CONICET (UNAM-IBS-CONICET): El Laboratorio de Citogenética General y Monitoreo Ambiental inició sus actividades en el año 1989, dedicándose a estudios citogenético-evolutivos en diferentes modelos animales con principal énfasis en peces neotropicales de agua dulce. Luego se incorporó la línea de citogenética humana, prestando servicios de diagnóstico en convenio entre la UNaM y el Instituto de Previsión Social de la Provincia de Misiones (IPS). Y a partir del año 1993 se incluyó la línea de mutagénesis y monitoreo ambiental, que tiene como objetivos estudiar el impacto de contaminantes en ambientes naturales y en bioensayos de laboratorio a través de técnicas citogenético-moleculares. Actualmente, el laboratorio está constituido por su director, el Dr. Alberto Fenocchio; docentes-investigadores: Dra. Jacqueline Caffetti y Lic. Héctor Roncati; El laboratorio en su línea de investigación referida a agentes contaminantes acuáticos, aplica técnicas específicas de genética toxicológica, como el Ensayo Cometa, a diferentes organismos expuestos a muestras de agua tomadas de los Ríos Paraná y Uruguay. En el área de

Monitoreo Ambiental del laboratorio se analiza el daño genético en organismos acuáticos y su impacto sobre la salud humana, como consecuencia de la exposición a contaminantes urbanos e industriales presentes en los ríos y arroyos de la Provincia de Misiones.

Para ello se aplican técnicas específicas de genética toxicológica, como el Ensayo Cometa, uno de los "biomarcadores genéticos" [1] que son útiles como señales de alerta temprana en cursos de agua contaminados y, por lo tanto, su análisis resulta de interés predictivo en evaluaciones de estado de calidad de cursos hídricos y manejo de cuencas.

El Ensayo Cometa es una técnica donde se cuentan 100 células por cada individuo (de un total de 10-15 individuos por tratamiento). Estas 100 células se clasifican en distintos subtipos de acuerdo al daño que presenten en su ADN. Las 5 clases, que se muestran en la figura 1, dependen de la intensidad de fluorescencia y el largo de las "colas de los cometas" que son equivalentes a la cantidad de ADN fragmentado o con daño: Clase 0 (sin daño, es decir, no tiene cola); Clase 1 (tamaño de la cola hasta una vez el diámetro de la cabeza); clase 2 (tamaño de la cola hasta dos veces el diámetro de la cabeza); clase 3 (tamaño de la cola hasta tres veces el diámetro de la cabeza) y clase 4 (casi todo el ADN aparece fragmentado en la cola).

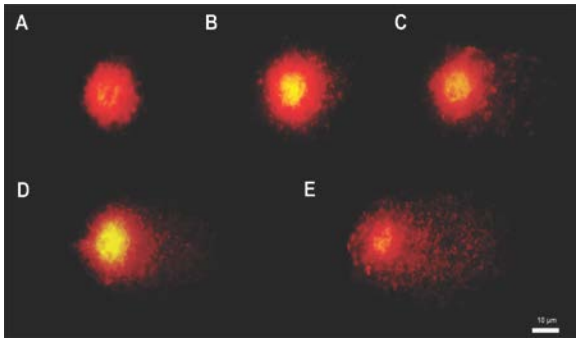


Figura 1. Clasificación de células en el ensayo cometa. A: Clase 0; B: clase 1; C: Clase 2; D: Clase 3; E: Clase 4.

El registro y clasificación de estos tipos celulares se hace de manera artesanal a través de procedimientos visuales y luego se realiza un registro manual. Una vez que se clasifican todas las células en las 5 clases, debe calcularse un índice de daño promedio (ID) o escore de daño por cada individuo y, a su vez, por cada tratamiento (promediando los datos de los 10-15 individuos).

El registro manual que lleva adelante el Laboratorio, incluye anotaciones en cuadernos del número de células en cada clase y luego pasar los datos a una planilla de Excel para el cálculo del score o ID. Recién a partir de esta instancia los datos son cargados al programa Infostat® para su análisis estadístico.

De la bibliografía consultada se pueden citar un gran número de técnicas y desarrollos que se encuentran en constante perfeccionamiento, entre ellos la clasificación de una imagen tratándola como un objeto [2]; la segmentación por color [3]; la extracción de contornos en imágenes digitales es una operación que facilita también los procesos de segmentación e identificación de patrones [4], ya sea para tareas de reconocimiento e interpretación, como también de clasificación de objetos [5]. Las imágenes se capturan con cámara fotográfica digital y microscopio en el Laboratorio de Citogenética General y Monitoreo Ambiental. El objetivo de este trabajo es el procesamiento digital de dichas imágenes, para extraer información útil que pueda parametrizarse según la técnica de Ensayo Cometa, a fin de clasificar las moléculas ADN y realizar un dataset válido a fin de tener registros históricos de los datos recabados y lograr nueva información y nuevos conocimientos.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Actualmente, las líneas de investigación están basadas en la utilización de algoritmos matemáticos y computacionales, redes neuronales [6], [7] y algoritmos neurofuzzy [5], [8] que nos permiten manejar las características del material biológico, logrando:

- Almacenar cada imagen en un archivo para su posterior procesamiento digital.
- Identificación y detección de imágenes a través de algoritmos de aprendizaje supervisado [8], [9].
- Determinar las diferentes zonas: Fondo, cabeza y cola a través de filtros digitales, de segmentación de color que se logra al

transformar las imágenes por medio de funciones matemáticas y algoritmos computacionales combinados con algoritmos neuro fuzzy y redes neuronales

- Para la mejora del análisis de los resultados obtenidos mediante ensayo cometa (captura, clasificación, conteo y gestión de los resultados). Se usarán técnicas digitales que combinan algoritmos matemáticos y computacionales que incluyen fuzzy logic [10] y redes neuronales [11],[12].

RESULTADOS OBTENIDOS

Actualmente se estudiaron y analizaron más de 200 imágenes obtenidas a partir de células de diferentes especies luego de la aplicación del Ensayo cometa, logrando realizar primeramente un dataset que nos permite tener un histórico para nuevos estudios y una clasificación a través de un prototipo creado y validado con las imágenes de prueba.

Con la aplicación de técnicas de clasificación de imágenes que tienen un muy bajo margen de error en cuanto a la detección de clases (a través de la detección y distinción de fondo, cabeza y cola, su medición y posterior clasificación) se logra menor intervención humana a la hora de procesarlos. En este caso el desarrollo del prototipo funcional se realiza en Matlab y se puede ejecutar en Octave. La posibilidad de ejecución de los algoritmos en Octave, software libre de distribución gratuita, es lo que permite avanzar con el prototipo y aportar una mejor performance en la clasificación y los recursos más económicamente convenientes a la hora de evaluar la implementación en el laboratorio de Citogenética (UNaM-IBS- CONICET).

El desempeño final del sistema se evalúa mediante inspección visual y métodos de validación cruzada de datos. Utilizando *K* grupos (K-Fold) de datos en los que estén representados diferentes tipos de clases y de diferentes especies en cada grupo. En (K-Fold) donde *K* representa las interacciones, el error se calcula de la siguiente manera:

$$Ei = \frac{numero\ de\ clasificaciones\ incorrectas}{total\ de\ datosdel\ grupo\ test}$$

El error final de la validación es:

$$Error = \frac{1}{K}\sum_{i=1}^{k} \quad E_i$$

Se obtiene el siguiente error promedio: **Error 7.4%**

Conformado de la siguiente manera en cuanto a la precisión:

Clase	**Error de Precisión**
Clase 0	0,00%
Clase 1	8.18%
Clase 2	14.3%
Clase 3	14,00%
Clase 4	6.2%

Se espera finalizar la investigación con el desarrollo de una aplicación para celulares que nos permita capturar, gestionar, clasificar y guardar las imágenes para lograr una manipulación de imágenes que se adapte a las especificaciones del Laboratorio de Citogenética General y Monitoreo Ambiental (UNaM-IBS-CONICET) en la detección de daños en el material genético, planteando como línea de investigación futura la posibilidad de trabajar en otros proyectos que actualmente se llevan a cabo en dicho Laboratorio.

FORMACIÓN DE RECURSOS HUMANOS

Este proyecto de investigación forma parte del desarrollo de la tesis de posgrado, correspondiente a la carrera de Magíster en Tecnologías de la Información dictada por la Universidad Nacional de Misiones y la Universidad Nacional del Nordeste. La tesis tiene como estudio de caso una técnica de análisis utilizada en el Laboratorio, el Ensayo Cometa, que se encuentra bajo la dirección de la Dra. Maria Inés Pisarello y se articula con la línea de investigación de la Docente Investigadora en el Instituto de Biología Subtropical (UNaM-IBS-CONICET) que lleva adelante la codirección de la misma. Esta tesis de maestría no solo aporta soluciones reales al laboratorio de citogenética en su línea de investigación y a las cátedras donde se incluye el aprendizaje de la técnica de ensayo cometa; sino que además en el campo profesional aportan formación de recursos humanos especializados en técnicas de inteligencia artificial.

Bibliografia:

[1] G. N. A. Furnus, J. D. Caffetti, E. M. García, M. F. Benítez, M. C. Pastori, y A. S. Fenocchio, «Frequências basais de micronúcleos e anormalidades nucleares

[2] A. Kaplan y M. Haenlein, «Rulers of the world, unite! The challenges and opportunities of artificial intelligence», *Bus. Horiz.*, 2019.

[3] P. Kickingereder *et al.*, «Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study», *Lancet Oncol.*, vol. 20, n.º 5, pp. 728-740, 2019.

[4] D. Sebasti Comas y G. J. Meschino, «Segmentación de Imágenes mediante Reconocimiento de Patrones», *Esc. y Work. Argenino en Ciencias las Imagenes*, 2014.

[5] S. K. Oh, W. Pedrycz, y B. J. Park, «Self-organizing neurofuzzy networks based on evolutionary fuzzy granulation», *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans.*, vol. 33, n.º 2, pp. 271-277, mar. 2003.

[6] A. P. Boyle, J. Guinney, G. E. Crawford, T. S. Furey, y A. Valencia, «F-Seq: a feature density estimator for high-throughput sequence tags», *Bioinforma. Appl. NOTE*, vol. 24, n.º 21, pp. 2537-2538, 2008.

[7] «Redes neuronales profundas - Tipos y Características - Código Fuente». [En línea]. Disponible en: https://www.codigofuente.org/redes-neuronales-profundas-tipos-caracteristicas/. [Accedido: 20-dic-2019].

[8] D. Ciregan, U. Meier, y J. Schmidhuber, «Multi-column deep neural networks for image classification», en *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642-3649.

[9] «Learning machine, los usos del aprendizaje supervisado». [En línea]. Disponible en: https://blog.es.logicalis.com/analytics/learning-machine-los-usos-del-aprendizaje-supervisado. [Accedido: 20-dic-2019].

[10] M. Landín, R. C. Rowe, y P. York, «Advantages of neurofuzzy logic against conventional experimental design and statistical analysis in studying and developing direct compression formulations», *Eur. J. Pharm. Sci.*, vol. 38, n.º 4, pp. 325-331, nov. 2009.

[11] C. Quintero, F. Merchán, A. Cornejo, y J. S. Galán, «Uso de Redes Neuronales Convolucionales para el Reconocimiento Automático de Imágenes de Macroinvertebrados para el Biomonitoreo Participativo», *KnE Eng.*, vol. 3, n.º 1, p. 585, feb. 2018.

[12] M. Polycarpou, A. C. P. L. F. de Carvalho, J.-S. Pan, M. Woźniak, H. Quintian, y E. Corchado, Eds., *Hybrid Artificial Intelligence Systems*, vol. 8480. Cham: Springer International Publishing, 2014.

Agentes Inteligentes para Recuperación de Información y Analítica Visual en Big Data

José Federico Medrano[1]

jfmedrano@fi.unju.edu.ar

[1]VRAIn / Visualización y Recuperación Avanzada de Información / Facultad de Ingeniería

Universidad Nacional de Jujuy (UNJu) - Ítalo Palanca 10, +54 (388) 4221587

RESUMEN

La creciente cantidad de datos generados en la web es cada vez más evidente y preocupante. El procesamiento y análisis de dichos datos en forma manual resulta una tarea realmente ardua, más aún, si la tendencia es que la mayoría de esos datos sean semi-estructurados o no-estructurados. Por ello las herramientas para trabajar con grandes datos ofrecen una alternativa viable para tal propósito, pero un par de cuestiones a resolver y que poco se habla al hablar de Big Data son: primero el origen de los datos, cómo, cuándo y dónde realizar la búsqueda y extracción de estos, y segundo que hacer con los datos una vez procesados, es decir, de qué modo presentárselos al usuario final. Este proyecto se enfoca por un lado en el diseño e implementación de agentes inteligentes capaces de recuperar información de medios digitales, para luego ser procesados mediante técnicas de Procesamiento del Lenguaje Natural y Aprendizaje Automático que permitan entregar información relevante y relacionada a diferentes ámbitos del conocimiento. Y por otro lado, el segundo enfoque es el tratamiento de la información masiva tanto desde el punto de vista de la representación visual como del de la interacción, a través de interfaces visuales que permitirán sintetizar información y derivar *insight* de los mismos.

Palabras clave: *Recuperación de Información; Agentes Inteligentes; Visualización de Información; Analítica Visual; Big Data*

CONTEXTO

La línea de investigación aquí presentada se encuentra enmarcada dentro del Proyecto Bianual 2020-2021 D/B035 denominado "*Agentes Inteligentes para Recuperación de Información y Analítica Visual en Big Data*", aprobado y financiado por la Secretaría de Ciencia y Técnica y Estudios Regionales de la Universidad Nacional de Jujuy (SeCTER – UNJu).

Este proyecto es llevado a cabo por el grupo de investigación Visualización y Recuperación Avanzada de Información (VRAIn) de la Facultad de Ingeniería de la UNJu.

1. INTRODUCCIÓN

El crecimiento exponencial de la información generada en la web, así como sus características de datos distribuidos, alta volatilidad, datos sin estructurar, redundantes y muy heterogéneos, han introducido nuevos problemas en los procesos de recuperación de información. Específicamente en lo que se refiere a exhaustividad y relevancia, dos cualidades contradictorias, ya que ningún sistema de información es perfecto. En la práctica, la mayor exhaustividad conduce a una cierta tasa de ruido (información inútil), y la mayor relevancia conduce a una cierta tasa de silencio (información útil no recuperada). Esta relación es parecida a la que se produce en la entropía. Así como se considera que no existe una maquina tan perfecta que no pierda una parte de su energía en forma de calor, se considera que no existe la posibilidad de tener un sistema de recuperación de información que sea, a la vez, perfectamente exhaustivo y relevante (Abadal & Codina, 2018). Una aproximación a esta problemática es la que pretende obtener resultados mucho más precisos, aunque sin perseguir altas tasas de exhaustividad, basándose en el uso de agentes inteligentes que rastreen la red según las necesidades informativas del usuario (Berrocal, Figuerola, Zazo, & Rodríguez, 2003).

Un agente se refiere a un componente de software y/o hardware que es capaz de actuar con precisión para realizar tareas en nombre de su usuario (Nwana & Ndumu, 1998). En este aspecto, un agente inteligente es un agente que emplea algún mecanismo o técnica de Inteligencia Artificial (IA) para realizar su propósito.

Los agentes han sido ampliamente utilizados para las tareas de Recuperación de Información (Brewington, y otros, 1999; Pilato, Vitabile, Vassallo, Conti, & Sorbello, 2003; Bach, Vlahovic, & Knezevic, 2005; Voorhees, 1994; Russell & Norvig, 2016), con el objetivo de automatizar la mayor parte de las tareas del proceso de recuperación debido principalmente a las grandes cantidades de información, a la diversidad y heterogeneidad de sus fuentes y formatos. Por ejemplo, en la recuperación de contenido desde la web el empleo de un *web crawler* es necesario para recolectar el contenido estructural de un sitio web para luego obtener las porciones o secciones relevantes del mismo.

Un problema recurrente basado en estas grandes cantidades de datos disponibles en la web, es el de encontrar información útil, dado que el usuario no es capaz de preparar una estrategia de búsqueda adecuada. Además, el tiempo que éste perdía años atrás recorriendo bibliotecas, lo pierde ahora buscando en una y otra base de datos, y recorriendo páginas obtenidas a través de buscadores, en búsqueda de información útil. Este exponencial crecimiento de información lleva al problema que los usuarios no son capaces de encontrar la información que buscan en una forma eficiente y simple, y frecuentemente no ven satisfechas sus necesidades de información. Si bien los usuarios no tienen por qué conocer técnicas de recuperación y extracción de información, se mejorarían los resultados de su búsqueda si el usuario fuera apoyado en este proceso de búsqueda por un sistema basado en la tecnología de agentes. Es por esto que un proceso de recuperación de información debe ser capaz no solo de extraer información de la web, sino también de procesarla, relacionarla y presentarla al usuario de un modo adecuado.

Mostrar información relevante y relacionada aporta ciertas mejoras sobre los sistemas tradicionales basados en búsqueda textual o

sintáctica (Medrano, 2018), donde los resultados tienen que ver con un posicionamiento, tal es el caso del *PageRank* (Page, Brin, Motwani, & Winograd, 1999) o con la existencia de los términos, palabras o frases buscados dentro de los campos de metadatos de los registros almacenados. Es aquí donde cobran relevancia las técnicas de Procesamiento del Lenguaje Natural (PLN) y de Aprendizaje Automático (Machine Learning) para sacar partido a partir del procesamiento inteligente de los datos. Esta etapa, la de transformación, es crucial puesto que aquí se reducirán, se filtrarán, se desecharán, se normalizarán datos para contar con un conjunto manejable, entendible, homogéneo, no redundante y menos voluminoso. Otra aproximación de este enfoque es la empleada para la detección de elementos duplicados, desambiguación, detección de spam, resúmenes automáticos, clasificación y organización de información, entre las más relevantes y estudiadas (Zhou, Pan, Wang, & Vasilakos, 2017; Martins, 2011; Hartrumpf, Vor Der Brück, & Eichhorn, 2010; Atkins, Nwala, Weigle, & Nelson, 2018).

Una vez que los datos son procesados, es necesario presentárselos al usuario de algún modo entendible, para este propósito se cuenta con la analítica visual (*Visual Analytics*) que es la ciencia del razonamiento analítico facilitado por interfaces visuales interactivas, es un campo multidisciplinario que incluye técnicas de razonamiento analítico, representaciones y transformaciones adecuadas de los datos y técnicas de representación visual y de interacción, lo que comúnmente se conoce como Visualización de Información (*InfoVis-Information Visualization*) (Heer, Card, & Landay, 2005). *InfoVis* busca aumentar el conocimiento humano mediante el aprovechamiento de las capacidades visuales humanas para dar sentido a la información abstracta (Card, 1999), proporcionando los medios por los cuales los seres humanos mediante sus capacidades perceptivas, pueden lidiar con el constante aumento de la cantidad de datos disponibles, es así que la analítica visual emplea técnicas de visualización de información para aumentar el conocimiento sobre un conjunto de datos a partir de representaciones visuales interactivas, es decir, se apoya en las capacidades cognitivas del ser humano para ampliar la información brindada, permitiendo al usuario sacar conclusiones, plantear hipotesis, realimentar el conocimiento previo entre otras.

En este contexto, el objetivo planteado es el tratamiento de la información masiva tanto desde el punto de vista de la representación visual como del de la interacción, a través de interfaces visuales que permitirán sintetizar información y derivar *insight* de datos masivos, dinámicos, ambiguos y a menudo contradictorios. Tomando como origen de datos la información extraída de la web a partir de agentes inteligentes diseñados específicamente para lidiar con las complejidades y variabilidades del entorno digital-cibernético.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La presente investigación se enfocará en dos aspectos claves al hablar de Big Data (grandes datos), la Recuperación de Información y la Visualización de Información. Los aspectos mencionados, son áreas de la Inteligencia Artificial en constante desarrollo debido al continuo aumento de las cantidades de datos generados y almacenados en la web. Este proyecto plantea el diseño de agentes inteligentes capaces de recuperar y obtener

información semi-estructurada o no estructurada de manera automática proveniente de diversos orígenes (distintos sitios web). La implementación de estos dispositivos (crawlers, arañas, spiders), piezas de software, conlleva una tarea de diseño a medida ya que atienden a la estructura de un sitio web en particular. La información obtenida, al ser textual y estar expresada en lenguaje natural, será procesada con técnicas de Procesamiento del Lenguaje Natural y Aprendizaje Automático para hallar patrones, tendencias y relaciones entre el contenido almacenado, es decir, los datos serán transformados una vez recolectados para luego ser presentados al usuario mediante visualizaciones interactivas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Una de las aplicaciones inmediatas de la solución que se plantea en este proyecto, será la de ofrecer noticias semánticamente relacionadas a partir de la cosecha/extracción/recolección de estas desde diferentes portales de noticias online. Esto permitirá identificar contenido duplicado (posible plagio), del mismo modo permitirá encontrar en un solo lugar (aplicación o sitio web) noticias relacionadas a partir de una búsqueda inicial, en otras palabras, se trata de un sistema de recomendación basado en el contenido (Ricci, Rokach, & Shapira, 2011) donde las recomendaciones o sugerencias serán noticias (Seroussi, 2010).

Paralelamente se ofrecerán visualizaciones de las búsquedas realizadas a modo de mapa conceptual, grafos y distintas representaciones que se diseñarán y se adecuarán al ámbito en cuestión. En el mismo sentido se plantea una aplicación novedosa al procesar la información de avisos clasificados tanto de portales de noticias como de redes sociales, en este aspecto, la búsqueda se realizará en base a características propias del elemento buscado (casa, automóvil, producto). Aquí nuevamente la visualización brindará otra óptica al poder representar las características extraídas de cada elemento y permitirle al usuario interactuar con la misma.

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes del grupo de investigación VRAIn llevan adelante esta línea de investigación desde hace unos años. En este sentido, cada año se incorporan al proyecto alumnos avanzados de las carreras Ingeniería en Informática y Licenciatura en Sistemas de la UNJu, quienes trabajan en temas relacionados. Del mismo modo, los docentes del equipo participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Ingeniería de la UNJu, UCSEDASS y UNLP.

5. BIBLIOGRAFÍA

Abadal, E., & Codina, L. (2018). Taxonomía, evolución y uso de los sistemas de información científica. In G. A. Torres Vargas, *Estudios de la información: teoría, metodología y práctica.* (pp. 161-180). UNAM.

Atkins, G. C., Nwala, A., Weigle, M. C., & Nelson, M. L. (2018). *Measuring News Similarity Across Ten US News Sites.* arXiv preprint.

Bach, M. P., Vlahovic, N., & Knezevic, B. (2005). Public data retrieval with software agents for business intelligence. *Proceedings of the 5th*

WSEAS Int. Conf. on Applied Informatics, (pp. 15-17).

Berrocal, J. L., Figuerola, C. G., Zazo, Á. F., & Rodríguez, E. (2003). Agentes inteligentes: recuperación automática de información en la web. *Revista española de documentación científica, 26*(1), 11-20.

Brewington, B., Gray, R., Moizumi, K., Kotz, D., Cybenko, G., & Rus, D. (1999). Mobile agents for distributed information retrieval. *Intelligent information agents*, 355-395.

Card, M. (1999). *Readings in information visualization: using vision to think.* Morgan Kaufmann.

Hartrumpf, S., Vor Der Brück, T., & Eichhorn, C. (2010). Semantic duplicate identification with parsing and machine learning. *International Conference on Text, Speech and Dialogue* (pp. 84-92). Springer, Berlin, Heidelberg.

Heer, J., Card, S. K., & Landay, J. A. (2005). Prefuse: a toolkit for interactive information visualization. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 421-430). ACM.

Martins, B. (2011). A supervised machine learning approach for duplicate detection over gazetteer records. *International Conference on GeoSpatial Sematics* (pp. 34-51). Springer, Berlin, Heidelberg.

Medrano, J. F. (2018). Filtrado basado en contenido para artículos académicos en repositorios institucionales. *In XXIV Congreso Argentino de Ciencias de la Computación*, (pp. 22-31). La Plata.

Nwana, H. S., & Ndumu, D. T. (1998). A brief introduction to software agent technology. *Agent technology*, 29-47.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The pagerank citation ranking: Bringing order to the web.* Stanford InfoLab.

Pilato, G., Vitabile, S., Vassallo, G., Conti, V., & Sorbello, F. (2003). A neural multi-agent based system for smart HTML pages retrieval. *IEEE/WIC International Conference on Intelligent Agent Technology*, (pp. 233-239).

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. *Recommender systems handbook*, 1-35.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach.* Malaysia: Pearson Education Limited.

Seroussi, Y. (2010). Utilising user texts to improve recommendations. *International Conference on User Modeling, Adaptation, and Personalization* (pp. 403-406). Springer.

Voorhees, E. M. (1994). Software agents for information retrieval. *Working notes of the AAAI spring symposium on Software agents*, (pp. 126-129).

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing, 237*, 350-361.

SUPERVISIÓN Y CONTROL DE PROCESOS

Enrique E. Tarifa[1,4], Sergio L. Martínez[1], Samuel Franco Domínguez[1], Susana A. Chalabe[1], Luis E. Ituarte[1], Álvaro F. Núñez[1], Jorgelina F. Argañaraz[1], Adolfo N. Riveros Zapata[2], Julieta Martínez[2], Juan P. Gutierrez[2,4], Lara V. Lescano Farias[3], Ubaldo J. M. Aramayo[1]

(1) Facultad de Ingeniería / Universidad Nacional de Jujuy / CONICET
Ítalo Palanca Nº 10 / S. S. de Jujuy / Provincia de Jujuy
(2) Facultad de Ingeniería / Universidad Nacional de Salta
Av. Bolivia 5150 / Salta Capital / Provincia de Salta
(3) Facultad de Agronomía y Agroindustrias / Universidad Nacional de Santiago del Estero
Avda. Belgrano (Sud) N° 1912 / Capital / Provincia de Santiago del Estero
(4) Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

eetarifa@fi.unju.edu.ar; smartinez@fi.unju.edu.ar; sfdominguez@yahoo.com.ar; susana.chalabe@gmail.com; luisituarte@hotmail.com; afnunez@fi.unju.edu.ar; jfarganaraz@hotmail.com; ariveros@unsa.edu.ar; julemartinez@unsa.edu.ar; gutierrezjp@unsa.edu.ar; lescanolaravaleria@gmail.com; uba.aramayo@gmail.com

Área temática: Agentes y Sistemas Inteligentes (ASI)

RESUMEN

El actual contexto productivo exige la optimización tanto del diseño como de la operación de las plantas industriales. Sin embargo, tradicionalmente, la Ingeniería de Procesos se enfocó más en el diseño que en la operación. Basados en ese enfoque, en el proyecto "Supervisión y Control de Procesos", se desarrollarán, adaptarán y aplicarán herramientas propias de la Operabilidad de Procesos (flexibilidad, controlabilidad, confiabilidad, robustez). Los procesos que tendrán prioridad son los que están implementados en la planta piloto de la Facultad de Ingeniería de la UNSa (extracción líquida-líquida, absorción gas-líquida, producción de vapor, pasteurización, reacción, entre otros), y los procesos vinculados al gas, al petróleo, al litio y a las energías no convencionales —especialmente a la energía solar—. Para alcanzar los objetivos del proyecto, se emplearán técnicas del campo de la Ingeniería de Procesos (simulación, optimización, control, diseño) y de la Inteligencia Artificial (sistemas expertos fuzzy, redes neuronales, minería de datos).

Palabras clave: Operabilidad, Flexibilidad, Controlabilidad, Confiabilidad, Robustez.

CONTEXTO

El grupo responsable de la realización del proyecto "Supervisión y Control de Procesos", es el grupo de investigación *IngProAr* (Ingeniería de Procesos Argentina), con lugar de trabajo en la Facultad de Ingeniería de la UNJu (Universidad Nacional de Jujuy), conjuntamente con el grupo de Petroquímica de la Facultad de Ingeniería de la UNSa (Universidad Nacional de Salta). El grupo *IngProAr* fue creado en 1995 por el Dr. Enrique Tarifa (investigador de CONICET y de la UNJu), y está constituido por docentes investigadores, profesionales y estudiantes de diversas áreas disciplinares. Desde sus inicios, este grupo trabajó en estrecha colaboración con el instituto de desarrollo y diseño INGAR (CONICET,

Santa Fe), el instituto de catálisis INCAPE (CONICET, Santa Fe) y el grupo de petroquímica de la Facultad de Ingeniería de la UNSa.

El proyecto al que hace referencia este trabajo, es continuación del proyecto anterior denominado "Desarrollo de Herramientas para la Operación de Procesos". En el presente proyecto, se propone continuar con la línea de investigación emprendida en el proyecto anterior, concentrándose en la supervisión y el control de procesos. Para ello, se desarrollarán, adaptarán y aplicarán herramientas propias de la Operabilidad de Procesos (flexibilidad, controlabilidad, confiabilidad, robustez). Estas herramientas serán aplicadas en los procesos implementados en la planta piloto de la Facultad de Ingeniería de la UNSa. Por otra parte, debido a la importancia que tienen los sistemas de energía no convencionales para Jujuy y para la UNJu, en este proyecto se considerarán también los sistemas de aprovechamiento de energía solar.

A fin de asegurar el cumplimiento del plan de trabajo que se propone, durante el proyecto anterior, los integrantes del grupo de investigación cursaron y aprobaron varios cursos de capacitación vinculados a la temática elegida: PLC, SCADA, HMI, control. Además, para la parte educativa, se tomaron cursos de capacitación docente, y se incorporó al grupo a la Dra. Jorgelina Argañaraz, Profesora en Ciencias de la Educación y Doctora en Ciencias Sociales. Por otra parte, se realizaron trabajos en colaboración con investigadores de INCAPE (CONICET, Santa Fe), el Dr. Carlos Román Vera y la Dra. Mariana Busto, las cuales proseguirán en el marco del proyecto propuesto. Por último, para proveer datos experimentales, al grupo *IngProAr*, se sumaron los siguientes investigadores de universidades de otras provincias: Adolfo Néstor Riveros Zapata (UNSa), Julieta Martínez (UNSa), Juan Pablo Gutierrez (UNSa) y Lara Valeria Lescano Farias (UNSE).

El proyecto de referencia es financiado por la Secretaría de Ciencia y Técnica y Estudios Regionales (SeCTER) y la Facultad de Ingeniería de la Universidad Nacional de Jujuy, y fue reconocido por la SeCTER-UNJu con el código D/0164 y por el Programa de Incentivos de la Secretría de Políticas Universitarias del Ministerio de Educación de la Nación.

1. INTRODUCCIÓN

El proyecto de referencia, propone el desarrollo, adaptación y aplicación de herramientas propias de la Operabilidad de Procesos. La operabilidad de un proceso productivo comprende las siguientes propiedades:

Flexibilidad: Implica la capacidad de un sistema de operar en estado estacionario estable para un rango de condiciones inciertas que pueden encontrarse durante la operación de la planta [1].

Controlabilidad: Puede ser definida como la capacidad de la planta para alcanzar y mantener un estado deseado [2].

Confiabilidad: Hace referencia a la capacidad de todos los elementos que conforman la planta, de funcionar continuamente satisfaciendo un conjunto de especificaciones y/o condiciones [3].

Robustez: Es la capacidad de la planta de hacer mínima la variación de la medida de calidad de los productos ante variaciones de las condiciones de operación [4].

Es de destacar que los actuales estudios de operabilidad involucran cada vez más el uso de técnicas de Inteligencia Artificial (redes neuronales, sistemas expertos con lógica fuzzy, entre otras), siendo los sistemas de control inteligentes los mejores exponentes de esta situación [5].

Para poder desarrollar herramientas destina a la operación de un proceso, es necesario contar con un conocimiento acabado del comportamiento estacionario y dinámico que el mismo presenta. Por ello, se desarrollará un modelo estacionario y otro dinámico para cada proceso a estudiar. Para lograr esto, se modelará el proceso escogido utilizando HYSYS —un simulador dinámico de propósito general— o, cuando

sea necesario, se desarrollarán modelos orientados a ecuaciones utilizando utilitarios matemáticos apropiados como MatLab, Simulink, Mathcad, Berkeley Madonna. En los casos que se justifique, se programará utilizando el lenguaje de programación más adecuado: Delphi, Visual Basic, C++.

Las licencias de las aplicaciones citadas tienen un costo considerable; por ese motivo, se hará un esfuerzo por sustituirlas con las opciones libres y de código abierto que brindan funcionalidades equivalentes: Lazarus, Octave, Scilab, Julia, wxMaxima. La evaluación de aplicaciones libres con código abierto es un tema que está cobrando importancia por sí mismo [6], y será un tema a analizar en el proyecto propuesto.

Cuando sea necesario desarrollar modelos propios, se adoptará el formato de modelo de espacio de estados. Los modelos a desarrollar serán modelos, en lo posible, de parámetros concentrados. De ser necesaria la utilización de modelos con parámetros distribuidos, se emplearán los siguientes métodos para su resolución en orden de preferencia: aproximación con una serie de modelos de parámetros concentrados y diferencias finitas.

Una vez que se cuente con el simulador del proceso seleccionado, se procederá a realizar un conjunto de estudios con el fin de evaluar la operabilidad del proceso (flexibilidad, controlabilidad, confiabilidad y robustez). También, se llevarán a cabo estudios de estabilidad y de optimización. Los modelos de optimización serán implementados en LINGO. La información obtenida de los estudios realizados se empleará para determinar las modificaciones que deberán efectuarse en el diseño, en la operación y en la gestión del proceso con el fin de optimizar su desempeño.

Para realizar algunos estudios de operabilidad, se recurrirá al empleo de simuladores estocásticos gratuitos (ModelRisk, JaamSim, HPSim), y se desarrollarán simuladores específicos para los casos en que sea necesario.

Para la supervisión del proceso, se diseñará el sistema de instrumentación y de procesamiento de datos que permita proveer al operador toda la información necesaria para la correcta operación del proceso. El método que se empleará para diseñar el sistema de instrumentación fue desarrollado por el grupo *IngProAr*. Otro método que se empleará también será la reconciliación de datos.

En cuanto a los sistemas de control, se desarrollarán sistemas de control avanzado empleando técnicas de Inteligencia Artificial (e.g., sistemas expertos fuzzy, redes neuronales, redes neuronales wavelets). Un punto importante a desarrollar serán los controladores MIMO (Multiple-input, Multiple-output) y los controladores predictivos. Para la implementación de técnicas de Inteligencia Artificial, se empleará preferentemente MatLab con licencia académica y, eventualmente, sus alternativas libres y abiertas.

Para el procesamiento de los datos, se emplearán los programas libres SciDAVis y PSPP, y el programa comercial, pero con licencia académica gratuita, RapidMiner Studio. Este último software es un potente entorno para realizar minería de datos, el cual no sólo permite obtener información a partir de datos, sino también "conocimiento". Este conocimiento es parte fundamental para el desarrollo de sistemas expertos fuzzy que serán empleados en el control del proceso seleccionado.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto de referencia se enmarca en las siguientes líneas prioritarias de la UNJu (Res. CS N°168/93):

Desarrollo Económico–Social Regional Sustentable: Las herramientas a desarrollar en el proyecto servirán para aumentar la operabilidad de los procesos productivos de la región. En consecuencia, se favorecerá el desarrollo económico de la región.

También, el proyecto se enmarca principalmente en la siguiente línea prioritaria de la Facultad de Ingeniería (Res. CAFI N°393/13):

Línea 3 - El estudio de procesos específicos o integrados que contribuyan a la cadena de valor de los productos obtenidos: Esta línea comprende los siguientes temas: Análisis, diseño y síntesis óptima de procesos; Simulación y optimización de procesos; Operación y mantenimiento de procesos; Supervisión y control de procesos; Sistemas de apoyo a la toma de decisiones en procesos; Integración de masa y energía; Análisis de la Performance de procesos. Los resultados a obtener con el proyecto propuesto están vinculados a todos los temas citados.

El proyecto en consideración además está vinculado con las siguientes líneas prioritarias de la Facultad de Ingeniería:

Línea 1 - La exploración, manejo, valorización, obtención, transformación y aprovechamiento de recursos naturales renovables, no renovables y energéticos. La Gestión ambiental relacionada: La vinculación con esta línea se presenta debido a que las herramientas a desarrollar en el presente proyecto pueden ser aplicadas a cualquier tipo de proceso productivo. Por ejemplo, en un proyecto anterior, las herramientas desarrolladas se aplicaron en la gestión óptima de fincas en la cuenca de la Quebrada de Humahuaca. Además, entre los procesos a estudiar, están los que involucran petróleo, gas, energía solar y litio; los cuales son los recursos energéticos más ampliamente utilizados en la actualidad.

Línea 5 - El desarrollo de la Informática, los sistemas y la tecnología derivados de ésta: Algunos de los productos del proceso son aplicaciones informáticas —simuladores, optimizadores, sistemas de control—, para las cuales se deben desarrollar nuevos algoritmos numéricos o técnicas de Inteligencia Artificial.

Línea 6 - El desarrollo de la calidad educativa y el estudio de los procesos de enseñanza y aprendizaje: Algunos productos del proyecto —simuladores, optimizadores, sistemas de control, publicaciones, apuntes— tienen impacto directo sobre la calidad educativa.

3. OBJETIVOS Y RESULTADOS

El proyecto en cuestión, tiene como objetivo general desarrollar herramientas para la operación de procesos. Para ello se diseñarán, desarrollarán, adaptarán y aplicarán herramientas propias de la Operabilidad de Procesos.

Los objetivos particulares del proyecto propuesto son los siguientes: 1) Desarrollar métodos para determinar y aumentar la flexibilidad de procesos; 2) Desarrollar sistemas de supervisión; 3) Desarrollar sistemas de control avanzados; 4) Desarrollar material para la enseñanza de Ingeniería Química (simuladores, optimizadores, sistemas de control, publicaciones y apuntes). Para alcanzar los objetivos propuestos, se emplearán técnicas del campo de la Ingeniería de Procesos (simulación, optimización, diseño) y de la Inteligencia Artificial (sistemas expertos fuzzy, redes neuronales, minería de datos).

Los procesos que tendrán prioridad son los que están implementados en la planta piloto de la Facultad de Ingeniería de la UNSa (extracción líquida-líquida, absorción gas-líquida, producción de vapor, pasteurización, reacción, entre otros), y los procesos vinculados al gas, al petróleo, al litio y a las energías no convencionales —especialmente, a la energía solar—.

Para el incipiente nivel de desarrollo del proyecto, aún no se cuenta con resultados destacables.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo que forma el grupo de investigación *IngProAr* cuenta con profesores, auxiliares, graduados y pasantes de la Facultad de Ingeniería, quienes desarrollan diversas actividades en la formación de recursos humanos, durante la ejecución del proyecto. Particularmente, los docentes investigadores del grupo *IngProAr* están a cargo de diversas cátedras en distintas carreras de la Facultad de Ingeniería de la UNJu: *"Simulación y Optimización de Procesos"*, *"Ingeniería de Procesos"*, *"Electricidad y Electrónica"* e *"Introducción a la*

Informática" —en Inge-niería Química—; *"Métodos de Simulación"*, *"Inteligencia Artificial"*, *"Técnicas y Estruc-turas Digitales"*, *"Metodología de la Progra-mación"* —en Ingeniería Informática—; *"Electrotecnia"* en Ingeniería Industrial; *"Geología Ambiental"* y *"Cartografía especial"* en Ingeniería de Minas. Es de destacar que los contenidos de dichas materias están directamente vinculados con los temas considerados en este proyecto. Como actividades de formación de recursos humanos, se pueden destacar:

Tesis de posgrado

- Tarifa E. E., codirección de la tesis doctoral *"Secado en lecho de chorro bidimensional para la deshidratación de proteínas del plasma y porción globina de sangre bovina"*, Ing. Lara Valeria Lescano Farías, Directora: Dra. Eve Liz Coronel. Res. F.A.A. N°641/2013, UNSE, Santiago del Estero.
- Tarifa E. E., dirección de la tesis doctoral *"Control Inteligente con Algoritmos Híbridos Optimizados aplicados a Modelos de Procesos Productivos"*, Ing. Sergio L. Martínez, Codirector: Dr. Juan P. Gruer (UNT), Facultad de Ciencias Exactas y Tecnología, Universidad Nacional de Tucumán, Res. FACET 620/2017, desde may/2017.
- Tarifa E. E., dirección de la tesis doctoral *"Diseño óptimo de un lecho fluidizado cónico para la inactivación y deshidratación del grano entero de soja"*, Ing. Gustavo Salcedo, UNSE, beca doctoral de CONICET, desde abr/2020.

Tesis de grado

- Martínez S. L., Dirección de la tesis de grado *"Desarrollo de una Herramienta con Interfaz Gráfica para Implementación de Redes Neuronales Feedforward"*. Tesista: David Llusco, codirector Ing. Jorge J. Gutiérrez, Carrera de Ingeniería Informática, UCSE-DASS, San Salvador de Jujuy. En ejecución.

5. BIBLIOGRAFÍA

1. Reay D., Ramshaw C.; Harvey A., Process Intensification: Engineering for efficiency, sustainability and flexibility. Butterworth-Heinemann, 2013.
2. Ramos W. B., Figueirêdo M. F., Brito K. D., Ciannella S., Vasconcelos L. G. S., Brito R. P., "Effect of Solvent Content and Heat Integration on the Controllability of Extractive Distillation Process for Anhydrous Ethanol Production", Ind. Eng. Chem. Res., 55 (43), pp 11315–11328, 2016.
3. Roy A., Srivastava P., Sinha S., "Risk and reliability assessment in chemical process industries using Bayesian methods", Reviews in Chemical Engineering, 40(5), 479-499, 2014.
4. Kharrazi A., Kumar P., Saraswat C., Avtar R., Mishra B. K., "Adapting Water Resources Planning to a Changing Climate: Towards a Shift from Option Robustness to Process Robustness for Stakeholder Involvement and Social Learnings", Journal of Climate Change, vol. 3, no. 2, pp. 81-94, 2017.
5. Santos M., "Aplicaciones Exitosas de Control Inteligente a Casos Reales", Revista Iberoamericana de Automática e Informática Industrial, V0-0, pp. 1–8. 2011.
6. Sharma N., Gobbert M. K., A comparative evaluation of Matlab, Octave, FreeMat, and Scilab for research and teaching. Technical Report HPCF–2010–7, UMBC High Performance Computing Facility, University of Maryland, Baltimore County, 2010.

Aplicaciones Inteligentes sobre Internet de las Cosas y Grandes Volúmenes de Datos: Un Enfoque Riguroso

Fernando Asteasuain –Federico D'Angiolo - Manuel Dubinsky – Fernando Pazos – Ivan Kwist - Matias Loiseau- David Contreras – Federico Calonge
Contacto: fasteasuain@undav.edu.ar
Ing. en Informática – Dpto. Tecnología y Administración – Universidad Nacional de Avellaneda

Resumen

Día tras días se generan millones y millones de nuevos datos y la cantidad de información a procesar es un desafío creciente. Entre los más destacados podemos mencionar el crecimiento exponencial de la "Internet de las cosas" o Internet of Things (IoT) en inglés en los sistemas modernos. En este sentido, no sólo el almacenamiento constante de información es un problema a resolver, sino también la extracción inteligente de información y su posterior análisis para introducir mejoras en los sistemas.

Este problema ha sido atacado desde la Inteligencia Artificial y la Optimización Combinatoria, pero se han encontrado algunas debilidades como la falta de modelado y diseño y de la aplicación rigurosa de técnicas de Ingeniería de Software. Esto hace que el problema se ataque de una manera "ad-hoc", por lo cual es necesario consolidar un enfoque de mayor formalidad en todas las etapas del proceso.

La presente investigación busca aplicar técnicas formales de Ingeniería de Software como Modelado y Model Checking al manejo y análisis de grandes volúmenes de datos.

Como caso de estudio concreto se trabajará con sensores para el Mantenimiento de parámetros del ambiente (como humedad o temperatura) del Laboratorio de Redes y Sistemas de Computación mediante protocolos de IoT.

Palabras claves: *Big Data, Procesamiento de Datos, Inteligencia Artificial, Métodos Rigurosos de Ingeniería de Software*.

Contexto

El presente proyecto se encuentra enmarcado dentro del proyecto UNDAVCYT denominado "Aplicaciones Inteligentes sobre Internet de las Cosas y Grandes Volúmenes de Datos: Un Enfoque Riguroso", dirigido por el Dr. Fernando Asteasuain y codirigido por el Ingeniero Federico D'Angiolo. El proyecto formó parte de la convocatoria UNDAVCY2019 y actualmente está en estado de revisión. El grupo de investigación está conformado por los autores del presente trabajo, combinando investigadores consolidados y en formación y estudiantes avanzados. El proyecto tiene una duración de dos años y está comenzando en este año. El proyecto estaría financiado 100% por la Universidad Nacional de Avellaneda en caso de ser otorgado. El proyecto forma parte de un objetivo de investigación más grande dentro de la carrera de Ingeniería en Informática de la UNDAV ya que propone el lanzamiento del Laboratorio de Inteligencia Artificial. En este sentido la carrera ha proporcionado una importante cantidad de recursos para tal fin, destacándose una computadora de alto poder de procesamiento en vistas a consolidar no sólo la investigación de la carrera sino también la posibilidad de transferencia y servicios al sector industrial y productivo. Las líneas de Investigación del proyecto tienen impacto en áreas prioritarias del *Plan Nacional de Ciencia y Técnica 2020* como AgroIndustria, Biodiversidad e Innovación Productiva.

1. Introducción

En los últimos años el desarrollo de sistemas para el procesamiento y análisis de grandes

volúmenes de datos (denominado Big Data) se ha transformado en un área trascendente para la Ingeniería de Software [1-8]. Como tal, la comunidad científica busca intensamente desarrollar herramientas y aplicaciones para extraer el máximo valor posible de los datos disponibles. Además, día tras días se generan nuevas fuentes de información, y a esto se le suma el hecho del crecimiento exponencial de la "Internet de las cosas" o Internet of Things (IoT) en inglés, la cual se define como la conexión de distintos equipos a una red para lograr la transmisión y recepción de datos, con el objetivo de gestionar distintos eventos [12]. Ejemplos de esto puede ser la automatización de refrigeración de un ambiente, siendo para esto, muy importante la utilización de sensores que permitan obtener datos de temperatura, humedad, presión o cualquier otra variable que sea necesaria.

No sólo el almacenamiento constante de información es un problema a resolver, sino también la extracción inteligente de información y su posterior análisis para introducir mejoras en los sistemas. Este problema ha sido atacado desde la Inteligencia Artificial [16-18,22] y la Optimización Combinatoria [13-15]. Propuestas como [22] exploran el análisis de Datos Exploratorio (EDA por sus siglas en inglés) basándose fuertemente en la minería de datos. En [16-18] presentan distintos enfoques para Big Data basado en redes neuronales y otras aplicaciones basadas en aprendizaje automático y redes neuronales.

Sin embargo, se han encontrado algunas debilidades como la falta de modelado y diseño y de la aplicación rigurosa de técnicas de Ingeniería de Software [1-8]. En trabajos como [4,6,8] se aplican técnicas formales de validación de software al proceso de BigData. Al estar basados en lógicas temporales hay lugar para introducir mejoras buscando lenguajes de especificación más expresivos y flexibles [11,4]. En trabajos como [1-4] se señalan debilidades en todo el proceso de desarrollo, buscando acercar la brecha entre la Ingeniería de Software y el análisis y procesamiento de BigData. Si bien estos enfoques han tenido su impacto, la mayoría de estos desafíos continúan abiertos. También desde un punto de vista de arquitectura de software es necesario brindar soluciones que brinden un desarrollo de mayor calidad [9,10]. Por ejemplo, son necesarias soluciones que cuenten con una logística dedicada al procesamiento distribuido y considerando conceptos relacionados al Cloud Computing.

Esto hace que el problema se ataque de una manera "ad-hoc", por lo cual es necesario consolidar un enfoque de mayor rigurosidad desde todas las etapas del proceso, el cual constituye el objetivo principal de la presente investigación.

El proyecto presentado buscar combinar métodos y técnicas rigurosas y formales de la Ingeniería de Software al procesamiento de grandes volúmenes de datos conocido como Big Data. El desafío involucra una sinergia entre diversas áreas como Inteligencia Artificial, Redes Neuronales, procesamiento dinámico de información incluyendo imágenes, sonido y diversas fuentes multimediales, procesamiento distribuido y Cloud Computing, Validación y Verificación Formal o Arquitecturas de Software.

Como se menciona en [1] el problema esencial de "Big Data" es que se está acumulando todo tipo de información a una velocidad superior a la cual puede ser procesada, y esta tendencia está creciendo día a día. Hay "cuatro V" que caracterizan a Big DATA:

1) Volumen: la cantidad creciente de información obtenida.

2) Variedad: Diversas y heterogéneas fuentes y tipos de dato.

3) Velocidad: para la manipulación y adquisición de datos, streaming en tiempo real, y datos de tiempo variable.

4) Veracidad: confianza en los datos obtenidos y su procesamiento.

Una solución a los mencionados problemas tendrá impacto en esta área consolidando métodos, técnicas y herramientas de la

Ingeniería de Software atacando cada una de las "cuatro V" que caracterizan a Big Data.

Se buscará en esta investigación aplicar conceptos de metodologías ágiles [19] al proceso de desarrollo y de investigación [20,21].

Como caso de estudio concreto, se trabajará con sensores para el Mantenimiento de parámetros del ambiente (como humedad o temperatura) del Laboratorio de Redes y Sistemas de Computación mediante protocolos de IoT. Este caso de estudio permitirá aportar contribuciones en un área cada vez más preponderante denominada "Ciencia Ambiental", la cual estudia modelos y herramientas de software aplicadas a una mejor conservación de los recursos energéticos de nuestro planeta [25,26]. El grupo de investigación ya viene de obtener resultados en el análisis del mencionado caso de estudio con técnicas de Inteligencia Artificial [27,28, 29].

2. Líneas de Investigación y Desarrollo

En el presente trabajo se explorarán las siguientes líneas de investigación:

- Aplicación de métodos formales de Ingeniería de Software al proceso de aprendizaje automático, redes neuronales, Big Data e IoT.
- Mejorar y aplicar nuevas variables (como por ejemplo la presión, además de la temperatura y humedad) a la aplicación del algoritmo *KNN* para el análisis de datos del caso de estudio [27, 28, 29].
- Analizar el problema de *overfitting* y *underfitting* en algoritmos de Regresión. En particular explorar si existe un polinomio que aproxima mejor a los datos obtenidos [27-29].
- Protocolos, modelado y verificación formal en IoT.
- Exploración de Redes Neuronales para la predicción de series temporales. Por ejemplo, redes que puedan realizar una predicción de cómo se puede comportar la temperatura (o humedad) del caso de estudio a utilizar.
- Análisis y Estudio del impacto de la utilización de aires acondicionados en el caso de estudio a utilizar. Dicha información puede ser relevante en diferentes entornos para detectar fallas o mal funcionamiento de los equipos que controlan la temperatura del cuarto refrigerado. También puede ofrecer un informe sobre la energía utilizada y recursos con el fin de estimar la huella ambiental que tiene la actividad desarrollada. Puede también extender la vida útil de los servidores manteniendo un registro controlado de operación dentro del rango de temperatura aconsejado por el fabricante.
- Exploración y utilización de Redes Neuronales Convolucionales.

3. Resultados Obtenidos/Esperados

El objetivo principal del presente proyecto de Investigación es aplicar métodos y técnicas rigurosas y formales de la Ingeniería de Software al procesamiento de grandes volúmenes de datos conocido como Big Data.

El desafío involucra combinar diversas áreas como Inteligencia Artificial, Redes Neuronales, procesamiento dinámico de información incluyendo imágenes, sonido y diversas fuentes multimediales, procesamiento distribuido y Cloud Computing, Validación y Verificación Formal o Arquitecturas de Software.

Como objetivos específicos podemos mencionar:

1. Investigar técnicas actuales para el manejo de grandes volúmenes de datos, incluyendo texto, imagen y sonido.

2. Analizar modelos de software y matemáticos que sean aplicables al procesamiento y análisis de información.

3. Relevar técnicas y herramientas utilizadas para el procesamiento y análisis de datos.

4. Aplicar herramientas y técnicas formales de la Ingeniería de Software para el desarrollo de sistemas de Big Data.

5. Consolidar técnicas para el procesamiento de imágenes y sonido.

6. Explorar Arquitecturas de Software que puedan llevar adelante procesos adecuados de BigData, considerando entornos distribuidos y de Cloud Computing.

7. Aplicar los modelos y conceptos teóricos desarrollados para el Mantenimiento de parámetros del ambiente del Laboratorio de Redes y Sistemas de Computación mediante protocolos de IoT.

8. Incorporar el procesamiento de imágenes obtenidas desde las cámaras dentro del Laboratorio de Redes y Sistemas de Computación.

9. Incorporación de más fuentes de información además de humedad y temperatura.

10. Extrapolar los resultados intermedios y finales para el mencionado Laboratorio a otras áreas del Dpto. de Tecnología y Administración, como otros Laboratorios, aulas y otras dependencias.

11. Extrapolar los resultados obtenidos más allá de la Universidad Nacional de Avellaneda, y buscar que sean aplicables para cualquier contexto y ubicación.

12. Divulgar los resultados de la investigación en congresos y revistas científicas de interés para los temas de la investigación.

13. Consolidar los recursos humanos en inicios de las tareas de investigación como estudiantes avanzados de la carrera, mediante la participación activa en el proyecto o mediante la realización de tesis de final de carrera de grado.

14. Profundizar el área de investigación dentro del Dpto. de Tecnología y Administración.

Los resultados esperados incluyen:

- Lanzamiento del Laboratorio de Inteligencia Artificial de la carrera de Ingeniería en Informática.

- Consolidar y profundizar técnicas formales de la Ingeniería de Software en un área de aplicación de vanguardia como Big Data e Internet de las Cosas. Como caso concreto de estudio se aplicarán los resultados en el Laboratorio de Redes y Sistemas de Computación, buscando optimizar parámetros como humedad, temperatura, y explorando nuevos parámetros a través de imágenes de cámara web y sonidos de ambiente. Esto dará la posibilidad directa de aplicar y validar los contenidos teóricos desarrollados en un caso real de importancia. Luego, es esperable el análisis de extrapolación de los resultados obtenidos a un contexto que trascienda el caso de estudio

- Potenciar la transferencia desde el grupo de investigación al sector productivo e industrial al proveer servicios y experiencia en el manejo de grandes volúmenes de datos, redes neuronales y métodos formales de Ingeniería de Software.

- Dirección de tesis de licenciatura y supervisión de prácticas profesionales (PPS).

4. Formación de Recursos Humanos

El equipo de investigación cuenta con dos investigadores formados (director y codirector), y dos investigadores con dedicación exclusiva en proceso de formación. Además, incluye 4 estudiantes avanzados que están realizando su tesis de investigación dentro del marco del presente proyecto. Dichas tesis de grado se encuentran en diferentes estadíos, desde más avanzadas a algunas todavía en fase de exploración.

5. Bibliografía

[1] Embley, David W., and Stephen W. Liddle. "Big data—conceptual modeling to the rescue."International Conference on Conceptual Modeling. Springer, Berlin, Heidelberg, 2013.
2) van der Aalst, Wil, and Ernesto Damiani. "Processes meet big data: Connecting data science with process science."IEEE Transactions on Services Computing 8.6 (2015): 810-819.
3) Arndt, Timothy. "Big Data and software engineering: prospects for mutual enrichment."Iran Journal of Computer Science 1.1 (2018): 3-10.
4) Camilli, Matteo. "Coping with the State Explosion Problem in Formal Methods: Advanced Abstraction Techniques and Big Data Approaches." (2015).
5) Hummel, Oliver, et al. "A collection of software engineering challenges for big data system development."2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). IEEE, 2018.
6) Bellettini, C., Camilli, M., Capra, L., & Monga, M. (2016). Distributed CTL model checking using MapReduce: theory and practice. Concurrency and Computation: Practice and Experience, 28(11), 3025-3041.
7) Kumar, Vijay Dipti, and Paulo Alencar. "Software engineering for big data projects: Domains, methodologies and gaps."2016 IEEE International Conference on Big Data (Big Data). IEEE, 2016..
8) Camilli, Matteo. "Formal verification problems in a big data world: towards a mighty synergy."Companion Proceedings of the 36th International Conference on Software Engineering. ACM, 2014.
9) Bakshi, Kapil. "Considerations for big data: Architecture and approach."2012 IEEE Aerospace Conference. IEEE, 2012.
10) Marz, Nathan, and James Warren. Big Data: Principles and best practices of scalable real-time data systems. New York; Manning Publications Co., 2015.
11) Asteasuain, Fernando, and Víctor Braberman. "Declaratively building behavior by means of scenario clauses."Requirements Engineering 22.2 (2017): 239-274.
12) Atzori, Luigi, Antonio Iera, and Giacomo Morabito. "The internet of things: A survey."Computer networks 54.15 (2010): 2787-2805.
13) Hu, C., Xu, Z., Liu, Y., Mei, L., Chen, L., & Luo, X. (2014). Semantic link network-based model for organizing multimedia big data. IEEE Transactions on Emerging Topics in Computing, 2(3), 376-387.;
14) Bolin, G., & Andersson Schwarz, J. (2015). Heuristics of the algorithm: Big Data, user interpretation and institutional translation.
15) Big Data & Society, 2(2), 2053951715608406.; Xu, Z., Liu, Y., Yen, N., Mei, L., Luo, X., Wei, X., & Hu, C. (2016). Crowdsourcing based description of urban emergency events using social media big data. IEEE Transactions on Cloud Computing.
16) Kasun, L. L. C., Zhou, H., Huang, G. B., & Vong, C. M. (2013). Representational learning with extreme learning machine for big data. IEEE intelligent systems, 28(6), 31-34;
17) Chen, X. W., & Lin, X. (2014). Big data deep learning: challenges and perspectives. IEEE access, 2, 514-525; Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015).
18) Deep learning applications and challenges in big data analytics. Journal of Big Data, 2(1), 1;Aggarwal, C. C. (Ed.). (2014). Data classification: algorithms and applications. CRC press.
19) Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., ... & Kern, J. (2001). Manifesto for agile software development.
20) ITAS: Highsmith, J. A., & Highsmith, J. (2002). Agile software development ecosystems. Addison-Wesley Professional;
21) Cockburn, A., & Highsmith, J. (2001). Agile software development: The people factor. Computer, (11), 131-133
22) Yu, Chong Ho. "Exploratory data analysis."Methods 2 (1977): 131-160.
23) F. Asteasuain, F. Calonge, F. Diaz, F. Dangiolo, and P. Gamboa. Expressing early behavior specifications with branching visual scenarios. In CONAIISI, 2018.
24) F. Asteasuain, F. Calonge, and M. Dubinsky. Exploring specification pattern based behavioral synthesis with scenario clauses. In 2018 CACIC ISBN 978-950-658-472-6, pages 22,34. CACIC, 2018.
25) Simm, William Alexander, et al. "Se in es: Opportunities for software engineering and cloud computing in environmental science."Proceedings of the 40th International Conference on Software Engineering: Software Engineering in Society. ACM, 2018.
26) Hřebíček, Jiří, et al., eds. Environmental Software Systems. Computer Science for Environmental Protection: 12th IFIP WG 5.11 International Symposium, ISESS 2017, Zadar, Croatia, May 10-12, 2017, Proceedings. Vol. 507. Springer, 2018
[27] Algoritmos de Regresión Lineal aplicados al mantenimiento de un Datacenter. Federico Gabriel D'Angiolo, Iván Federico Kwist, Matias Loiseau, David Exequiel Contreras, Fernando Asteasuain. Universidad Nacional de Avellaneda, Avellaneda, Argentina. Departamento de Tecnología y Administración. Ingeniería en Informática. CACIC 2019.
[28] Algoritmo de KNN Aplicado al Mantenimiento de un Datacenter. Federico Gabriel D'Angiolo, Iván Federico Kwist, Matias Loiseau, David Exequiel Contreras, Gregorio Oscar Glas. CONAIISI 2019
[29] Monitorización de un Datacenter mediante Protocolos de IoT. Iván Federico Kwist, Matias Loiseau, David Exequiel Contreras, Federico Gabriel D'Angiolo, Roberto Osvaldo Mayer. CONAIISI 2019.

Inteligencia Artificial aplicada a IoT

Azar Miguel Augusto[1], García Jorge Luis[2], Bernal Sergio[3], Aleman Liliana Judith[4], Tolaba Matías[5]
Facultad de Ingeniería - Universidad Nacional de Jujuy
Italo Palanca Nº 10 (+54 0388 4221587) – San Salvador de Jujuy - Jujuy
augusto.azar@gmail.com[1], garciajorge4612@gmail.com[2], sergio.b196@gmail.com[3], jud2301@gmail.com[4], diego95mati@gmail.com[5]

Resumen

El concepto de Internet de las Cosas (IoT) actualmente comprende una interrelación de varias disciplinas. Si bien en sus primeras etapas se limitaba al conexionado de "dispositivos electrónicos dotados de sensores y conectados a Internet" [1], actualmente involucra el conocimiento de dominios tales como la física newtoniana, la física electrónica, la minería de datos, la big data y sobre todo la inteligencia artificial.

En general el campo de la inteligencia artificial (AI) se aplica en diversos ámbitos en los cuales se registra la existencia de tareas repetitivas, alta precisión, manejo de grandes volúmenes de información, riesgo de vida, extrema complejidad en la resolución de problemas, entre otros [2].

Dado que existe un incremento notable de la bibliografía que da cuenta de la simbiosis entre IoT y AI, el presente proyecto detalla tanto los elementos involucrados como sus interrelaciones dentro de una investigación que plantea el desarrollo de algoritmos inteligentes en sistemas IoT (sensores, módulos de entrada/salida, circuitos electrónicos, drivers, bibliotecas, microcontroladores, firmware, periféricos de comunicación, protocolos, sistemas operativos, servers cloud computing, sistemas machine/deep learning y metodologías de desarrollo).

Palabras clave: Artificial Intelligence, Embeded Systems, Sensors, Internet of Things, Smart City.

Contexto

Inicialmente la investigación se concentró en el desarrollo de algoritmos de inteligencia artificial que permitan inferir conocimiento acerca de los sensores conectados a sistemas embebidos. En el mismo la I/D/I se realizó en base a una convocatoria de la Secretaría de Ciencia y Técnica de Facultad de Ingeniería (UNJu) a través del Expediente S-7710/17 y mediante las Resoluciones C.A.F.I. Nº 270/17 y C.A.F.I. Nº 271/17. La convocatoria se estableció para docentes incorporados al sistema de incentivos de la Secretaría de Políticas Universitarias, con categorías IV y V. También se enmarcó para dar cumplimiento a los compromisos de mejoramiento desde la Resolución CONEAU Nº 1230/12 de acreditación de las Carreras de Ingeniería Informática y Licenciatura en Sistemas. La propuesta fue aprobada mediante Resolución C.A.F.I. Nº 661/17.

Actualmente por resolución C.A.F.I. Nº 449/19 (Facultad de Ingeniería - UNJu), el enfoque del nuevo proyecto se concentra en los sistemas inteligentes involucrados en arquitecturas IoT. El mismo dió inicio el 1 de agosto de 2019 y se extiende por el término de un año.

Introducción

Algunos autores definen IoT como una extensión de Internet hacia el ámbito físico [3], mientras que otros lo definen como el colectivo de sensores colocados en "cosas" y dentro de infraestructuras cibernéticas [4]. La relevancia de este concepto se ha intensificado con el transcurso del tiempo debido, entre otras razones, al creciente uso de sensores en smartphones y automóviles [5].

Precisamente la principal línea de investigación de este proyecto se concentra en el desarrollo de un sistema que permita la detección y geo-localizacion de averias viales; tanto dentro de las arterias de una ciudad como en rutas.

Estado del arte

Existen proyectos similares que plantean diferentes enfoques, pero en su mayoría no contemplan el uso de inteligencia artificial como sistema de apoyo para la detección de baches.

En [6], los autores se concentran en diferentes algoritmos de detección de baches basados en los datos del acelerómetro de smartphones con sistema operativo Android y realizan una clasificación de baches (pequeños, grupo de baches y grandes). De este modo logran una identificación de los daños en calles en forma exitosa, de hasta un 90%.

En el trabajo presentado por [7], los autores se enfocan en un análisis técnico profundo y desarrollan un sistema de adquisición de datos que son procesados y sometidos al cálculo de un índice de rugosidad compuesto en base a un modelo matemático utilizado para extraer información de las condiciones de la superficie de una carretera.

Otra propuesta es la desarrollada por [8] en la que se diseña una combinación de hardware y software compuesta por acelerómetro, unidad de GPS, microcontrolador, módulo Zigbee [9] y servicio en la nube.

En [10] se emplea un conjunto hardware consistente en una unidad de GPS, acelerómetro, Arduino y un servidor que es actualizado mediante el uso de un módulo Bluetooth. El enfoque para detectar baches está basado en funciones definidas en [6].

El trabajo presentado por [11] emplea una aplicación móvil que además del acelerómetro utiliza el giroscopio. A partir de los datos recolectados se entrenan modelos SVM (Support Vector Machines) logrando una precisión superior al 90%.

Otros trabajos similares a los ya citados son los estudiados por [12]-[14]. En [15] se describe de manera muy simplificada la publicación de una patente de un sistema embebido que emplea acelerómetro, GPS y microprocesador.

Mohideen y colegas [16], captura datos a partir de un acelerómetro colocado en una bicicleta y propone a futuro instalarlo en vehículos.

El trabajo de [17] utiliza GPS, acelerómetro angular, sensor geomagnético y zigbee. Para el análisis de los datos utiliza power spectral density (PSD) para calcular el international roughness index (IRI).

En [18] se utilizan los sensores del celular para recopilar información de las

calles y su ubicación. También emplea Spectral analysis y SVM para clasificar las mediciones en 3 grupos.

Los autores de [19] plantean un sistema idéntico a [18] pero utiliza Power Spectral Density Analysis y Gaussian Modeling.

Los diferentes modelos que se pueden aplicar para el análisis de irregularidades de la superficie de la carretera se plantean en [20].

En [21] se realiza un estudio de los modelos no supervisados y aprendizaje profundo para series temporales que podría adaptarse para los datasets de la presente investigación.

Otro desarrollo similar [22] utiliza los sensores del celular para captura de los datos de profundidad de baches. El análisis lo hace con wavelets, redes neuronales autorregresivas y transformada de Hilbert.

Un trabajo muy reciente [23] detecta baches a través de imágenes, los datos lo saca de Google Street View y usa una arquitectura de redes residuales (Resnet).

En el caso de [24] se utiliza la arquitectura SqueezeNet y se analiza el coeficiente de fricción para identificar el tipo de carretera en la cual se está circulando.

Otro trabajo también reciente [25] se enfoca en detectar la rugosidad del pavimento, pero trata los baches como anomalías, usa Z-Peak algorithm (para altas velocidades) Z-Sus algorithm (en velocidades bajas).

En [26] se utiliza acelerómetro, giroscopio y GPS conectados a una Raspberry PI. Muestra diferentes algoritmos de ML para el análisis; siendo Random Forest el que obtuvo un 86% de precisión.

Los autores de [27] desarrollan un artículo en donde se utiliza sensor de proximidad, giroscopio, sensor de efecto hall, instalados en una motocicleta con el mismo fin de determinar magnitudes de baches.

En [28] se utiliza nodo sensor con un acelerómetro, y detecta los baches con una fórmula que calcula el umbral a partir del cual se considera la lectura del sensor.

Por último, en el trabajo de Hsu y colegas [29] se recolectan datos a través del sensor IMU (Inertial Measurement Unit), laser Rangefinder y una cámara. Analiza de forma individual cada caso y de forma conjunta.

Justificación del proyecto

Existe una creciente demanda de productos tanto en el campo de IoT como de AI. Se espera que la inversión en IoT se eleve a 1.5 billones de dólares para 2020 [30]. También una encuesta destaca la importancia de las tecnologías del futuro se concentra principalmente en dos campos, IoT y AI [31].

Metodología

Técnicamente se trata del paradigma Ignite [32]. Dicho enfoque, fue propuesto por Slama y colegas en 2015 y plantea subdividir el desarrollo y construcción del sistema en dos fases: Iniciación de la estrategia y Entrega de soluciones. En paralelo el equipo se encuentra trabajando en base a esta metodología.

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación principal es el desarrollo de sistemas adaptables a automóviles que permita identificar aquellos baches presentes en rutas y

ciudades para su reparación urgente. Además se prevé que el sistema a traves de realidad aumentada permite mostrar al conductor donde se encuentran los próximos baches y a que velocidad se recomienda transitar de manera de atenuar los posibles daños mecánicos.

Ello requiere el uso de placas de desarrollo, sensores, módulos, conectividad 4G, servicio en la nube y sistema machine learning que interprete los datos leídos.

En cuanto a las placas de desarrollo se dispone de 4 computadoras industriales EDU-CIAA, 4 Raspberry PI (diferentes versiones) y sensores varios. Actualmente el proyecto se encuentra en etapa de pruebas mediante Raspberry PI en combinación con el módulo acelerómetro y módulo GPS.

Resultados y Objetivos

El objetivo general es desarrollar algoritmos inteligentes en base a plataformas embebidas (online y standalone) para el tratamiento de señales provistas por sensores.

Objetivos particulares

✔ Analizar y evaluar las tecnologías existentes dentro del ámbito de los sensores y actuadores.

✔ Estudiar las alternativas de hardware/firmware/software para el tratamiento de los datos que los sensores proveen.

✔ Diseñar estructuras algorítmicas destinadas a mejorar las técnicas de inteligencia artificial convencionales.

✔ Seleccionar e incorporar una metodología de desarrollo adaptable a cada arquitectura y despliegue de hardware/software.

Formación de Recursos Humanos

Actualmente el grupo de investigación está formado por un director y cuatro estudiantes de la carrera Ingeniería Informática. La orientación del director ha permitido a los estudiantes tener contacto con las placas de desarrollo EDUCIAA en su mayoría adquiridas con fondos propios del director. El trabajo de los estudiantes se centra también en el desarrollo de firmware en lenguaje C, la conectividad de diferentes sensores y por otro lado la investigación en el campo de la Inteligencia Artificial que trabaje en base a los datos adquiridos por los sensores.

Referencias

[1] Ashton K. That 'Internet of Things' Thing: In the real world, things matter more than ideas. RFiD Journal. RFID. 2009.

[2] Azar M.A., Tapia M.A., García J.L. y Pérez A.J.M. Inteligencia Artificial de las Cosas. Workshop de Investigadores de Ciencias de la Computación (WICC). 2019.

[3] Tsirmpas C., Anastasiou A., Bountris P. y Koutsouris D. A New Method for Profile Generation in an Internet of Things Environment: An Application in Ambient-Assisted Living. IEEE Internet of Things Journal. Volume: 2, Issue: 6. Page(s): 471 – 478. ISSN: 2327-4662. 2015.

[4] Nahrstedt K., Li H., Nguyen P., Chang S. y Vu L. Internet of mobile things: Mobility-driven challenges, designs and implementations. In Proceedings - 2016 IEEE 1st International Conference on Internet-of-Things Design and Implementation, IoTDI 2016 (pp. 25-36). [7471348] Institute of Electrical and

Electronics Engineers Inc. DOI: 10.1109/IoTDI.2015.41. 2016.

[5] Wearable Sensors Market by Type - Global Forecast to 2022. Online. http://www.marketsandmarkets.com/Market-Reports/wearable-sensor -market-158101489.html. Consultado el 03 de febrero de 2019.

[6] A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, and L. Selavo. Real time pothole detection using android smartphones with accelerometers. In IEEE DCOSS. 2011.

[7] G. Alessandroni, L. C. Klopfenstein, S. Delpriori, M. Dromedari, G. Luchetti, B. D. Paolini, A. Seraghiti, E. Lattanzi, V. Freschi, A. Carini and A. Bogliolo. SmartRoadSense: Collaborative Road Surface Condition Monitoring. In Proc. of UBICOMM. 2014.

[8] H. A. Jamakhandi and K. G. Srinivasa. Internet of things based real time mapping of road irregularities. Assistive Devices Group M. S. Ramaiah Institute of Technology, Proceedings of International Conference on Circuits, Communication, Control and Computing. I4C. 2014.

[9] Zigbee alliance, https://www. zigbee.org/what-is-zigbee/494-2/

[10] S. Dimple, V. Monica, A. Anirudh and C. Adarsh, "Monitoring Of Road Irregularities Using IOT", International Journal of Management and Applied Science (IJMAS), pp. 127-130, Special Issue. 2016.

[11] U. Bhatt, S. Mani, E. Xi and Z. Kolter. Intelligent Pothole Detection and Road Condition Assessment. Data For Good Exchange. 2017.

[12] A. S. El-Wakeel, J. Li, M. T. Rahman, A. Noureldin and H. S. Hassanein, "Monitoring road surface anomalies towards dynamic road mapping for future smart cities", Signal and Information Processing (GlobalSIP) 2017 IEEE Global Conference on, pp. 828-832. 2017.

[13] B. Lanjewar et al., "Road bump and intensity detection using smartphone sensors", Int. J. Innov. Res. Comput. Commun. Eng., vol. 4, no. 5, pp. 9185-919. 2016.

[14] H. Song, K. Baek and Y. Byun. Pothole Detection using Machine Learning. Advanced Science and Technology Letters Vol.150 (AST 2018), pp.151-155. 2018.

[15] H. A. Troemel and M. R. Stelts, “Road Health (Pothole) Detection over Wireless Infrastructure”, U.S. Patent 15/662,753, Jul. 28. 2017.

[16] Haja Mohideen, A. J., Rosli, M. F., Mohamad Hanif, N. H. H., Mohd Zaki, H. F., Husman, M. A., Abdul Muthalif, A. G., & Kumar, D. Pavement Condition Analysis Via Vehicle Mounted Accelerometer Data. IIUM Engineering Journal, 21(1), 73 - 84. 2020.

[17] Du, Y., Liu, C., Wu, D., & Li, S. (2016). Application of Vehicle Mounted Accelerometers to Measure Pavement Roughness. International Journal of Distributed Sensor Networks. 2016.

[18] Li, X., & Goldberg, D. W. Toward a mobile crowdsensing system for road surface assessment. Computers, Environment and Urban Systems, 69, 51–62. 2018.

[19] Harikrishnan, & Gopi, V. P. Vehicle Vibration Signal Processing for Road Surface Monitoring. IEEE Sensors Journal, 17(16), 5192–5197. 2017.

[20] Nguyen, T., Lechner, B. & Wong, Y.D. Response-based methods to measure road surface irregularity: a state-of-the-art review. Eur. Transp. Res. Rev. 11, 43 2019.

[21] Längkvist, M., Karlsson, L., & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. Pattern

Recognition Letters, 42, 11–24. doi:10.1016/j.patrec.2014.01.008. 2014.

[22] Kanarachos, S., Christopoulos, S.-R. G., Chroneos, A., & Fitzpatrick, M. E. Detecting anomalies in time series data via a deep learning algorithm combining wavelets, neural networks and Hilbert transform. Expert Systems with Applications, 85, 292–304. 2017.

[23] Alipour, M., Harris, D.K. A big data analytics strategy for scalable urban infrastructure condition assessment using semi-supervised multi-transform self-training. J Civil Struct Health Monit 10, 313–332. 2020.

[24] Šabanovič, Žuraulis, Prentkovskis, & Skrickij. Identification of Road-Surface Type Using Deep Neural Networks for Friction Coefficient Estimation. Sensors, 20(3), 612. 2020.

[25] L. Janani, V. Sunitha & Samson Mathew. Influence of surface distresses on smartphone-based pavement roughness evaluation, International Journal of Pavement Engineering. 2020.

[26] Bansal, K., Mittal, K., Ahuja, G., Singh, A., & Gill, S. S. DeepBus: Machine Learning based Real Time Pothole Detection System for Smart Transportation using IoT. Internet Technology Letters. 2020.

[27] Muhammad Hanif, H., Lie, Z. S., Astuti, W., & Tan, S. Pothole detection system design with proximity sensor to provide motorcycle with warning system and increase road safety driving. IOP Conference Series: Earth and Environmental Science, 426, 012039. 2020.

[28] De Silva, G. D., Perera, R. S., Laxman, N. M., Thilakarathna, K. M., Keppitiyagama, C. I., & De Zoysa, K. Automated pothole detection system. In Proc. Int. Conf. Adv. ICT Emerg. Regions. 2013.

[29] Y. Hsu, J. Perng y Z. Wu. Diseño e implementación de un sistema inteligente de detección de carreteras con integración multisensor. Conferencia Internacional 2016 sobre Aprendizaje Automático y Cibernética (ICMLC), Jeju. pp. 219-225. 2016.

[30] Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016. https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says- 8-billion-connected-things-will-be-in- use-in-2017-up-31-percent-from-2016.

[31] Top 10 Tech Trends from CES 2019. https://www.emarketer.com/content/top-10-tech-trends-from-ces-2019.

[32] Slama D., Puhlmann F., Morrish J. y Bhatnagar R.M. Enterprise IoT, Strategies & Best Practices for Connected Products & Services. O'Reilly. 2015.

Aprendizaje Automático Profundo y Visión por Computadora. Aplicaciones en el Reconocimiento de Lengua de Señas e Imágenes Astronómicas

L. Lanzarini[1,2], F. Ronchetti[1,2,3], F. Quiroga[1,2,4], , G. Rios[1,3], U. Cornejo Fandos[1,5], K. Canaza[1] , Pedro Dal Bianco[1], Ivan Mindlin[1], F. Ravettino[1], A. Rosete[6], R. Gamen[7], D. Puig Valls[9], J. Torrents-Barrena[9] , Y. Aidelman[7,8], C. Escudero[8].

[1] Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina.*
[2] Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina
[3] Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)
[4] Becario postgrado UNLP [5] Becario CIN
[6] Universidad Tecnológica de La Habana "José Antonio Echeverría" (CUJAE), La Habana, Cuba
[7] Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, La Plata, Argentina
[8] Instituto de Astrofísica de La Plata (IALPCONICET) , La Plata, Argentina
[9] Universitat Rovira i Virgili (URV), Tarragona, España
* Centro asociado de la Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)

Contacto: laural@lidi.info.unlp.edu.ar

CONTEXTO

Esta presentación corresponde a las tareas de investigación que se llevan a cabo en el III-LIDI en el marco del proyecto F025 "Sistemas inteligentes. Aplicaciones en reconocimiento de patrones, minería de datos y big data" perteneciente al Programa de Incentivos (2018-2021).

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes para la resolución de problemas de reconocimiento de patrones en imágenes y video, utilizando técnicas de Aprendizaje Automático clásicas, junto con Redes Neuronales Convolucionales y Aprendizaje profundo. El trabajo presentado describe diferentes casos de aplicación en visión por computadora.

Una de las líneas de investigación principales que se continúa desarrollando es el reconocimiento de lengua de señas. Este es un problema complejo y multidisciplinar, que presenta diversos subproblemas a resolver como el reconocimiento del intérprete, la segmentación de manos, la clasificación de diferentes configuraciones y de un gesto dinámico, entre otros.

En esta área se está estudiando la forma de reconocer formas de mano de la Lengua de Señas con conjuntos de datos de tamaño reducido, dada la falta de datos de entrenamiento para este dominio.

Además, se están comenzando a utilizar Redes Generativas Adversarias (GANs) para aumentar bases de datos de formas de mano, con el objetivo de complementar desde otro enfoque el entrenamiento de modelos para su clasificación.

Por otro lado, se está estudiando la forma en que las redes neuronales codifican la invarianza a las transformaciones y otras propiedades transformacionales, con el objetivo de poder analizar y comparar estos modelos, y finalmente mejorarlos. De esta forma se espera poder mejorar los modelos de clasificación de objetos transformados, en particular, de formas de mano.

Siguiendo con la línea de reconocimiento de patrones en imágenes, se está llevando a cabo una colaboración con investigadores de la Facultad de Astronomía y Geofísica de la UNLP para crear modelos de clasificación de imágenes de objetos celestes. Además, se está desarrollando un sistema para recuperar la información de placas espectrográficas astronómicas antiguas, con el objetivo de recuperar los espectrogramas que contienen.

Por último, se está diseñando un prototipo para crear experiencias interactivas multimedia que incorpore técnicas de aprendizaje profundo tanto para la entrada de datos por parte de los usuarios, como para la generación de una salida en forma de imágenes, videos, sonido y texto.

Palabras clave: Aprendizaje Automático, Visión por Computadoras, Lengua de Señas, Redes Generativas Adversarias, Invarianza, Equivarianza, Imágenes Astronómicas, Interacción Inteligente.

1. INTRODUCCION

El Instituto de Investigación en Informática LIDI (III-LIDI) tiene una larga trayectoria en el estudio, investigación y desarrollo de Sistemas Inteligentes basados en distintos métodos de Aprendizaje Automático y Redes Neuronales

Como resultado de estas investigaciones se han diseñado e implementado técnicas originales aplicables a la clasificación y el análisis de características de objetos en imágenes, generación de imágenes para aumentación de datos, y estudio del funcionamiento de las redes neuronales. En relación con esta línea, actualmente se están desarrollando los siguientes temas:

1.1. Clasificación de formas de mano de la Lengua de Señas

Las lenguas de señas utilizan un conjunto finito de formas de mano que, en combinación con movimientos de las manos y el cuerpo, y expresiones faciales, se utilizan para señar.

Una etapa fundamental en el reconocimiento de la lengua de señas es la clasificación de estas formas de mano, y por ende un área de investigación prioritaria para mejorar el reconocimiento.

Por otro lado, las Redes Neuronales Convolucionales (CNN, por sus siglas en inglés) son actualmente el estado del arte en el área. En un estudio previo, se realizaron experimentos comparando diversas arquitecturas de CNN para clasificar formas de manos, pero sólo con dos bases de datos, LSA16 y RWTH (ver Figura 1). Los resultados de estas investigaciones fueron publicados en [1] advirtiendo que el desempeño de los modelos utilizados estuvo limitado por la falta de datos.



Figura 1. Ejemplos de las bases de datos de formas de mano RWTH y LSA16.

Por ende, se realizó una comparación de distintos modelos de Redes Neuronales diseñados específicamente para trabajar con conjuntos de datos de tamaño reducido [2], extendiendo también el análisis a otras bases de datos. Estos modelos específicos demostraron una mejora significativa respecto a los modelos tradicionales. Actualmente, se está continuando este trabajo incrementando la variedad de bases de datos y modelos a evaluar.

1.2. Redes Generativas Adversarias para la Lengua de Señas.

Recientemente, las Redes Generativas Adversarias han mostrado resultados satisfactorios al crear imágenes artificiales a partir de datos aleatorios, o de una imagen para conseguir una transformación de la misma [6]. Actualmente, se está comenzando a investigar cómo funcionan estas redes para generar videos artificiales.

Por otro lado, las bases de datos de Lengua de Señas, dada la naturaleza compleja que poseen, o bien poseen pocos datos o los datos que las forman no presentan la suficiente diversidad. Además, la creación de datos artificiales no es trivial como en otros dominios, lo que hace muy difícil aplicar técnicas clásicas como *Data Augmentation*.

Por tal motivo, en el III LIDI se están comenzando a utilizar Redes Generativas Adversarias (GANs) para generar imágenes y

videos artificiales relacionados con la lengua de señas. Este tipo de redes permitirá aumentar las bases de datos de formas de mano, con el objetivo de complementar desde otro enfoque el estudio de modelos y algoritmos de clasificación para bases de datos con pocos datos etiquetados. Estas investigaciones son llevadas a cabo en marco de una tesis doctoral financiada por la UNLP a través de una beca de postgrado.

1.3. Métricas de Equivarianza

Una de las propiedades deseables en un modelo de clasificación de formas de mano es que sea invariante a la rotación y otras transformaciones afines, debido a que la forma de mano se considera la misma independientemente de la orientación que tenga, su tamaño, y su posición.

En los últimos años, varios modelos fueron propuestos para añadir invarianza a la rotación y otras transformaciones en CNNs [3]. No obstante, no está claro como estos modelos impactan en el aprendizaje usual de los pesos de la red.

Por este motivo, se ha trabajado en definir y evaluar diversas métricas de las propiedades Invarianza y Auto-Equivarianza, casos especiales de la Equivarianza. Estas métricas permiten la caracterización de la forma en que los distintos modelos invariantes o equivariantes codifican dichas propiedades [4]. Se está trabajando en mejorar el desempeño y la usabilidad de las métricas para popularizar su uso entre investigadores del área.

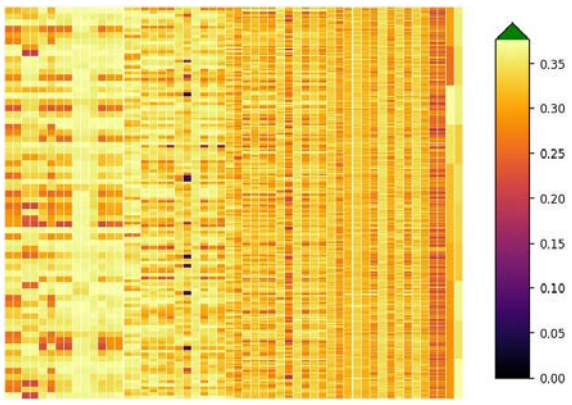


Figura 2. Invarianza por capas y unidades de una CNN con arquitectura ResNet.

1.4. Análisis de Imágenes Astronómicas

En los últimos años, la cantidad de información astronómica disponible se ha multiplicado de forma exponencial. En consecuencia, diversas tareas que anteriormente se realizaban de forma manual o semi-manual deben ahora automatizarse aún más. En ocasiones, también requieren de sistemas inteligentes para realizar tareas cuyos criterios no son objetivos, como la clasificación o identificación de objetos celestes.

En este ámbito, en el III-LIDI se están desarrollando dos proyectos. El primero consiste en determinar modelos de clasificación a partir de información fotométrica de estrellas Be [5]. Estas estrellas son estrellas no-supergigantes cuyo espectro exhibe emisión en línea Hα. La técnica que generalmente se utiliza para detectar estrellas Be es el uso de diagramas fotométricos de dos colores. No obstante, estos diagramas pueden identificar otras fuentes astrofísicas, como por ejemplo estrellas supergigantes o estrellas Wolf-Rayet, entre otras. Por otro lado, como ya se mencionó, la cantidad de información recolectada por los telescopios actuales es enorme, lo que lleva a un tiempo extremadamente alto para analizar de forma manual estos diagramas.

En este sentido, se están implementando modelos clásicos de Aprendizaje Automático que permitan una detección de estrellas Be candidatas, para luego ser evaluadas por un experto [7]. Para esto se está comparando la información proveniente del catálogo público de MohrSmith, el cual se encuentra etiquetado, con otros catálogos no etiquetados. Además, se está investigando qué variables son realmente relevantes para analizar estos cuerpos celestes.

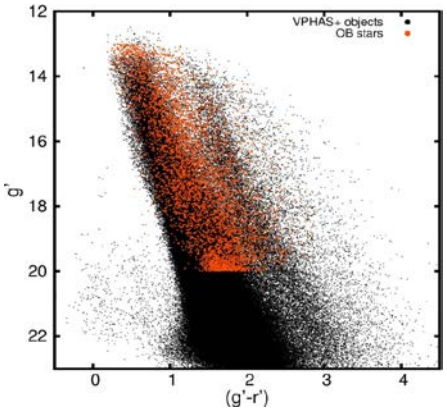


Figura 3. Objetos celestes en la región Carina's Arm.

El segundo proyecto que se está desarrollando en el III-LIDI se refiere al diseño e implementación de un sistema de procesamiento automático de placas espectrográficas astronómicas antiguas. La espectroscopia tiene un uso especial en astronomía ya que se puede usar para averiguar muchas propiedades de estrellas y galaxias distantes. En el caso particular de las placas espectrográficas, se están utilizando técnicas de procesamiento de imágenes inteligente para corregir algunos defectos propios de la tecnología de la época. De esta forma, se podrán recuperar los espectrogramas que contienen información nunca antes analizada ni tampoco capturada, ya que no existían otros telescopios mirando la misma porción del cielo en esas épocas.

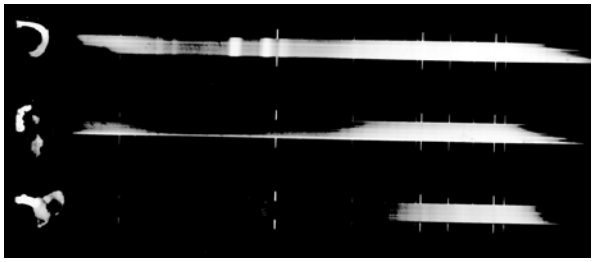


Figura 4. Placa espectrográfica digitalizada.

1.5. Multimedia interactiva con Aprendizaje Profundo

En el marco del proyecto de investigación, desarrollo e innovación "Generación de contenido multimedia mediante IA para entornos interactivos" evaluado y financiado por la Facultad de Informática de la UNLP y continuando la línea de trabajos anteriores de reconocimiento y localización de objetos en video, se está trabajando en el desarrollo de aplicaciones que combinen herramientas de inteligencia artificial como el Aprendizaje Profundo e interacción multimedial.

De esta forma, se busca utilizar la salida de modelos de reconocimiento de acciones y emociones, localización de personas y otros, como entradas a un sistema de generación de imágenes, videos y sonido de forma interactiva.

Para llevar esto a cabo, se están analizando diferentes modelos de generación de texto, audio e imágenes, basados en redes recurrentes y generativas [8][9].

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

- Definición de métricas de invarianza y equivarianza en modelos de redes neuronales
- Análisis de modelos de redes neuronales y sus características en términos de invarianza y auto-equivarianza.
- Representación y clasificación de configuraciones de manos para el lenguaje de señas.
- Generación de imágenes de formas de mano con GANs.
- Modelos de clasificación de imágenes de estrellas.
- Desarrollo de un sistema de procesamiento automático de placas espectrográficas astronómicas antiguas.
- Sistemas interactivos multimediales inteligentes.

3. RESULTADOS OBTENIDOS/ESPERADOS

- Comparación de modelos especializados para bases de datos con pocas muestras para la clasificación de formas de mano.

- Desarrollo de métricas de invarianza y auto-equivarianza para redes neuronales
- Análisis de los mecanismos de las redes neuronales para adquirir invarianza y autoequivarianza a las transformaciones afines.
- Redes generativas para la creación de datos artificiales en la Lengua de Señas.
- Modelo de clasificación de estrellas Be generalizable a distintos conjuntos de datos.
- Sistema de procesamiento automático de placas espectrográficas astronómicas.
- Prototipo de interacción multimedia utilizando IA.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por: 1 profesor con dedicación exclusiva, 1 investigador CIC-PBA, 1 becario de posgrado de la UNLP con dedicación docente, 1 becario CIC, 1 becario CIN, 5 tesistas, 3 profesores extranjeros, y 3 investigadores externos.

Dentro de los temas involucrados en esta línea de investigación, en los últimos dos años se han finalizado 2 tesis de doctorado, 1 tesis de especialización, y 7 tesinas de grado de Licenciatura.

Actualmente se están desarrollando 1 tesis de doctorado, 1 tesis de especialista, 3 tesinas de grado de Licenciatura y 4 trabajos finales de Ingeniería en Computación. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

[1] Quiroga, F., Antonio, R., Ronchetti, R., Lanzarini, L., Rosete, A. A Study of Convolutional Architectures for Handshape Recognition applied to Sign Language, publicado en el XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017) (pp. 13-22).

[2] Cornejo Fandos, U., Rios, G., Ronchetti, F, Quiroga, F., Hasperué, W., Lanzarini, L. Recognizing Handshapes using Small Datasets, publicado en el XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019, Rio Cuarto) (pp. 105-114).

[3] Quiroga F., Ronchetti F., Lanzarini L., Fernandez-Bariviera A. Revisiting Data Augmentation for Rotational Invariance in Convolutional Neural Networks. International Conference on Modeling and Simulation in Engineering, Economics and Management (MS'2018 GIRONA) .

[4] Quiroga, F., Torrents-Barrena, J., Lanzarini, L., & Puig, D. (2019, June). Measuring (in) variances in Convolutional Networks. In Conference on Cloud Computing and Big Data (pp. 98-109). Springer, Cham.

[5] Aidelman Y., Escudero C. Using Machine-Learning to obtain Be star candidates. IX La Plata International School (LAPIS) on Astronomy and Geophysics. Febrero 2020.

[6] Goodfellow I. J., Pouget-Abadie j., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Networks. NIPS'14 Proceedings of the 27th International Conference on Neural Information Processing Systems. v2. pp 2672-2680. 2014.

[7] Jaschek M., Slettebak A., Jaschek C. *Be star terminology*. Be Star Newsletter. 1981.

[8] Vondrick C., Pirsiavash H., Torralba A. *Generating Videos with Scene Dynamics*. 29th Conference on Neural Information Processing Systems (NIPS 2016). 2016.

[9] Zhu Y., Lan Z., Newsam S., Hauptmann A. G.. *Hidden Two-Stream Convolutional Networks for Action Recognition*. Accepted at ACCV 2018. arXiv:1704.00389. 2017.

Analíticas para Sistemas de Atención con Grandes Volúmenes de Eventos

Esteban Schab [1,3], Carlos Casanova[1,3]y Fabiana Piccoli [1 y 2]

[1] Universidad Autónoma de Entre Ríos, Concepción del Uruguay
[2] LIDIC- Univ. Nacional de San Luis, San Luis
[3] Univ. Tecnológica Nacional, Facultad Regional Concepción del Uruguay, Concepción del Uruguay
Argentina
mpiccoli@unsl.edu.ar
{schabe, casanovac}@frcu.utn.edu.ar

RESUMEN

Cualquier proceso de negocio, hoy en día, exige tomar decisiones rápidas, las cuales se deben ir adaptando a los cambios. Una posibilidad es resolver el proceso computacionalmente. Para ello es necesario, en base a los datos, producir analíticas, las cuales buscan transformarlos en conocimiento para la toma de decisiones. Existen varios tipos de analíticas, particularmente en este trabajo proponemos una línea de investigación enfocada en la analítica prescriptiva, como nivel más avanzado, capaz de calcular acciones a ser ejecutadas en el momento (decisiones operativas) o en el futuro (decisiones tácticas para corto y mediano plazo, decisiones estratégicas para largo plazo) para lograr un objetivo deseado. El cálculo de las acciones involucra otros aspectos, uno de ellos derivado de la necesidad de bajos tiempos de respuesta, la computación de alto desempeño, particularmente involucrada en el procesamiento de *datastreams*.

Palabras clave: Inteligencia Computacional. Analíticas. Big Data. Computación de Alto Desempeño.

CONTEXTO

Esta propuesta de trabajo se lleva a cabo dentro de los proyectos de investigación: "Tecnologías Avanzadas aplicadas al Procesamiento de Datos Masivos" (LIDIC, UNSL), "Cómputo de Altas Prestaciones aplicado a la Solución de Grandes Problemas" (UADER) y "Descubrimiento de conocimiento en bases de datos" (GIBD, UTN).

1. INTRODUCCIÓN

La mejora continua y adaptativa de los procesos de negocio resulta clave para mantener la competitividad de las organizaciones. La digitalización de los procesos, así como el incremento en las tecnologías de monitoreo, han llevado a producir una enorme cantidad de datos, los cuales tienen un gran potencial para la mejora de los procesos conducida por analíticas [3] [5] [15].

Las analíticas buscan transformar los datos en conocimiento para la toma de decisiones [4], pudiendo distinguirse cuatro tipos de analíticas según el nivel de automatización del proceso [7]. Primero, la descriptiva intenta responder qué ha pasado o está pasando y la diagnóstica

por qué ha pasado o está pasando. Ambas analizan datos históricos. Luego, la analítica predictiva busca responder qué sucederá. Aplica el conocimiento para predecir nuevos datos sobre el presente o el futuro (pronóstico). Ninguno de estos enfoques sugiere acciones concretas, sino que descansan en el juicio subjetivo y las habilidades analíticas del usuario para deducir acciones de mejora. Finalmente, la analítica prescriptiva responde qué debería hacerse para lograr un objetivo. Determina acciones a ser ejecutadas en el momento (decisiones operativas) o en el futuro (decisiones tácticas para corto y mediano plazo, decisiones estratégicas para largo plazo).

A pesar de los avances tecnológicos en general, las analíticas de procesos existentes dentro de la industria actual no aprovechan completamente el conocimiento oculto en los grandes volúmenes de datos con los que cuentan debido a las siguientes limitaciones[3]:

a. No hacen uso de técnicas prescriptivas para transformar los resultados del análisis en acciones de mejora concretas, dejando este paso completamente a criterio del usuario.
b. Hacen un uso intensivo de datos de sistemas en producción, generando un deterioro en el desempeño de las herramientas de software que soportan los procesos.
c. La optimización es conducida *ex post*, después de completado el proceso, en contraste a la mejora proactiva durante la ejecución del proceso.

Dentro de *Big Data* puede, también, identificarse como área emergente el procesamiento de *datastreams*, o llamado *Data Stream Mining* [11] [17]. Un *datastream* es una representación digital y transmisión continua de datos que describen una clase de eventos relacionada [13]. Su procesamiento permite lograr respuestas en tiempo real a los eventos en forma de toma de decisiones.

En los sistemas de atención, particularmente los *datastreams* abren nuevas y amplias oportunidades para la creación de valor en las organizaciones. Un caso de particular interés son los bancos o cualquier otros sistemas de atención al público [9]. En ellos la generación de *datastreams* es causada por los sistemas de gestión de la atención que implementan. Estos sistemas en general utilizan modelos relacionales de bases de datos y, si bien permiten la elaboración de analíticas, puede resultar inapropiada su implementación debido a que el desempeño del sistema de atención, especialmente en un contexto de *Software as a Service* (SaaS)[16], puede verse colapsado por las continuas consultas necesarias para realizar el monitoreo. Por tal motivo, la generación y el procesamiento de los eventos en forma de *datastreams* como una componente paralela al sistema de atención puede ser la tecnología de base para permitir un monitoreo eficiente sin el deterioro de la performance del propio sistema. En la siguiente sección se describen las características básicas de la línea de investigación a seguir.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

La línea de investigación se basa en la idea de la Optimización de procesos de negocio por recomendación.

Los grandes volúmenes de datos generados por los sistemas de atención pueden ser utilizados en algoritmos de aprendizaje supervisado o no supervisado para la generación de analíticas, particularmente las descriptivas y predictivas. Existe, sin embargo, una dificultad en lo relativo a las analíticas

prescriptivas: no cuentan con un "profesor" que les diga qué acción tomar en cada circunstancia. Un tipo de aprendizaje que no necesita de un profesor es el llamado aprendizaje por refuerzo. En este esquema es el propio agente quien es capaz de juzgar y criticar sus acciones con base en sus percepciones y de alguna medida de aptitud, recompensa o refuerzo. La tarea del aprendizaje por refuerzo es usar recompensas observadas para aprender una política óptima (o aproximadamente óptima) del entorno, sin asumir ningún conocimiento *a priori* [13]. Esta política le dice al agente qué hacer en cada estado posible a alcanzar. Por lo tanto, se juzga al aprendizaje por refuerzo como una herramienta muy útil para la elaboración de analíticas prescriptivas, ya que ambos comparten el mismo objetivo.

Esta línea propone, en consecuencia, la composición de modelos de analítica prescriptiva que superen los inconvenientes descritos. Estos modelos formarían una parte esencial de un proceso de mejora continua basado en la recomendación de acciones operativas y tácticas destinadas a mantener los indicadores de rendimiento del sistema dentro de los valores deseados.

Para la construcción de modelos prescriptivos, cuya principal función resulta en la determinación de las acciones a llevar a cabo, se debe hacer uso de modelo predictivos para explorar los futuros cercanos, modelos descriptivos para calcular la aptitud de dichos estados, y un lenguaje de especificación que permita determinar valores deseados y estrictos de aptitud, es decir, una medida de calidad de sistema. En consecuencia, para complementar el aprendizaje por refuerzo se apela a técnicas provenientes de la inteligencia computacional: redes neuronales como modelos, teoría de conjuntos difusos como lenguaje de especificación, y métodos numéricos y metaheurísticos para entrenamiento de tales modelos [2] [14] [17].

Como las analíticas se encuentran inmersas dentro de un proceso de negocio dinámico, que exige tiempos de respuestas rápidos, y dadas las características propias de cada una de las técnicas de inteligencia computacional antes mencionadas, es mandatorio pensar en la aplicación de modelos/paradigmas de computación de alto desempeño[1][6][10], principalmente en el entrenamiento y ejecución de analíticas. Este proceso debe ser lo suficientemente rápido como para procesar los *datastreams* que el sistema genera de manera continua y brindar resultados en tiempo real.

3. RESULTADOS OBTENIDOS/ESPERADOS

En cuanto a las analíticas, se espera construir un modelo prescriptivo dirigido por los datos que constituya una parte esencial de un proceso de mejora continua. Este modelo estará basado en la recomendación automática y proactiva de acciones operativas y tácticas destinadas a mantener los indicadores de rendimiento de un sistema de atención dentro de los valores deseados, y en un contexto con grandes volúmenes de eventos.

En cuanto al procesamiento, se espera construir y probar algoritmos de procesamiento aplicando computación paralela capaz de reducir los tiempos derivados de entrenar y ejecutar el modelo. Para ello es deseable facilitar el desarrollo de una solución paralelas portable, de costo predecible, capaz de explotar las ventajas de modernos ambientes HPC a través de herramientas y "frameworks de computación" de alto nivel.

4. FORMACIÓN DE RECURSOS HUMANOS

Los resultados esperados respecto a la formación de recursos humanos son el desarrollo de 1 tesis de de doctorado, 2 de maestría y de varias tesinas de grado en las universidades intervinientes.

5. BIBLIOGRAFÍA

[1] Mercedes Barrionuevo, Mariela Lopresti, Natalia Miranda, and María Fabiana Piccoli. "Solving a big-data problem with gpu: the network traffic analysis". Journal of Computer Science and Technology, 15(01): p.30–39, Apr. 2015.

[2] A. Ebrahimnejad and J. L. Verdegay. "Fuzzy sets-based methods and techniques for modern analytics". Springer International Publishing. 2018.

[3] Christoph Gröger, Holger Schwarz, and Bernhard Mitschang. "Prescriptive analytics for recommendation-based business process optimization". In *International Conference on Business Information Systems*, pages 25–37. Springer, 2014.

[4] Clyde Holsapple, Anita Lee-Post, and Ram Pakath. "A unified foundation for business analytics. Decision Support Systems", 64:130{141, 2014.

[5] Mandeep Kaur Saggi and Sushma Jain. "A survey towards an integration of big data analytics to big insights for value-creation". Information Processing & Management, 54(5):758{790, 2018.

[6] S. Kurgalin and S. Borzunov, "A Practical Approach to High-Performance Computing". Springer. 2019.

[7] Michael Minelli, Michele Chambers, and Ambiga Dhiraj. "Big data, big analytics: emerging business intelligence and analytic trends for today's businesses". Volume 578. John Wiley & Sons, 2013.

[8] Albert Bifet and Jesse Read. "Ubiquitous artificial intelligence and dynamic data streams". In Proceedings of the 12th ACM International Conferenceon Distributed and Event-Based Systems, DEBS '18, Pp. 1–6, Association for Computing Machinery. New York, USA, 2018.

[9] T. Olanrewaju. "The rise of the digital "bank.https://www.mckinsey.com/business-functions/digital-mckinsey/ourinsights/the-rise-of-the-digital-bank/. Accessed: 2020.

[10] Peter Pacheco. "An Introduction to Parallel Programming", 1st ed., San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.

[11] Federico Pigni, Gabriele Piccoli, and Richard Watson. "Digital data streams: Creating value from the real-time flow of big data". California Management Review, 58(3):5–25, 2016.

[12] Sergio Ramírez-Gallego, Bartosz Krawczyk, Salvador García, Michal Wozniak, and Francisco Herrera. "A survey on data preprocessing for data stream mining: Current status and future directions". Neurocomputing, 239:39 – 57, 2017.

[13] Stuart J Russell and Peter Norvig. "Inteligencia Artificial: un enfoque moderno". 2004.

[14] Nazmul Siddique and Hojjat Adeli. "Computational intelligence: synergies of fuzzy logic, neural networks and evolutionary computing". John Wiley & Sons, 2013.

[15] Usarat Thirathon, Bernhard Wieder, Zoltan Matolcsy, and Maria-Luise Ossimitz. "Impact of big data analytics

on decision making and performance". In International Conference on Enterprise Systems, Accounting and Logistics, 2017.

[16] M. Turner, D. Budgen and P. Brereton, "Turning software into a service". In Computer, vol. 36, no. 10, pp. 38-44, Oct. 2003.

[17] Lotfi A. Zadeh. "Fuzzy logic, neural networks, and soft computing". Commun. ACM 37, 3 , 77–84. DOI: https://doi.org/10.1145/175247.175255. March 1994.

CIUDADES INTELIGENTES, EFICIENTES Y SOSTENIBLES

Villagra A.[1], Errecalde M.[1,2], Pandolfi D[1]., Mercado V[1], Torres M. [1], Valdéz J. [1], Molina D.[1], Guzmán J. [1], López M. [1]

[1]Laboratorio de Tecnologías Emergentes (LabTEm)
Instituto de Tecnología Aplicada (ITA) - Unidad Académica Caleta Olivia
Universidad Nacional de la Patagonia Austral

[2]Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)
Departamento de Informática - Universidad Nacional de San Luis

avillagra@uaco.unpa.edu.ar, merreca@unsl.edu.ar, {dpandolfi, vmercado}@uaco.unpa.edu.ar, marianagalois@gmail.com, {jcvaldez, dmolina, jguzman, mlopez}@uaco.unpa.edu.ar

RESUMEN

Una ciudad inteligente (CI) y sostenible trata de espacios basados en la eficiencia y el bienestar que utilizan la tecnología como una herramienta fundamental para mejorar las condiciones de vida de la ciudadanía. En este tipo de ciudades se establece una relación interactiva entre los ciudadanos.

Los nuevos retos de la ciudad y las limitaciones, tanto de nuestro entorno, como de los recursos de todo tipo que necesita de forma continuada, en la mayoría de los casos provenientes de fuentes finitas (energéticos, infraestructuras, económicos, etc.), hace que la disposición de éstos, su utilización y empleo de la forma más eficiente posible sea una de las claves si se quiere conseguir niveles adecuados de sostenibilidad.

Una CI es la visión holística de una ciudad que aplica las Tecnologías de la Información y Comunicación (TIC) para la mejora de la calidad de vida y la accesibilidad de sus habitantes y asegura un desarrollo sostenible económico, social y ambiental en mejora permanente.

Esta línea de trabajo se presenta como un aporte en el camino hacia el desarrollo sostenible de una cuidad inteligente. En particular, se propone identificar, resolver y desarrollar prototipos de aplicaciones y servicios en tres ejes movilidad, sociedad y entorno.

Palabras clave: Ciudades Inteligentes, Metaheurísticas, Big Data, Transito.

CONTEXTO

La línea de trabajo se lleva a cabo en el Laboratorio de Tecnologías Emergentes (LabTEm), Instituto de Tecnología Aplicada (ITA) de la Unidad Académica Caleta Olivia Universidad Nacional de la Patagonia Austral, en el marco del Proyecto de Investigación 29/B225 "Soluciones inteligentes para el desarrollo urbano sostenible". Este proyecto se desarrolla en cooperación con el LIDIC de la UNSL, y el Grupo NEO de la UMA (España).

1. INTRODUCCIÓN

Actualmente, aproximadamente el 55% de la población mundial vive en zonas urbanas, y se espera que el número aumente hasta un 66% para 2030 (Naciones Unidas, 2016). A medida que la ciudad crece, surgen nuevos problemas (por ejemplo, congestión del tráfico, gestión de residuos, contaminación, etc.) y los recursos son escasos. Por lo tanto, trabajar en la adaptación de la ciudad a las necesidades actuales (y futuras) es una prioridad de todos y en particular para los investigadores.

El concepto CI surgió hace unos años como una combinación de "ideas sobre cómo las tecnologías de la información y la comunicación podrían mejorar el funcionamiento de las ciudades" [BA+12]. El término de CI sigue ganando popularidad (cada día más ciudades se etiquetan como inteligente), se considera que el concepto sigue siendo un tema en progreso ([MC+14], [CD+11],[Coc14]).

Según [GF+07] una CI es aquella que realiza actividades en al menos una de las seis áreas de acción inteligente: *Smart Economy*, *Smart People*, *Smart Governance*, *Smart Mobility*, *Smart Environment* y *Smart Living*. La definición coincide con las ideas de B. Cohen [Coh12], quien en 2012 propuso una rueda donde se integran los ejes mencionados con otras actividades para su puesta en marcha. A pesar de estos avances en la definición, conocimiento y aparente penetración del concepto de ciudades inteligentes, este dominio apenas está despegando.

Hablar de CI no es solo hablar de tecnología sino en algo más complejo y humano. Es necesario utilizar mejor los recursos públicos y explotar los activos naturales de forma consciente y responsable. Transformar "ciudades tradicionales" en Ciudades Inteligentes, o *Smart Cities* es una demanda cada vez más importante.

A medida que el paradigma de Internet de las cosas (en inglés *Internet of Things - IoT*) se va extendiendo, las tecnologías de la información y la comunicación (TIC) tienen un cometido fundamental que desempeñar para incrementar la eficiencia en todos los sectores industriales y permitir innovaciones tales como los sistemas de transporte inteligentes y la gestión "inteligente" del agua, de la energía y de los residuos, entre otros.

El concepto IoT, por lo tanto, tiene como objetivo hacer que Internet sea aún más inmersivo y omnipresente. Además, al permitir un fácil acceso y la interacción con una amplia variedad de dispositivos tales como, electrodomésticos, cámaras de vigilancia, sensores, actuadores, pantallas, vehículos, etc., la IoT fomenta el desarrollo de una serie de aplicaciones que hacen uso de la cantidad potencialmente enorme y variedad de datos generados por tales objetos para proporcionar nuevos servicios a los ciudadanos, empresas y a las organizaciones con aplicación en diferentes ámbitos ([BC+13],[SD+16], [GS+17] y [Kun14]).

La expansión de *Big Data* y la evolución de las tecnologías de IoT juegan un rol importante en la factibilidad de las iniciativas para ciudades inteligentes. *Big Data* ofrece potencial para que las ciudades obtengan conocimiento de valor de grandes cantidades de información colectada de varias fuentes y la IoT permite la integración de sensores, identificación de radiofrecuencias, y *bluetooth* en el entorno del mundo real utilizando servicios en red. La combinación del IoT y Big Data es un área de investigación poco explorada que ha traído nuevos e interesantes retos para alcanzar el objetivo de futuras ciudades inteligentes [RA+16], [BD+16], [JJ+15]. El fenómeno de Big Data desde hace tiempo se ha caracterizado por el volumen, la velocidad y una variedad de tipos de datos que se incrementa día a día. Big Data ofrece el potencial para que la ciudad obtenga información valiosa de una cantidad considerable de datos recogidos a través de diversas fuentes.

Crear aplicaciones inteligentes en una ciudad inteligente es uno de los mayores desafíos a los que podemos enfrentarnos. Los problemas en las ciudades inteligentes son variados (economía, movilidad, gobernanza, personas, vida y entorno), multidisciplinares por naturaleza, aparentemente inconexos y requieren un vasto aparato científico. Los desafíos son varios e involucran la combinación de técnicas de optimización, aprendizaje de máquina y análisis de datos para obtener soluciones usables y perdurables en ciudades inteligentes y sostenibles.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En esta sección se describen las líneas de investigación que se llevan a cabo en el proyecto.

Como se comentó anteriormente una CI es la que realiza actividades en al menos una de las seis áreas de acción inteligente: Economía inteligente, Sociedad inteligente, Gobernanza inteligente, Movilidad inteligente, Medioambiente inteligente, Modo de vida inteligente. Algunas aplicaciones que se pueden definir en estas áreas y que son de interés en este trabajo de investigación son: movilidad, sociedad y entorno.

Con respecto a la movilidad inteligente, estamos trabajando en problemas de tráfico y en particular en el control de ciclos semafóricos.

Actualmente, una ciudad con tráfico congestionado y muchos embotellamientos conduce a más contaminación. Por lo tanto, con la gestión inteligente del tráfico, el consumo de combustible y las emisiones contaminantes pueden ser reducidas. A diferencia de técnicas tales como semáforos inteligentes que requieren nueva infraestructura, ubicación de sensores y modificaciones en obras civiles ya construidas, el uso de técnicas de inteligencia artificial en la optimización de los ciclos de semáforos se presenta como una herramienta viable, rápida, eficiente y de bajo costo [GF+14]. En esta línea nos centramos en dos versiones de un Algoritmo Genético Celular (cGA)[AB09], aplicado al problema de planificación de los ciclos de los semáforos. Las soluciones obtenidas por los cGAs son simuladas por el popular micro-simulador SUMO (*Simulator of Urban Mobility*) [KB+06]. Se evalúan grandes escenarios cercanos a la realidad de vastas áreas urbanas.

En cuanto a sociedad inteligente, se está trabajando en temas de *Big Data* y Minería de Datos en análisis de autoría (AA) [Sta09], en particular la determinación del perfil del autor (DPA), es decir, aquella que identifica patrones compartidos por un grupo de gente y que aborda problemas de clasificación de los usuarios de la Web de acuerdo a la edad, género, orientación política, etc. La DPA, un sub-campo del área más general conocida como análisis de autoría (AA), es un tema muy importante de investigación por sus potenciales (y actuales) aplicaciones en problemas de seguridad nacional e inteligencia, lingüística forense, análisis de mercados e identificación de rasgos de personalidad, entre otros. Particularmente, se ha tomado como caso de estudio primario los documentos de periodistas con diversas orientaciones políticas (oficialista vs opositor) con el objetivo de realizar con los mismos el Análisis de Autor y la Determinación/Caracterización del perfil del autor.

Por otro lado, en cuanto al entorno inteligente el trabajo realizado en este proyecto abrió una línea de investigación y este año se ha presentado un nuevo proyecto de investigación abocado por completo a la recolección de residuos tecnológicos. Además, una nueva línea que se está comenzando a trabajar es con respecto a realidad aumentada. La realidad aumentada es una tendencia tecnológica de gran interés dentro del desarrollo tecnológico, porque se basa en la interacción con el entorno del usuario. La podemos ver como un mundo de representaciones generadas en tiempo real por un programa informático. Esta realidad se puede percibir únicamente dentro de los diferentes sistemas que la hacen posible, es decir, puede generar sensaciones desde una perspectiva física pero sólo es una simulación que se encuentra activa frente al usuario, depende de niveles de inmersión que permite al usuario interactuar con los diferentes objetos generados por el mundo virtual. Generalmente esta realidad busca activar los sentidos humanos, pero por el momento y por la tecnología que se tiene y sus costos, únicamente se centra en los sentidos de vista y oído [HF+11], [AA17], [CB+17].

3. RESULTADOS OBTENIDOS/ESPERADOS

La solución de problemas a través de los ejes de desarrollo en ciudades inteligentes permitirá hacer prototipos reales, aprender más y

producir un mayor impacto en Ciencia e Industria. Dada la amplia visión se espera generar oportunidades para nuevos proyectos y socios.

En particular, en cuanto a tránsito hemos abordado el problema de programación de semáforos, para una serie de objetivos. El primer objetivo proponer un algoritmo competitivo basado en el cGA aplicado para resolver escenarios cercanos a la realidad. El segundo objetivo realizar un análisis en profundidad de las soluciones obtenidas para determinar cómo afectan la disminución de la contaminación ambiental y detectar patrones y características. Se utilizan dos versiones de cGA (síncrono y asíncrono) para resolver instancias grandes y reales. Nuestros algoritmos superan las técnicas de vanguardia y las configuraciones de expertos. Se llevaron a cabo varios análisis en profundidad de los resultados, estudio genotípico y fenotípico (para mayor detalle ver [VA+20]). La investigación futura se centrará en el número de evaluaciones con el objetivo de reducir el esfuerzo computacional en términos de tiempo de procesamiento. Además, planeamos definir una función de aptitud más precisa que explore el espacio de búsqueda de manera más eficiente.

Sobre la determinación/caracterización del perfil del autor se ha generado un nuevo corpus de la alineación política de periodistas argentinos, con un análisis exhaustivo que incluyó estudios estadísticos, modelado de temas y análisis de sentimientos, y una comparación de textos basados en categorías el sistema *Linguistic Inquiry and Word Count* (LIWC). Como resultado de este análisis, se han detectado algunos patrones interesantes en la distribución de polaridad, qué temas son los que hablan y, sus valores para muchas categorías del sistema LIWC (para mayor detalle ver [MV+19]). Como trabajo futuro, se plantea usar los diferentes tipos de información obtenida en la representación de documentos supervisados (clasificación) y tareas no supervisadas (agrupamiento) utilizando características basadas en LIWC y en el método *Latent Dirichlet Allocation* (LDA). Además, se los comparará con enfoques clásicos (bolsa de palabras) y más recientes como redes neuronales profundas con incrustaciones de palabras.

En cuanto a realidad aumentada se pretende desarrollar aplicaciones orientadas en particular a mejorar la accesibilidad y la información para los nuevos miembros de la comunidad universitaria de la Universidad Nacional de la Patagonia Austral. Se identificarán las necesidades y posteriormente a través de esta tecnología (realidad aumentada y sus diversas herramientas) se generan soluciones que las satisfagan.

Finalmente, se pretende colaborar con los gobiernos locales para la implementación de políticas y acciones inteligentes y sostenibles que impacten en la calidad de vida de los ciudadanos.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo se encuentra formado por tres Doctores en Ciencias de la Computación, dos Magisters en Ciencias de la Computación, un Magister en Matemática Avanzada, cinco Ingenieros en Sistemas y tres estudiantes de la Carrera Ingeniería en Sistemas de la UNPA.

Este proyecto de investigación proporcionará un marco propicio para la iniciación y/o finalización de estudios de posgrado de los integrantes docentes. De igual forma, será un ámbito adecuado para la realización de tesis de grado. Actualmente dos integrantes están desarrollando sus tesis de Maestría y dos sus tesis de Doctorado. Además, se cuenta con dos becarios de grado.

5. BIBLIOGRAFIA

[AA17] Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. Educational Research Review, 20, 1-11.

[AB09] Alba, E., & Dorronsoro, B. (2009). Cellular genetic algorithms (Vol. 42). Springer Science & Business Media.

[BA+12] Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., ... & Portugali, Y. (2012). Smart cities of the future. The European Physical Journal Special Topics, 214(1), 481-518.

[BC+13] Bellavista, P., Cardone, G., Corradi, A., & Foschini, L. (2013). Convergence of MANET and WSN in IoT urban scenarios. IEEE Sensors Journal, 13(10), 3558-3567.

[BD+16] Botta, A., De Donato, W., Persico, V., & Pescapé, A. (2016). Integration of cloud computing and internet of things: a survey. Future Generation Computer Systems, 56, 684-700.

[CD+11] Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). Smart cities in Europe. Journal of urban technology, 18(2), 65-82.

[CB+17] Chatzopoulos, D., Bermejo, C., Huang, Z., & Hui, P. (2017). Mobile augmented reality survey: From where we are to where we go. Ieee Access, 5, 6917-6950.

[Coc14] Cocchia, A. (2014). Smart and digital city: A systematic literature review. In Smart city (pp. 13-43). Springer, Cham.

[Coh12] Cohen Boyd, What Exactly Is A Smart City? http://www.fastcoexist.com/1680538/what-exactly-is-a-smart-city, 2012

[GF+07], Giffinguer, R.; Fertner, C; Kramar, H.; Meijers, E. (2007) Ranking of European Medium-Size Cities. Centre of Regional Science, Univ. Tecnológica de Viena.

[GF+14] Garcia-Nieto, J., Ferrer J., & Alba E. (2014). Optimising traffic lights with metaheuristics: Reduction of car emissions and consumption. In Neural Networks (IJCNN), 2014 International Joint Conference on, pages 48-54. IEEE, 2014.

[GS+17] Gharaibeh, A., Salahuddin, M. A., Hussini, S. J., Khreishah, A., Khalil, I., Guizani, M., & Al-Fuqaha, A. (2017). Smart Cities: A Survey on Data Management, Security and Enabling Technologies. IEEE Communications Surveys & Tutorials. 775-780). IEEE.

[HF+11] Hugues, O., Fuchs, P., & Nannipieri, O. (2011). New augmented reality taxonomy: Technologies and features of augmented environment. In Handbook of augmented reality (pp. 47-63). Springer, New York, NY.

[JJ+15] Jaradat, M., Jarrah, M., Bousselham, A., Jararweh, Y., & Al-Ayyoub, M. (2015). The internet of energy: Smart sensor networks and big data management for smart grid. Procedia Computer Science, 56, 592-597.

[KB+06] Krajzewicz, D., Bonert, M., & Wagner, P. (2006). The open source traffic simulation package SUMO. RoboCup 2006 Infrastructure Simulation Competition, 1, 1-5.

[Kun14] Kunzmann, K. R. (2014). Smart cities: a new paradigm of urban development. Crios, 4(1), 9-20.

[MC+14] Manville, C., Cochrane, G., Cave, J., Millard, J., Pederson, J. K., Thaarup, R. K., ... & Kotterink, B. (2014). Mapping smart cities in the EU.

[MV+19] Mercado, V., Villagra, A., & Errecalde, M. L. (2019). Exploratory Analysis of a New Corpus for Political Alignment Identification of Argentinian Journalists. In XXV Congreso Argentino de Ciencias de la Computación (CACIC)(Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019

[RA+16] Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016). Urban planning and building smart cities based on the internet of things using big data analytics. Computer Networks, 101, 63-80.

[SD+16] Sarasola, B., Doerner, K. F., Schmid, V., & Alba, E. (2016). Variable neighborhood search for the stochastic and dynamic vehicle routing problem. Annals of Operations Research, 236(2), 425-461.

[Sta09] Stamatatos, E. (2009). A survey of modern authorship attribution methods. Journal of the American Society for information Science and Technology, 60(3), 538-556.

[VA+20] Villagra, A., Alba, E., & Luque, G. (2020). A better understanding on traffic light scheduling: New cellular GAs and new in-depth analysis of solutions. Journal of Computational Science, 41, 101085.

Inteligencia Colectiva para una Recuperación Sostenible de Residuos Tecnológicos

D. Pandolfi[1], J. Rasjido[1], A. Villagra[1], S. Orozco[1], D. Perez[1], V. Varas, M. Bilbao, G. Leguizamon[12]
{dpandolfi, jrasjido,avillagra,sorozco,dperez, vvaras,mbilbao}@uaco.unpa.edu.ar, legui@unsl.edu.ar

[1] Laboratorio de Tecnologías Emergentes (LabTEm)
Instituto de Tecnología Aplicada (ITA) - Unidad Académica Caleta Olivia
Universidad Nacional de la Patagonia Austral

[2] Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)
Departamento de Informática - Universidad Nacional de San Luis

Resumen

La creciente demanda en el uso las tecnologías de la información y comunicación (TIC) han revolucionado nuestra vida cotidiana, sin embargo, han contribuido a aumentar las preocupaciones ambientales y sociales. La aceleración de factores como el agotamiento y escasez de recursos, daños ambientales, uso de agua y energía, y un incremento insostenible de desechos tecnológicos en zonas urbanas, se ha convertido en una preocupación para el gobierno. El sostenimiento del ritmo de consumo depende de estrategias que permitan el reciclaje de los desechos y el fortalecimiento de una economía circular realizando un cambio sistémico, en la investigación y la innovación, tanto tecnológica como logística

Este trabajo consiste en generar rutas recolección dado una cantidad de clientes por atender, un conjunto de vehículos de recogida, permitiendo minimizar ciertos factores que ayuden a la empresa a obtener beneficios. La programación de la logística de recolección de residuos tecnológicos corresponde al problema de enrutamiento de vehículos (*Vehicle Routing Problem*, VRP) que presentan una extensa disponibilidad bibliográfica de los problemas de enrutamiento verde de vehículos (GVRP).

En particular, se propone identificar, resolver y realizar un prototipo de planificación eficiente para este problema aplicando principalmente técnicas de inteligencia colectiva y otras inspiradas en computación evolutiva.

Palabras clave: Ciudades Inteligentes, Residuos Tecnológicos, Problema de enrutamiento de vehículos, Inteligencia Computacional.

1. CONTEXTO

La línea de trabajo se lleva a cabo en el Laboratorio de Tecnologías Emergentes (LabTEm), Instituto de Tecnología Aplicada (ITA) de la Unidad Académica Caleta Olivia Universidad Nacional de la Patagonia Austral, en el marco del Proyecto de Investigación 29/B252 "Rutas inteligentes en la recolección de residuos tecnológicos". Este proyecto se desarrolla en cooperación con el LIDIC de la UNSL, y el Grupo NEO de la UMA (España).

2. RUTAS INTELIGENTES PARA RECOLECCION DE WEEE

Una ciudad inteligente es una ciudad donde el desempeño de sus componentes fundamentales (economía, movilidad, medioambiente, salud, gobernanza, etc.), están construidas sobre la base de nuevas capacidades donde

sus ciudadanos pueden afrontar de manera inteligente los desafíos futuros de sostenibilidad urbana. El crecimiento demográfico urbano constante de los próximos años hace prever que la sociedad y los gobiernos locales se enfrentarán a nuevos retos ante el constante incremento de la generación de residuos tecnológicos.

La creciente demanda en el uso las tecnologías de la información y comunicación (TIC) han revolucionado nuestra vida cotidiana, sin embargo, han contribuido a aumentar las preocupaciones ambientales y sociales. La aceleración de factores como el agotamiento y escasez de recursos, daños ambientales, uso de agua y energía y un incremento insostenible de desechos tecnológicos principalmente en zonas urbanas, se ha convertido en una preocupación para la industria y el gobierno. El sostenimiento del ritmo de consumo a mediano y largo plazo depende de estrategias que permitan el reciclaje de los desechos y el fortalecimiento de una economía circular que realice un cambio sistémico, en la investigación y la innovación, tanto tecnológica como logística. La recuperación de residuos tecnológicos requiere de estrategias inteligentes de recolección que minimicen los costos asociados, maximicen el reuso de materias primas y la reparabilidad de los bienes para la prolongación de su vida útil.

Los equipos eléctricos y electrónicos (EEE) se convierten en uno de los grupos más importantes de residuos que contienen materiales fáciles de reciclar. Los desechos electrónicos son conocidos como RAEE (en español, Residuos de Aparatos Eléctricos y Electrónicos), o *e-waste*, o bien WEEE (en inglés, *Waste Electrical and Electronic Equipment*). Los WEEE, contienen muchas sustancias que son tóxicas y potencialmente peligrosas para el medio ambiente y la salud humana (Oguchi et al., 2012). La reducción, reciclaje y eliminación de los desechos finales de los WEEE se ha convertido en uno de los principales objetivos para el desarrollo sostenible de la creciente industria tecnológica. La tasa de WEEE está creciendo a un ritmo alarmante, especialmente en los países desarrollados, donde los mercados están saturados con grandes cantidades de nuevos productos electrónicos (Widmer et al., 2005). En 2013, un informe de StEPI (*Solving the e-waste Problem Initiative*), una iniciativa financiada por Naciones Unidas alertó que la cifra de WEEE se había elevado hasta cerca de los 49 millones de toneladas, un promedio de 7 kilos por cada uno de los 7.000 millones de habitantes del planeta Tierra. A la luz de estos datos, no resulta muy difícil hacerse una idea de la importancia que adquiere el reciclaje de este tipo de residuos.

Las empresas de recolección de residuos aplican diversos métodos de recolección, y pueden dividirse en estacionarios y móviles (Baxter et al., 2016). El objetivo tradicional de la gestión de distribución o recolección de productos consiste de minimizar los costos de todo el sistema. *Green Logistics* (GL) se ha convertido en una tendencia en la gestión de la distribución de bienes y la recolección de productos al final de su vida útil. Con su enfoque en maximizar el valor económico y ambiental mediante el reciclaje y el control de emisiones, GL contribuye al desarrollo sostenible de la industria, pero también requiere un esquema de transporte más completo cuando se realizan servicios de logística (Lin et al. 2014 (a) (b)).

Boussaïd et al. 2013, desarrollan un estudio de varios algoritmos de optimización inspirados en las metáforas del comportamiento de enjambre en la naturaleza, conocida como Inteligencia Colectiva, (SI, *Swarm Intelligent*). Ejemplo de SI son Optimización de Colonias de Hormigas (ACO, *Ant Colony Optimization*), la Optimización de Cúmulo de Partículas (PSO, *Particle Swarm Optimization*), la Optimización de Forrajeo Bacteriano (BFOA, *Bacterial Foraging Optimization Algorithm*), Optimización de Colonias de Abejas (BCO, *Bee Colony Optimization*), Sistemas Inmunes Artificiales (AIS, *Artificial Immune Systems*) y Optimización Basada en Biogeografía (BBO, *Biogeography-Based Optimization*). En particular, ACO fue propuesta

fue propuesta por Dorigo et al (1996, 1991) como una metaheurística inspirada en el comportamiento de una colonia de hormiga en el proceso de forrajeo y aplicada principalmente para la solución de problemas de optimización combinatoria.

Las Metaheurísticas (MHs) son métodos que integran de diversas maneras, procedimientos de mejora local y estrategias de alto nivel para crear un proceso capaz de escapar de óptimos locales y realizar una búsqueda robusta en el espacio de búsqueda. En su evolución, estos métodos han incorporado diferentes estrategias para evitar la convergencia a óptimos locales, especialmente en espacios de búsqueda complejos. En otras palabras, las MHs proveen de un marco general que permite crear nuevos híbridos a través de la combinación de conceptos derivados de: heurísticas clásicas, inteligencia artificial, evolución biológica, sistemas naturales, mecánica estadística, etc.

Los cGAs son una subclase de metaheurísiticas con una población estructurada espacialmente, es decir, los individuos de la población pueden aparearse solo con sus vecinos. Los cGAs, se diseñaron inicialmente para trabajar en máquinas paralelas, formadas por muchos procesadores que ejecutaban simultaneamente las mismas instrucciones sobre diferentes datos (maquinas SIMD - El primer modelo de cGA fue propuesto por Robertson en 1987 (Alba y Dorronsoro 2009) y fueron inicialmente desarrollados para trabajar en máquinas paralelas, y más tarde se fueron adaptando para funcionar también en máquinas de procesamiento secuencial.

El problema de recolección estacionaria se lleva a cabo en puntos de recolección específicos de desechos, ubicados en lugares populares o en los alrededores de los lugares de residencia. La elección de la ubicación de dichos centros y la elección de un recorrido óptimo representa distintos problemas de decisión. Así mismo, la recolección móvil, propone nuevos medios de recolección de WEEE a pedido y principalmente está asociado a residuos tecnológicos medianos o grandes (Nowakowski et al., 2017).

3. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta línea de trabajo se enfoca en la generación de rutas recolección que permitan minimizar ciertos factores económicos y ambientales y maximizar los beneficios referidos a al reciclaje y tratamiento de WEEE. De los objetivos económicos podemos mencionar: minimizar el tiempo de recolección, kilómetros recorridos, maximizar el ahorro de combustible en los vehículos, minimizar la cantidad de vehículos, todo lo cual llevaría a obtener menores costos. Un objetivo económico adicional está dado por la valoración de la recuperación las materias primas tales como material ferroso, cobre, y metales preciosos, etc. De los beneficios ambientales podemos mencionar: la disminución en la emisión de gases efecto invernadero (CO_2, CH_4 y N_2O) y los gases de contaminación ambiental (CO, SO_2, PM, NOx). Tanto los objetivos económicos como ambientales promueven una mejor calidad de servicio e imagen, para las empresas y los gobiernos locales.

4. RESULTADOS OBTENIDOS/ESPERADOS

Como objetivo general se espera modelar, resolver e implementar un amplio y variado conjunto de servicios inteligentes de la ciudad y producir un mayor impacto en Ciencia e Industria.

Como objetivos específicos se pretende: (a) Estudiar el estado de tecnologías y técnicas de Inteligencia Computacional para resolver problemas de VRP. (b) Proponer instancias de problemas relacionados a recolección de WEEE. (c) Proponer y seleccionar elementos de medición relacionados a los problemas identificados. (d) Diseñar y construir prototipos de software relacionados a los problemas.

Actualmente se han generado cuatro instancias de pruebas para para 75, 105, 150 y 210 locaciones de recogida de WEEE de dos localidades. Las instancias se han probado con dos algoritmos de inteligencia computacional (ACO y cGA) los resultados preliminares fueron publicados en Pandolfi et al. 2019.

Para elegir los vehículos más adecuados para la colección, analizamos las siguientes variables: total de km recorridos, Costo total de uso vehículo, Total de Combustible consumido, Total de CO2 emitido y Tiempo Total de horas de Trabajo.

En general, con las instancias iniciales y teniendo en cuenta la cantidad de km recorridos y la cantidad de horas de trabajo, los resultados indican que los vehículos de alta carga con un mayor proceso de automatización tienen un mejor rendimiento. Para la variable de consumo de combustible y, por lo tanto, la variable de emisión de CO2, los vehículos más pequeños sin automatización han mostrado mejores resultados. Finalmente, solo para la variable de costo del vehículo, el vehículo con una capacidad de 22 m3 de carga es la mejor opción.

En trabajos futuros analizaremos en distintas ciudades y se propondrán nuevas funciones de optimización teniendo en cuenta las variables consideradas en este estudio.

5. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo se encuentra formado por cuatro investigadores con distintos niveles de posgrado, tres Doctores en Ciencias de la Computación, dos Magíster en Ciencias de la Computación, tres Ingenieros en Sistemas y un estudiante de la Carrera Ingeniería en Sistemas.

Esta línea de investigación proporcionará un marco propicio para la iniciación y/o finalización de estudios de posgrado de los integrantes docentes. De igual forma, será un ámbito adecuado para la realización de tesis de grado. En ese sentido, dos integrantes de este proyecto de investigación está desarrollando su Tesis de Maestría en temáticas afines y un integrante está desarrollando su Tesis de doctorado. Además, se cuenta con un becario alumno de la carrera de Ingeniería en Sistemas.

BIBLIOGRAFÍA

ALBA, E., y DORRONSORO, B. (2009). Cellular genetic algorithms (Vol. 42). Springer Science & Business Media.

J. BAXTER, K.A., LYNG, C. ASKHAM, O.J. HANSSEN (2016). "High-quality collection and disposal of WEEE: environmental impacts and resultant" issues. Waste Manage. 57, 17–26.

I BOUSSAÏD, J. LEPAGNOT, P. SIARRY, A survey on optimization metaheuristics, Information Sciences, Volume 237, 2013, Pages 82-117, ISSN 0020-0255.

M. DORIGO, V. MANIEZZO, A. COLORNI, (1991) Positive Feedback as a Search Strategy, Technical Report 91-016, Dipartimento di Elettronica, Politecnico di Milano, Milan, Italy, 1991.

M. DORIGO, V. MANIEZZO, A. COLORNI, (1996) The ant system: optimization by a colony of cooperating agents, IEEE Transactions on Systems, Man, and Cybernetics Part B26 29–41.

C. LIN, K.L. CHOY, G.T.S. HO, T.W. Ng (a), A Genetic Algorithm-based optimization model for supporting green transportation operations, Expert Systems with Applications, Volume 41, Issue 7, 2014, Pages 3284-3296, ISSN 0957-4174.

C. LIN, K.L. CHOY, G.T.S. HO, S.H. CHUNG, H.Y. LAM (b), Survey of Green Vehicle Routing Problem: Past and future trends, Expert Systems with Applications, Volume 41, Issue 4, Part 1, 2014, Pages 1118-1138, ISSN 0957-4174.

P. NOWAKOWSKI, A. KRÓL, B. MRÓWCZNSKA (2017), Supporting mobile WEEE collection on demand: A method for multi-criteria vehicle routing, loading and cost optimisation, Waste Management, Volume 69, 2017, Pages 377-392, ISSN 0956-053X.

M. OGUCHI, H. SAKANAKURA, A. TERAZONO (2012). Toxic metals in WEEE: characterization and substance flow analysis in waste treatment processes. Sci. Total Environ.

D. PANDOLFI, J. RASJIDO, A. VILLAGRA, G. LEGUIZAMON (2019), Following the green footprint of technological waste: A smart and sustainable collection, RPIC - Reunión de Trabajo en Procesamiento de la Información y Control. ISBN 9789871648.

R. WIDMER, H. OSWALD-KRAPF, D. SINHA-KHETRIWAL, M., SCHNELLMANN, H., BÖNI (2005). Global perspectives on ewaste. Environmental Impact Assessment Review 25 (5), 436–458.

Minería de texto y Deep Learning aplicados a determinar la pertenencia de las consultas realizadas a un metabuscador a cada área temática dentro de las ciencias de la computación

Kuna H.D[1]., Rambo A.R.[1], Canteros A.[1], Rey, M.[1], Zamudio, E.[1], Martini, E.[1], Pautsch, G. [1], Biale, C. [1], Krujoski, S. [1], Rauber, F. [1]

[1]Depto. de Informática, Facultad de Ciencias Exactas Químicas y Naturales (FCEQyN), Universidad Nacional de Misiones (UNaM) Posadas, Misiones 3300/Argentina.

hdkuna@gmail.com

RESUMEN

Al trabajar en un metabuscador que permita identificar cuáles son las mejores recomendaciones en un área temática específicas según las consultas ingresadas por el usuario existen diversos problemas que abordar. En el presente trabajo se realiza el análisis de la pertenencia de una consulta escrita en lenguaje coloquial a una o varias áreas de temáticas dentro de las ciencias de la computación. Para lo cual se podrán tomar varios criterios y se realiza un relevamiento del estado de arte específico con la finalidad de determinar las mejores estrategias tanto para el tratamiento del texto como para la clasificación de los artículos. En este trabajo se presenta el relevamiento realizado para abordar y resolver esta problemática.

Palabras clave: *minería de texto, metabuscadores, recuperación de información, pertenencia a áreas temáticas, recomendación.*

CONTEXTO

Esta línea de investigación articula el Programa de Investigación en Computación (PICom) de la Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones (FCEQyN/UNaM) con el grupo de investigación Soft Management of Internet and Learning (SMILe) de la Universidad de Castilla-La Mancha, España, y con el Departamento de Matemáticas de la Universidad de Sonora, México.

1 INTRODUCCION

Dentro de las actividades científicas se encuentra la búsqueda de antecedentes y datos científico-tecnológicos que permitan determinar el estado del arte en el área temática bajo análisis. En la actualidad no solo el volumen de información generada es un problema si no también contar con herramientas que permitan filtrar y cruzar si no todas al menos la gran parte de fuentes confiables teniendo en cuenta estándares predefinidos y consensuados por el cuerpo científico. Entre las diferentes herramientas que ayudan al investigador a recuperar datos de calidad y relevantes se pueden mencionar: los buscadores científicos, los repositorios digitales, redes sociales del ambiente y soluciones de procesamiento de grandes bases de datos (Tang, 2016), (Ortega y Aguillo, 2014), (Lee et. al., 2006).

En el presente proyecto de investigación acreditado “Diseño y construcción de procesos de explotación de información para el área de las ciencias de la computación”, aprobada por RES RECTORAL 601/18 y RES CD 274/18 se han abordado diferentes problemáticas como el análisis sobre los nombres de los autores permitiendo la desambiguación de los mismos (Kuna et. al. 2019). La búsqueda y definición de expertos tema de una tesis de grado (Cantero, 2019), la evaluación y recomendación de autores definiendo diferentes métricas tema de una tesis de grado (Canteros et. al., 2018b), (Canteros A., 2020). Para la prueba e integración de estas propuestas elaboradas y siguiendo la línea de

investigación se han abordado algunos de los problemas mencionados y se ha desarrollado un Sistema de Recuperación de Información (SRI), concretamente un metabuscador, que opera sobre documentos científicos del área de Ciencias de la Computación (Kuna et. al., 2013), (Kuna et. al., 2014). En el desarrollo del SRI se han planteado diferentes avances en torno a la conformación de una base de datos (BD) interna para el metabuscador que permitiera la generación de soluciones como las mencionadas en el apartado anterior (Kuna et. al, 2017), (Canteros et. al., 2018b).

2 LÍNEAS DE INVESTIGACION, DESARROLLO E INNOVACIÓN

Si bien se ha trabajado y logrado avances en varias cuestiones sobre el metabuscador, la presente línea de investigación propone abordar otra de las problemáticas detectadas al momento de generar las búsquedas hacia la estructura del SRI, el desarrollo de una técnica para determinar la pertinencia de la búsqueda a determinadas áreas específicas, lo que sería otorgar un nivel de simplicidad y especificidad para el metabuscador.

Sobre esta línea de trabajo se detectan a su vez que dado el conjunto de datos con los que contamos en la estructura propuesta, lo que nos sería de utilidad plantear para la presente implementación sería por un lado definir el método que permita determinar a qué área temática dentro de las ciencias de la computación, según la taxonomía de la *Association for Computing Machinery* - ACM (ACM 2008), (ACM, 2012), pertenece cada consulta realizada, replicar similar criterio para los autores y para los artículos, tomando como referencia previa los artículos analizados y guardados.

3 RESULTADOS Y OBJETIVOS

En el presente proyecto se prevé tomar el almacenamiento, del metabuscador desarrollado, las consultas y los perfiles de los autores en la base de datos según se observa en la fig. 1. Los datos de consultas almacenados en formato JavaScript Object Notation, (notación de objeto de JavaScript) JSON y ejecutadas con el metabuscador, los datos de perfiles de autores almacenados en la base de datos, los artículos devueltos y los resultados seleccionados por los usuarios.

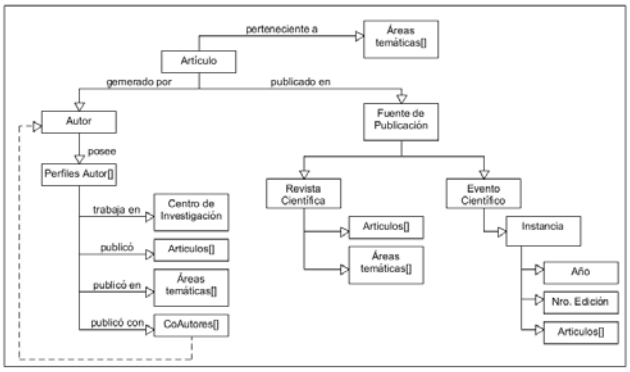


Figura 1: Esquema de artículos, autores y fuentes de publicación en el metabuscador

Uno de los temas principales en la minería de texto es la representación de texto, que es fundamental e indispensable para el procesamiento de información inteligente basado en texto. En general, la representación de texto incluye dos tareas: indexación y ponderación, por lo tanto, el primer abordaje es definir el criterio para la representación de las frases a ser analizadas. Los espacios vectoriales semánticos los cuales se basan en la idea de que el significado de una palabra puede ser aprendido de un entorno lingüístico, el de las áreas de la ciencias de la computación en nuestro caso, y poseen dos enfoques, la semántica distribucional y la semántica composicional (Torres López & Arco García, 2016).

Se ha estudiado comparativamente TF-IDF *(term frequency inverse document frequency)* y LSI *(latent semantic indexing)* como representación de varias palabras para texto. Donde se ha demostrado que, en la categorización de texto, LSI tiene un mejor rendimiento que otros métodos. Donde se observa que LSI tiene una calidad semántica y estadística favorable y es diferente con la afirmación de que LSI no puede producir poder discriminatorio para la indexación (Zhang et. al, 2011). Una aplicación de este mismo esquema de representación se observa en (Xamena et.al., 2019) donde se utiliza bolsa de palabras básica y una representación TF-IDF para los documentos históricos y para la tarea de modelado de temas, se empleó el método *LDA (Linear discriminant analysis)*.

El modelo Fast Text propuesto por (Bojanowski et. al., 2017) donde según este enfoque, basado en el modelo de skipgram, cada palabra se representa como una bolsa de caracteres n-gramas. Una representación vectorial está asociada a cada carácter n-gramo; palabras representadas como la suma de estas representaciones. El método denominado Fast Text demuestra ser rápido y permite calcular representaciones de palabras para palabras que no aparecieron en los datos de entrenamiento.

Otros trabajos (Mikolov et. al., 2013a) presentan varias extensiones que mejoran la calidad de los vectores y la velocidad de entrenamiento para el modelo continuo de Skip-gram (Mikolov et. al, 2013b). Por submuestreo de las palabras frecuentes obtienen una aceleración significativa y también se logra que aprenda representaciones de palabras más regulares. También describen una alternativa simple al softmax jerárquico llamado muestreo negativo donde logran buenas representaciones para los vectores de aprendizaje para millones de frases.

De esta manera se obtendrán representaciones de las secuencias de texto, para alimentar esto a arquitecturas Deep Learning de Clasificación de Textos que actualmente se encuentran en revisión. Como es la propuesta del estudio que toma como base la Wikipedia[1], teniendo en cuenta que este es un gran volumen de datos y proponen la creación del corpus que define una categoría mediante técnicas de Procesado de Lenguaje Natural (PLN) que analizan sintácticamente los textos a clasificar. El resultado final del sistema propuesto presenta un alto porcentaje de acierto, incluso cuando se compara con los resultados obtenidos con técnicas alternativas de Aprendizaje Automático (Quinteiro-González et. al, 2011).

(Zhang et. al, 2015), como el trabajo donde presentan una nueva arquitectura (*Very Deep Convolutional Neuronal Networks* - VDCNN) para el procesamiento de texto que opera directamente a nivel de personaje y usos solo pequeñas convoluciones y operaciones de agrupamiento. Demostrando que el rendimiento de este modelo aumenta con la profundidad en clasificaciones de textos públicos (Conneau et.al., 2016).

[1] Wikipedia. La Enciclopedia Libre. Ultima visita: 03/03/2020. Disponible en: https://www.wikipedia.org/

4 FORMACION DE RECURSOS HUMANOS

Este proyecto es parte de las líneas de investigación del "Programa de Investigación en Computación" de la FCEQyN de la UNaM, con cuatro integrantes relacionados con las carreras de Ciencias de la Computación de la UNaM. De los cuales dos están realizando su tesis de pos-grado, dos se encuentran realizando tesis de grado.

5 BIBLIOGRAFIA

1. Association for Computing Machinery: "CS2008 Curriculum Update: The Computing Curricula Computer Science Volume is complete and approved", 2008.
2. Association for Computing Machinery: "Computer Science Curricula 2013 Strawman Draft (February 2012)", 2012.
3. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, 135-146.
4. Cantero, L. (2019) "Desarrollo de un Proceso de Recomendación de Autores para un Sistema de Recuperación de Información de Ciencias de la Computación". Tesis de Grado, Licenciatura en Sistemas de Información. F.C.E.Q.yN. – Universidad Nacional de Misiones. Argentina.
5. Canteros A., Rey M., Kuna H., (2018a) "Clasificación de autores para un proceso de recomendación integrado a un metabuscador científico," in XXIV Congreso Argentino de Ciencias de la Computación - Libro de Actas, Tandil, Argentina.
6. Canteros A., Zamudio E., Kuna H. D., (2018b) "Desambiguación de autores para un sistema de recuperación de expertos en un contexto académico," in XIX Simposio Argentino de Inteligencia Artificial (ASAI)-JAIIO 47, CABA, Argentina.
7. Canteros, A. (2020) "Desambiguación de Autores para un Sistema de Recuperación de Información de Ciencias de la

Computación". Tesis de Grado, Licenciatura en Sistemas de Información. F.C.E.Q.yN. – Universidad Nacional de Misiones. Argentina.
8. Conneau, A., Schwenk, H., Barrault, L., & Lecun, Y. (2016). Very deep convolutional networks for text classification. arXiv preprint arXiv:1606.01781.
9. Kuna H., Rey M., Martini E., Solonezen L., Podkowa L., (2013) "Desarrollo de un Sistema de Recuperación de Información para Publicaciones Científicas del Área de Ciencias de la Computación," Rev. Latinoam. Ing. Softw., vol. 2, no. 2, pp. 107–114.
10. Kuna H., Rey M., Podkowa L., Martini E., Solonezen L., (2014) "Expansión de Consultas Basada en Ontologías para un Sistema de Recuperación de Información," XVI Workshop de Investigadores en Computación. RedUNCI. Argentina.
11. Kuna H. et al., (2017) "An Entity Profile Schema for Data Integration in an Academic Metasearch Engine," in Proceedings of the 2017 International Conference on Artificial Intelligence, Las Vegas, USA, pp. 281–285.
12. Kuna H. Rambo A., Zamudio E., Cantero A., Canteros A., Biale C., Martini R., Rauber F., Pautsch G. Krujoski S., Rey M. (2019) "Avances en el Desarrollo de Métodos de Desambiguación y Recomendación de Autores Científicos para un Metabuscador de las Ciencias de la Computación", XXI Workshop de Investigadores en Ciencias de la Computación, San Juan, RedUNCI - USJ. Argentina.
13. Li, H., Councill, I., L, e W.-C, Giles , C. L., (2006) "CiteSeerx: An Architecture and Web Service Design for an Academic Document Search Engine," in Proceedings of the 15th International Conference on World Wide Web, New York, NY, USA, pp. 883–884.
14. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013a). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).
15. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013b). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. ICLR Workshop.
16. Ortega J. L., Aguillo I. F., (2014) "Microsoft academic search and Google scholar citations: Comparative analysis of author profiles," J. Assoc. Inf. Sci. Technol., vol. 65, no. 6, pp. 1149–1156.
17. Quinteiro-González, J. M., Martel-Jordán, E., Hernández-Morera, P., Ligero-Fleitas, J. A., & López-Rodriguez, A. (2011). Clasificación de textos en lenguaje natural usando la Wikipedia. RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação, (8), 39-52.
18. Tang J., (2016) "AMiner: Mining Deep Knowledge from Big Scholar Data," in Proceedings of the 25th International Conference Companion on World Wide Web, Geneva, Switzerland, pp. 373–373.
19. Torres López, Carmen, & Arco García, Leticia. (2016). Representación textual en espacios vectoriales semánticos. Revista Cubana de Ciencias Informáticas, 10(2), 148-180. Recuperado en 30 de marzo de 2020, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992016000200011&lng=es&tlng=es.
20. Xamena, E., Marmanillo, W. G., & Mechaca, A. L. (2019). Rebuilding the Story of a Hero: Information Extraction in Ancient Argentinian Texts. In V Simposio Argentino de Ciencia de Datos y GRANdes DAtos (AGRANDA 2019)-JAIIO 48 (Salta).
21. Zhang, W., Yoshida, T., & Tang, X. (2011). A comparative study of TF* IDF, LSI and multi-words for text classification. Expert Systems with Applications, 38(3), 2758-2765.
22. Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In Advances in neural information processing systems (pp. 649-657).

Agente conversacional inteligente como herramienta de ayuda al proceso de atención al aspirante de las carreras de la Facultad de Tecnología y Ciencias Aplicadas (FTyCA) de la Universidad Nacional de Catamarca (UNCa).

María V. Póliche[1], Hernán C. Ahumada[1], Daniel A. Rivas[1], Oscar E. Quinteros[1], Cecilia E. Gallardo[1], Nelson A. Contreras[1], Marta del V. Miranda[1], Javier Favore[1]

[1]Fac. de Tecnología y Ciencias Aplicadas - Universidad Nacional de Catamarca

Maximio Victoria 55, 4700 Catamarca, Argentina

{ vpoliche, hcahumada, darivas, oequinteros, ceciliagallardo, nacontreras, mvmiranda, jfavore,}@tecno.unca.edu.ar

Resumen

La Inteligencia Artificial (IA) es la simulación de la inteligencia humana por parte de una máquina que le permite solucionar problemas similares a como lo realiza un ser humano. Una de las principales áreas en donde se aplica la IA es en el desarrollo de bots conversacionales o chatbots, los cuales son sistemas inteligentes que se utilizan para simular una conversación con una persona utilizando un lenguaje natural ya sea hablado o escrito. Los bots conversacionales tienen la capacidad de retener información para ser utilizada en otro momento de la conversación. Un bot conversacional es un software que puede imitar una conversación con una persona utilizando un lenguaje natural.

Una de las principales dudas de todo interesado a iniciar el cursado de una carrera es conocer los requisitos para inscribirse. Obtener esta información puede consumirle un tiempo significativo al interesado como a los encargados de la atención al público de proporcionarle esta información.

Este proyecto, tiene por finalidad desarrollar un prototipo de agente conversacional inteligente como herramienta de ayuda al proceso de atención al aspirante de las carreras de la Facultad de Tecnología y Ciencias Aplicadas (FTyCA) de la Universidad Nacional de Catamarca (UNCa).

Palabras clave: inteligencia artificial - bot conversacional - proceso de atención al aspirante

Contexto

Esta línea de trabajo se viene llevando a cabo como un proyecto de investigación acreditado por SECyT - UNCA dentro del Programa de Incentivos mediante Resolución Rectoral Nº 0713/19, denominado "*Desarrollo de un agente conversacional inteligente como herramienta de ayuda al proceso de atención al aspirante de las carreras de la Facultad de Tecnología y Ciencias Aplicadas (FTyCA) de la Universidad Nacional de Catamarca (UNCa),* que se desarrolla en el Departamento de

Informática de la Facultad de Tecnología y Ciencias Aplicadas (Universidad Nacional de Catamarca).

La temática del presente proyecto requiere y, a la vez permite, la definición y consolidación de una línea de investigación que interrelacione distintas disciplinas de las Ciencias de la Computación. Es por ello que el equipo del proyecto está formado por docentes investigadores y alumnos de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca y docentes-investigadores de 2 Universidades de Colombia: de la Corporación Universitaria Lasallista (Caldas) y del Grupo de Investigación en Ingeniería del Software (Grupo GIISTA) de la Institución Universitaria Tecnológico de Antioquia (Medellín).

Introducción

La Inteligencia Artificial (IA) es una rama de las Ciencias de la Computación que intenta encontrar esquemas generales de representación del conocimiento, y formalizar procesos de razonamiento coherentes, que permitan resolver problemas difíciles en dominios de aplicación concretos. Para esto, define cuatro categorías: sistemas que piensan como humanos, sistemas que actúan como humanos, sistemas que piensan racionalmente y sistemas que actúan racionalmente. Estos últimos, intentan emular en forma racional el comportamiento humano; siendo este, el caso de los agentes inteligentes (AI).

Un AI, es una entidad capaz de percibir su entorno, procesar tales percepciones, y actuar en consecuencia de manera racional sobre el mismo. En los últimos años, con el objetivo de resolver problemas de alta complejidad, se ha incrementado la investigación y el desarrollo de este tipo de herramientas. Una temática recurrente ha sido la implementación de sistemas que permitan a una computadora mantener una conversación en lenguaje natural.

En el contexto del procesamiento de lenguaje natural, un chatbot llamado también agente conversacional inteligente, es una aplicación computacional multiplataforma que tiene una interfaz diseñada para reconocer aquello que le escribimos y que es capaz de contestar de una forma coherente al contexto y al tema. Estos sistemas están diseñados para establecer un diálogo basado en reglas y técnicas de procesamiento de lenguaje natural, como los corpus lingüísticos, ya sea hablado o leído, que transforman la interacción entre el usuario y el sistema mediante interfaces cuasi-conversacionales con la finalidad de mejorar la experiencia usuario.

Los corpus anotados lingüísticamente, a menudo, son usados como base del conocimiento de sistemas inteligentes y se construyen extrayendo conversaciones naturales de entre los usuarios del futuro chat. Disponer de corpus de la lengua anotados proporciona una base sobre la que aplicar métodos de adquisición automáticos de restricciones selectivas así como métodos de desambiguación semántica.

Podemos distinguir dos grupos de chabots. El primer grupo son los “asistentes personales generales”, que pueden mantener una conversación sobre diversos temas. Ejemplo de ellos son Cortana (Microsoft), Siri (Apple), Alexa (Amazon), Now(Google) y Messenger (Facebook). En el otro grupo están los “asistentes personales específicos” que

operan en un dominio muy específico o que ayudan en actividades puntuales.

Las aplicaciones de este tipo de sistemas son actualmente muy variadas ya que pueden implementarse para cualquier servicio que se pueda resolver con una lógica conversacional. Algunos chatbots conversacionales tienen la capacidad de retener información para ser utilizada en otro momento de la interacción con el usuario. Otra de las ventajas del uso de esta tecnología, es que puede integrarse no solo en plataformas web y móviles, sino también dentro de plataformas que han sido creadas para ello como Slack, Alexa, Skype, Telegram, Discord e incluso en redes sociales como Twitter y Messenger (Facebook), razón por la cual se encuentran en sitios web comerciales, en los cuales tienen la función de promocionar productos, ayudar en la navegación al usuario y obtener información del cliente a través del diálogo. Debido a la gran cantidad de aplicaciones, los chatbots conversacionales se han convertido en herramientas muy populares en el ámbito educativo-administrativo.

La tecnología de chatbots requiere la confluencia de varias disciplinas: Ontología, Semántica Representación del conocimiento, Procesamiento del Lenguaje Natural, Lingüística Computacional Aprendizaje Automático, Lenguajes de programación para aplicaciones web.

En la actualidad existe una gran variedad de herramientas informáticas para desarrolladores que facilitan la implementación de un chatbot. Sin embargo estos son todavía relativamente rudimentarios y solo son capaces de reconocer un número reducido de comandos simples y de palabras claves. Por lo tanto, todavía hay muchas maneras de innovar a los efectos de que los chatbots sean capaces de proporcionar una interfaz más natural e intuitiva para que los usuarios interactúen con ellos.

Debido a la gran cantidad de aplicaciones, los bots conversacionales se han convertido en herramientas muy populares en el ámbito educativo-administrativo.

En los últimos años el número de interesados a iniciar el cursado de carreras de la FTyCA ha aumentado considerablemente, más aún, teniendo en cuenta la incorporación de la carrera de Arquitectura en el año 2018. Por otra parte se observa también un número elevado de interesados proveniente del interior de la provincia y provincias vecinas.

Actualmente, la FTyCA, brinda información acerca de las carreras y los requisitos para inscribirse a las mismas a través de la página web de la Facultad o personalmente en la Dirección de Asuntos Académicos.

Por todo lo expuesto con anterioridad, se considera pertinente el desarrollo de una herramienta que permita proporcionarle a todo interesado información acerca carreras que se dictan en la FTYCA y del proceso de preinscripción en forma inmediata y a toda hora del día de manera online, de modo equivalente al que normalmente recibiría a través del diálogo con un humano experto.

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación aquí presentada trata de incorporar un novedoso enfoque para la enseñanza de nociones básicas de programación a estudiantes de Ingeniería en Informática.

En el caso de estudio del presente proyecto, el desarrollo de un prototipo de agente conversacional inteligente

permitirá, la interacción con él de manera amigable y eficiente, teniendo en cuenta las características de Usabilidad y Accesibilidad establecidas en el estándar de calidad ISO/IEC 25010.

Desde el punto de vista institucional, en caso de ser implementado, el prototipo, permitirá brindar un servicio de atención a interesados en iniciar estudios universitarios con disponibilidad las 24 horas del día, los 7 días de la semana durante todo el año.

Cabe señalar, que en la práctica los aspirantes acuden varias veces al humano experto para aclarar sus dudas, pero éste algunas veces se encuentra ocupado realizando otras actividades ya que cuenta con un horario establecido para atender a los interesados. El prototipo que proponemos busca minimizar la cantidad de veces que el potencial alumno acude al experto y estar accesible las 24 horas desde .cualquier lugar a través de internet.

La principal ventaja es que el interesado podrá obtener la información que desee sobre el proceso de inscripción desde cualquier lugar y en cualquier momento mientras cuente con una conexión a internet, sin tener la necesidad de acudir a las instalaciones en donde se encuentre el humano experto.

Por otro lado, nuestro prototipo proporcionará la posibilidad de reducir una parte de la carga de trabajo del experto y le permitirá realizar otras tareas de igual importancia.

Resultados y Objetivos

Objetivo General

Desarrollar un prototipo de agente conversacional inteligente (chabot) como herramienta de ayuda (asistencia) a el proceso de atención al aspirante de las carreras de la Facultad de Tecnología y Ciencias Aplicadas (FTYCA) de la Universidad Nacional de Catamarca (UNCa).

Objetivos Específicos

- Relevar el estado del arte de métodos y tecnologías asociadas al desarrollo e implementación de un agente conversacional inteligente.
- Comprender la secuencia de etapas del proceso bajo estudio y representar el conocimiento de cada una de ellas.
- Interpretar y seleccionar apropiadamente las tecnologías y técnicas computacionales que mejor se adapten para el caso de estudio.
- Representar adecuadamente el flujo de conversaciones asociadas al proceso del caso de estudio.
- Desarrollar el modelo conversacional asociado al proceso.
- Obtener las funciones de procesamiento de cada nodo del modelo conversacional.
- Implementar a prueba el prototipo de agente conversacional inteligente desarrollado.
- Obtener métricas de desempeño del chatbot y opinión de los usuarios.

Resultados esperados

La temática del presente proyecto requiere y, a la vez permite, la definición y consolidación de una línea de investigación que interrelacione distintas disciplinas de las Ciencias de la Computación. Es por ello que el equipo del proyecto está formado por docentes investigadores y alumnos de la Facultad de Tecnología y Ciencias Aplicadas y

también docentes-investigadores de 2 Universidades de Colombia, de la Corporación Universitaria Lasallista (Caldas) y del Grupo de Investigación en Ingeniería del Software (Grupo GIISTA) de la Institución Universitaria Tecnológico de Antioquia (Medellín).

Dada la diversidad de ramas del conocimiento involucradas en el desarrollo del presente proyecto, se acuerda la conformación del Grupo de Investigación y Desarrollo en "Generación y Gestión Computacional de Conocimiento" (Grupo I+D+2Ge+2Co) en el ámbito de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca.

Asimismo contribuirá a enriquecer el dictado de las asignaturas del plan de estudio de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca, con los temas de investigación del presente proyecto.

Formación de Recursos Humanos

Los conocimientos generados y experiencia adquirida mediante este proyecto permitirán luego una ampliación y aplicación de los mismos en otros procesos y dominios. Por lo tanto, es un tema que cuenta con un gran potencial de transferencia tecnológica al entorno socio-productivo tanto a nivel nacional como internacional.

Por otra parte, este proyecto, aspira a entre otras cosas a la formación de profesionales investigadores, y también a ser fuente de origen de trabajos de investigación de graduación y postgraduación.

Los integrantes de este Proyecto de Investigación, tienen previsto en los dos años que dura el mismo, participar de eventos científicos con una doble finalidad de capacitarse en temáticas específicas concernientes al proyecto por un lado y además para dar a conocer los resultados de las investigaciones realizadas. Respecto de ésto, se prevé para todos los integrantes del proyecto la participación y asistencia a Eventos Científicos Nacionales e Internacionales

Referencias

Abu, S. B., E.Atwell. (2007). "Chatbots: Are theyReallyUseful", JournalforLanguageTechnology and ComputationalLinguistics (JLCL), vol 22, pp. 29-49.

Blas, M. J., Ferreyra, N. E. D., Sarli, J. L., Gutierrez, M., & Roa, J. M. (2012). aiStudent: Agente inteligente para la interacción en Lenguaje Natural. In 41 JAIIO - EST 2012.

Dale, R. (2016). Thereturn of thechatbots. Natural LanguageEngineering, 22(5), 811-817.

Espinosa Rodríguez, R., Pérez de Celis Herrero, C., Muñoz, L., del Carmen, M., Somodevilla García, M. J., & Pineda Torres, I. H. (2018). Chatbots en redes sociales para el apoyo oportuno de estudiantes universitarios con síntomas de trastorno por déficit de la atención con hiperactividad. TE& ET.

Moine, J. M., Gordillo, S. E., &Haedo, A. S. (2011). Análisis comparativo de metodologías para la gestión de proyectos de Minería de Datos. In Congreso Argentino de Ciencias de la Computación (Vol. 17).

Nigro,O. et al (2004). KDD (Knowledge Discovery in Databases): Un proceso centrado en el usuario. In VI Workshop de Investigadores en Ciencias de la Computación.

Wirth, R and Hipp, J.. (2000). Crisp-dm: Towards a standard processmodelfor data mining. In Proceedings of the 4th internationalconferenceonthepracticalapplications of knowledgediscovery and data mining, pages 29–39.

Castellón, I., &Lloberes, M., &Fisas, B., &Julià, A., &Rigau, G., &Climent, S., &Coll-Florit, M. (2010). Anotación semántica de los sustantivos del corpus SenSem. Procesamiento del Lenguaje Natural, (45), 315-316.

Evolución Diferencial para Problemas Restringidos Dinámicos

Sebastián A. Hernández[*1], Guillermo Leguizamón[**1,2], and Efrén Mezura-Montes[***3]

[1]Universidad Nacional de la Patagonia Austral, Argentina
[2]Universidad Nacional de San Luis, Argentina
[3]Universidad Veracruzana, México

1. Resumen

La resolución de problemas (optimización de funciones, cálculo de raíces) se ha vuelto una rama interesante dentro de las aplicaciones de la inteligencia computacional y en particular, en el campo de las metaheurísticas. Tradicionalmente, los problemas matemáticos se dividieron (a grandes rasgos) en dos partes bien diferenciadas: lineales y no lineales. Para los lineales existen muchas técnicas capaces de resolverlos eficazmente en poco tiempo de cómputo (método *simplex*, *Gauss-Jordan*, entre los más conocidos). Pero para los no lineales la dependencia del concepto de derivada es muy fuerte (*Newton Raphson*, *steepest descendent*, por nombrar algunos). Esto hace que resolver un problema no lineal se torne muy complejo cuando la cantidad de variables crece.

La aparición de las metaheurísticas permitió resolver, con un grado alto de precisión, los problemas no lineales de una manera simple. Entre las principales metaheurísticas poblacionales se encuentran los algoritmos de Evolución Diferencial (ED), siendo uno de sus atractivos principales, además de la potencia de cálculo, la sencillez de implementación [17].

ED ha permitido resolver problemas complejos, incluyendo problemas altamente dimensionales, con restricciones y dinámicos, entre muchos otros. En esta comunicación, se presenta una propuesta de investigación para el desarrollo y aplicación de versiones híbridas de ED para la resolución de problemas dinámicos con restricciones (DCOPs, siglas en Inglés para Dynamic Constrained Optimization Problems). Para la puesta a punto y evaluación de los algoritmos desarrollados, se considerarán *benchmarks* usados en la literatura.

[*] email: sebastian.unpa@gmail.com
[**] email: legui@unsl.edu.ar
[***] email: emezura@uv.mx

2. Contexto

La linea de investigación propuesta es parte de los objetivos generales del Proyecto de Investigación "PROICO 03-1018: Inteligencia Computacional para Resolución de Problemas Complejos" el cual es financiado por la Universidad Nacional de San Luis. El objetivo principal linea de investigación es la consecución de una tesis doctoral a ser desarrollada por el Mg. Sebastián Hernández, quien además es integrante del proyecto de referencia. Cabe destacar además, que junto al Dr. Mezura Montes de la Universidad Veracruzana (México), se vienen manteniendo desde hace muchos años diversas colaboraciones tanto en trabajos de investigación, como en el asesoramiento conjunto de tesis de posgrado, entre ellas la dirección conjunta de la tesis concluida de Maestría del Mg. Hernández por parte del Dr. Mezura Montes y el Dr. Leguizamón. Entendemos en consecuencia, que las condiciones están dadas para llevar a cabo la investigación propuesta.

3. Introducción

La aplicación de ED para resolver exitosamente problemas de optimización permitió que los investigadores comenzaran a ampliar los horizontes y por ello, trabajar con problemas de optimización dinámica (DOPs - Dynamic Optimization Problems). Un DOP es un problema de optimización que posee uno o mas parámetros dependientes del tiempo que cambian su valor a medida que evoluciona, por lo tanto tiene un valor óptimo cambiante en el tiempo. Debido a que gran cantidad de problemas del mundo real presentan parámetros cambiantes en el tiempo, es importante contar con herramientas capaces de darle solución a un DOP. En el caso de que se agreguen restricciones, se transforma en un problema restringido de optimización dinámica (DCOP - Dynamic Constrained Optimization Problems). Una de las expresiones matemáticas más comunes de un DCOP es:

$$\text{mín } f(\vec{x}, t), \vec{x} = (x_1, x_2, \ldots, x_D) \in S, L_i < x_i < U_i$$

sujeto a:

$$\begin{cases} g_j(\vec{x},t) \leq 0, j = 1,2,\ldots,l \\ h_j(\vec{x},t) = 0, j = l+1, l+2,\ldots,m \end{cases}$$

donde t es la instancia de tiempo discreto, $\vec{x}$ es el vector de decisión, x_i es la i-ésima variable de decisión, L_i y U_i son los límites interior y superior de x_i, respectivamente, D es el número de variables de decisión, $S = \prod_{i=1}^{D} [L_i, U_i]$ es el espacio de decisión o espacio de búsqueda, $f(\vec{x},t)$ es la función objetivo, $g_j(\vec{x},t)$ es la j-ésima restricción de desigualdad, $h_j(\vec{x},t)$ es la j-ésima restricción de igualdad, y l y $(m-l)$ son los números de restricciones de desigualdad e igualdad respectivamente.

El enfoque tradicional, Seguimiento Óptimo con el Tiempo (de aquí en mas TMO por sus siglas en Inglés para "Tracking Moving Optimum"), supone que los cambios actuales son pequeños, entonces se torna importante conocer lo que ha ocurrido en el pasado [5],[18]. De esta forma, se busca un nuevo óptimo global cada vez que se modifican las condiciones de la función objetivo y/o las restricciones. Este enfoque presenta algunos inconvenientes al aplicarse en problemas del mundo real que poseen cambios no tan pequeños como los supuestos. Además, si la velocidad de los cambios es rápida, es posible que se consigan valores óptimos *atrasados* en el tiempo.

Los problemas reales rara vez son estáticos, por lo que es de gran importancia poder resolver problemas dinámicos con diferentes características y complejidades. En tal sentido, varios autores vienen realizando propuestas para lograr aportar soluciones a esta problemática. A los efectos de brindar una revisión general de la literatura, se presenta a continuación una breve reseña de algunos trabajos del estado del arte relativos al campo de investigación del tema propuesto:

- Cuando el dinamismo está presente, el algoritmo de optimización tiene la necesidad de proveer mecanismos que permitan adaptarse a los cambios del entorno durante el proceso de búsqueda. Por ejemplo, la estrategia sugerida en [5] radica en utilizar memoria sobre la evolución (información de aprendizaje). El trabajo propone diversos tipos de memoria para adaptar la búsqueda cuando se produzcan los cambios.

- La hipotésis de que los algoritmos creados para problemas estáticos no resuelven correctamente los problemas dinámicos es verificada en [12]. Más aún, el dinamismo en las restricciones genera nuevos escenarios dentro del problema a resolver. La propuesta es un algoritmo basado en la reparación de GENOCOP III, para identificar áreas promisorias y además mantener la diversidad.

- La detección de cambios dota a un algoritmo evolutivo de la capacidad de adaptarse adecuadamente para mantener su desempeño en un ambiente dinámico [18]. Dos propuestas de detección de cambios: 1) basada en la población (testeo estadístico para ver sus cambios de características) y 2) en la relación entre las detecciones efectuadas y los sensores empleados de acuerdo a la dimensión del problema.

- En [13] se proponen mecanismos para manejar el dinamismo en optimización. El objetivo es combinar las ventajas de la optimización dinámica (inmigrante aleatorio e hipermutación) con las estrategias del manejo de restricciones (método de reparación) y adaptarlos a los problemas dinámicos con restricciones.

- En [14], los autores identifican las principales características de DCOPs del mundo real, planteando la discusión de cómo estas propiedades difieren con respecto a DOPs (Dynamic Optimization Problems). Luego plantean la adaptación de estrategias de optimización dinámica y manejo de restricciones para resolver DCOPs, analizando debilidades y fortalezas.

- El enfoque tradicional para resolver DCOPs es utilizar un algoritmo y luego una técnica de manejo de restricciones separada del mismo (reparación o funciones de penalidad). Sin embargo, la reparación con elitismo atentan contra la diversidad de la población y las funciones de penalidad no resuelven bien cuando hay regiones factibles disconexas. La propuesta presentado en [15] es un algoritmo basado en Búsqueda Gravitacional con una modificación del método de reparación, que incorpora uso de información y reinicio en base al conocimiento actual de la función.

- Para resolver DCOPs, los autores adaptan ED basándose en un mecanismo para detectar cambios en la función objetivo o en las restricciones, permitiendo diversidad en la población [2]. Esto se logra combinando dos variantes de ED y utilizando memoria de las mejores soluciones. Además, se agregan inmigrantes aleatorios a la población en cada ejecución, junto con un buscador local basado en *hill climbing*.

- En [3] se presenta un nuevo método de reparación que no requiere de soluciones factibles como referencia, inspirado en la mutación diferencial. Se utiliza en simultáneo con dos variantes de ED.

- A los efectos de mejorar la idea de rastreo en múltiples regiones factibles, se utilizan diversas subpoblaciones [6]. Se propone un conjunto de

estrategias de localización y seguimiento de regiones viables para manejar diferentes tipos de restricciones dinámicas.

- En [4] se presenta un estudio empírico de la evolución diferencial dinámica del algoritmo con variantes combinadas con un método de reparación (DDECV + Reparación) en la solución de problemas de optimización dinámica restringida. El estudio se centra en cuatro ítems: el rol de cada uno de sus elementos; sensibilidad ante diferentes frecuencias de cambio; capacidad para detectar un cambio y recuperarse después de él y su desempeño con dinamismo presente en diferentes partes del problema.

- Cuatro de las técnicas de manejo de restricciones más populares para problemas restringidos (estáticos) son estudiadas en [1] con el fin de aplicarlos a problemas restringidos dinámicos (DCOPs).

- En [16] los autores proponen un generador de funciones capaz de crear entornos dinámicos restringidos al extender el generador de funciones de referencia con picos móviles o *"Moving Peaks Benchmark"* (MPB). Además, realizan un análisis de los entornos resultantes producidos por el generador a través de *Fitness Landscape Analysis* (FLA).

- En [19] se analiza el comportamiento de tres algoritmos evolutivos diseñados para resolver DCOPs. Uno de ellos, DyCODE, es la propuesta de los autores. A partir del benchmark analizado, se explicitan fortalezas y debilidades de todos los algoritmos analizados.

- Con el obejtivo de emular el dinamismo existente en restricciones de problemas del mundo real, en [7] se utilizan cambios de tipo lineal en dichas restricciones. Se propone además, un *framework* capaz de crear *benchmarks* de entornos dinámicos con cambios en la restricciones dinámicas.

4. Lineas de Investigación y Desarrollo

Si bien la aplicación de ED a problemas con restricciones ha sido ampliamente estudiada, aun existe una cantidad importante de posibles lineas de investigación y desarrollo cuando se considera la inclusión de dinamismo a nivel de función objetivo y/o de las restricciones. Mas aún, cuando la dimensión de los problemas crece, la complejidad de las estrategias a investigar y desarrollar se van incrementando de manera considerable. En consecuencia, se plantea como principales lineas de investigación el diseño de estrategias para manejo de restricciones en DCOPs, el uso de buscadores locales que puedan ayudar en el proceso de exploración y además, diferentes operadores sensibles a los cambios que se puedan producir, tanto en la función objetivos, como en las restricciones de los problemas a estudiar.

5. Resultados Obtenidos y Esperados

La investigación propuesta, tiene como base los trabajos previos realizados por el Mg. Sebastián Hernández como parte de su tesis de Maestría y, tal cual se indicará en la próxima sección, se pretende avanzar en la consecución de un futuro plan de tesis doctoral. A los fines de mostrar la experiencia del Mg. Hernández y el resto de los autores, se indican algunos de los principales trabajos realizados de manera conjunta [8][11][10][9]. En función de la experiencia previa, se espera lograr diseñar algoritmos eficientes basados en ED que sean competitivos con aquellos del estado del arte y además, lograr presentar los resultados en congresos y revistas de nivel internacional.

6. Formación de RRHH

La presente investigación se centra principalmente en el desarrollo de la tesis doctoral del Mg. Sebastián Hernández de la Universidad de la Patagonia Austral, cuyo tema y plan de tesis se encuentra en proceso de elaboración para su posterior evaluación, siendo sus asesores, el Dr. Guillermo Leguizamón y el Dr. Efrén Mezura-Montes. Sin embargo, dada la amplitud del tema y diversidad de líneas de trabajo, existe la posibilidad de incorporar de incorporar becarios para realizar trabajos finales de grado y/o tesis de posgrado.

Referencias

[1] Maria-Yaneli Ameca-Alducin, Maryam Hasani-Shoreh, Wilson Blaikie, Frank Neumann, and Efren Mezura-Montes. A comparison of constraint handling techniques for dynamic constrained optimization problems. In *2018 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, jul 2018.

[2] Maria-Yaneli Ameca-Alducin, Efren Mezura-Montes, and Nicandro Cruz-Ramirez. Differential evolution with combined variants for dynamic constrained optimization. In *2014 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, jul 2014.

[3] María-Yaneli Ameca-Alducin, Efrén Mezura-Montes, and Nicandro Cruz-Ramírez. A repair method for differential evolution with combined variants to solve dynamic constrained optimization problems. In *Proceedings of the 2015 on Genetic and Evolutionary Computation Conference - GECCO15*. ACM Press, 2015.

[4] María-Yaneli Ameca-Alducin, Efrén Mezura-Montes, and Nicandro Cruz-Ramírez. Dynamic differential evolution with combined variants and a repair method to solve dynamic constrained optimization problems: an empirical study. *Soft Computing*, 22(2):541–570, sep 2016.

[5] J. Branke. Memory enhanced evolutionary algorithms for changing optimization problems. In *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*. IEEE, 1999.

[6] Chenyang Bu, Wenjian Luo, and Lihua Yue. Continuous dynamic constrained optimization with ensemble of locating and tracking feasible regions strategies. *IEEE Transactions on Evolutionary Computation*, 21(1):14–33, feb 2017.

[7] Maryam Hasani-Shoreh, Maria-Yaneli Ameca-Alducin, Wilson Blaikie, Frank Neumann, and Marc Schoenauer. On the behaviour of differential evolution for problems with dynamic linear constraints. In *2019 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, jun 2019.

[8] Sebastián Hernández, Guillermo Leguizamón, and Efrén Mezura-Montes. Hibridación de evolución diferencial utilizando hill climbing para resolver problemas de optimización con restricciones. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2012.

[9] Sebastián Hernández, Efrén Mezura Montes, and Guillermo Leguizamón. De with random vector based mutation for high dimensional problems. In *XXV Congreso Argentino de Ciencias de la Computación (CACIC)(Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019)*, 2019.

[10] Sebastián Alejandro Hernández, Guillermo Leguizamón, and Efrén Mezura-Montes. A hybrid version of differential evolution with two differential mutation operators applied by stages. In *Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2013, Cancun, Mexico, June 20-23, 2013*, pages 2895–2901. IEEE, 2013.

[11] Sebastián Alejandro Hernández, Guillermo Leguizamón, and Efrén Mezura-Montes. Hybridization of differential evolution using hill climbing to solve constrained optimization problems. *Inteligencia Artif.*, 16(52):3–15, 2013.

[12] Trung Thanh Nguyen and Xin Yao. Benchmarking and solving dynamic constrained problems. In *2009 IEEE Congress on Evolutionary Computation*. IEEE, may 2009.

[13] Trung Thanh Nguyen and Xin Yao. Solving dynamic constrained optimisation problems using repair methods, 2010.

[14] Trung Thanh Nguyen and Xin Yao. Continuous dynamic constrained optimization—the challenges. *IEEE Transactions on Evolutionary Computation*, 16(6):769–786, dec 2012.

[15] Kunal Pal, Chiranjib Saha, Swagatam Das, and Carlos A. Coello Coello. Dynamic constrained optimization with offspring repair based gravitational search algorithm. In *2013 IEEE Congress on Evolutionary Computation*. IEEE, jun 2013.

[16] Gary Pamparà and Andries P. Engelbrecht. A generator for dynamically constrained optimization problems. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion on - GECCO19*. ACM Press, 2019.

[17] Kenneth Price, Rainer Storn, and Jouni Lampinen. *Differential evolution - A practical approach to global optimization*. Springer-Verlag, 2005.

[18] Hendrik Richter. Detecting change in dynamic fitness landscapes. In *2009 IEEE Congress on Evolutionary Computation*. IEEE, may 2009.

[19] Yong Wang, Jian Yu, Shengxiang Yang, Shouyong Jiang, and Shuang Zhao. Evolutionary dynamic constrained optimization: Test suite construction and algorithm comparisons. *Swarm and Evolutionary Computation*, 50:100559, nov 2019.

Arquitectura, Redes y Sistemas Operativos

PROGRAMA DE INVESTIGACIÓN, DESARROLLO Y ENSEÑANZA EN COMPATIBILIDAD ELECTROMAGNÉTICA E INTEGRIDAD DE SEÑAL FCEIA/UNR

Roatta, Santiago - sroatta@fceia.unr.edu.ar - Digital/Departamento de Electrónica/FCEIA/UNR.
Geninatti, Sergio – foco@fceia.unr.edu.ar – Departamento de Electrónica/FCEIA/UNR.
Gennai, Gerardo – ggennai@fceia.unr.edu.ar Digital/Departamento de Electrónica/FCEIA/UNR.

RESUMEN

Este trabajo presenta el programa de investigación y desarrollo del Departamento de Electrónica de la Facultad de Ingeniería de la Universidad Nacional de Rosario sobre compatibilidad electromagnética e integridad de señal en el proceso de desarrollo de sistemas electrónicos.

El aumento de las frecuencias de reloj, la reducción de las tensiones de alimentación, y la popularidad de las comunicaciones inalámbricas, fuentes conmutadas y equipamiento electrónico en general, plantean exigencias crecientes en estos campos que difícilmente pueden ser abordadas a partir de conocimientos de ciencias y tecnologías básicas y/o de tecnologías aplicadas no específicas.

En este contexto, el Departamento de Electrónica de la FCEIA UNR, está abocado al estudio y desarrollo de medios que permitan integrar estos conocimientos en los procesos de formación de grado y posgrado.

Como parte de las actividades se ha implementado una asignatura electiva de grado, un curso de posgrado, un proyecto de vinculación tecnológica con el entramado productivo local, la adquisición de instrumental, sondas y antenas; y un proyecto de investigación.

Palabras Clave: EMC, Integridad de Señal, EMI.

CONTEXTO

La investigación se encuentra inscripta en el proyecto de investigación y desarrollo de la UNR: “Laboratorio para la enseñanza de compatibilidad electromagnética e integridad de señal: desarrollo, aplicación y optimización”, Código 80020190300026UR.
Además del anterior, se encuentra en su primera etapa de ejecución el proyecto PFIP ESPRO 2017 COFECYT: “Impulso para la mejora competitiva regional en diseño y producción de equipamiento electrónico”, declarado de interés por la Provincia de Santa Fe y financiado por el Cofecyt, la UNR y las siguientes empresas: Krets SA; Ingeniería electrónica argentina SRL; Chiarella diseño y tecnología; Ohana SRL; Sensor automatización agrícola SA. y Texel SRL. Este último proyecto, permitirá la adquisición de instrumental, sondas y antenas para el equipamiento de un laboratorio destinado a la enseñanza de EMC e IS y ensayos de precumplimiento.

1. INTRODUCCIÓN

A pesar de la gran demanda sobre cuestiones de compatibilidad electromagnética (EMC) e integridad de señal (IS) en el sector industrial, históricamente, las carreras de ingeniería no han abordado estos temas. A partir la década del 80 el ámbito académico fue reconociendo la necesidad de enseñanza de estas temáticas y las incorpora progresivamente. Algunas universidades lo hacen tempranamente [1], por ejemplo: Institutos federales de tecnología de Lausana y Zurich; Universidad técnica de Eindhoven; Instituto real de tecnología de Suecia; Universidades de Florencia, Turín y Roma; Instituto real de tecnología de Melbourne; Universidad de Kansas; Universidad de California-Berkeley y otras. Dentro del grupo destaca la Universidad de Kentucky con el trabajo pionero de Clayton Paul, autor del libro *“Introduction to Electromagnetic Compatibility”*, primer texto pensado para clases de EMC en el ámbito universitario. Esta universidad presenta un curso integrado y electivo ubicado en el último año de la carrera. El curso fue electivo como solución de compromiso para lidiar con el conocido problema de currícula sobrecargada de las carreras de ingeniería. A pesar de que no se podía asegurar que todos los estudiantes optaran por un curso electivo, se logra que

Roatta, Santiago - sroatta@fceia.unr.edu.ar - Digital/Departamento de Electrónica/FCEIA/UNR.
Geninatti, Sergio – foco@fceia.unr.edu.ar – Departamento de Electrónica/FCEIA/UNR.
Gennai, Gerardo – ggennai@fceia.unr.edu.ar Digital/Departamento de Electrónica/FCEIA/UNR.

fuera muy convocante con la ayuda de experiencias prácticas y visitas a laboratorios[2], [3]. La Universidad Estatal de Dakota del Norte, también aplica esta metodología [4]. La "University of Liverpool", opta por la enseñanza de EMC de manera distribuida, durante todo el transcurso de la carrera, orientando cursos obligatorios como electromagnetismo hacia la faceta de la EMC. Aún con una metodología distinta para encarar la sobrecarga de la currícula utilizan experiencias prácticas para reforzar el interés de los estudiantes y la fijación de conocimientos [5]. Las mismas conclusiones se ven en la experiencia en la Universidad de Rolla-Missouri [6]. Un caso desarrollado en India plantea un curso a distancia para mayor alcance y disminución de costos [7], con énfasis en la importancia de las mediciones, en este caso impartidas a través de videos.

Siguiendo con la cronología, la "San Francisco State University", deja de lado la dicotomía entre curso integrado y enseñanza distribuida y aplica ambos conceptos sobre su currícula [8]. Incorpora principios y aplicaciones relacionados con EMC en dos cursos de electromagnetismo y uno de comunicaciones y agrega un curso específico de EMC. Todos respaldados por experiencias prácticas en un laboratorio de alta frecuencia. En el mismo sentido, la Universidad estatal de Utah comienza con la enseñanza de EMC por primera vez en 2003 [9], introduciendo los conceptos de EMC durante el transcurso de la carrera para finalmente integrarlos con un curso final. Aquí se remarca la importancia de las experiencias prácticas y se recomienda el "EMC Education Manual" de la IEEE EMC Society.

La Universidad de Sidney Occidental remarca que, si bien, los conceptos de EMC pueden introducirse a lo largo de la carrera, es indiscutible la importancia de un curso de integración final. Este trabajo [10] también describe brevemente seis experiencias prácticas y sugiere que durante el cursado los estudiantes deberían tener total acceso a los laboratorios. También propone la utilización de Internet para la investigación de casos de Interferencia electromagnética (EMI).

Las Universidades Normal de Nanjing y Tecnológica de Singapur plantean el diseño de un circuito impreso (PCB) por parte de los estudiantes en dos versiones: una al principio del curso, sin conocimiento de EMC y otra al final del curso de EMC para finalmente comparar las emisiones radiadas de ambas [11]. Se realza la importancia del diseño y de la utilización del instrumental.

El Instituto de Tecnología Rose-Hulman en Indiana, presenta 2005 un curso sobre EMC e IS en la carrera de ingeniería informática [12]. El objetivo era el diseño digital de alta velocidad, por lo que se centraba en los aspectos de IS y EMC de PCB para sistemas digitales de alta velocidad. Además del programa sintético se describen ocho experiencias prácticas.

Una interesante experiencia de laboratorio ilustra los efectos del acoplamiento de campo cercano usando dos PCB separadas [13], conceptos que se aplican al diseño de PCB multicapa.

La Universidad McGill, Montreal, también focaliza principalmente sobre IS [14]. En este caso, la enseñanza se realiza utilizando simuladores de circuitos y experiencias prácticas especialmente diseñadas para que los estudiantes se familiaricen con los instrumentos de medición: analizador de espectro, analizador de redes, sondas de campo cercano y antenas.

Tres interesantes experiencias de laboratorio son utilizadas en el curso de EMC e IS de la carrera de Ingeniería Informática del Instituto de Tecnología Rose-Hulman, Indiana [15].

El proceso de enseñanza de EMC de la Universidad de Hong Kong comienza en 1994

PROGRAMA DE INVESTIGACIÓN, DESARROLLO Y ENSEÑANZA EN COMPATIBILIDAD ELECTROMAGNÉTICA E INTEGRIDAD DE SEÑAL FCEIA/UNR

Roatta, Santiago - sroatta@fceia.unr.edu.ar - Digital/Departamento de Electrónica/FCEIA/UNR.
Geninatti, Sergio – foco@fceia.unr.edu.ar – Departamento de Electrónica/FCEIA/UNR.
Gennai, Gerardo – ggennai@fceia.unr.edu.ar Digital/Departamento de Electrónica/FCEIA/UNR.

con un curso de grado que posteriormente devino en uno de los cursos de postgrado más populares en dicha universidad [16]. El mismo incluye diversas actividades pero pone énfasis en las pruebas de laboratorio e instalaciones de EMC.
Algunos expertos remarcan la importancia de la enseñanza práctica presentando algunos experimentos de laboratorio, el curso de Universidad de Ciencias Aplicadas JAMK en Jyväskylä, Finlandia es un ejemplo de ello. A decir de los autores, este curso es muy práctico y sin sobrecarga teórica [17].
Las implementaciones de experiencias de laboratorio son enormemente enriquecedoras en conceptos de EMC para electrónica de potencia [18].
Las herramientas de simulación son muy útiles para el tratamiento de IS. [19]
La Universidad de Reno, Nevada, USA desarrolla en 2016 un programa de renovación de sus populares cursos de grado y posgrado de EMC [20]. El documento describe los temas del curso, experiencias prácticas de laboratorio, un ejercicio de simulación numérica y conferencias de expertos de EMC invitados de la industria. También contiene un análisis sobre las opiniones de los estudiantes que tomaron el curso.

Todo este material referenciado (acotado en este informe por razones de brevedad), visibiliza problemas, soluciones y conflictos que se encuentran durante el desarrollo de diversos procesos de enseñanza de EMC e IS. Conflictos y soluciones que también fueron experimentados nuestra carrera de ingeniería electrónica: sobrecarga de horas en la currícula, dicotomía entre curso integrado o enseñanza distribuida, curso obligatorio o electivo y enseñanza teórica intensa o no. Una actividad curricular electiva denominada “Compatibilidad Electromagnética” aumentó notablemente sus inscriptos cuando pasó a llamarse “Diseño avanzado de circuitos impresos”.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

El PID UNR “Laboratorio para la enseñanza de compatibilidad electromagnética e integridad de señal: desarrollo, aplicación y optimización” busca, a partir de este año, el desarrollo de experiencias de laboratorio para su utilización en los cursos de grado y posgrado como así también el desarrollo de sondas de bajo costo.
Se proponen como objetivos generales incorporar conceptos de EMC e IS a lo largo de la carrera; implementar un curso integrado al final de la carrera de grado; ofrecer cursos de posgrado y soporte a la industria local; constituir un laboratorio que permita realizar dicho soporte donde los estudiantes accedan al instrumental específico; difundir, promover y fomentar el interés el conocimiento sobre estas temáticas; y la formación de recursos humanos en la temática.

3. RESULTADOS OBTENIDOS/ESPERADOS

Por el momento, pueden mencionarse como resultados:

Aprobación del proyecto PFIP ESPRO, actualmente en su primeros meses de ejecución, que permitirá no solo la adquisición del instrumental inicial para el laboratorio y el entrenamiento en cuestiones de EMC e IS de los equipos de desarrollo de las empresas intervinientes, sino también la visibilización dentro del entramado productivo local de la existencia de un núcleo de investigación y vinculación sobre estas temáticas.

Lanzamiento en 2018 de la asignatura electiva “Diseño avanzado de circuitos impresos” y del curso de posgrado “Compatibilidad Electromagnética aplicada al diseño electrónico”.

Roatta, Santiago - sroatta@fceia.unr.edu.ar - Digital/Departamento de Electrónica/FCEIA/UNR.
Geninatti, Sergio – foco@fceia.unr.edu.ar – Departamento de Electrónica/FCEIA/UNR.
Gennai, Gerardo – ggennai@fceia.unr.edu.ar Digital/Departamento de Electrónica/FCEIA/UNR.

Participación en la SASE 2019 con el tutorial: "Diseño de Circuitos impresos EMC" y el workshop: "Principios para el desarrollo de impresos con baja emisión y alta inmunidad".

Como resultados específicos esperados podemos agregar la constitución del laboratorio de EMC e IS, la implementación de las experiencias prácticas para que los alumnos tanto de grado como de posgrado puedan tener acceso y entrenamiento con instrumental acorde y la posibilidad de brindar soporte al entramado productivo local.

Como objetivo general obtenido y esperado, incrementar en cantidad y calidad los recursos humanos disponibles, para la aplicación integral de EMC e IS en el diseño electrónico.

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de las cátedras de Sistemas Digitales III y Sistemas Digitales IV integran el equipo de trabajo inicial (un profesor titular, un asociado, dos adjuntos y un Jefe de trabajos prácticos). En un principio, la capacitación del equipo fue realizada por seminarios internos a cargo del director y codirector, con lo que se logró cierta uniformidad de conocimientos básica, que posteriormente fue perfeccionada por las actividades de investigación.

Por otro lado, es necesario recordar que nuestra línea de I/D está principalmente centrada en el desarrollo de las condiciones para impulsar la formación de recursos humanos en la temática. Prueba de ello es la institucionalización de los cursos de grado y posgrado además de la especial formación que se imparte a los equipos de desarrollo de las empresas participantes en el PFIP ESPRO.

Adicionalmente, el equipo de investigación tradicionalmente ha dirigido entre 5 y 6 Proyectos de Ingeniería (tesinas de grado) por año. En los últimos años, independientemente de la temática del trabajo, las consideraciones de EMC e IS se han incrementado progresivamente. Especialmente en aquellos que incluyen el desarrollo de un circuito impreso. De este modo, aunque las temáticas abordadas en los trabajos son diversas, las cuestiones de EMC e IS se encuentran embebidas en ellos.

5. BIBLIOGRAFÍA

[1] Hubing, T., & Orlandi, A. (2005). A brief history of EMC education. In *16th International Zurich Symposium and Technical Exhibition on EMC* (pp. 95-98). Unknown Publisher.

[2] Paul, C. R. (1986, September). An undergraduate course in electromagnetic compatibility. In *1986 IEEE International Symposium on Electromagnetic Compatibility* (pp. 1-5). IEEE.

[3] Paul, C. R. (1990). Establishment of a university course in electromagnetic compatibility (EMC). *IEEE Transactions on Education*, *33*(1), 111-118.

[4] Nelson, R. M. (1991, July). The EMC course: Bridging the gap between academic preparation and industrial needs. In *IEEE 1991 International Symposium on Electromagnetic Compatibility* (pp. 319-320). IEEE

[5] Austin, B. A. (1992, September). Teaching EMC engineering in an overloaded curriculum. In *Eighth International Conference on Electromagnetic Compatibility, 1992.* (pp. 263-266). IET.

[6] Drewniak, J. L., Hubing, T. H., Van Doren, T. P., & Sha, F. (1995, August). Integrating electromagnetic compatibility laboratory exercises into undergraduate electromagnetics. In *Proceedings of International Symposium on Electromagnetic Compatibility* (pp. 35-40). IEEE.

[7] Deb, G. K. (1999, December). Importance of EMC education. In *Proceedings of the International Conference on Electromagnetic*

Roatta, Santiago - sroatta@fceia.unr.edu.ar - Digital/Departamento de Electrónica/FCEIA/UNR.
Geninatti, Sergio – foco@fceia.unr.edu.ar – Departamento de Electrónica/FCEIA/UNR.
Gennai, Gerardo – ggennai@fceia.unr.edu.ar Digital/Departamento de Electrónica/FCEIA/UNR.

Interference and Compatibility (pp. 160-167). IEEE.

[8] Pantic-Tanner,Z.(2001). Integrated EMC Education at San Francisco State University.

[9] Jost, R. J. (2003, August). Introducing EMC into the curriculum for the first time: experiences in achieving depth and breadth. In *2003 IEEE Symposium on Electromagnetic Compatibility. Symposium Record (Cat. No. 03CH37446)* (Vol. 1, pp. 170-172). IEEE.

[10] Hellany, A., & Nagrial, M. H. (2003, August). Project based engineering education: a case for teaching EMC/EMI. In *Asia-Pacific Conference on Applied Electromagnetics, 2003. APACE 2003.* (pp. 88-91). IEEE.

[11] Zhao, Y., & See, K. Y. (2004). A practical approach to EMC education at the undergraduate level. *IEEE Transactions on Education*, *47*(4), 425-429.

[12] Song, J., Voltmer, D. R., & Wheeler, E. (2005, August). A required EMC course for computer engineering undergraduates. In *2005 International Symposium on Electromagnetic Compatibility, 2005. EMC 2005.* (Vol. 1, pp. 13-18). IEEE.

[13] Foo, W. J., & Chee, J. (2006). Teaching near field coupling with PCB layout. In *2006 17th International Zurich Symposium on Electromagnetic Compatibility* (pp. 557-560). IEEE.

[14] Abhari, R. (2007, June). Teaching Electromagnetic Compatibility with emphasis on signal integrity in designing simulation and laboratory experiments. In *2007 IEEE Antennas and Propagation Society International Symposium* (pp. 5383-5386). IEEE

[15] Hoover, K. E., Song, J., Wheeler, E. D., & Drewniak, J. L. (2009). Three Practical and Effective RF and EMC Experiments for a Computer Engineering Course on Electromagnetics and EMC.

[16] Leung, S. W., & Chan, K. H. (2012). Development of electromagnetic compatibility courses at the City University of Hong Kong.*IEEE Electromagnetic Compatibility Magazine*, *1*(1), 50-54.

[17] Väänänen, O. (2013, September). EMC course for electronics bachelor degree students at JAMK University of applied sciences. In *2013 International Symposium on Electromagnetic Compatibility* (pp. 896-899). IEEE.

[18] Tarateeraseth, V. (2014). Educational laboratory experiments on EMC in power electronics. *IEEE Electromagnetic Compatibility Magazine*, *3*(3), 55-60.

[19] Balaji, U. (2014, May). Teaching electromagnetic compatibility to engineering technology students. In *IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014* (pp. 1-5). IEEE.

[20] Chatterjee, I. (2016, April). Electromagnetic Compatibility lecture and laboratory course for undergraduate and graduate students. In *2016 10th European Conference on Antennas and Propagation (EuCAP)* (pp. 1-3). IEEE.

Explorando las redes definidas por software (SDN)

Diego R. Rodríguez Herlein, Carlos A. Talay, Claudia N. González

Campus Universitario – oficina 18/ Dpto. Ciencias Exactas e Informática UARG / UNPA
(dherlein, ctalay, cgonzalez)@uarg.unpa.edu.ar

Luis A. Marrone

L.I.N.T.I. – Universidad Nacional de La Plata
Calle 50 y 115 – 1er. Piso – Edificio Bosque Oeste
lmarrone@linti.unlp.edu.ar

Resumen

Este trabajo prende dar una visión de las redes definidas por software (SDN), los fundamentos que llevaron a su desarrollo y evaluar los controladores que se utilizan para el despliegue de las mismas.

Palabras clave: SDN, controladores, Interfaces

Contexto

El presente trabajo está enmarcado en el PI 29/A451-1 "Análisis de performance del protocolo TCP utilizado en redes móviles" radicado de la UNPA-UARG y actualmente se encuentra en su etapa inicial de desarrollo. El proyecto está compuesto por cuatro investigadores de la UNPA, uno de la UNLP y un alumno becado con beca de iniciación en la investigación para estudiantes de la UNPA-UARG, dirigido por el Sr. Carlos A. Talay y codirigido por el Sr. Luis A. Marrone perteneciente a la UNLP. Este proyecto se financia íntegramente con fondos destinados a proyecto de investigación de la UNPA-UARG.

1. Introducción

En la actualidad la complejidad de las redes es cada vez mayor. Además de su creciente tamaño, requieren el despliegue dinámico de servicios. Sin embargo, las redes tradicionales no poseen la flexibilidad para probar nuevas ideas sin interrumpir los servicios que se estén ejecutando.

Paralelamente, la configuración y mantenimiento de las redes siguen siendo complejas: un administrador de red accede mediante la interfaz de línea de comandos a los dispositivos [1], lo que presenta dificultades en el caso de trabajar con dispositivos de distintos fabricantes.

Surgen también nuevas exigencias por parte de los usuarios, como una gran movilidad, seguridad, o aplicaciones 'Big Data'.

Es así que la capacidad de las redes de comunicación de datos está siendo rebasada, llegando a ser insuficientes para poder atender los grandes volúmenes de información que demandan estos nuevos servicios. No solo se trata del volumen, sino también de la gran variedad de información disponible en dispositivos móviles, GPS, contenidos multimedia de alta definición, sensores eléctricos, medidores, etc., que además deben comunicarse a una gran velocidad.

Las redes convencionales consisten de dispositivos cerrados, donde cada dispositivo tiene que ser configurado de manera individual, utilizando comandos específicos que tiene cada proveedor de equipos de red. Esto se ha convertido en un obstáculo, al desplegar nuevas versiones de protocolos existentes, actualizaciones o protocolos nuevos en las redes ya instaladas y operando.

El concepto SDN surgió como la arquitectura de red, donde el plano de control se desacopla del plano de datos. En lugar de hacer cumplir políticas y protocolos que se ejecutan en dispositivos dispersos, la red se limita a realizar el simple reenvío de información y el controlador de la red toma las decisiones [2], agilizando los flujos de información en la red.
La idea fundamental de las Redes Definidas por Software (SDN) es clara: trasladar el plano de control de los dispositivos físicos (enrutadores y conmutadores) unificándolo en un solo elemento externo a la red física, el controlador [3].
La arquitectura de SDN se puede observar en la siguiente figura (Figura 1) [4], se puede ver como se intercambia información entre las diferentes capas de la arquitectura: aplicación, control e infraestructura.

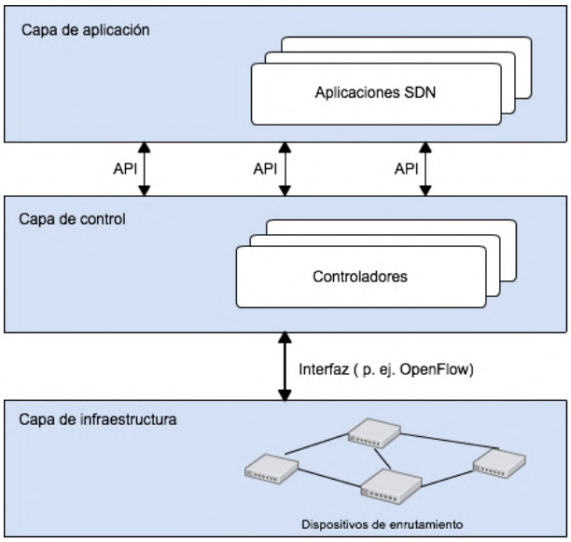


Fig.1: Arquitectura SDN

En las redes tradicionales, el plano de reenvío, control, administración y servicios está integrado en cada dispositivo de red. Por el contrario, el modelo SDN implementa la mayoría de las decisiones de administración, servicios y selección de ruta como funciones avanzadas dentro del controlador o en aplicaciones [5].

De esta manera, en una SDN se observa [6]:

- El plano de control y de datos están desacoplados. La funcionalidad de control se elimina de los dispositivos físicos de red que se convertirán en elementos de reenvío de paquetes.
- Las decisiones del reenvío de paquetes se toman en función a flujos, en lugar de basarse en las direcciones destino de los paquetes.
- La lógica de control se mueve a una entidad externa, llamado el controlador de la SDN o sistema operativo de red.
- La red es programable a través de aplicaciones en ejecución en la parte superior del controlador que interactúa con los dispositivos en el plano de datos subyacente.
- La arquitectura de una SDN contempla tres capas funcionales, el plano de datos, el plano de control y el plano de aplicación.

Interfaces

Toda la lógica de control de la red se implementa como una aplicación en la parte superior del sistema operativo de red que interactúa con las API de red que ofrece el controlador. [7].
El controlador cuenta con dos interfaces principales, una que interactúa con el plano de gestión llamada Northbound (NBI), y otra que interactúa con el plano de datos, llamada Southbound (SBI) (Figura 1). Así mismo, y con la finalidad de establecer comunicación con otros controladores en la red, se cuenta con las interfaces Eastbound (EBI) y Westbound (WBI) [8].
El objetivo de estas interfaces es proporcionar una apropiada abstracción de estados y eventos relacionados con la red subyacente y un canal de comunicaciones con características apropiadas, para apoyar la necesaria funcionalidad requerida entre los clientes en cualquier extremo del protocolo.

Interfaz Southbound: Formaliza la manera en que los elementos de control y plano de datos interactúan [6]. Esto permite la comunicación entre el controlador SDN y los dispositivos de Red en la capa de Infraestructura. Básicamente, su función es permitir al Controlador comunicarse con los elementos de conmutación de la red y programar la lógica de las comunicaciones en el hardware. Hay varias implementaciones de estas APIs, entre ellas OpenFlow que es el elemento principal en esta interfaz para implementar las soluciones SDN y VxLAN (más utilizadas en virtualización de redes, para definir capas intermedias u overlays).
OpenFlow es el protocolo empleado para comunicación entre el controlador (plano de control) y el switch (plano de datos). Se define como un protocolo emergente y abierto de comunicaciones y es la primera interfaz de comunicaciones estándar definida entre los planos de control y datos de una arquitectura de SDN.

Interfaz Northbound: Permite el intercambio de datos entre el controlador y la aplicación. El tipo de información que se intercambia, su forma y su frecuencia depende de cada aplicación. No hay estandarización para esta interfaz [9].

Interfaces Eastbound y Westbound: Permiten la interconexión de redes convencionales con las SDN, además sirven como conducto de información entre varios planos de control de diferentes dominios de SDN. [9].

Plano de Aplicación

Las aplicaciones SDN comunican sus requisitos a la red a través de una API que conecta con la capa de control. Están diseñadas para satisfacer las necesidades de los usuarios. De esta manera, a la capa de aplicación la componen las aplicaciones de usuarios. Presentan una interfaz común para el desarrollo de aplicaciones y permiten la interacción entre las aplicaciones externas y el controlador SDN. Actualmente los controladores de red ofrecen una variedad de APIs Northbound, como APIs RESTful, sistemas de archivos y lenguajes de programación. Incluye aplicaciones tales como enrutamiento, cortafuegos, balanceadores de carga, seguimientos y entre otros. En esencia, una aplicación de gestión define las políticas, que se traducen en última instancia en instrucciones específicas para la SouthBand Interface que programan el comportamiento de los dispositivos de reenvío [6].

Plano de Control

La capa intermedia la forma el Controlador SDN, quien tiene una visión global de la red. Al igual que en un sistema operativo tradicional, la plataforma de control abstrae los detalles de bajo nivel de la conexión y la interacción con dispositivos de reenvío, es decir materializa las políticas de red [6]
El controlador es el componente más importante de la arquitectura SDN ya que gestiona la capa de aplicación e infraestructura mediante sus interfaces, una con cada plano adjunto. El controlador es un elemento crítico en una arquitectura SDN, ya brinda el apoyo a la lógica de control para generar la configuración de la red con base en las políticas definidas.
Mediante la comunicación con el plano de infraestructura, se recoge el estado de la red y, según las exigencias de las aplicaciones, actualiza, en los dispositivos, las reglas de reenvío.
Por otra parte, se comunica con las aplicaciones SDN con un lenguaje de alto nivel, teniendo opciones como lenguajes existentes (Python, Java, C++), librerías en un kit software de desarrollador (SDK) como OnePK de Cisco, o lenguajes nuevos como Flowbased Management Language (FML), Frenetic o Nettle. Además, del estado de la

red que se recoge de la capa de infraestructura.
Aunque el modelo del plano de control es centralizado, la implementación puede ser distribuida para la escalabilidad y redundancia
Se necesita que un controlador sea capaz de comunicarse con otros debido a que un solo controlador puede ocasionar problemas de congestión ya que se trata de un punto crítico.
Se puede tener controladores de 'backup'. La solución más citada en este sentido es HyperFlow, que proporciona una vista sincronizada y consistente entre múltiples controladores.

Plano de Datos

La capa inferior es la capa de infraestructura y hace referencia a los dispositivos de red como los hosts, switches/routers físicos y/o virtuales, y los medios de transmisión.
El plano de datos proporciona el servicio fundamental de reenvío de datos.
Se utiliza el término conmutador SDN (SDN switch) para referirse al elemento de reenvío en el plano de datos. El término conmutador SDN no debe considerarse como una limitación de las funciones del elemento de red [8], si el dispositivo es un enrutador (Router), conmutador (Switch), o cortafuegos (Firewall). Estos dispositivos recogen el estado de la red (topología o estadísticas de tráfico) y se lo comunica al controlador, el cual a su vez le indica las reglas de reenvío de paquetes.
Es aquí donde cambia totalmente la filosofía respecto a las redes tradicionales.
Las aplicaciones definen el uso que se va a dar a la red, lo comunican a través de las Northbound APIs al controlador SDN, el cuál toma las decisiones oportunas y las comunica a la infraestructura de red mediante las SouthBound APIs. Estas APIs pueden ser abiertas o propietarias.

2. Líneas de Investigación, Desarrollo e Innovación

En el centro de la arquitectura de las SDN se encuentra la capa Control con el controlador SDN, que es quien gestiona los flujos. El controlador equivale al sistema operativo de la red que controla todas las comunicaciones entre las aplicaciones y los dispositivos. El controlador SDN se encarga de traducir las necesidades o requisitos de la capa Aplicación a los elementos de red, y de proporcionar información relevante a las aplicaciones SDN, pudiendo incluir estadísticas y eventos. Es por ello que se intentará realizar un estudio y análisis de la capa de control.

3. Resultados y Objetivos

El objetivo de esta primera parte es el estudio y análisis de los diferentes controladores de red disponibles en el mercado y poder realizar comparaciones cualitativas y cuantitativas de sus funciones y alcances.

4. Formación de Recursos Humanos

La formación de recursos humanos es una preocupación permanente. En este proyecto se tienen dos investigadores en plena formación y un estudiante avanzado que se encuentra usufructuando una beca de investigación.

5. Referencias

[1] Juniper Networks, "What's behind network downtime?" (2008). https://www-935.ibm.com/services/au/gts/pdf/200249.pdf (Consulta: 6 de agosto de 2015).

[2] Egilmez, H. E., & Dane, S. T. "OpenQoS: An OpenFlow controller design for

multimedia deliverywith end-to-end Quality of Service over Software-Defined Networks". Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE, 1–8. (2012). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6411795

[3] Hakiri, A., Gokhale, A., Berthou, P., Schmidt, D. C., & Gayraud, T. Software-Defined Networking: Challenges and research opportunities for Future Internet. Computer Networks, 75, 453–471. (2014). http://doi.org/10.1016/j.comnet.2014.10.015

[4] Xia, W.; Weng, Y.; Foh, C. H.; Niyato, D.; Xie, H. "A Survey on Software-Defined Networking". (2015). http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6834762

[5] Centeno, Alejandro & Rodriguez Vergel, Carlos Manuel & Anías Calderón, Caridad & Camilo, Frank & Bondarenko, Casmartiño & Uci,. (2014). "Controladores SDN, elementos para su selección y evaluación". Telemática. 13. 10-20.

[6] Kreutz, D., Ramos, F. M. V, Verissimo, P., Rothenberg, C. E., Azodolmolky, S., Uhlig, S., Ramos, F. Software-Defined Networking: A Comprehensive Survey. IEEE, 103(1), 14–76. (2014). https://ieeexplore.ieee.org/document/6994333

[7] Tijare, Poonam & Vasudevan, Deepika. (2016). The Northbound APIs of Software Defined Networks. 10.5281/zenodo.160891.

[8] DeCusatis, C. (Ed.). Handbook of fiber optic data communication: a practical guide to optical networking (pp. 427–445). Academic Press. (2013).

[9] ONF Open Networking Foundation. ONLINE: ttps://www.opennetworking.org/ [Dic 2019].

MODELO DE COMUNICACIÓN BASADO EN INTERNET DE LAS COSAS PARA LA GESTIÓN DE CATÁSTROFES

Salinas, Sergio Ariel[1,2,*]., Huespe, Josefina[1,2*]., Vargas, Pablo[3*], Moneti, Julio.,[1,3*] Murcia, Enzo[4,*]

[1]Departamento de Ingeniería en Sistemas de Información
Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579
[2]Facultad de Ingeniería- Universidad Nacional de Cuyo
Centro Universitario. M5502JMA. Mendoza, Argentina +54 261 4135000
[3]Instituto Tecnológico Universitario - Universidad Nacional de Cuyo
Centro Universitario. M5502JMA. Mendoza, Argentina +54 261 405 4884
[4]Facultad de Ciencias Económicas - Universidad Nacional de Cuyo
Centro Universitario. M5502JMA. Mendoza, Argentina +54 261 4135000 Int.2411

RESUMEN

Internet de las Cosas es una infraestructura que promete mejorar la eficiencia del desarrollo de actividades humanas mediante el uso de dispositivos comunicados por medio de Internet.
Esta infraestructura constituye una plataforma de desarrollo de nuevas aplicaciones basadas en las comunicaciones.
En este proyecto se propone utilizar dispositivos IoT para construir un sistema de comunicación autónomo de Internet.
Este sistema puede ser utilizado en situaciones de catástrofes en las cuales los servicios de comunicación permanecen fuera de servicio.
El proyecto incluye una línea de investigación principal enfocada en el desarrollo del modelo de comunicación y dos iniciativas complementarias. La primera está relacionada con el relevamiento de la situación local respecto a la gestión de catástrofes. La segunda se enfoca en la evaluación y prueba de dispositivos IoT.

Los principales resultados esperados del proyecto son los siguientes. En primer lugar proponer un modelo de comunicación autónomo basado en IoT. En segundo lugar conocer la situación local respecto a la gestión de catástrofes y reconocer los dispositivos IoT. En tercer lugar se espera desarrollar y potenciar habilidades de investigación de los participantes del proyecto como así también fortalecer vínculos entre distintas universidades del medio.

Palabras clave: IoT, Modelos de Comunicación en contingencias, Catástrofes, Redes ad hoc, Tolerancia a fallas.

*
s4salinas@gmail.com,
josefina.huespe@gmail.com,
pablo.vargas@itu.uncu.edu.ar,
julio.monetti@gmail.com
enzo.murcia.79@gmail.com.

CONTEXTO

La Universidad Nacional de Cuyo (UNCuyo) trabaja activamente en el incentivo de la investigación, promoviendo no solo la formación de recursos humanos sino también la articulación con instituciones, con el objetivo de aportar soluciones a problemáticas demandadas por la sociedad.

Cada dos años se realiza una convocatoria de proyectos fundamentada por la Ordenanza Nº 25/2016.

Este documento resume el proyecto aprobado por la UNCuyo denominado "Desarrollo de un modelo de comunicación basado en Internet de las Cosas".

El equipo de trabajo está formado por cinco docentes provenientes de la UNCuyo y de la Universidad Tecnológica Nactional - Facultad Regional Mendoza (UTN-FRM).

La motivación del equipo de trabajo para desarrollar el proyecto es la siguiente: IoT es una tecnología que requiere de un número significativo de dispositivos para su implementación. Los dispositivos de IoT proveen servicios de comunicación ya que este es un aspecto clave para esta infraestructura. El elevado número y densidad de dispositivos en IoT ofrece una oportunidad para crear sistemas de comunicaciones independientes de Internet. Mendoza al igual que otros lugares del país es una zona sísmica donde un escenario de catástrofe natural es factible. Utilizando los dispositivos de IoT es posible crear un sistema de comunicación de emergencias para casos de catástrofes.

La ejecución del proyecto comenzó en agosto de 2019, con una duración de dos años.

Actualmente, estamos en la etapa inicial de ejecución del proyecto y con expectativas de generar resultados de aplicación regional.

Al finalizar el proyecto se espera despertar el interés de actores en el medio local para continuar el desarrollo de modelo.

1. INTRODUCCIÓN

IoT permite una gestión inteligente de recursos heterogéneos para realizar actividades humanas de manera eficiente [4]. Estas actividades dependerán de la disponibilidad de los servicios de comunicación para su correcto desarrollo [5]. Sin embargo, esta disponibilidad puede ser afectada por fallas tales como interrupciones en el suministro de energía eléctrica, congestión de la red de datos o fallas de hardware.

En casos extremos, fenómenos naturales tales como terremotos, inundaciones, huracanes o desplazamiento de tierra pueden interrumpir el normal funcionamiento de los servicios de comunicación. Si bien, la ocurrencia de estos fenómenos es eventual y depende de la geografía donde se ubique cada ciudad, las consecuencias pueden ser graves.

Por ejemplo, en 2010 un sismo de 8,8 grados Richter afectó el centro sur de Chile. Numerosos detalles de este evento pueden encontrarse en Internet en donde se destaca el colapso del sistema de comunicaciones.

En una situación de catástrofe los servicios de comunicación desempeñan un papel importante para tareas de gestión de catástrofe.

Considerando lo expuesto, el problema que aborda el proyecto es: “la vulnerabilidad del servicio de comunicaciones de IoT en un zonas

urbanas ante fallas parciales en la infraestructura de comunicación".

En este contexto, se enmarca objetivo general del proyecto es "desarrollar un modelo de comunicación autónomo basado en dispositivos de IoT".

La hipótesis de trabajo del proyecto sostiene que "es factible desarrollar un modelo de comunicación autónomo de la infraestructura de comunicación utilizando dispositivos de IoT."

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

El eje central de investigación es el desarrollo de un modelo de comunicación autónomo basado en IoT. Sin embargo, el proyecto tiene dos líneas de investigación complementarias. Por un lado, el relevamiento preliminar de la situación local respecto a una posible situación de catástrofe en la provincia de Mendoza. Por otro lado, el análisis, evaluación y prueba de dispositivos de IoT para conocer sus características y limitaciones.

El resultado de ambas iniciativas contribuye a definir requerimientos de diseño que debe cumplir el modelo.

La línea de investigación principal contempla aspectos técnicos y no técnicos, áreas de conocimientos relacionadas con sistemas de comunicación y la factibilidad técnica de implementación.

Respecto a los aspectos técnicos es necesario observar que una catástrofe afecta a los sistemas de comunicación interrumpiendo el servicio por periodos prolongados. Un sistema alternativo de comunicación es clave en este escenario y el mismo debería ser capaz de proveer servicios mínimos de comunicación de forma confiable.

En cuanto a los aspectos no técnicos es importante considerar la factibilidad de acceso a los dispositivos. Un actor fundamental en la gestión de catástrofes es el gobierno. Los dispositivos IoT ubicados en espacios públicos se encuentran bajo el dominio del mismo de manera que es factible obtener acceso a los mismos. Para ello es necesario involucrar al gobierno en la iniciativa.

Además de los aspectos técnicos y no técnicos el proyecto contempla soluciones existentes para el desarrollo del modelo.

Las áreas de conocimientos relacionadas con las comunicaciones son extensas. Luego de un análisis preliminar el modelo considerará las redes ad hoc [6] y estrategias de tolerancia a fallas [7] de sistemas de comunicaciones existentes en la literatura. En ambos casos se deberá considerar las funcionalidades que proveen los dispositivos para IoT [8].

La factibilidad técnica para implementar el modelo es importante para su desarrollo. Los modelos que describe la literatura asumen la existencia de determinadas condiciones para su implementación. Por ejemplo, la factibilidad de comunicar dispositivos heterogéneos y la existencia de un número suficiente de dispositivos para comunicar un área geográfica determinada. En este sentido, es factible crear un modelo de comunicación autónoma si se consideran los siguientes dos aspectos.

Por un lado, los dispositivos de IoT proveen servicios de comunicación inalámbrica que permite intercambiar mensajes entre dispositivos geográficamente cercanos. Actualmente, es posible comunicar dos dispositivos sin mediar una infraestructura de comunicación subyacente si se encuentran en un radio entre 30 y 100 metros.

Por otro lado, la densidad de dispositivos en dominios de IoT será elevada, se estima que para el año 2025 existirán 7 x 109 dispositivos heterogéneos

comunicados a través de Internet [9]. Esta densidad de dispositivos permitiría el intercambio de mensajes entre distintos puntos geográficos dentro de un área determinada [10].

Como se mencionó anteriormente el proyecto contempla dos líneas secundarias de investigación.

La primera línea secundaria está relacionada con el relevamiento de la situación actual en la provincia de Mendoza para conocer los mecanismos, protocolos y planes de acción disponibles distintos actores sociales que intervienen en la gestión de una catástrofe. En este proceso, también se espera identificar los dispositivos IoT disponibles localmente o que se espera utilizar en el futuro.

La segunda línea secundaria está enfocada en el estudio de dispositivos IoT. Estos dispositivos son heterogéneos en dimensiones, peso, funciones, autonomía energética y en los protocolos de comunicación que implementan. Por este motivo es importante evaluar e identificar los posibles dispositivos factibles de utilizar en la creación de un sistema de comunicación basado en IoT.

Realizar pruebas sobre los dispositivos permitirá determinar su rendimiento y limitaciones en distintos escenarios de estrés propios de una situación de catástrofe.

Las distintas líneas de investigación del proyecto son complementarias y tienen como objetivo abordar el desarrollo un modelo que trascienda el enfoque meramente académico y sea factible de ser implementado. De esta manera se espera despertar el interés de actores locales para su implementación.

3. RESULTADOS ESPERADOS

Los resultados esperados al finalizar el proyecto son los siguientes:

1. Desarrollar un diseño de un modelo de comunicación autónomo factible de ser utilizado en escenarios de catástrofes.
2. Obtener conocimiento relacionado con propuestas recientes de redes ad hoc y estrategias de tolerancia a fallas en sistemas de comunicaciones.
3. Reconocer, evaluar y realizar pruebas utilizando diferentes tecnologías que intervienen en IoT.
4. Difundir conocimiento sobre la tecnología de IoT en el medio académico local.
5. Consolidar un equipo interdisciplinario de trabajo de investigación transformando los diferentes perfiles profesionales en una fortaleza.
6. Conocer la situación local respecto a los recursos disponibles para la gestión de catástrofes a nivel local.
7. Validar el modelo de comunicación en medios académicos y científicos mediante la publicación de avances en el proyecto a nivel nacional y provincial.

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes del proyecto provienen de áreas de conocimiento distintas: física, economía, programación, comunicaciones y sistemas de comunicación. Además, recientemente se incorporaron dos estudiantes de la tecnicatura en programación de la UTN-FRM. Por este motivo, la formación de recursos humanos es diferente para cada perfil.

En el equipo de trabajo se identifican tres perfiles: docentes con experiencia en

investigación, docentes que se inician en investigación y estudiantes.
Para aquellos docentes con experiencia en investigación se espera profundizar los conocimientos en los modelos de comunicación.
Los docentes que comienzan a recorrer el camino de la investigación es una oportunidad no sólo para adquirir conocimientos técnicos sino también aspectos relacionados con la metodología de la investigación.
Para los estudiantes representa una oportunidad de obtener experiencia práctica en un tema atractivo desde la perspectiva tecnológica.
Se espera para el equipo de trabajo integrar las fortalezas de cada disciplina en las cuales han sido formados.
Los diferentes perfiles permiten crear sinergia en las distintas etapas de desarrollo del proyecto.
Además, de la formación de recursos humanos a través del proyecto se busca generar y fortalecer vínculos entre actores de las universidades UNCuyo y UTN-FRM.
Es importante al finalizar la ejecución del proyecto generar lazos que permitan la continuidad del desarrollo de la iniciativa más allá de la fecha de conclusión del proyecto.

5. BIBLIOGRAFÍA

[1] Wang, Feng et al. "A survey from the perspective of evolutionary process in the internet of things." International Journal of Distributed Sensor Networks 2015 (2015): 9.

[2] Borgia, Eleonora. "The Internet of Things vision: Key features, applications and open issues." Computer Communications 54 (2014): 1-31.

[3] Gubbi, Jayavardhana et al. "Internet of Things (IoT): A vision, architectural elements, and future directions." Future Generation Computer Systems 29.7 (2013): 1645-1660

[4] Miorandi, Daniele et al. "Internet of things: Vision, applications and research challenges." Ad Hoc Networks 10.7 (2012): 1497-1516.

[5] Li, Shancang, Li Da Xu, and Shanshan Zhao. "The internet of things: a survey." Information Systems Frontiers 17.2 (2015): 243-259.

[6] Tie Qiu, Ning Chen, Keqiu Li, Daji Qiao, Zhangjie Fu,Heterogeneous ad hoc networks: Architectures, advances and challenges, Ad Hoc Networks, Volume 55, 2017,Pages 143-152,ISSN 1570-8705.

[7] Samira Chouikhi, Inès El Korbi, Yacine Ghamri-Doudane, Leila Azouz Saidane, A survey on fault tolerance in small and large scale wireless sensor networks, Computer Communications, Volume 69, 2015, Pages 22-37, ISSN 0140-3664.

[8] Al-Fuqaha, Ala et al. "Internet of things: A survey on enabling technologies, protocols, and applications." IEEE Communications Surveys & Tutorials 17.4 (2015): 2347-2376.

[9] Vincentelli, Alberto Sangiovanni. "Let's get physical: Adding physical dimensions to cyber systems." Low Power Electronics and Design (ISLPED), 2015 IEEE/ACM International Symposium on 22 Jul. 2015: 1-2.

[10] Internet of Things Journal: http://iot.ieee.org/iot-scenarios.html.

Análisis, simulación y estudio experimental del comportamiento de métricas de QoS y QoE de streamings de video multicast IPTV – Caso de Estudio en la red de la UTN - Mendoza

Higinio Facchini, Santiago Pérez, Fabian Hidalgo, Pablo Varela
CeReCoN (Centro de Investigación y Desarrollo en Computación y Neuroingeniería)
Facultad Regional Mendoza, Universidad Tecnológica Nacional
Rodriguez 273, Mendoza, Argentina - *0261-5244576*
(higiniofac,santiagocp)@frm.utn.edu.ar

RESUMEN

La necesidad de comunicación institucional es un elemento importante en toda organización. Las formas de dicha comunicación son variadas según los ámbitos involucrados y las posibilidades técnicas existentes. Una de las formas más actuales y de acceso masivo es el formato audiovisual, en sus distintas representaciones, como en pantallas gigantes distribuidas en los diferentes ambientes, en dispositivos ubicados en salas, aulas y/o laboratorios (pantallas tipo LED, SmartTV, etc), y a través de dispositivos individuales, como pueden ser smartphones, tablets, notebooks, etc. Mediante este trabajo se pretende analizar el comportamiento del formato IPTV (Televisión por IP), y proponer alternativas que faciliten el acceso masivo, adicionando un modelo de comunicación audiovisual. Para el desarrollo experimental se utilizará un servidor de streaming en distintos formatos (usando videos desarrollados y/o un generador sintético), considerando como base un sistema de Televisión digital por IP, emitiendo en distintos canales, en tráfico multicast. Se propone encontrar mecanismos para administrar los canales y su uso, y el desarrollo de un cliente para la recepción de la señal de IPTV, multiplaforma, y en dispositivos, tanto cableados como inalámbricos, en sus distintas presentaciones. Para que el sistema funcione en forma óptima en su conjunto, se realizarán las pruebas, las mediciones y los análisis de los requerimientos de QoS (Calidad de Servicio) y, posteriormente, las mediciones y análisis de QoE (Calidad de Experiencia de Usuario). Se propone que se prevea un canal en formato de streaming hacia Internet.

Los integrantes del Proyecto tienen experiencia en el análisis de tráfico de redes de datos y, especialmente, en el tráfico de video sobre distintos tipos de redes. Y uno de ellos desarrolla una tesis de Maestría afín al mismo. La UTN Regional Mendoza facilitará al tesista los recursos hardware y software para el desarrollo experimental, y la autorización para realizar las pruebas sobre su red.

Palabras clave: IPTV, multicast, codecs, tráfico de video

CONTEXTO

La línea de investigación está inserta en dos proyectos de análisis de tráfico multimedia, llevados adelante en el ámbito del Centro UTN CeReCoN (Centro de Investigación y Desarrollo en Computación y Neuroingeniería), del Departamento Ingeniería en Electrónica, de la Facultad Regional Mendoza, de la Universidad Tecnológica Nacional.

Los proyectos están dirigidos a investigar sobre: 1) Análisis de tráfico de video IPTV en redes cableadas e inalámbricas reales bajo modalidad multicast, y 2) Análisis de tráfico de video en redes Wi-Fi bajo distintas normas 802.11 con el análisis correspondiente de QoS.

1. INTRODUCCIÓN

Introducción a IPTV

Literalmente, IPTV significa Internet Protocol TV, o Televisión por Protocolo de Internet. Se piensa que la televisión por protocolo Internet (IPTV) se limita exclusivamente a la transmisión de programas de televisión por medio de ese protocolo. Los motivos son históricos ya que, a mediados de los años 90 se transmitieron por primera vez programas de televisión (en modo serie) por protocolo Internet, y se introdujo la abreviatura IPTV. Es una de las principales aplicaciones de las redes de la próxima generación (NGN), que se basan en el IP y ofrecen numerosas capacidades y oportunidades a los proveedores de servicios gracias a la integración y convergencia de los servicios (Figura 1).

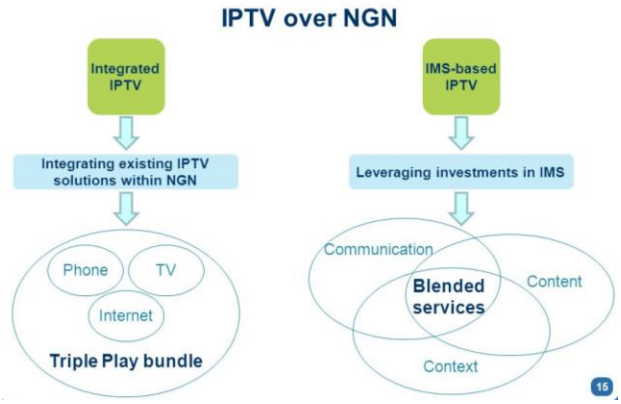


Figura 1. Convergencia IPTV sobre NGN.

Actualmente, el principal motor de la IPTV es el vídeo a la demanda (VoD), un servicio en el cual el consumidor puede seleccionar contenido de vídeo cuándo y cómo lo desea, y controlar el ritmo al cual ve el contenido (por ejemplo, puede arrancar la reproducción, ponerlo en pausa, avanzar rápidamente o rebobinar). Los proveedores de este servicio VoD pueden utilizar dos modelos, a saber, centralizado y distribuido. En el primero, todo el contenido está almacenado en un servidor central, y está principalmente configurado para servicios VoD limitados. Para la arquitectura distribuida se necesita una tecnología de distribución de contenido más inteligente. Ambas arquitecturas tienen ventajas e inconvenientes.

Para proporcionar servicios IPTV se utilizan códecs de audio y vídeo, y dispositivos de cifrado. Para VoD se utilizan por ejemplo los códecs MPEG-2 y MPEG-4. Después, los servicios IPTV llegan al consumidor a través de un cable, una línea de abonado digital (DSL) o una conexión a Internet por fibra óptica (con módems o adaptadores multimedios).

Varias organizaciones de normalización trabajan sobre la IPTV pero, hasta que el Sector de Normalización de las Telecomunicaciones de la UIT (UIT–T) comenzó a trabajar seriamente sobre el tema, no se disponía de normas mundialmente aceptadas que abarcaran todos los aspectos de una solución de extremo a extremo, es decir del proveedor al consumidor. En abril de 2006, el UIT–T creó un Grupo Temático sobre IPTV encargado de "coordinar y promover la elaboración de normas IPTV mundiales, teniendo en cuenta los trabajos en curso en las Comisiones de Estudio de la UIT, foros y consorcios".

La primera tarea importante del Grupo Temático consistió en elaborar una definición mundialmente aceptable de la IPTV, a fin de indicar sus capacidades actuales y futuras para una solución de extremo a extremo.

Definición de IPTV de la UIT-T

El UIT–T define la IPTV como sigue: "servicios multimedios, tales como televisión, vídeo, audio, texto, gráficos y datos, transmitidos por redes basadas en el IP y gestionadas para ofrecer el nivel requerido de calidad de servicio (QoS), calidad percibida (QoE), seguridad, interactividad y fiabilidad". Lo que se destaca en la definición anterior es que se insiste en el término "gestionadas", que distingue a la IPTV de otros servicios tales como los canales de televisión por Internet gratuitos o "YouTube", con los que cualquiera puede mirar programas de televisión o telecargar y/o mirar cualquier vídeo libremente, en cualquier momento y sin garantías de

calidad. Las especificaciones de la IPTV definidas y preparadas por los organismos de normalización ayudan a las empresas de explotación, los proveedores de cable y los vendedores a proporcionar servicios seguros y rentables, generadores de ingresos, comparables a los actuales, o incluso mejores. La arquitectura IPTV abarca cuatro dominios distintos, a saber, el proveedor de contenido, entidad que posee contenido o activos de contenido, o dispone de una licencia para venderlos, el proveedor de servicio, que proporciona servicios de telecomunicaciones a los consumidores, el proveedor de red, organización que mantiene y explota los componentes de red necesarios para las funciones IPTV y, por último, el consumidor, que utiliza los productos o servicios.

Arquitecturas de IPTV

El UIT–T eligió tres arquitecturas funcionales IPTV que permiten la prestación de esos servicios:

- "Arquitectura funcional no NGN IPTV", basada en los componentes de red y protocolos o interfaces existentes".
- "Arquitectura funcional basada en la NGN, no IMS IPTV", que utiliza componentes de la arquitectura de referencia NGN para permitir la prestación de servicios IPTV.
- "Arquitectura funcional IPTV basada en NGN IMS", que utiliza componentes de la arquitectura NGN, incluido el componente IMS, para permitir la prestación de servicios IPTV, junto con otros servicios IMS si son necesarios. (IMS significa "subsistema multimedios IP", arquitectura que transmite multimedios IP a dispositivos móviles).

La arquitectura de la IPTV no depende del acceso, es decir que éste puede efectuarse, por ejemplo, por línea fija, cable o sistemas inalámbricos, y los servicios se pueden transmitir a aparatos fijos o móviles. Desde el punto de vista de funcionamiento, hay cuatro elementos a tener en cuenta:

- Señal de video y sus respectivos servidores.
- Distribución de contenido.
- Software necesario.
- Suscripción y equipo de acceso

Además hay que considerar, que esta tecnología puede tener ciertos inconvenientes dependiendo de las características y funcionamiento de la red. Se pueden tomar muchos parámetros para la verificación de un funcionamiento óptimo, como ancho de banda, velocidad de enlaces, tipos de dispositivos clientes, etc. Pero una de las mejores opciones es trabajar con QoE. La calidad de experiencia (QoE, Quality of Experience) se define como la aceptabilidad global de una aplicación o servicio, tal y como se percibe subjetivamente por el usuario final. Incluye la totalidad de efectos del sistema extremo a extremo (cliente, terminal, red, servicios de infraestructura, etc) y puede verse influenciada por las expectativas de los usuarios y el contexto. Esto tiene como consecuencia que la QoE se mida subjetivamente y pueda diferir de un usuario a otro.

La calidad de la experiencia (QoE) para contenidos multimedia como IPTV también se encuentra definida por la organización de estándares de la industria ETSI TISPAN (European Telecommunications Standards Institute Telecommunications and Internet converged Services and Protocols for Advanced Networking) en su norma TR 102 479, como la experiencia que tiene el usuario al momento de hacer uso de servicios de comunicaciones o de aplicaciones proporcionadas por los CSP (Communication Service Provider), describiendo cómo le parece el servicio y si este satisface sus expectativas seriamente sobre el tema, no se disponía de normas mundialmente aceptadas que abarcaran todos los aspectos de una solución de extremo a extremo, es decir del proveedor al consumidor. En abril de 2006, el UIT–T creó un Grupo Temático sobre IPTV encargado de "coordinar y promover la elaboración de normas IPTV mundiales, teniendo en cuenta los trabajos en curso en las Comisiones de Estudio de la UIT, foros y consorcios".

2. OBJETIVOS

Objetivo General

Desarrollar, simular e evaluar experimentalmente QoS y QoE de un sistema de servicio de streaming de video y audio, en formato IPTV, con tráfico tipo multicast, en diferentes canales, emitiendo desde un servidor de streaming, y recepción usando un software cliente multiplataforma, desde cualquier dispositivo cableado y/o móvil.

Objetivos Específicos

a. Estudiar y comprender en profundidad el sistema de Televisión por protocolo IP (IPTV). Consiste en la lectura de las normas internacionales específicas para IPTV, y en documentos y trabajos de investigación asociados a la temática.
b. Desarrollar y simular el servicio de streaming de video apoyándose en frameworks específicos. Consiste en el desarrollo de trabajos experimentales del software básico de streaming con el uso de frameworks de manejo y streaming de video (del tipo ffmpeg) en el servidor.
c. Estudiar y comparar diferentes softwares clientes para recepción de IPTV. Consiste en la búsqueda, prueba y comparación de diferentes softwares clientes para recepción de IPTV para distintos dispositivos.
d. Desarrollar, simular y evaluar el comportamiento del sistema de IPTV en la red LAN de la UTN Regional Mendoza. Una vez que se hayan realizado todas las implementaciones y pruebas en una red de laboratorio, pasar el sistema a una red piloto de pruebas para medir parámetros de QoS y de QoE.
g. Desarrollar, simular y evaluar el comportamiento del sistema de IPTV como salida a Internet a través de YouTube.

3. METODOLOGÍA

En base a los objetivos propuestos de la tesis, se plantea la siguiente forma de trabajo:

1. La recopilación de información y revisión del marco teórico correspondiente.
2. La búsqueda y análisis de productos de software, librerías de streaming de video y APIs, para finalizar con la propuesta, diseño y desarrollo del prototipo de un framework de streaming de video.
3. El análisis y diseño del soporte de la infraestructura de red y hardware asociado para la implementación del prototipo del framework de video, en un escenario piloto de laboratorio controlado.
4. La implementación y prueba del framework en el laboratorio piloto.
5. El desarrollo de la aplicación cliente multiplataforma.
6. Pruebas de las aplicaciones en laboratorio piloto.
7. Ajustes, implementación y pruebas de sistema en la red en producción de la facultad.
8. Toma de muestras y análisis de QoS y QoE del sistema en funcionamiento

4. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los temas y líneas de investigación, que se tratarán durante el desarrollo del proyecto son:

- Comportamiento de streaming IPTV.
- Tráfico multicast.
- Direccionamiento multicast en IPv6.
- Protocolos de ruteo multicast en IPv6.
- Códecs de video para IPTV.
- Generación y análisis de métricas resultantes directas e indirectas. Análisis estadístico del tráfico de video.

5. ESTADO DE AVANCE

En proyectos anteriores de análisis de distintos tipos de tráficos, y específicamente de tráfico de video, se obtuvieron datos importantes en cuanto al funcionamiento y rendimiento de multicast frente a unicast, en diferentes escenarios, contemplando redes

cableadas e inalámbricas bajo el protocolo IPv4 e IPv6.

Para el tráfico de video IPTV se utilizarán videos pregrabados, o tráfico de video generado sintéticamente, con el objeto de obtener los siguientes datos primarios:

- Cantidad de bytes y paquetes por códec,
- Tasa de bits,
- Tamaño de paquetes promedio,
- Espacio intertrama,
- Distribución estadística de paquetes por orden de llegada y de espacio intertrama, y
- Comportamiento frente a requisitos mínimos de QoS.

6. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está integrado por docentes investigadores, y becarios graduados y alumnos del Centro UTN CeReCoN (Centro de Investigación y Desarrollo en Computación y Neuroingeniería) de la Facultad Regional Mendoza de la Universidad Tecnológica Nacional. Entre estos se encuentra un tesista de Maestría, que el año anterior presentó su pretesis de Maestría afín a este proyecto. Además está la relación de la materia Proyecto Final de la carrera de Ingeniería en Electrónica, en la cual se incentiva que los proyectos finales de los alumnos estén enmarcados dentro de los proyectos de investigación y desarrollo del CeReCoN. Las actividades se llevan a cabo en el ámbito de las instalaciones del Centro, que cuenta con sus propias áreas de trabajo, 1 oficina técnico-administrativa, 2 Laboratorios con 11 computadoras cada uno, con material y con el siguiente equipamiento:

- 4 Routers CISCO 2811,
- 6 Routers CISCO 1721,
- 3 Switchs CISCO 2950,
- 2 Switchs CISCO 2960,
- 2 Switchs CISCO 3560,
- 1 ASA CISCO 5505,
- 2 routers Mikrotik,
- 4 Access Point Cisco y 2 Mikrotik,
- Placas inalámbricas de red,
- 2 cámaras de video IP con soporte de streaming multicast IPv4/IPv6,
- 1 Servidor de streaming de video,
- 22 Computadoras con Sistemas Operativos Linux y Windows 7,
- Software IP Traffic de ZTI – Generador de tráfico IPv4/IPv6 unicast/multicast/broadcast y Medidor de performance (throughput, cantidad de paquetes, jitter, número de errores, tanto enviados como recibidos),
- Software Analizador de tráfico Wireshark,
- Hardware Air Pcap para captura de tráfico wireless,
- Conexión a Internet por IPv4 e IPv6,
- Servidor HP Proliant con Linux base y Máquinas Virtuales.

7. BIBLIOGRAFÍA Y REFERENCIAS

- https://ieeexplore.ieee.org/document/5446107/
- https://ieeexplore.ieee.org/document/5256148/
- https://pdfs.semanticscholar.org/de3b/223091076794e9947a0faae4803fba44e56b.pdf
- https://dl.acm.org/citation.cfm?id=2211492
- https://www.thinkmind.org/index.php?view=article&articleid=content_2012_1_50_60082
- https://www.semanticscholar.org/paper/Reliable-Application-Layer-Multicast-Over-Combined-Kobayashi-Nakayama/10e7903c83b0b2771bf72f156bd49dfb2783ecc6
- https://revistas.udistrital.edu.co/ojs/index.php/REDES/article/view/7179
- https://www.fujitsu.com/global/documents/about/resources/publications/fstj/archives/vol44-3/paper17.pdf
- https://opensource.com/article/17/6/ffmpeg-convert-media-file-formats
- https://www.labnol.org/internet/useful-ffmpeg-commands/28490/
- http://www.etsi.org/deliver/etsi_tr/102400_102499/102479/01.01.01_60/tr_102479v010101p.pdf
- https://www.itu.int/md/T05-FG.IPTV-IL-0050/es

Propuesta de métricas para comparación de Framewoks IoT

Diego A. Godoy [a], Hernán Bareiro[b], Fabian Favret[c], Juan Benitez[d], Guillermo Colloti[e], Juan Pablo Blariza[f]

Centro de Investigación en Tecnologías de la Información y Comunicaciones (C.I.T.I.C.)
Departamento de Ingeniería y Ciencias de la Producción-Universidad Gastón Dachary
Av. López y Planes 6519- Posadas, Misiones, Argentina. Teléfono: +54-376-4438677

[a] diegodoy@citic.ugd.edu.ar, [b] hbareiro@citic.ugd.edu.ar, [c] efabianfavret@citic.ugd.edu.ar, [d]juan.benitez@citic.edu.ar, [e] gcolloti@citic.ugd.edu.ar, [f] jblariza@citic.ugd.edu.ar

Resumen

En este trabajo se presenta un proyecto de investigación denominado "Tecnologías para Desarrollos Sostenibles de Ciudades Inteligentes". Particularmente en este artículo se presentan los avances realizados en relación a "Determinar el Framework de IoT más adecuado en cuanto a métricas de software para diseñar una solución que contribuya a la eficiencia energética en organizaciones". Para ello se realizarán comparaciones entre los frameworks de IoT más difundidos actualmente en base a las métricas propuestas.

Palabras claves: Home automation, Internet of Things (IoT), Framework IoT, energy efficiency.

Contexto

El trabajo presentado en este artículo tiene como marco el proyecto de investigación denominado "Tecnologías para Desarrollos Sostenibles de Ciudades Inteligentes", registrado actualmente en la Secretaría de Investigación y Desarrollo de la Universidad Gastón Dachary (UGD) con el número Código IP A10002/19 y radicado en el Centro de Investigación en Tecnologías de la Información y Comunicaciones de dicha universidad.

El mismo fue incorporado como proyecto aprobado en el llamado a presentación interna de la UGD de proyectos de investigación N°10 mediante la Resolución Rectoral (R.R). 44/A/2019 y es una continuidad de los Proyectos Simulación en las TICs: Diseño de Simuladores de Procesos de Desarrollo de Software Ágiles y Redes de Sensores Inalámbricos para la Industria y la Academia. R.R. UGD N° 07/A/17 y Simulación como herramienta para la mejora de los procesos de software desarrollados con metodologías ágiles utilizando dinámica de sistemas, R.R. UGD N° 18/A/14 y R.R. UGD N° 24/A/15.

Entre las líneas con mayores resultados dentro del proyecto referido, se encuentran las de: Construcción de una plataforma de gestión y simulación de datos de redes de sensores inalámbricos, una interfaz web para el simulador de WSN Shawn, Sistemas de gestión de residuos de la ciudad de posadas con tecnologías de Internet de la cosas, Sistema de monitoreo de la temperatura en el proceso de secado del Té.

Introducción

Las organizaciones modernas, cuentan con una gran cantidad de oficinas, cada una con numerosos dispositivos de iluminación y refrigeración (aires acondicionados). "Diversos organismos, comprometidos con el uso eficiente de la energía y la conservación de nuestro medio ambiente, han reportado que los

edificios son responsables por el consumo del 40% o más de toda la energía primaria producida a nivel mundial" [1]. De acuerdo a esto, es razonable afirmar que la energía consumida un día de clases en horario pico, resulta elevado e ineficiente en su utilización.

Es ineficiente dado que en el uso diario de las instalaciones, las tareas de control energético; por ejemplo encendido, apagado y regulación de temperatura en aires acondicionados e iluminación, se efectúa de forma manual. Esto significa que cada vez que una oficina se utiliza, una persona debe encender las luces y los aires acondicionados (en caso de ser necesario), y apagarlos en el momento que no se los requiera. Además, si tenemos en cuenta que las personas que realizan estas tareas son las mismas encargadas de muchas otras tareas y en movilidad constante, es probable que en reiteradas ocasiones queden diversos dispositivos encendidos sin necesidad, durante varias horas.

Así mismo, se observa que cada persona o grupo de personas tiene distintos requerimientos de iluminación. La temperatura claramente dependerá de las condiciones climáticas del día: con temperaturas más extremas (frío o calor) necesitarán mayor uso de aires acondicionados.

Por lo dicho, resulta de suma importancia encontrar un método para administrar de forma eficiente el consumo energético. No solo lograr una forma de reducir los gastos de una institución, sino también para contribuir a preservar recursos finitos y de esta forma atenuar el impacto ambiental por su uso innecesario.

De acuerdo a todo esto, y en vista que conforme avanza la tecnología, esta se introduce cada vez más a los procesos que se utilizan diariamente, se pretende diseñar una solución al problema mediante la utilización de un Framework de IoT, que permita a cada usuario configurar de forma independiente su perfil deseado de iluminación y temperatura para la oficina a utilizar, y que este sea aplicado de forma automática en el tiempo justo y necesario, sólo en el momento que se encuentre utilizando efectivamente la oficina.

Dado este contexto, resulta de crucial importancia elegir un framework adecuado. El mismo debería contar con documentación oficial en su sitio web, permitir integración con el hardware necesario, el desarrollo y despliegue local, de preferencia, con licencia gratuita para su utilización y distribución. Para esto, se deberá realizar una exploración de los frameworks existentes, para luego seleccionar el más adecuado considerando parámetros definidos a partir de los requerimientos y mediante métricas comparativas de software.

Línea de Investigación

En esta línea de investigación se han planteado varios objetivos. El objetivo principal de la misma corresponde es:

Determinar el Framework de IoT más adecuado en cuanto a métricas de software para diseñar una solución que contribuya a la eficiencia energética en organizaciones.

Como objetivos específicos se propusieron los siguientes: 1) Analizar bibliografía y trabajos existentes sobre IoT aplicada a la eficiencia energética. 2) Definir y aplicar un proceso de selección de dos frameworks de IoT basado en las facilidades de implementación, seleccionados de entre cinco de los más usados. 3) Diseñar un prototipo de aplicación que contribuya a la eficiencia energética en ambientes organizacionales a implementarse en los dos frameworks seleccionados. 4) Elaborar diversos escenarios que

podrían ocurrir y realizar pruebas en éstos de implementación de los prototipos con datos estáticos previamente elaborados. 5) Determinar el framework de IoT que mejor se adapte al escenario propuesto, considerando las métricas de software de los prototipos desarrollados.

Resultados

Para llegar a una solución tecnológica al problema planteado se definieron los siguientes componentes arquitectónicos.

- Hardware: posibilita las funciones de censar el ambiente, actuar sobre los dispositivos que consumen energía y transmitir datos a través de una red.
- Aplicación web: es el software que permite al usuario configurar el comportamiento de la iluminación y el acondicionamiento de temperatura mediante perfiles. Esta aplicación contiene toda la lógica del funcionamiento de los perfiles y preferencias de los usuarios y, por ende, los datos de cuándo un dispositivo debe encenderse o apagarse. No controla el hardware por sí misma, sino que se vale de un middleware que realiza la interconexión con el hardware.
- Middleware: sirve como capa de comunicación entre la aplicación web y el componente hardware. Éste es propiamente el framework de IoT, cuya funcionalidad de comunicación ya se encuentra resuelta y generalmente brindan una API REST y un SDK para su utilización. Este middleware recibirá datos, los almacenará y los mantendrá disponibles para su utilización.

La arquitectura de la solución se muestra la Figura 1:

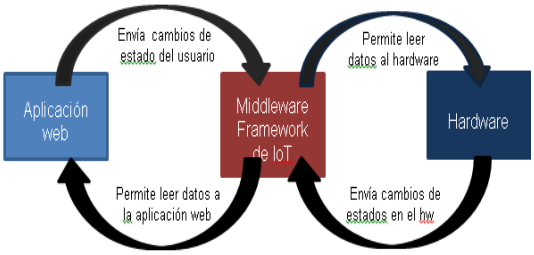


Figura 1 – Arquitectura de la solución propuesta.

Cada flecha del gráfico anterior, representa un request HTTP que realizan la aplicación web y el hardware respectivamente. En base a esto, tenemos cuatro situaciones:

1. La aplicación web envía datos de cambios de estados al framework de IoT. Estos cambios de estados se refieren a cambios en los perfiles según las preferencias de usuario.

2. Los datos disponibles son consumidos por el Hardware. Este último lee los datos, mediante un request HTTP, y se limita a ejecutar la acción.

3. Existen ciertos momentos donde el hardware precisa alimentar de información al sistema, por ejemplo, con el sensor de temperatura. En este caso, tomará la temperatura del ambiente y la enviará hacia el framework de IoT, para que ésta se encuentre disponible para quien la requiera.

4. Por último, existirán casos donde la aplicación Web requiera retroalimentarse de información del sistema, como ser, por ejemplo, cuando el hardware informa la temperatura

actual de determinada oficina. De esta manera, la aplicación Web sabrá qué información producir y enviar al framework de IoT nuevamente.

Métricas

Una métrica de software es una medida de características de software que son medibles o contables [1], esto ayuda a determinar el mejor framework del proyecto en relación al escenario y cronograma. Para la especificación de las métricas que están a continuación se tuvo en cuenta el alcance del proyecto.

Facilidad de implementación: Un framework es útil si es fácil de implementar, o en su defecto, si está bien documentado. Soluciones que impliquen dificultad en su implementación o tengan documentaciones engorrosas o incompletas, tienden a ser poco utilizadas.

Interoperabilidad: Es la capacidad que tienen los dispositivos inteligentes para comunicarse y trabajar juntos con dispositivos e interfaces que no son iguales. Otra característica que se debe considerar es el soporte para la construcción de interfaces adaptables a teléfonos inteligentes [2].

Soporte de integración: Debe proporcionar acceso a las operaciones importantes y los datos que deben ser expuestos desde la plataforma IoT [3].

Conectividad: Se refiere a la capacidad de conexión de los frameworks, debido a que los diferentes dispositivos tienen diferentes requisitos de conectividad.

Escalabilidad: Son los requisitos memoria y de velocidad de cálculo de ejecución del framework. Mientras más bajos sean los requerimientos mayores son las ventajas ya que vamos a tener en cuanto a un menor costo y tamaño físico de hardware, como así también un menor consumo de energía [2].

Seguridad: La conexión entre los dispositivos y el framework debería ser cifrada e implementar autenticación y listas de control de acceso [3].

Frameworks de IoT analizados.

Para conformar una base de selección inicial se tuvo en cuenta la popularidad de utilización por parte de los usuarios. Que cuente con documentación oficial actualizada. Que también permita integración con hardware de IoT. Desarrollo y despliegue local, de preferencia con licenciamiento gratuito de uso y distribución. Teniendo en cuenta estos criterios se seleccionaron los siguientes frameworks:

1. **SiteWhere:** Es un framework IoT de código abierto y ofrece para distintos dispositivos un sistema que facilita el almacenamiento, procesamiento y e integración de los datos [4].
2. **Eclipse Kura:** Proporciona una plataforma para construir puertas de enlace IoT. Es un contenedor de aplicaciones inteligentes que permite la administración remota de dichas puertas de enlace y ofrece una API que permite escribir e implementar nuestra propia aplicación IoT [5].
3. **Kaa:** Es un middleware open source, flexible y multiuso, para

implementar soluciones IoT de extremo a extremo, conectando aplicaciones y productos inteligentes [6].

4. **ThingSpeak:** Este framework es una plataforma IoT donde se puede almacenar, visualizar y analizar flujos de datos en la nube y desarrollar aplicaciones IoT. Para que los dispositivos se comuniquen con la plataforma deben ser compatibles con los protocolos TCP/IP, HTTP o MQTT [9].
5. **Ubidots:** Se especializa en soluciones conectadas de hardware y software para monitorear, controlar y automatizar de forma remota los procesos. En el año 2018, Ubidots desplegó la plataforma Ubidots for Education ofreciendo a los entusiastas y estudiantes de IoT un lugar donde construir, desarrollar, probar, aprender y explorar el futuro de las aplicaciones y soluciones conectadas a Internet [9].

Formación de Recursos Humanos

El equipo de trabajo se encuentra formado por tres investigadores, un Doctor en Tecnologías de la Información; un Maestreando en Tecnologías de la Información, un Ingeniero en Informática y ocho estudiantes en período de realización de trabajos finales de grado de Ingeniería en Informática de la UGD. Actualmente, el número de tesinas de grado aprobadas en el contexto de este proyecto, es de dos, y otras dos en proceso de desarrollo.

Bibliografía

[1] E.O Sosa, D.A Godoy, J. Benítez, and M.E. Sosa, "Eficiencia Energética y Ambientes Inteligentes. Investigación y Desarrollo Experimental en la UNaM," Posadas, Misiones, Argentina, 2015.

[2] Javatpoint. (2019) Software Engineering | Software Metrics - javatpoint. [Online]. https://www.javatpoint.com/software-engineering-software-metrics

[3] Erik Nimmermark and Alexander Larsson, "Comparison of IoT frameworks for the smart home," 2016.

[4] Miyuru Dayarathna. (2019, July) dzone.com. [Online]. https://dzone.com/articles/iot-software-platform-comparison

[5] SiteWhere. (2019) SiteWhere Open Source Internet of Things Platform. [Online]. https://sitewhere.io/es/

[6] Eclipse Fundation. (2019) Kura - The extensible open source Java/OSGi IoT Edge Framework. [Online]. https://www.eclipse.org/kura/

[7] Kaa. (2019) Kaa IoT platform. [Online]. https://www.kaaproject.org/

[8] (2019) ThingSpeak. [Online]. https://thingspeak.com/prices

[9] Ubidots. (2019) Ubidots: IoT platform | Internet of Things. [Online]. https://ubidots.com/

Análisis de componente de hardware para la utilización con Frameworks de IoT

Diego A. Godoy [a], Hernán Bareiro[b], Fabian Favret[c], Edgardo Belloni[d], Guillermo Colloti[e], Juan Pablo Blariza[f]

Centro de Investigación en Tecnologías de la Información y Comunicaciones (C.I.T.I.C.)
Departamento de Ingeniería y Ciencias de la Producción-Universidad Gastón Dachary
Av. López y Planes 6519- Posadas, Misiones, Argentina. Teléfono: +54-376-4438677

[a] diegodoy@citic.ugd.edu.ar, [b] hbareiro@citic.ugd.edu.ar, [c] fabianfavret@citic.ugd.edu.ar, [d]ebelloni@ugd.edu.ar, [e] gcolloti@citic.edu.ar, [f] jblariza@citic.ugd.edu.ar

Resumen

En este artículo se presenta un proyecto de investigación denominado "Tecnologías para Desarrollos Sostenibles de Ciudades Inteligentes". Particularmente se presentan los avances realizados en relación a "Determinar el Framework de IoT más adecuado en cuanto a métricas de software para diseñar una solución que contribuya a la eficiencia energética en organizaciones". Particularmente en ese trabajo se expone un breve análisis de distintas alternativas de hardware a ser utilizadas en el proyecto.

Palabras claves: Home automation, Internet of Things (IoT), Framework IoT, energy efficiency.

Contexto

El trabajo presentado en este artículo tiene como marco el proyecto de investigación denominado "Tecnologías para Desarrollos Sostenibles de Ciudades Inteligentes", registrado actualmente en la Secretaría de Investigación y Desarrollo de la Universidad Gastón Dachary (UGD) con el número Código IP A10002/19 y radicado en el Centro de Investigación en Tecnologías de la Información y Comunicaciones de dicha universidad. El mismo fue incorporado como proyecto aprobado en el llamado a presentación interna de la UGD de proyectos de investigación N°10 mediante la Resolución Rectoral (R.R). 44/A/2019 y es una continuidad de los Proyectos Simulación en las TICs: Diseño de Simuladores de Procesos de Desarrollo de Software Ágiles y Redes de Sensores Inalámbricos para la Industria y la Academia. R.R. UGD N° 07/A/17 y Simulación como herramienta para la mejora de los procesos de software desarrollados con metodologías ágiles utilizando dinámica de sistemas, R.R. UGD N° 18/A/14 y R.R. UGD N° 24/A/15.

Introducción

Las organizaciones modernas, cuentan con una gran cantidad de oficinas, cada una con numerosos dispositivos de iluminación y refrigeración (aires acondicionados). "Diversos organismos, comprometidos con el uso eficiente de la energía y la conservación de nuestro medio ambiente, han reportado que los edificios son responsables por el consumo del 40% o más de toda la energía primaria producida a nivel mundial" [1]. De acuerdo a esto, es razonable afirmar que la energía consumida en un día de clases normal en horario pico, resulta elevado e ineficiente en su utilización.

Esto resulta bastante normal teniendo en cuenta el uso y el movimiento diario de personas, como así también resulta normal que el control de temperatura (prendido, apagado y regulación de temperatura en aires acondicionados) e iluminación (apagado y prendido de luces) en las oficinas, se efectúa de forma manual. Esto significa que, cada vez que una oficina se utiliza, una persona debe encender las luces y los acondicionadores de aire (en caso de ser necesario), y apagarlos en el momento que no se los requiera. Además, si tenemos en cuenta que las personas que realizan estas tareas son las mismas personas, encargados de muchas otras tareas y en movilidad constante, es probable que en reiteradas ocasiones queden diversos dispositivos encendidos sin necesidad, durante varias horas.

Así mismo, se observa que cada persona o grupo de personas tiene su requerimiento de iluminación. La temperatura claramente dependerá de las condiciones climáticas del día: con temperaturas más extremas (frío o calor) necesitarán mayor uso de aires acondicionados.

Por lo dicho, resulta de suma importancia encontrar el método de administrar de forma eficiente el consumo energético, no solo para reducir los gastos de la institución, sino también para contribuir a preservar recursos finitos y de esta forma atenuar el impacto ambiental por su uso innecesario.

De acuerdo a todo esto, y en vista que conforme avanza la tecnología, esta se introduce cada vez más a los procesos que se utilizan diariamente, se pretende diseñar una solución al problema mediante la utilización de un Framework de IoT, que permita a cada usuario configurar de forma independiente su perfil deseado de iluminación y temperatura para la oficina a utilizar, y que este sea aplicado de forma automática en el tiempo justo y necesario, sólo en el momento que el éste se encuentre utilizando efectivamente la oficina.

Dado este contexto, resulta de crucial importancia elegir el Framework adecuado y el hardware asociado al mismo. Para esto, se deberá realizar una exploración del hardware existente, para seleccionar el más adecuado considerando parámetros definidos a partir de los requerimientos y mediante métricas comparativas propuestas.

Línea de Investigación

En esta línea de investigación se han planteado varios objetivos. El objetivo general de la misma corresponde a:

Determinar el Framework de IoT más adecuado en cuanto a métricas de software para diseñar una solución que contribuya a la eficiencia energética en organizaciones. Como uno de las cuestiones específicas surge también la necesidad de analizar y comparar hardware que se pueda utilizar conjuntamente con los frameworks.

Resultados

En el desarrollo de esta línea de investigación, se han analizado el siguiente hardware: Arduino Uno R3; NodeMCU ESP8266 y RaspberryPI 3B+.

Arduino Uno: Es una placa de microcontrolador de código abierto basado en el microchip ATmega328P.1 La placa está equipada con conjuntos de pines de E/S digitales y analógicas, que permiten la conexión a otros circuitos como así también la expansión de funcionalidades utilizando “shields” o placas de expansión. Puede ser alimentado por el cable USB o por una batería externa de 9 voltios, aunque acepta voltajes entre 7 y 20 voltios.

Arduino Uno (Figura 1) tiene 6 entradas analógicas, etiquetadas de A0 a A5, cada una de las cuales proporciona 10 bits de resolución (es decir, 1024 valores diferentes). Por defecto, miden desde tierra hasta 5 voltios, aunque es posible cambiar el extremo superior de su rango utilizando el pin AREF y la función analogReference() [2].

Características técnicas:

- Voltaje de funcionamiento: 5 voltios
- Voltaje de entrada: 7 a 20 voltios
- Pines de E/S digitales: 14 (de los cuales 6 proporcionan salida PWM)
- Pines de entrada analógica: 6
- Corriente DC por Pin de E/S: 20 mA
- Corriente CC para Pin de 3.3V: 50 mA
- Memoria Flash: 32 KB de los cuales 0.5 KB utilizados por el gestor de arranque
- SRAM: 2 KB
- EEPROM: 1 KB
- Velocidad del reloj: 16 MHz
- Longitud: 68.6mm
- Ancho: 53,4mm
- Peso: 25g

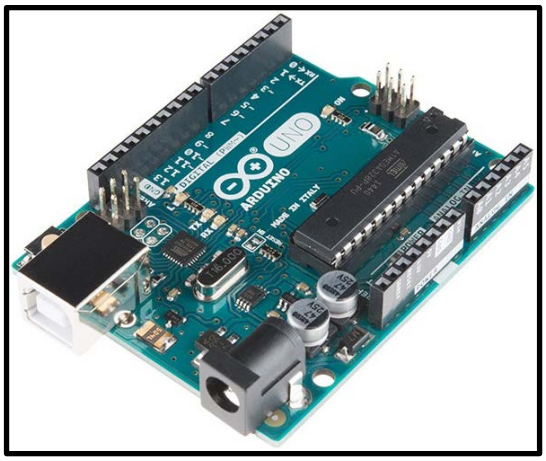


Figura 1: Placa Arduino Uno [3] [4]

NodeMCU ESP8266: Es una plataforma IoT (Figura 2) de código abierto. Incluye el firmware que se ejecuta en el SoC (Sistema en chip) WiFi ESP8266 de Espressif Systems y el hardware que se basa en el módulo ESP-12 [5].

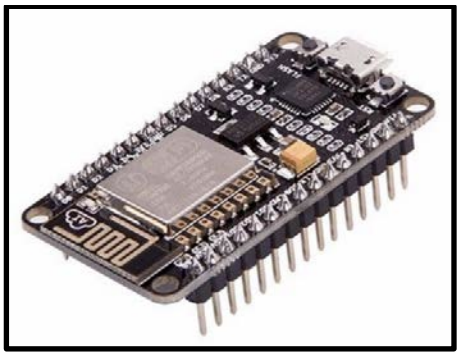


Figura 2: NodeMCU ESP8266 [6]

El ESP8266 es un chip de bajo costo WiFi con una pila TCP/IP completa y un microcontrolador.

El firmware puede ser programado utilizando el lenguaje de scripts Lua, aunque actualmente el IDE de Arduino también soporta programar firmware para dicho dispositivo en lenguaje C [7].

Características de la placa

- CPU RISC de 32-bit: Tensilica Xtensa LX106 a un reloj de 80 MHz
- RAM de instrucción de 64 KB, RAM de datos de 96 KB.
- Capacidad de memoria externa flash QSPI - 512 KB a 4 MB* (puede soportar hasta 16 MB)
- IEEE 802.11 b/g/n Wi-Fi
- Soporte de autenticación WEP y WPA/WPA2
- 16 pines GPIO (Entradas/Salidas de propósito general)
- SPI, I²C,
- Interfaz I²S con DMA (comparte pines con GPIO)
- Pines dedicados a UART, más una UART únicamente para transmisión que puede habilitarse a través del pin GPIO2
- 1 conversor ADC de 10-bit

Raspberry Pi: Es una computadora (Figura 3) de placa simple (SBC) de bajo costo desarrollado en el Reino Unido por la Fundación Raspberry Pi,

con el objetivo de estimular la enseñanza de informática en las escuelas.

En todas sus versiones incluye un procesador Broadcom, una memoria RAM, una GPU, puertos USB, HDMI, Ethernet, 40 pines GPIO y un conector para cámara. No incluye almacenamiento, pero permite la inserción de una memoria microSD para tal fin [8].

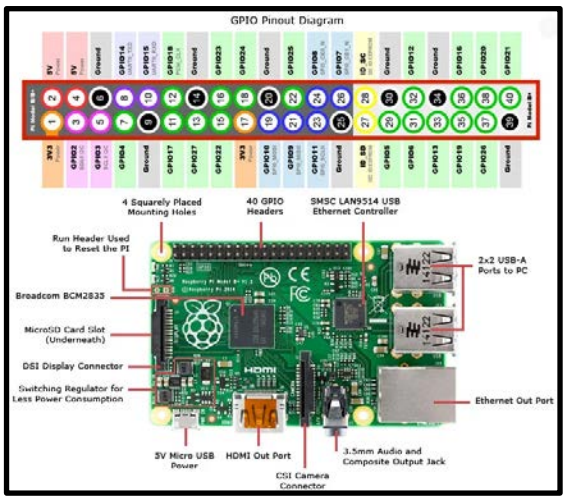


Figura 3: Configuración de pines de la Raspberry PI 3 [8].

En base a análisis que se expone a continuación para el presente trabajo, se optó por utilizar únicamente las placas NodeMCU ESP8266.

Si bien la RaspBerry PI 3 es un dispositivo altamente utilizado debido a su gran capacidad, para este trabajo en particular, permite hacer exactamente lo mismo que la NodeMCU ESP8266, pero a un costo muy superior. La NodeMCU ESP8266 no es tan potente como la Raspberry Pi, ya que la Pi es una computadora real y la NodeMCU es solo un SoC con microprocesador incorporado. Las únicas ventajas que tiene un NodeMCU sobre un Pi son su precio y tamaño, lo que le permite ser una opción obvia para aplicaciones específicas de IoT (por ejemplo: enviar valores de humedad a la nube), principalmente donde poner un Pi sería excesivo y costoso.

Arduino Uno, de igual forma, fue desplazado por placas NodeMCU ESP8266. Al iniciar el proyecto, se tenía desconocimiento del dispositivo, y por ende tampoco se sabía que la NodeMCU ESP8266 podía cumplir la misma función.

La razón de desplazar el Arduino y utilizar placas NodeMCU ESP8266, se debe principalmente a cuatro factores:

- **Tamaño**: El tamaño de la placa NodeMCU ESP8266 es tres veces inferior, siendo más fácil ubicarla en lugares pequeños.
- **Conectividad**: La NodeMCU, a diferencia del Arduino, se basa en la conectividad WiFi con el módulo ESP8266. Esta puede ser incorporada en Arduino, como "Shield" (add-on), pero también puede funcionar de manera independiente.
- **Costo**: Tomando como referencia Mercado Libre, el costo promedio de un Arduino Uno es de U$S 10, mientras que un NodeMCU ESP8266 cuesta U$S 4.50 en promedio. Para poder utilizar Arduino Uno con conectividad WiFi, se deben sumar ambos precios, resultando más del doble.
- **Rendimiento**: Tanto en lectura y escritura en los pines digitales, como así también en velocidad de procesamiento, la NodeMCU ESP8266 es muy superior a Arduino.

Formación de Recursos Humanos

El equipo de trabajo se encuentra formado por tres investigadores, un Doctor en Tecnologías de la Información; un Maestreando en Tecnologías de la Información, un Ingeniero en Informática y ocho estudiantes en período de realización de trabajos finales de grado de Ingeniería en Informática de la UGD. Actualmente, el número de tesinas de grado aprobadas en el contexto de este proyecto, es de dos, y otras dos en proceso de desarrollo.

Bibliografía

[1] E.O Sosa, D.A Godoy, J. Benítez, and M.E. Sosa, "Eficiencia Energética y Ambientes Inteligentes. Investigación y Desarrollo Experimental en la UNaM," Posadas, Misiones, Argentina, 2015.

[2] Arduino. (2019) Arduino UNO. [Online]. https://store.arduino.cc/usa/arduino-uno-rev3

[3] Reichelt. [Online]. https://www.reichelt.com/de/en/arduino-uno-rev-3-atmega328-usb-arduino-uno-p119045.html

[4] (2019) ArduinoJson. [Online]. https://arduinojson.org/v6/assistant/]

[5] NodeMcu. (2019) NodeMcu. [Online]. https://www.nodemcu.com/index_en.html

[6] Seeedstudio. [Online]. https://www.seeedstudio.com/NodeMCU-v2-Lua-based-ESP8266-development-kit.html

[7] Programar facil. [Online]. https://www.programarfacil.com/esp8266/como-programar-nodemcu-ide-arduino/

[8] Jameco. [Online]. https://www.jameco.com/Jameco/workshop/circuitnotes/raspberry-pi-circuit-note.html

Implementación del protocolo AODV en el simulador de redes Shawn

Fabián Favret [a], Hernán Bareiro[b], Diego Godoy[c], Juan Bentiez[d],
Mauricio Selzer[e], Aníbal Correa[f], Eduardo Sosa[g]

Centro de Investigación en Tecnologías de la Información y Comunicaciones (C.I.T.I.C.)
Departamento de Ingeniería y Ciencias de la Producción-Universidad Gastón Dachary
Av. López y Planes 6519- Posadas, Misiones, Argentina. Teléfono: +54-376-4438677

{efabianfavret[a], hbareiro[b], diegodoy[c], juan.benitez[d]}@citic.ugd.edu.ar;
{ maurichampy[e], anibalfranciscocorrea[f]}@gmail.com; eososa@unam.edu.ar[g]

Resumen

Las Wireless Sensor Network (WSN) conforman un tipo único de red que están formadas por un gran número de nodos con recursos limitados. Existe un interés creciente sobre la metodología y supuestos utilizados en la simulación de una WSN, hardware idealizado, protocolos simplificados y modelos de radios poco realistas que conducen a resultados erróneos. Es por ello que en este trabajo se propone la implementación del protocolo Ad hoc On demand Distance Vector (AODV) en el simulador de red Shawn [1].

Palabras claves: Protocolo Ad hoc On demand Distance Vector, Simulador Shawn, Wireless Sensor Network

Contexto

El trabajo presentado en este artículo tiene como marco el proyecto de investigación denominado “Tecnologías para Desarrollos Sostenibles de Ciudades Inteligentes”, registrado actualmente en la Secretaría de Investigación y Desarrollo de la Universidad Gastón Dachary (UGD) con el número Código IP A10002/19 y radicado en el Centro de Investigación en Tecnologías de la Información y Comunicaciones de dicha universidad.

El mismo fue incorporado como proyecto aprobado en el llamado a presentación interna de la UGD de proyectos de investigación N°10 mediante la Resolución Rectoral (R.R). 44/A/2019 y es una continuidad de los Proyectos Simulación en las TICs: Diseño de Simuladores de Procesos de Desarrollo de Software Ágiles y Redes de Sensores Inalámbricos para la Industria y la Academia. R.R. UGD N° 07/A/17 y Simulación como herramienta para la mejora de los procesos de software desarrollados con metodologías ágiles utilizando dinámica de sistemas, R.R. UGD N° 18/A/14 y R.R. UGD N° 24/A/15.

Entre las líneas con mayores resultados dentro del proyecto referido, se encuentran: Construcción de una plataforma de gestión y simulación de datos de redes de sensores inalámbricos, una interfaz web para el simulador de WSN Shawn, sistemas de gestión de residuos de la ciudad de Posadas con tecnologías de Internet de la cosas, sistema de monitoreo de la temperatura en el proceso de secado del Té.

Introducción

Las WSN se utilizan en varios campos del mundo real, estas son capaces de detectar de manera eficiente diferentes

magnitudes (temperatura, humedad, luminosidad, etc.) con alta precisión y con un consumo bajo de energía. Los avances y reducción de costos en dispositivos electrónicos y de comunicación inalámbrica, hicieron posible el desarrollo de estos dispositivos [2].

Esto conlleva a que el estudio de las WSN se ha convertido en un área de investigación de gran interés y de rápido desarrollo. Sin embargo, antes de cualquier despliegue de nodos o una optimización/modificación realizada a algún algoritmo de la WSN se debe evaluar el impacto de este cambio mediante la simulación. Esta proporciona un método rentable de evaluar la conveniencia antes de la implementación real.

Existen varios simuladores de redes ampliamente utilizados incluyendo: NS-2[1], NS-3[2], OPNET[3], MATLAB[4] y OMNET++[5], solo por nombrar algunos. La velocidad de procesamiento para redes grandes o pequeñas es limitada, aunque el aspecto funcional de los simuladores es importante, no se puede descuidar su velocidad de ejecución. Los sensores en un simulador se basan en operaciones independientes, y por lo tanto, un sensor usa un hilo de procesamiento. Por esta razón, un aumento en el número de nodos ralentiza significativamente la velocidad de simulación [3].

Una WSN está compuesta por una gran cantidad de nodos (cientos o miles), mientras que los escenarios futuros anticipan redes de varios miles de millones de nodos, lo cual compromete la elección del simulador. Una de las características más importantes que poseen los nodos en redes WSN es la habilidad de crear redes autónomas, donde todos los nodos son de gran importancia ya que se encargan del encaminamiento de todos los mensajes de control e información. Un nodo no solo monitorea y mide ciertos fenómenos, sino también entrega datos recopilados. Por lo tanto, cada nodo debe ser un enrutador inalámbrico además de ser un simple dispositivo de detección [4].

Debido al progreso dinámico en el campo de las WSN se están realizando importantes investigaciones con los protocolos de encaminamiento, ya que no todos nos ofrecen las mejores condiciones [5]. Específicamente el protocolo AODV es uno de los más investigados. En AODV la eficiencia energética es un problema crítico para los dispositivos móviles alimentados por batería en las redes. Donde el enrutamiento basado en parámetros relacionados con la energía se utiliza para extender la vida útil de la red [6].

La creciente complejidad y las limitaciones de baja potencia de las actuales WSN requieren metodologías eficientes para la simulación de redes y el análisis del rendimiento del software integrado de los nodos. En [7] la optimización del protocolo AODV realizado por los autores, utilizan el simulador OMNET++ en un escenario con pocos nodos debido a la problemática mencionada en los simuladores, debido al costo computacional que implica realizar las simulaciones y los requerimientos de hardware de alto rendimiento. Como es de esperar, existen muchos escenarios

[1] NS-2 - The Network Simulator - https://www.isi.edu/nsnam/ns/

[2] NS-3 - Ns Network Simulator 3 is a open source software - http://ns3-code.com/ns-network-simulator-3/

[3] Opnet – Optimun Network Performance - http://opnetprojects.com/

[4] Matlab - MATrix LABoratory, https://la.mathworks.com/products/matlab.html

[5] OMNeT++ - Discrete Event Simulato - https://omnetpp.org/

donde estas redes consistirán en miles de nodos, por lo que un simulador debe ser capaz de operar con tantos nodos, para solucionar la problemática y lograr aumentar la velocidad de la simulación. El simulador Shawn propone la sustitución de modelos de bajo nivel con modelos abstractos e intercambiables, es decir, simular el efecto causado por un fenómeno y no el fenómeno en sí mismo, con este enfoque se logra simular una red grande en tiempos razonables. Shawn fue desarrollado con el objetivo de contar con una escalabilidad superior, donde es posible ejecutar con éxito simulaciones con más de 100.000 nodos en computadoras con hardware estándar [1]. Pero este simulador aún no cuenta con la implementación del protocolo AODV.

Línea de Investigación y Desarrollo

En esta línea de investigación se propone implementar el protocolo de enrutamiento AODV en el motor del simulador Shawn.

Las simulaciones por computadora son un medio prometedor para abordar la tarea de la ingeniería de los algoritmos y protocolos para WSN. Hay varias herramientas de simulación disponibles que reproducen los efectos del mundo real dentro del entorno de simulación, tal como las propiedades de propagación de radio y las influencias ambientales. Estas herramientas disminuyen los esfuerzos necesarios, en comparación a las implementaciones en el mundo real y, por lo tanto, puede ayudar a aumentar el tamaño de la red, entre otras ventajas.

Sin embargo, las simulaciones detalladas de estos sistemas pueden requerir una cantidad excesiva de tiempo de Central Processing Unit (CPU). Por ejemplo, no es inusual que en las simulaciones de grandes redes se requiere cientos de horas o incluso días para obtención de resultados finales. Como consecuencia, el desarrollo de métodos para acelerar las simulaciones ha recibido recientemente un gran interés. La creciente demanda de simulación de sistemas grandes y complejos provocó nuevos desafíos en la comunidad de simulación de eventos discretos paralelos y distribuidos. Que requieren no solo la extensión y avance en las metodologías actuales de simulación paralela y distribuida, sino también el descubrimiento de enfoques y técnicas innovadoras para hacer frente a las expectativas de rápida expansión de los diseñadores de redes inalámbricas[8].

Simulador Shawn

Shawn es un simulador de eventos discretos, de código abierto, para WSN escrito en el lenguaje de programación C++. Brinda soporte para el desarrollo y prueba de algoritmos de alto nivel, así como los protocolos de red distribuidos. El objetivo principal de éste simulador es reemplazar los efectos a bajo nivel de una red, con modelos abstractos intercambiables de manera que se pueda utilizar en la simulación de grandes redes dentro de un tiempo razonable. Es por ello que un aspecto crítico en el diseño de Shawn es el de poder realizar exitosamente simulaciones de una red con más de 100.000 nodos.

El significado del término “simulación” puede variar en gran medida entre investigadores y publicaciones científicas. Es importante distinguir los distintos enfoques. El primer enfoque proporciona una imagen precisa de lo que ocurre en las redes reales y cómo los protocolos interactúan entre sí a costa de los recursos que demandan su simulación, lo que lleva a problemas de escalabilidad. Dicho de otra manera, se basan en la simulación de fenómenos físicos, tales como características de propagación de señal de radio y

protocolos de capa Open System Interconnection (OSI), por ejemplo, control de acceso al medio. El segundo enfoque se centra en los aspectos de los algoritmos realizando abstracciones de las capas de bajo nivel. Emplea modelos abstractos del mundo real, en lugar de simular hasta el nivel de bits. Su propósito es el análisis de la estructura de la red, así como el diseño y evaluación de algoritmos (y no de los protocolos) [1].

El enfoque central de Shawn es el segundo mencionado anteriormente, es decir, simular el efecto causado por un fenómeno, no el fenómeno en sí. Dado su diseño en la mayoría de los aspectos en Shawn, se puede elegir entre una implementación de modelo abstracto, simplificado o realista [1]. Un ejemplo de esto es, que en lugar de simular una capa MAC completa que incluye un modelo de propagación de radio, en Shawn únicamente se modelan sus efectos, es decir, la pérdida de paquetes y la corrupción. Un usuario puede seleccionar la granularidad y el comportamiento y, por lo tanto, puede adaptar la simulación a sus necesidades específicas.

Este tipo de modelado se adapta específicamente al paradigma de una WSN, donde el foco de la investigación de un problema de software está en entender la estructura fundamental de la red. Una tarea que a menudo está a un nivel por encima de los detalles técnicos referentes a los nodos individuales y los efectos de bajo nivel.

Shawn difiere de otras herramientas de simulación, en que no compite con estos en el área de simulación de pila de red. El objetivo principal es apoyar los pasos necesarios para lograr una implementación completa de protocolos o algoritmos. Brindando un comportamiento de red que permita obtener rendimiento y velocidad de desarrollo [1].

Arquitectura de Shawn

La arquitectura de Shawn se puede dividir en tres partes principales:

- los modelos,
- el secuenciador, y el
- ambiente de simulación

Todas las partes de Shawn son influenciadas por uno o más modelos, que son la clave de su flexibilidad y escalabilidad. Shawn hace una distinción entre lo que es un modelo y su respectiva implementación.

Un modelo es una interfaz que utiliza Shawn para controlar el proceso de simulación sin ningún tipo de conocimiento de cómo puede ser la implementación para un propósito específico. Shawn posee un repositorio de implementaciones de modelos básicos, que se pueden utilizar en la simulación general, para generar un comportamiento específico.

El secuenciador de Shawn es el encargado de controlar la simulación en general, prepara el mundo en el que viven los nodos simulados, crea instancias y parámetros para la implementación los modelos según las entradas de configuración.

El ambiente de simulación es el mundo virtual en el cual todos los objetos de la simulación residen. Dos de esos objetos que son de principal interés para el usuario son los nodos y los procesadores. Los usuarios pueden desarrollar sus propias implementaciones y utilizar los procesadores a tal fin.

Un nodo puede contener n procesadores, permitiendo de esta manera que múltiples aplicaciones puedan ser combinadas en una sola corrida de simulación, sin cambiar sus implementaciones internas. Por ejemplo, un procesador podría implementar un protocolo específico,

mientras que otro procesador recopila datos estadísticos.

Resultados y objetivos

Objetivo General

Implementar el protocolo de ruteo AODV para ser utilizado con el motor de simulación Shawn.

Objetivos específicos

- Comprender el funcionamiento interno del protocolo de ruteo AODV y del simulador Shawn,
- Modelar e implementar protocolo de ruteo AODV en Shawn a nivel de usuario o capa de aplicación.
- Validar el funcionamiento de la implementación teniendo en cuenta el Request for Comments (RFC) 3561[6].

Formación de Recursos Humanos

El equipo de trabajo se encuentra formado por tres investigadores, un Doctor en Tecnologías de la Información; un Maestreando en Tecnologías de la Información, un Ingeniero en Informática y ocho estudiantes en período de realización de trabajos finales de grado de Ingeniería en Informática en la UGD. Actualmente, el número de tesinas de grado aprobadas en el contexto de este proyecto, es de dos, y otras dos en proceso de desarrollo.

Bibliografía

[1] S. F. A. Kroeller , D. Pfisterer, C. Buschmann, S. Fekete, "Shawn: A new approach to simulating wireless sensor networks," *arXiv preprint cs/0502003*, 2005.

[2] A. G. G. Mercado, R. Borgo, "RED SIPIA: Red de Sensores Inalámbricos para Investigación Agronómica," pp. 37–41, 2011.

[3] A. N. Kang, H.-W. Kim, L. Barolli, and Y.-S. Jeong, "An Efficient WSN Simulator for GPU-Based Node Performance," *International Journal of Distributed Sensor Networks*, vol. 9, no. 10, p. 145863, Oct. 2013, doi: 10.1155/2013/145863.

[4] P. Z. B. Musznicki, "Survey of Simulators for Wireless Sensor Networks," *International Journal of Grid and Distributed Computing*, vol. 5, no. 3, pp. 23–50, 2012.

[5] S. Medina, "Comparativa de los protocolos AODV y OLSR con un emulador de redes Ad hoc," *Univ. Politec. Cataluña*, 2006.

[6] M. J. L. X. Jing, "Energy-aware algorithms for AODV in ad hoc networks," *Proceedings of Mobile Computing and Ubiquitous Networking*, pp. 466–468, 2004.

[7] H. B. J.D. Benitez, F. Favret, D. Godoy, "Ampliando la vida útil de las WSN por medio de los protocolos de ruteo, modificación de AODV," *Red de Universidades con Carreras en Informática (RedUNCI)*, pp. 152–157, 2017.

[8] J. O. L. Barriere, P. Fraigniaud, L. Narayanan, "Robust position-based routing in wireless ad hoc networks with irregular transmission ranges," *Wireless Communications and Mobile Computing*, vol. 3 Nro 2, pp. 141–153, 2003.

[6] IETF - Internet Engineering Task Force - https://tools.ietf.org/html/rfc3561

Implementación de un Servidor DNS Seguro basado en Pi-Hole utilizando un entorno virtualizado

Ernesto Sánchez, Daniel Arias Figueroa, Álvaro Ignacio Gamarra, José Nelson Mayorga

Universidad Nacional de Salta / Departamento de Informática / Facultad de Ciencias Exactas / Universidad Católica de Salta / Facultad de Ingeniería / Universidad Católica de Colombia
Av. Bolivia 5150, 3874255408 / Campus Castañares, 08105558227
esanchez@cidia.unsa.edu.ar, daaf@cidia.unsa.edu.ar, alvaroig@cidia.unsa.edu.ar, jnmayorga07@ucatolica.edu.co

Resumen

Desde su creación, el servicio de resolución de nombres proporcionado por el Sistema de Nombres de Dominio, se considera parte crítica para el funcionamiento de Internet. Esto lo convierte en el blanco de los más diversos ataques tales como redirección de consultas a sitios falsos, denegación de servicio, envenenamiento de cache, entre otros.

Como contramedida, se aunaron esfuerzos para el despliegue de extensiones de seguridad, que permitan autenticar, validar y hasta encriptar los mensajes intercambiados en el proceso de consulta/respuesta entre un cliente y un servidor DNS.

El presente trabajo, muestra una experiencia en el proceso de instalación y configuración de la solución de software Pi-Hole, para el despliegue de un Servidor DNS Resolver con extensiones de seguridad, en un ambiente virtualizado sobre VMWare y GNS3, posteriormente se realizaron capturas de tráfico de red, a fin de analizar funcionalidad y consumo de recursos.

Esta experiencia es de utilidad en la enseñanza del protocolo DNS en las asignaturas de grado y en los cursos de postgrado que organiza el Departamento de Informática de la Universidad Nacional de Salta.

Palabras clave: DNS Sumidero, DNSSEC, GNS3, Virtualización.

Contexto

La línea de investigación se encuentra apoyada por el C.I.D.I.A. (Centro de Investigación y Desarrollo en Informática Aplicada) que depende de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta, por la Facultad de Ingeniería de la Universidad Católica de Salta y por el Gobierno de la Provincia de Salta, por lo tanto, se cuenta con toda la infraestructura disponible para esta investigación. El proyecto contará con el financiamiento del CIUNSa – Consejo de Investigación de la Universidad Nacional de Salta y el financiamiento del Consejo de Investigación de la Universidad Católica de Salta.

Introducción

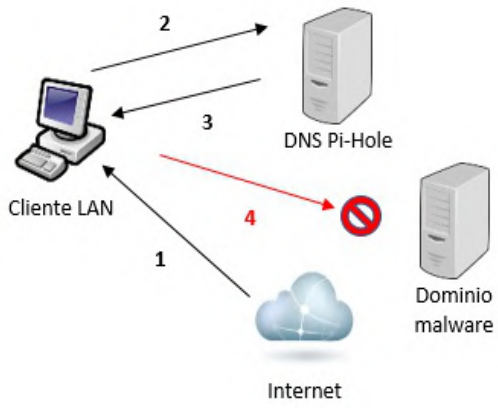
Desde su creación, el servicio de resolución de nombres proporcionado por el Sistema de Nombres de Dominio, se considera parte crítica para el funcionamiento de Internet, sin él, no sería posible el uso de aplicaciones de mensajería, redes sociales, comercio electrónico, redes privadas virtuales y tantas otras que hoy están presentes en Internet. Esto lo convierte en el blanco de los más diversos ataques tales como redirección de consultas a sitios falsos, denegación de servicio, envenenamiento de cache, entre otros. Como contramedida, se aunaron esfuerzos para el despliegue de extensiones de seguridad, que permitan autenticar, validar y encriptar los mensajes intercambiados en el proceso de consulta/respuesta entre un cliente y un servidor DNS.

La primera alternativa que logró un despliegue y adopción a escala global fue DNSSEC [1], sin embargo, este despliegue fue lento, debido a su complejidad y tareas de coordinación necesarias, dada la estructura jerárquica del sistema DNS.

Una vez superadas las etapas de pruebas y adopción, se desplegaron servicios para la resolución de nombres, proporcionados por Servidores Públicos tales como, Google DNS, OpenDNS, Cloudflare, Quad9, entre otros [2]. Tales Servidores proveen servicios extras como bloqueo de sitios con contenido malware, filtro de contenidos, optimización de tiempos de respuestas, así como aspectos de seguridad. Paralelamente, se desarrollaron soluciones de software open source, que automatizan y reducen los tiempos de implementación y despliegue de un Servidor DNS Resolver que hacen uso de los servicios proporcionados por los Servidores Públicos y brindan además un conjunto de herramientas, como el bloqueo de anuncios, gestión de "listas blancas" y "listas negras", extensiones de seguridad, todo administrable desde una interfaz web, que además permite monitorizar todo el tráfico de red intercambiado durante el proceso de resolución.

Una de estas soluciones de software es la que se presenta con el nombre de Pi-Hole, la cual se define como "*una aplicación para bloqueo de anuncios y rastreadores en Internet a nivel de red en Linux, que actúa como un sumidero de DNS, destinado para su uso en una red privada*" [3]

La siguiente figura muestra el principio de funcionamiento del servidor Pi-Hole.



El escenario anterior, presenta el caso en el que un Cliente conectado a una red LAN, solicita una página web que referencia a un dominio con contenido malware o recibe un mail conteniendo tal referencia. Se sigue la siguiente secuencia de pasos: [4]

1: Cliente LAN recibe mensaje que referencia a dominio con contenido malware.

2: Cliente LAN realiza con consulta DNS a Servidor DNS Local Pi-Hole.

3: Servidor Pi-Hole, identifica el dominio malicioso a partir de una "blacklist" y devuelve como respuesta una dirección IP loopback.

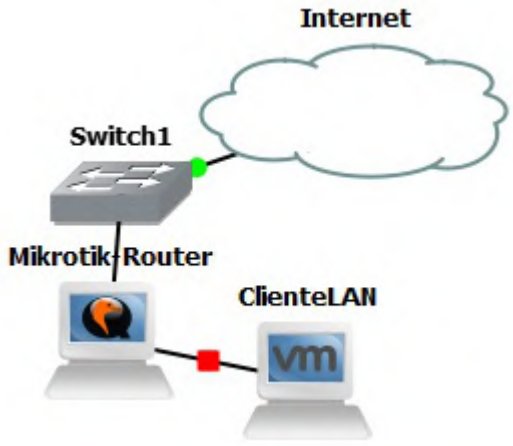
4: El cliente LAN no puede establecer conexión con el servidor con contenido malicioso.

Entorno de prueba utilizado.

A fin de poner a prueba los servicios proporcionados por el servidor Pi-Hole, se montó un entorno virtualizado utilizando las siguientes herramientas de software: [5]

- Plataforma de virtualización VMWare Workstation 15.
- Plataforma de simulación de redes GNS3 sobre Windows 10.
- Servidor Linux Ubuntu versión 16.04.6, virtualizado sobre VMWare, donde se configuró DNS Pi-Hole.

La topología creada sobre la plataforma GNS3, se muestra en la siguiente figura:



La siguiente tabla describe la configuración IP y funcionalidad de cada dispositivo:

Dispositivo	Dir. IP	Función
Nube		´Bridged a iface física
Mikrotik WAN	192.168.88.100	Enlace Internet
Mikrotik LAN	10.0.0.1	Gateway LAN
Cliente LAN	10.0.0.10	

El Servidor DNS se virtualizó sobre VMWare, con dirección IP: 192.168.88.20

Instalación y configuración DNS Pi-Hole

El proceso de instalación de la utilidad Pi-Hole se realiza en un único sencillo paso, utilizando un script disponible en el sitio oficial. Una vez instalado, se accede desde una interfaz web, a un completo panel de administración, que permite visualizar en tiempo real, cantidad de consultas procesadas, consultas bloqueadas y permitidas, entre otros datos estadísticos. Permite, además, la configuración de los servidores DNS públicos a utilizar y opciones adicionales de seguridad. [6]

Las opciones de configuración seleccionadas fueron; DNS público Quad9, y DNSSEC habilitado.

Líneas de Investigación, y Desarrollo

Los principales ejes temáticos que se están investigando son los siguientes:

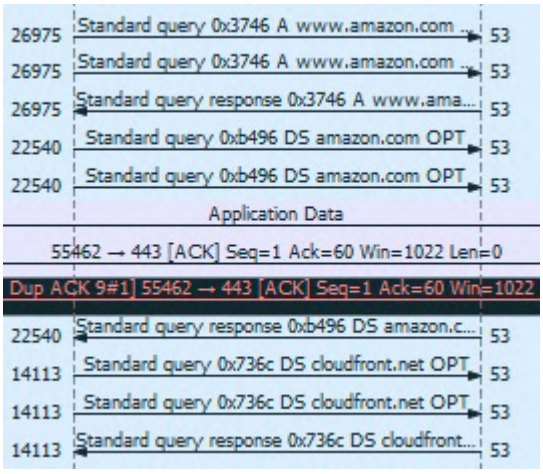
- Sistema de Nombres de Dominio.
- DNS Sumidero
- DNSSEC
- Simulación - Virtualización

Resultados Obtenidos/Esperados

Análisis de trazas consulta/respuesta DNSSEC.

Con el propósito de analizar la secuencia de los mensajes intercambiados en el proceso de resolución de una consulta DNS con extensiones de seguridad, se realizaron capturas de tráfico de red, utilizando la herramienta de software Wireshark, obteniendo los siguientes resultados.

A partir de la captura de una consulta al dominio www.amazon.com,se filtraron los mensajes correspondientes al protocolo DNSSEC, donde se observan los Registros de Recurso DS y RRSIG.



Ignorando los mensajes duplicados, la secuencia es la siguiente:

1) De cliente LAN a DNS Local Pi-Hole, consulta por dominio www.amazon.com.
2) De DNS Pi-Hole a DNS Público Quad9, consulta por dominio www.amazon.com. Se observa la presencia del bit DO activado, (Se aceptan Registros de Recursos DNSSEC).
3) Respuesta de DNS Público Quad9 a DNS Pi-Hole, indicando CNAME e IP.
4) En este punto, se inicia el proceso de verificación de la cadena de confianza, por lo que DNS Pi-Hole, realiza consulta de Registro DS (Delegation Signer) a DNS Público Quad9.
5) Servidor Quad9 responde con un Registro RRSIG, el cual contiene la firma sobre el dominio delegado, firmado por la entidad responsable del dominio .com.
6) Finalmente, el DNS Local responde al Cliente LAN, luego de haber comprobado la integridad y autenticidad de la respuesta recibida.

En conclusión, al habilitar las extensiones de seguridad, duplican la cantidad de mensajes y tiempos de respuesta, necesarios para el proceso de validación de la cadena de confianza DNSSEC, sin embargo, se consideran despreciables estos costos que aseguran una comunicación segura.

Análisis de bloqueo de anuncios

Con respecto a la funcionalidad principal del Servidor Pi-Hole, para el bloqueo de páginas con anuncios, se pudo observar que al recibir una consulta por un dominio que se encontrara en la "blacklist", Pi-Hole devuelve como respuesta un registro tipo A con dirección IP = 0.0.0.0.

A fin de conocer cuáles son los dominios bloqueados para el caso de la consulta por el dominio www.amazon.com, los dominios bloqueados son:

- fls-na.amazon.com
- s.amazon-adsystem.com
- c.amazon-adsystem.com

El listado anterior, así como otra información relevante a las consultas permitidas y bloqueadas, se obtienen a través de los reportes generados por Pi-Hole, accediendo desde su interfaz web.

En términos generales, se puede concluir que el despliegue de una arquitectura de Servidor DNS Resolver con extensiones de seguridad, requiere un mínimo de recursos, tanto de hardware como humanos. Así mismo, el proceso de instalación y puesta en marcha solo demanda unas cuantas horas.

Actualmente, existen dos implementaciones de esta arquitectura en entornos reales, una en el ámbito de la Universidad Nacional de Salta, y la otra en una dependencia del Gobierno de la Provincia de Salta. Como trabajo futuro, se pretende habilitar las opciones de extensiones de seguridad y realizar mediciones en cuanto a consumo de ancho de banda, tiempos de respuesta y consumo de recursos de procesamiento.

Formación de Recursos Humanos

El equipo de investigación se conforma con un director, dos docentes y un estudiante de intercambio de la Universidad Católica de Colombia con miras a en un futuro cercano firmar convenios de colaboración. Además se sumaran docentes de otras Universidades del país a través de convenios de cooperación que se firmarán oportunamente.

Bibliografía

[1] Tesis de Maestría, "Un estudio comparativo en Extensiones de Seguridad para el Sistema de nombres de Dominio", Mag. Ernesto Sánchez. http://sedici.unlp.edu.ar/handle/10915/63910.

[2] A List Of The Best Free and Public DNS Services https://makeawebsitehub.com/free-and-public-dns-services/

[3] Sitio oficial Pi-Hole. https://pi-hole.net/

[4] European Union Agency for Cyber Security. DNS Sinkhole. https://www.enisa.europa.eu/topics/csirts-in-europe/glossary/dns-sinkhole.

[5] Simulación de Redes con Mikrotik y GNS3. PCC en acción. MUM Argentina 2015. Ing. Álvaro Gamarra. https://mum.mikrotik.com/presentations/AR15/presentation_2878_1447676829.pdf.

[6] Pi Black Hole for Internet Advertisements. Rhythm Kr Dasgupta, JIS COLLEGE OF ENGINEERING. https://www.researchgate.net/publication/326319875

“Determinación de la eficiencia en el procesamiento sobre Arquitecturas Multiprocesador y Estrategias de Tolerancia a Fallos en HPC”

Jorge R. Osio[1,2], Diego Montezanti[1,4], Marcelo Cappelletti[1,2], Eduardo Kunysz[1], Martín Morales[1,3]

[1] Programa de Tecnologías de la información y la comunicación (TIC) en aplicaciones de interés social - IIyA - UNAJ
[2] Línea CeTAD - GCA - Instituto Leici –Fac. de Ingeniería -UNLP
[3]Codiseño HW SW para Aplicaciones en Tiempo Real - UTN - FRLP
[4]Instituto de Investigación en Informática LIDI - Fac. de Informática –UNLP

{josio, dmontezanti, mcappelletti, ekunysz,martin.morales}@unaj.edu.ar

Resumen

Dentro de la línea de investigación que se está desarrollando, existen varios enfoques. Por un lado se viene trabajando sobre la implementación de algoritmos de procesamiento de imágenes sobre dispositivos reconfigurables, utilizando una combinación de diferentes técnicas de concurrencia y paralelismo para tener en cuenta aspectos comunes de dichos algoritmos, y así mejorar la eficiencia en el procesamiento sobre las imágenes médicas. Por otra parte, debido a que el procesamiento paralelo requiere de la implementación de sistemas de múltiples procesadores, se ha trabajado en el desarrollo de metodología de tolerancia a fallos transitorios, que son cada vez más frecuentes en las arquitecturas paralelas (HPC), y que afectan especialmente a las aplicaciones de cómputo intensivo y ejecuciones de larga duración. Actualmente se está estudiando la detección y recuperación de errores en memorias y dispositivos de procesamiento sometidos a pulsos láser, técnica conocida como Láser Testing.

Palabras clave: *arquitecturas paralelas, procesamiento de imágenes, checkpoints de capa de sistema, tolerancia a fallos, sistemas multicores, Laser Testing, dispositivos reconfigurables.*

Contexto

Las líneas de Investigación descritas en este trabajo forman parte del Proyecto de Investigación Científico-Tecnológico “Tecnologías de la información y las comunicaciones mediante IoT para la solución de problemas en el medio socio productivo”, que se desarrolla en la Universidad Nacional Arturo Jauretche (UNAJ).

Parte de las líneas de investigación desarrolladas se encuentran enmarcadas en los convenios de colaboración en Actividades de Investigación firmados entre la UNAJ y el Instituto LIDI, de la Facultad de Informática, UNLP.

1. Introducción

En los últimos años, se ha buscado expandir el concepto del procesamiento paralelo con computadoras basadas en multicores hacia la utilización de plataformas de procesamiento más específicas. Para obtener mayor eficiencia, los fabricantes de computadoras de altas prestaciones, han introducido unidades FPGAs, (arreglo de compuertas programables en campo), en su diseño como soporte para el cómputo ([1-3]).

Si bien el estudio de sistemas paralelos con múltiples procesadores, es un campo bien desarrollado, la utilización de múltiples cores en sistemas reconfigurables es un terreno que tiene múltiples posibilidades de exploración [4]. En el presente proyecto se exploran mejoras en la implementación de los algoritmos mediante procesamiento paralelo de SW y Concurrencia en VHDL (lenguaje de descripción de hardware). Por otro lado, se investiga sobre la tolerancia a fallos de sistemas con gran poder de cómputo y basados en HPC. Adicionalmente, se está estudiando la detección y recuperación de errores en memorias y sistemas de procesamiento que han sido expuestos a experimentos de Láser Testing [5].

Plataformas FPGAs para procesamiento paralelo

La implementación de paralelismo en plataformas FPGAs consiste en el uso de procesadores embebidos para ejecutar aplicaciones y en la utilización de las características que provee la lógica programable para manejar las porciones de código que se ejecutan concurrentemente [4].

La facilidad de implementar procesadores embebidos en forma rápida, junto con la posibilidad de proveer concurrencia mediante la programación en HW permite combinar las FPGAs con el paralelismo obtenido mediante sistemas multicore para alcanzar la máxima eficiencia.

Entre las ventajas que proveen las FPGAs actuales para el cómputo paralelo se dispone de SoftCores (o procesadores embebidos) que permiten realizar las tareas de administración de datos. También, la posible conexión de IP-Cores (CoProcesadores) específicos que pueden realizar procesamiento de millones de MACs por segundo (operaciones de multiplicación / acumulación), poseen Delay Locked Loops (DLL) que permiten la multiplicación de la frecuencia de reloj, entre otras prestaciones.

Además, contienen bloques DRAM / SRAM de alta velocidad. Bloques RAM con capacidad de múltiples puertos para almacenar datos en paralelo.

Tolerancia a Fallos en sistemas multicore

El aumento en la escala de integración, con el objetivo de mejorar las prestaciones en los procesadores, y el crecimiento del tamaño de los sistemas de cómputo, han provocado que la confiabilidad se haya vuelto un aspecto crítico. En particular, la creciente vulnerabilidad a los fallos transitorios se ha vuelto altamente relevante, a causa de la capacidad de estos fallos de alterar los resultados de las aplicaciones.

El impacto de los fallos transitorios aumenta notoriamente en el contexto del Cómputo de Altas Prestaciones, debido a que el Tiempo Medio Entre Fallos (MTBF) del sistema disminuye al incrementarse el número de procesadores. En un escenario típico, en el cual cientos o miles de núcleos de procesamiento trabajan en conjunto para ejecutar aplicaciones paralelas, la incidencia de los fallos transitorios crece en el caso de que las ejecuciones tengan una elevada duración, debido a que el tiempo de cómputo y los recursos utilizados desperdiciados resultan mayores [11]. Esto último justifica la necesidad de desarrollar estrategias específicas para mejorar la confiabilidad y robustez en sistemas de HPC. En particular, el foco está puesto en lograr detección y recuperación automática de los fallos silenciosos, que no son detectados por ninguna capa del software del sistema, esto

significa que son capaces de corromper los resultados de la ejecución.

Análisis de los efectos únicos de fallos en FPGA

Cuando los dispositivos electrónicos trabajan en ambientes hostiles con presencia de partículas de alta energía (iones, electrones, fotones, etc.) pueden sufrir daños por ionización, la cual puede ocasionar efectos de naturaleza transitoria conocidos como efectos únicos de falla (SEE: Single Event Effects). Estos efectos de la radiación se manifiestan en los FPGA, cuando las partículas impactan sobre un área sensible de los dispositivos, generalmente como pulsos transitorios o cambios de nivel en uno o varios bits. Esto puede provocar desde la pérdida de datos de manera transitoria hasta permanente, interrupción de la funcionalidad y fallas electrónicas permanentes entre otros deterioros [12].

Antecedentes del Grupo de Trabajo

Los antecedentes del grupo de trabajo parten de la investigación acerca de procesamiento de imágenes sobre diferentes arquitecturas paralelas ([3],[4] y [7]) y de la investigación de la Tolerancia a fallos en sistemas multicore ([3], [9] y [10]). Respecto al análisis de fallos en FPGAs mediante la técnica de Laser Testing, el Dr. Cappelletti ha dirigido trabajos finales basados en esta técnica aplicada sobre Memorias EPROM y sobre microcontroladores ([5] y [13]).

El Director del proyecto participa en proyectos de investigación desde el año 2005 en la Facultad de Ingeniería de la UNLP, además dirige proyectos de investigación desde el año 2015 en la UNAJ. Actualmente, el grupo de trabajo investiga en los temas: Procesamiento Digital de Imágenes Médicas sobre plataformas FPGA, Procesamiento de Imágenes en sistemas HPRC y tolerancia a fallos pertenecientes a estudios de Maestrías y Doctorados que se realizaron en el marco de estudios de posgrado en la Facultad de Ingeniería, UNLP y acuerdos de colaboración entre el Instituto LIDI y la UNAJ a través de los respectivos proyectos de investigación.

Como resultado del trabajo realizado recientemente en la UNAJ se han publicado artículos en diferentes congresos nacionales de Ingeniería Informática y Ciencias de la Computación [3],[4] y [6-8].

2. Líneas de Investigación y Desarrollo

El grupo de investigación que se ha constituido recientemente en la UNAJ es multidisciplinario, y sus miembros cuentan con experiencia en sistemas multiprocesador, tolerancia a fallos en HPC, procesamiento de imágenes y análisis de los efectos únicos de fallos mediante Láser Testing.

Temas de Estudio e Investigación

- Implementación de un sistema multiprocesador en Dispositivos Lógicos Programables (FPGAs).
- Análisis y determinación del desempeño logrado en el procesamiento de imágenes mediante la combinación de cómputo paralelo y concurrencia.
- Diseño e implementación de una estrategia que permite al sistema detectar y recuperarse automáticamente de los errores producidos por fallos transitorios. La implementación está basada en replicación de software y es totalmente distribuida; está diseñada para operar en entornos de clusters de multicores donde se ejecutan aplicaciones paralelas científicas de paso de mensajes.
- Estudiar los efectos y aplicar la técnica de Laser Testing para poder predecir la ubicación espacial y temporal de los SEE en el interior del FPGA, con la finalidad de contribuir al diseño de dispositivos más resistentes a la radiación.

3. Resultados Obtenidos/Esperados

Investigación experimental

Hasta el momento se han obtenido resultados satisfactorios en relación a los objetivos principales:

- En cuando al procesamiento de imágenes sobre dispositivos FPGAs, se finalizó la Maestría, en donde se han implementado varios algoritmos en sistemas con dos procesadores microblaze, en donde el acceso a los datos se realizó mediante memoria compartida DDR2-SDRAM de alta velocidad y con múltiples puertos. El procesador master es el encargado de coordinar la lectura y procesamiento de los datos mediante pasaje de mensajes. Respecto a los resultados, el tiempo de procesamiento se disminuye casi al 50% al utilizar dos procesadores. Para minimizar este efecto se utilizaron CoProcesadores que ejecutan buena parte del algoritmo aprovechando la concurrencia que provee el lenguaje VHDL minimizando aún más el tiempo.

- En cuanto a la tolerancia a los fallos transitorios, se finalizó el doctorado, en donde se ha diseñado e implementado SEDAR, una metodología que permite detectar los fallos transitorios y recuperar automáticamente las ejecuciones, aumentando la fiabilidad y la robustez en sistemas en los que se ejecutan aplicaciones paralelas determinísticas de paso de mensajes, de una manera agnóstica a los algoritmos a los que protege. Aquí se aplica una estrategia de detección basada en la replicación de procesos y el monitoreo de las comunicaciones. Además, se desarrolló un mecanismo de recuperación basado en el almacenamiento de un conjunto de checkpoints distribuidos de capa de sistema. La recuperación es automática y se realiza regresando atrás al checkpoint correspondiente. Para la evaluación, se ha desarrollado un conjunto de casos de prueba. Además, se ha diseñado una estrategia de recuperación alternativa, para el caso en el que se cuente con checkpointing de capa de aplicación. En este caso, es posible almacenar sólo el último checkpoint luego de validarlo.

En la línea de tolerancia a fallos, se ha diseñado e implementado una metodología distribuida basada en replicación de software, diseñada específicamente para aplicaciones paralelas científicas de paso de mensajes, capaz de detectar los fallos transitorios que producirían resultados incorrectos y recuperar de manera automática las ejecuciones [9]. En este tipo de aplicaciones, cuyos procesos cooperan para obtener un resultado, la mayor parte de los datos relevantes son comunicados entre tareas. Por lo tanto, la estrategia de detección consiste en la validación de los contenidos de los mensajes a enviar y de los resultados finales. Esta solución representa un término medio entre un alto nivel de cobertura frente a fallos y la introducción de un bajo *overhead* temporal. De esta forma, la corrupción de los datos que utiliza un proceso de la aplicación se mantiene aislada en el contexto de ese proceso, evitándose la propagación a otros procesos. Así, no solamente se mejora la confiabilidad del sistema, sino que también disminuye la latencia de detección, y por lo tanto el tiempo luego del cual se puede relanzar la aplicación.

Para recuperar al sistema de los efectos del error, la propuesta se basa en restaurar la aplicación a un estado seguro previo a su ocurrencia. Para lograr este objetivo, se ha integrado la detección con un mecanismo basado en múltiples *checkpoints* coordinados en capa de sistema, construidos con la librería DMTCP (que proporcionan cobertura en el caso de que un *checkpoint* resulte afectado por un fallo), o la utilización de un único *checkpoint* no coordinado de capa de aplicación (construido ad-hoc, basándose en el conocimiento de la aplicación), que puede ser verificado para asegurar la integridad de sus datos.

El uso de estas estrategias posibilitará prescindir de la utilización de redundancia triple con votación para detectar y recuperar de fallos transitorios, proveyendo opciones que proporcionan flexibilidad a los

requerimientos del sistema. Además, como estos fallos no requieren reconfiguración del sistema, la recuperación puede realizarse mediante re-ejecución en el mismo core en el que ocurrió el fallo.

Respecto al análisis de los efectos únicos de fallos mediante Láser Testing en FPGAs, se comenzará desarrollando una plataforma de permita la detección y recuperación de errores.

4. Formación de Recursos Humanos

Dentro de la temática de la línea de I+D, todos los miembros del proyecto participan en el dictado de asignaturas de la carrera de Ingeniería Informática de la UNAJ dentro del Área Arquitectura, Sistemas Operativos y Redes.

En este proyecto existe cooperación a nivel nacional. Hay dos Doctores, un Magíster en temas de ejecución de algoritmos de procesamiento de imágenes en sistemas multiprocesador, un Investigador Doctorado en tolerancia a fallos en sistemas HPC y un investigador iniciando el Doctorado en temas relacionados con análisis de los efectos únicos de fallos mediante Láser Testing en FPGAs.

Adicionalmente, se cuenta con la colaboración de estudiantes avanzados.

5. Bibliografía

[1] O. Mencer, K. Tsoi, S.Craimer, T. Todman, W.Luk, Ming Wong and P. Leong ,"CUBE: a 512-FPGA Cluster", Dept. of Computing, Imperial College London , Dept. of Computer Science and Engineering The Chinese University of Hong Kong . (2009)

[2]Keith Underwood, "FPGAs vs. CPUs: Trends in Peak Floating-Point Performance", Sandia National Laboratories. (2011)

[3]J. Osio, D Montezanti, E. Kunysz, Morales M., "Análisis de eficiencia y tolerancia a fallo en Arquitecturas Multiprocesador para aplicaciones de procesamiento de datos ", UNNE, Corrientes, WICC 2018.

[4] J. Osio, J. Salvatore, E. Kunysz, V. Guarepi, M. Morales, "Análisis de Eficiencia en Arquitecturas Multiprocesador para Aplicaciones de Transmisión y Procesamiento de Datos", UNER, Ciudad de Concordia, WICC 2016.

[5] I. Garda, A. Cédola, M.A. Cappelletti, F. San Juan y E.L. Peltzer y Blancá, "Design and Implementation of a Measurement Unit for Laser Testing of Semiconductor Memories", IEEE Xplore Digital Library. ISBN 978-987-1907-44-1. Pages: 96-101 (2013).

[6] J. Osio, D. Montezanti, M. Morales, "Análisis de Eficiencia en Sistemas Paralelos", Ushuahia, Tierra del Fuego, WICC 2014

[7] E. Kunysz, J. Rapallini, J. Osio,"Sistema de cómputo reconfigurable de alta performance (Proyecto HPRC)", 3ras Jornadas ITE - 2015 -Facultad de Ingeniería – UNLP

[8]J. Osio, D Montezanti, E. Kunysz, Morales M., "Determinación de la eficiencia y Estrategias de Tolerancia a Fallos en Arquitecturas Multiprocesador para aplicaciones de procesamiento de datos", UNSJ, San Juan, WICC 2019.

[9]D. Montezanti, A. De Giusti, M. Naiouf, J. Villamayor, D. Rexachs, E. Luque, "A Methodology for Soft Errors Detection and Automatic Recovery", in Proceedings of the 15th International Conference on High Performance Computing & Simulation (HPCS). ISBN: 978-1-5386-3250-5/17. IEEE, 2017, pp. 434

[10] F. Cappello, A. Geist, W. Gropp, S. Kale, B. Kramer, and M. Snir, "Toward exascale resilience: 2014 update," Supercomputing frontiers and innovations, vol. 1, no. 1, pp. 5–28, 2014.

[11] Grama A, Gupta A, Karypis G, Kumar V. "Introduction to parallel computing".Pearson Addison Wesley, 2003.

[12] F. Lima Kastensmidt et al., "Laser testing methodology for diagnosing diverse soft errors in a nanoscale SRAM-based FPGA," IEEE Trans. Nuclear Science, vol. 61, no. 6, pp. 3130-3137, 2014.

[13] María Florencia Stanley, Beca de Estímulo a las Vocaciones Científicas CIN-UNLP. Duración 1 año. Período Octubre 2013-Septiembre 2014.

Entorno de integración continua para validación de sistemas embebidos de tiempo real

Waldo Valiente, Esteban Carnuccio, Mariano Volker, Graciela De Luca, Raúl Villca, Matías Adagio
Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza
Dirección: Florencio Varela 1703 – CP 1754 – {wvaliente, ecarnuccio, mvolker, gdeluca}@unlam.edu.ar, {raul.villcasd, mati.adagio}@gmail.com

RESUMEN

Cada año se desarrollan billones de sistemas embebidos. Por lo que el proceso de selección del dispositivo a utilizar se vuelve una tarea compleja. En parte por la cantidad de fabricantes y características en sus modelos, sumado a la necesidad del cumplimiento de requisitos no funcionales, tales como la limitación en el consumo energético y la criticidad de los tiempos de respuesta a eventos externos. Este trabajo se centrará en que esta tarea sea factible en tiempos razonables desde el punto de vista del negocio, como así también del proceso de selección/elaboración de un sistema embebido de tiempo real para lograr, incluso, su óptima replicación y fabricación. Por ese motivo se pretende desarrollar un entorno de mejora continua que facilite el desarrollo, pruebas y optimización de los sistemas embebidos. La mejora continua permite generar ciclos de medición y retroalimentación de las tres partes que forman el proceso de desarrollo. Además se pretende que el conjunto pueda ejecutar sobre distintos dispositivos, tanto real como virtual. Permitiendo así la correcta selección del dispositivo, incluso antes de ser adquirido.

Palabras clave: *Sistemas Embebidos, Tiempo Real, Optimización, Virtual, Mejora Continua.*

CONTEXTO

Nuestra Línea de Investigación es parte del proyecto "*Entorno de integración continua para validación de sistemas embebidos de tiempo real"*, dependiente de la Unidad Académica del *Departamento de Ingeniería e Investigaciones Tecnológicas*, perteneciente al programa de Investigaciones CyTMA2 de la Universidad Nacional de La Matanza, el cual es formado por docentes e investigadores de la carrera de ingeniería en informática. Este proyecto es continuación de los trabajos que viene realizando el grupo de investigación en sistemas operativos, computación de alto rendimiento y en el área de Internet de las cosas.

1. INTRODUCCIÓN

En sus inicio los Sistemas Embebidos (SE), se requerían para funcionalidades sencillas, con el fin de obtener programas de reducido tamaño, debido a las limitaciones en la memoria que disponía el hardware en esa época. Con el paso del tiempo y gracias a los avances tecnológicos, los microcontroladores incrementaron la cantidad de memoria. Por ende se volvieron más sofisticados, por lo que requirió mayor complejidad en la programación, hasta el punto de agregar funcionalidades complejas que pertenecen al ámbito de los Sistemas Operativos de Tiempo Real (RTOS). Con ello aparecieron múltiples modelos de dispositivos,

confeccionados por diferentes fabricantes, cada uno con su propia configuración y conjunto de herramientas de uso. Con el auge de Internet de las cosas apareció un creciente desarrollo de los SE, por año se crean 6 billones de SE nuevos [1], mientras que computadoras de escritorio, solo rondan los 100 millones [2]. Existen proyectos científicos que acompañan esta evolución. Por un lado la *National Science Foundation*, el año pasado finalizó un proyecto de optimización semántica para RTOS que utilizan SE [3]. Mientras que la Comunidad Europea, planteó un proyecto con el objetivo de bajar el consumo energético para RTOS con altas prestaciones que ejecutan en SE [4]. Así mismo desarrollar software para Sistemas Embebidos de Tiempo Real (RTSE) es una tarea compleja. Está afirmación es justificada por la experiencia adquirida de parte del equipo de investigación en los proyectos "Sistema de monitoreo y alarma para personas adultas mayores" y "Dispositivo de asistencia de personas mediante monitoreo y análisis de datos en la nube".

Problemática a investigar:

Los dispositivos utilizados en SE poseen múltiples diferencias, ya que cada uno cuenta con sus herramientas propias por lo que trabajar con ellos requiere de reunir constantemente habilidades y conocimientos técnicos para desenvolverse en dicha área. También existen dos complicaciones en la selección y el uso de cada microcontrolador, ya que cada uno posee una tecnología propia. La primera de ellas surge debido a la correcta elección del dispositivo, más allá de las especificaciones básicas, junto con el costo y la facilidad de adquisición en el mercado local. Cuando se plantean requisitos no funcionales estrictos, como el tiempo de respuesta ante distintos eventos, o que el consumo energético permita que el dispositivo funcione con batería asegurando la disponibilidad, existe una amplia variedad de modelos, que pueden estar sobre dimensionados o no llegar a satisfacer estos requerimientos. No es económicamente viable adquirir tantos modelos de dispositivos para elegir el que mejor se adapte a los requisitos. Por lo que se propone de utilizar un SE virtual para realizar las pruebas de rendimiento, verificando distintas optimizaciones, antes de llegar a adquirir el hardware del dispositivo. La segunda complicación, ya con el dispositivo seleccionado, surge por el amplio conjunto de herramientas suministrado por los fabricantes de SE. Estas se enfocan en el modelado, desarrollo, depuración e implementación del programa en el SE. Por lo que se busca armar un entorno de trabajo que permita realizar la ejecución de SE virtuales y reales, y medir las características de optimización que aseguren los requisitos no funcionales, en el uso de los recursos del SE.

Por otra parte se encontraron problemas comunes inherentes a la funcionalidad de los RTSE, estos se describen a continuación.

Problemas Comunes

Según [5], "En sistemas embebidos el objetivo principal de la programación del proceso es garantizar una respuesta rápida a eventos externos y cumplir el tiempo de ejecución del mismo". Para lo cual las funciones de código específicas pueden dividirse en tareas. Cada una puede tener una prioridad, esta dependerá de su periodicidad o si son activados como respuesta a eventos externos. En los SE suele haber procesadores de un solo núcleo, por eso lo que se suele multiplexar uso del procesador. Esto se efectúa para alternar el uso del procesador entre tareas, generando concurrencia. Para los SE complejos pueden tener procesadores de varios núcleos, logrando verdadera paralelización de tareas o realizando distintas tareas concurrentemente, lo que puede ocasionar retardos no buscados por problemas de

sincronización debido al uso de semáforos, memoria compartida, tuberías, entre otros.

Los sistemas operativos que al principio se ejecutaban en un SE, eran programas con funciones muy simples de control, que esperaban eventos desde el exterior para funcionar. Actualmente como la demanda de los sistemas multifuncionales aumenta, también lo hace la infraestructura donde se ejecutan. Por lo que se dispone de sistemas operativos cada vez más complejos, como los Smartphone que tienen casi tantas características computacionales como los de una computadora de escritorio. Algunos poseen memoria en el orden de los Gigabytes, procesadores con varios núcleos y pantallas de alta definición. Por lo que ya no se suele realizar diseños propios del sistema operativo para cada SE. Lo que se realiza es adaptar el sistema operativo tradicional, mediante configuraciones a estas arquitecturas, por ejemplo Android baso su versión de GNU/Linux o Microsoft Windows® para Microsoft Surface Go®.

En el último tiempo hubo un incremento exponencial de Sistemas Embebidos de Tiempo Real (RTOS) [6]. El requisito de Tiempo Real es necesario para completar su trabajo y entregar servicios regularmente, ya que tienen requisitos de tiempo estricto que deben cumplirse. Todos los días estos sistemas proporcionan tareas importantes, por ejemplo al conducir, al volar, monitoreo de pacientes, etc. La diferencia de estos sistemas con el de PC tradicional, es que los últimos ejecutan aplicaciones complejas que no requieren de tiempo real, como navegadores, procesadores de texto, entre otros. La vertiente actual busca utilizar los recursos de SE junto con funciones específicas de concurrencia que poseen los Sistemas Operativos tradicionales. RTOS es la opción que logra combinar ambos mundos, ya que están preparados para ejecutar tareas que poseen requerimiento de tiempos estrictos. Un RTOS debe responder inmediatamente a eventos externos con la menor latencia de interrupción. Además, debe completar los pedidos de servicio antes que se cumpla el tiempo límite. Para lograr cumplir con estos requisitos tienen las siguientes capacidades [5]: Menor latencia de interrupción, regiones críticas cortas, administrador de tareas apropiativo y algoritmos de planificación dinámicos. Los dos primeros requisitos tienen que ver con acortar los tiempos de ejecución de RTOS y que las tareas tengan mayor tiempo de procesador. Los otros dos requisitos tienen que ver como son administrados los tiempos de procesador, permitiendo a tareas de mayor prioridad cumplir su tiempo límite. Algunos de los RTOS que cumplen con estos requisitos son FreeRTOS, MicroC/OS, NuttX y QNX.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Por esta creciente necesidad de programas más robustos se tuvo que establecer un proceso de desarrollo para SE, este se compone de tres etapas [7], representadas en (Fig. 1):

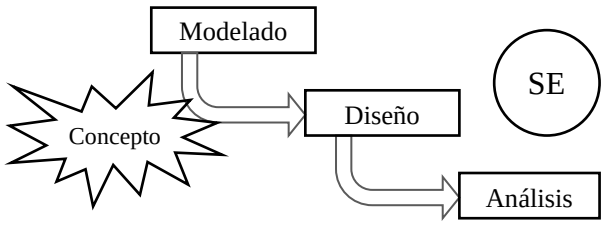


Fig. 1 Etapas del proceso de construcción de SE.

La concepción del SE parte desde una idea o concepto que se debe plasmar en el hardware.

1°) La primera etapa, la del Modelado, se centra en la definición de la dinámica del comportamiento de las partes que formarán el SE, como los eventos físicos serán medidos y convertidos en pulsos eléctricos por los sensores. El paradigma que representa mejor esta lógica es a través de máquinas de estados concurrentes, donde los sensores van

registrando eventos que accionan funcionalidades del SE para dar una salida por sus actuadores. Las herramientas que suelen utilizarse para el modelado son *Simulink*® y *LabVIEW*®. Estas son herramientas que permiten realizar pruebas, mediciones, controles y depuración en un entono simulado o real y embebido. Posibilita realizar todas las verificaciones antes de construir el hardware. Este programa puede conectarse al software de diseño electrónico llamado Proteus, a través de la interfaz emulador virtual serial [8], permitiendo así la simulación del microcontrolador. Cabe aclarar que estos programas son de licenciamiento pago.

2°) La etapa correspondiente al Diseño se centra en cómo se realiza la integración de las partes del SE y como obtener datos precisos de los sensores, ya sea por calibración o filtrado. Así como la funcionalidad que debe realizar el microcontrolador con esa información (sobre cómo el sistema interactúa con su entorno). En esta etapa se plantean las técnicas que debe emplearse para cumplir con requisitos específicos sobre el SE. Estas técnicas pueden requerir atender a varios sensores simultáneamente, como así también enfocar el diseño para lograr un menor consumo energético, para que funcione con baterías. Otro requisito, que también se diseña, tiene que ver con la forma de realizar cálculos complejos en el propio SE. Las herramientas que se utilizan en esta etapa son provistas por el fabricante de SE, ya que permite programarlo, muchas de ellas se basan en la interfaz de programación Eclipse. Para arquitecturas Arduino: Eclipse C++, para procesadores ARM: posee un conjunto de herramientas con licenciamiento pago *KIEL*® y la versión libre *System Workbench*, entre otros.

3°) La etapa de Análisis. Es utilizada para garantizar que el SE resultante, cumpla con la funcionalidad correcta, como con los requisitos planteados en las etapas de Modelado y Diseño. Debe incluir especificaciones, tales como seguridad, consumo, fiabilidad, entre otras, que deben ser definidas como propiedades en forma precisas, expresadas técnicamente, para evitar ambigüedad, facilitando la recolección y comparación de resultados. El análisis debe ser realizado cuantitativamente para lograr validar las definiciones de características o alcance del SE. En consecuencia existen técnicas para analizar tiempos dinámicamente [6], tales como: **Circuito Emulador:** Es un dispositivo de propósito especial que se comporta como un procesador particular, pero con mejores capacidades para depuración e inspección del programa que ejecuta. **Osciloscopios:** Son herramientas destinadas a medir con exactitud atributos de las señales eléctricas que recorren circuitos, por lo que pueden ser utilizados para medir tiempos de respuesta de las entradas y salidas de SE. **Analizadores lógicos:** Son dispositivos que permiten leer la memoria o el bus de direccionamiento y analizar las instrucciones que son leídas. **Dispositivo de rastreo:** Los puertos de rastreo son incorporados a los microcontroladores actuales. Estos permiten depurar el programa en ejecución e inspeccionar sus datos. **Temporizadores en el código:** Se puede agregar dentro del programa, funciones específicas que miden el intervalo de tiempo transcurrido en ejecutar partes del código. **Contadores de procesador:** Procesadores como X86, PowerPC o MIPs poseen contadores dentro del procesador, que se incrementan a medida que ejecutan instrucciones. **Herramientas de inspección de desempeño:** Obtienen el rendimiento del hardware, a través de código agregado por el compilador. **Simuladores:** Programa que trata de reproducir

la conducta de un microcontrolador real [2], a los fines de validar la funcionalidad, medir tiempos y atributos de la ejecución. **Emulador software:** Programa que permite ejecutar el programa SE como si ejecutara sobre la arquitectura real. Son de fácil acceso en las primeras etapas del desarrollo, aun sin contar con el hardware real.

3. RESULTADOS OBTENIDOS/ESPERADOS

Con el propósito de armar el entorno de trabajo de SE sé planifica el proyecto dividiéndolo en dos etapas. En la primera etapa se realizarán análisis sobre sistemas embebidos disponibles en el mercado, además de los tipos de RTOS que se pueden ejecutar sobre los mismos. Para lograr comparar sus características, ventajas y desventajas, se desarrollarán algoritmos y casos de prueba, que se almacenarán en repositorios para su utilización por parte de la comunidad. Estos se comparará sobre SE reales y virtuales, que serán configurados en las distintas pruebas. Se realizarán las compilaciones en el sistema embebido y se ejecutarán automáticamente con alguna herramienta de optimización de desarrollo, que serán evaluadas. Para el segundo año del proyecto se concentrara en el desarrollo de las pruebas automatizadas con herramientas disponibles con licencia GPL o similar. Estas herramientas están destinadas a mejorar el desarrollo de sistemas embebidos de tiempo real, depurando o midiendo rendimiento. Luego se realizarán análisis y pruebas sobre las optimizaciones que se puedan utilizar. De esta manera con la integración de las herramientas y las pruebas automáticas se podrá verificar y comparar el impacto de las optimizaciones sobre distintos SE. Permitiendo la elección del SE que se adecue a las necesidades.

4. FORMACIÓN DE RECURSOS HUMANOS

La presente línea de investigación dentro del departamento de Ingeniería e Investigaciones Tecnológicas forma parte del trabajo que uno de los investigadores se encuentra realizando para su maestría. Completan el grupo de investigación dos de docentes de categoría V y dos ingenieros en formación de investigador.

5. BIBLIOGRAFÍA

[1] X. Fan, Real-Time Embedded Systems: Design Principles and Engineering Practices, 1st ed., Newton, MA: Newnes, 2015.

[2] M. Barr y A. Massa, Programming Embedded Systems: With C and GNU Development Tools, 2 ed., Mumbai, India: O'Reilly Media, Inc., 2006.

[3] F. Eric, "Semantics of Optimization for Real Time Intelligent Embedded Systems (SORTIES)," National Science Foundation, Alexandria, Virginia 22314, USA , 2018.

[4] M. Bertogna, «High-Performance Real-time Architectures for Low-Power Embedded Systems,» 31 12 2018. [En línea]. Available: https://cordis.europa.eu/project/rcn/199161/factsheet/es. [Último acceso: 09 2019].

[5] K. C. Wang, Embedded and Real-Time Operating Systems, 1st ed., vol. 130, Springer Publishing Company, Incorporated, 2017, pp. 24, 36.

[6] I. Lee, J. Y.-T. Leung y S. H. Son, Handbook of Real-Time and Embedded Systems, 1st ed., Chapman & Hall/CRC, 2007.

[7] E. A. Lee y S. A. Seshia, Introduction to Embedded Systems: A Cyber-Physical Systems Approach, 2nd ed., The MIT Press, 2016.

[8] R. Singh, A. Gehlot, B. Singh y S. Choidhury, Arduino-Based Embedded Systems: Interfacing, Simulation, and LabVIEW GUI, CRC Press, 2017.

RED DE IoT: ESTIMACIÓN DE CONSUMO DE ENERGÍA

Matías Siracusa*, Carlos Taffernaberry*, Gustavo Mercado*,Diego Dujovne§, Sebastian Tobar*, Ana Lattuca*

*GridTICs – Grupo en Tecnologías de la Información y las Comunicaciones

Departamento de Electrónica / Facultad Regional Mendoza / UTN- Argentina

{matias.siracusa, carlos.taffernaberry,sebastian.tobar, ana.lattuca}@gridtics.frm.utn.edu.ar

§Escuela de Informática y Telecomunicaciones, Universidad Diego Portales, Chile

diego.dujovne@mail.udp.cl

RESUMEN

Este trabajo tiene como objetivo proponer, desarrollar y evaluar alternativas de solución Multicast, para redes de IoT, basadas en algoritmos y procesos que permitan una mejora en el uso de los recursos y su escalabilidad. La aplicación concreta de estas técnicas se puede aplicar en nodos sensores que deban enviar información a múltiples nodos destino, o cuando un nodo intenta difundir una orden (una consulta, una actualización de firmware, un comando, etc.) a un grupo de nodos de sensores. Se debe tener presente que la transmisión de datos es el factor que mayor consumo de energía produce en un nodo.

Palabras Clave: IoT, Multicast, consumo energía, 6LoWPAN, CoAP

CONTEXTO

El presente trabajo se desarrolla en el ámbito de un proyecto de investigación acreditado por la Universidad Tecnológica Nacional con código CCUTIME0005150TC denominado "Multi-IoT: Mejora del desempeño en redes 6LoWPAN utilizando técnicas de Multicast". El proyecto es llevado adelante por investigadores y becarios de la Facultad Regional Mendoza, con el apoyo de Investigadores formados de la Universidad Diego Portales.

Este trabajo es la evolución de distintos proyectos anteriores llevados a cabo por el grupo Gridtics[1] [2], mejorando y cambiando el hardware utilizado, como así también protocolos de comunicaciones y software, siempre buscando mejoras en eficiencia y autonomía del sistema.

1 INTRODUCCIÓN

6LoWPAN [3] es una pila (stack) de protocolos definido por la publicación 4944 (Request For Comments 4944) del Grupo de Trabajo de Ingeniería de Internet (IETF) que posibilita a redes inalámbricas de área personal (Wireless Personal Area Networks - WPAN) el acceso a Internet. Un tipo particular de WPAN son las redes de nodos sensores inalámbricos (WSN), por lo que, usando este protocolo, cada nodo puede interactuar, con cualquier otro dispositivo conectado a Internet, constituyendo actualmente el standard más relevante en la Internet de las Cosas (IoT). Uno de los campos de mayor interés reciente en esta área es mejorar la eficiencia en el uso de recursos limitados que tienen los nodos de estas redes (ancho de banda, energía, almacenamiento, capacidad de procesamiento) y la escalabilidad para el encaminamiento (routing) de los datos. En forma genérica, el uso de técnicas de encaminamiento Multicast

permiten reducir considerablemente el tráfico de una red cuando los contenidos son compartidos por un número importante de nodos. Sin embargo, en la actualidad, los algoritmos asociados a esta alternativa sólo responden a adaptaciones incompletas de sus variantes en redes jerárquicas y redes móviles ad hoc (MANET), sin considerar una solución integral a partir de los requerimientos y restricciones de redes WPAN.

2 LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Uno de los objetivos de este proyecto es "Desarrollar una metodología de medición de consumo energético de una red 6LoWPAN para evaluar bajo la misma base las distintas propuestas de algoritmos de ruteo y sus mejoras."

Para la implementación de esta metodología se decidió trabajar con TSCH, Contiki-NG y hardware de OpenMote, a continuación se describe su utilización.

- 2.1 TIME-SLOTTED CHANNEL HOPPING (TSCH)

Uno de los desafíos más grandes de la comunicación IoT reside en obtener alta confiabilidad con mínimo consumo energético. Se han propuesto numerosos protocolos de comunicación que intentan resolver este dilema, entre ellos el protocolo TSCH Time-Slotted Channel Hopping [4] el cual fue propuesto como un mecanismo de capa MAC en el protocolo 802.15.4e. TSCH utiliza salto de frecuencias para mejorar la confiabilidad, reduciendo así el efecto de la interferencia externa y el desvanecimiento de múltiples trayectos. Para disminuir el consumo se utiliza una planificación que indica a un nodo exáctamente cuándo debe transmitir o recibir, evitando así desperdiciar energía en períodos de competencia por la captura del canal. Además pueden los nodos desactivar su hardware de comunicación cuando no se encuentre en uso.

En redes TSCH, cada nodo utiliza una planificación común. Esta planificación se divide en ranuras de tiempo (*timeslots*), los cuales tienen una duración entre 10 a 15ms. El conjunto de todos las ranuras de tiempo en una ronda de transmisión se denomina *slotframe* y su longitud depende de la cantidad de nodos que formen parte de la red. TSCH contempla además un salto en frecuencia, con lo que a la coordenada de tiempo se le agrega una coordenada de frecuencia, resultando en una matriz bidimensional como la que se observa en la figura 1.

Una ranura de tiempo dentro de un timeslot puede ser de cuatro tipos: RX, TX, shared y off. RX y TX indican que un nodo deba recibir o transmitir respectivamente, mientras que shared permite el acceso de cualquier nodo mediante un mecanismo de competencia. Estos tipos de ranuras de tiempo son las que permiten la sincronización de la planificación entre nodos, así como la unión de nuevos nodos a la red. Conociendo la planificación, un nodo puede fácilmente determinar en qué intervalo de tiempo debe mantener su radio encendida y cuando puede apagarse para ahorrar energía.

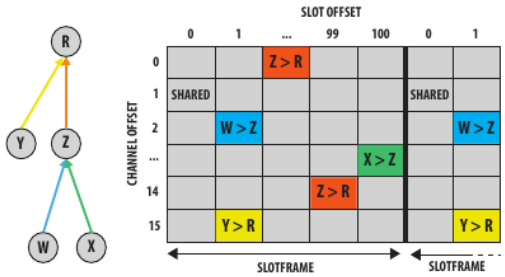


Figura 1: Ejemplo de planificación TSCH

Es importante destacar que el protocolo TSCH no define un mecanismo para iniciar y mantener la planificación, sino que los detalles de la implementación quedan a criterio del usuario.

Para habilitar TSCH en Contiki-NG, es necesario habilitarlo en el Makefile del proyecto

```
MAKE_MAC = MAKE_MAC_TSCH
```

Adicionalmente se debe realizar la configuración necesaria para que el hardware de los nodos soporte TSCH. Con lo anterior

queda implementada una función para el inicio y mantenimiento de la planificación.

- 2.2 CONTIKI-NG

El sistema operativo utilizado es Contiki-NG en la versión 4.4 [5]. Este sistema operativo se enfoca en comunicación de baja potencia en dispositivos de bajos recursos computacionales. Incluye soporte por defecto para protocoles estándar IETF, entre otros: IPv6/6LowPAN, 6TiSCH, RPL y CoAP.

Brevemente se exploró la posibilidad de utilizar OpenWSN[6], otro sistema operativo con énfasis en dispositivos de bajos recursos, sin embargo se encontró documentación confusa y a menudo obsoleta. En contraste, la documentación de Contiki-NG se mantiene actualizada y posee una comunidad dispuesta a evacuar dudas en la plataforma gitter.

Para la estimación del consumo, Contiki-NG utiliza un módulo de software llamado Energest. Provee una estimación basada en software aproximada para dispositivos motes de IoT. Al rastrear los diversos componentes de hardware, como la radio, cpu, etc. y conocer el consumo de energía cada componente, es posible estimar el consumo de energía.

El módulo Energest realiza un seguimiento de cuándo varios componentes se encienden y apagan. Al saber cuánto tiempo han estado los componentes en diferentes estados y el consumo de energía de cada uno de los componentes, es posible estimar el consumo de energía.

- 2.3 OPENMOTE

En el escenario presentado en este trabajo, el dispositivo utilizado es el *Openmote CC2538[7]*, un dispositivo modular que utiliza *open hardware*. Éste consta de CC2538 SoC (System on Chip), con un procesador ARM Cortex-M3 de 32 MHz y con 32 Kb de RAM y 256/512 Kb de memoria FLASH. Cuenta además con un trasceptor 802.15.4 de 2,4GHz y soporta múltiples periféricos como USB, NVCI, I2C, SPI, UUART entre otros. Presenta tres sensores digitales: temperatura/humedad, luminosidad y acelerómetro. El rango de transmisión es de 50 metros en interiores y de 200 metros en exteriores.

Se decidió trabajar con los mismos, debido a su existencia en el laboratorio GridTics.

3 RESULTADOS OBTENIDOS

La arquitectura ensayada al momento consta de 4 Openmote CC2538, los cuales forman un Destination Oriented Directed Acyclic Graph (DODAG) generado en base al protocolo de ruteo RPL-Lite.[8], como se puede observar en la Figura 2.

En la raíz del DODAG se ubica el router de borde, el cual corre el programa rpl-border-router, utilizando parte de los ejemplos provistos por Contiki-NG. La función de este nodo es actuar como punto de acceso para los nodos y enrutar las comunicaciones provenientes desde estos hacia Internet.

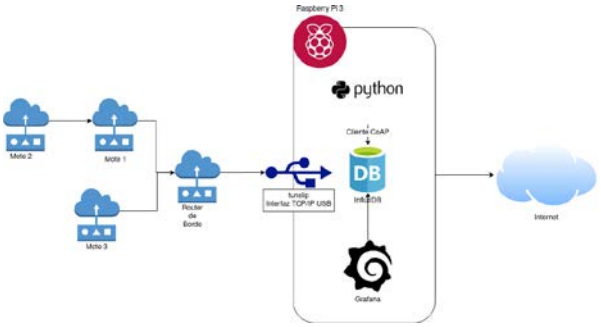


Figura 2: Arquitectura para ensallos

En el extremo de la red se ubica una Raspberry Pi, la cual sirve como plataforma para consultar a los nodos y almacenar información. Para poder conectar la *Raspberry Pi* al router de borde, es necesario utilizar el protocolo SLIPv6 (Serial Line Interface Protocol). Este protocolo crea un puente entre la red RPL y la Raspberry Pi, permitiendo que los nodos tengan acceso a Internet.

El resto de los nodos que forman parte del DODAG, ejecutan una aplicación servidor CoAP [9], el cual responde consultas con información de temperatura, humedad y luminosidad. La aplicación crea recursos propios de *Openmote CC2438* y un recurso adicional CoAP para consultar el consumo energético del nodo,entregado por el módulo

Energest.

CoAP (*Contrained Application Protocol*) es un protocolo de comunicación especializado para dispositivos de bajos recursos computacionales, definido en RFC 7252. Corre sobre UDP y es similar a HTTP.

Cada lectura tomada por los sensores del Openmote se representa como un recurso CoAP, el cual puede ser accedido en una URL propia. Si, por ejemplo, la dirección IPv6 de un nodo es 2001:0db8:85a3::8a2e:0370:7334, el recurso temperatura se puede leer consultando la URL: coap://[2001:0db8:85a3::8a2e:0370:7334]/sensors/temperature y, de manera similar, el recurso luminosidad se encontrará disponible en coap://[2001:0db8:85a3::8a2e:0370:7334]/sensors/light.

La mayor dificultad en la implementación la presentó el cliente que realiza consultas a los servidores CoAP en los nodos. Se decidió dejecutar en la placa *Raspberry Pi el* programa que haga las consultas, puesto que dispone de mayores recursos computacionales y energéticos que el router de borde, para poder mantener comunicaciones con todos los nodos. También en la placa *Raspberry Pi* se encuentra alojada la base de datos, por lo que el programa que realiza las consultas CoAP también se encarga de almacenar los resultados obtenidos en ésta última.

Adcionalmente, se eligió a *InfluxDB[10]* como tecnología para almacenar los datos provenientes de los servidores CoAP. *InfluxDB* es una base de datos orientada a series de tiempo (*Time series database*) optimizada para almacenar datos colectados a intervalos regulares y en grandes volúmenes, tales como lectura de sensores o telemetría.

La lectura de los sensores se realiza mediante consultas CoAP de manera muy similar a un HTTP GET Request. El código del cliente se desarrolló en Python 3 y se utilizó la librería aiocoap. Ésta tiene la particularidad de utilizar funciones no bloqueantes de Python de manera tal que se puedan consultar todos los nodos de la red de manera asincrónica. Es importante destacar que, de utilizar métodos síncronos, la consulta a un nodo en particular, bloquearía el programa hasta que éste no regrese una respuesta. En el caso que su batería se haya agotado, produciría un retardo importante, por los reintentos que haría el cliente.

El funcionamiento del programa cliente se detalla en el diagrama de secuencia en la Figura 3. La comunicación la inicia el cliente para descubrir los nodos que conforman el DODAG. Para esto se hace una consulta HTTP al router de borde. El firmware de éste último incluye un servidor HTTP que contiene la lista actualizada de los nodos de la red. El cliente los obtiene a partir de esta respuesta y crea una estructura de datos con los nodos y sus respectivas direcciones IPv6.

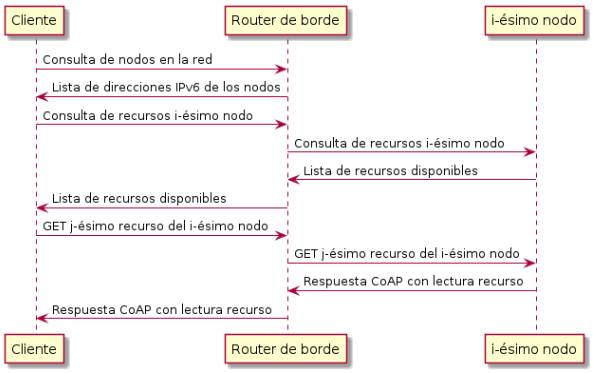


Figura 3: Secuencia de comunicación

El cliente luego itera por esta estructura y consulta el recurso =.well-known/core= de cada nodo. Éste es un endpoint en el cual el servidor CoAP registra los recursos disponibles. La lista de los recursos CoAP se agrega a una tabla hash de tal manera que la llave de dicha tabla es la dirección IPv6 de los nodos y el valor es una lista con los recursos disponibles.

Finalmente el cliente formatea una serie de peticiones que son enviadas de manera asíncrona a los nodos a través del router de borde. Las respuestas devueltas por los nodos son almacenadas en la base de datos, junto con etiquetas que identifican el recurso CoAP al cual se asocia el valor y el nodo del cual proviene. Opcionalmente se puede configurar a nivel de código etiquetas como

geolocalización o fecha y hora. Pero estos valores no son leídos desde los nodos, sino que son generados por software.
Justamente este punto nos encontamos en la actualidad. A futuro inmediato se comenzarán a medir los consumos de energía, para distintos algoritmos de ruteo multicast, para finalmente proponer las mejoras de los mismos.

4 FORMACIÓN DE RECURSOS HUMANOS

Desde la creación del GridTICs, uno de los objetivos a realizar fue la capacitación de los recursos humanos tanto del grupo como externos. Esta actividad de formación se viene cumpliendo desde el comienzo, junto a la divulgación de las tecnologías de redes de sensores. La meta como investigadores es fortalecer la capacidad para realizar investigación científica, generar conocimientos y facilitar la transferencia de tecnología que permita el desarrollo humano.
Este proyecto de investigación posibilita la colaboración inter-institucional y la ejecución de proyectos conjunto entre grupos I+D .
Los cantidad de integrantes que componen el presente proyecto son:
- Dos Investigadores formados
- Tres Investigadores de apoyo
- Dos Becarios graduado (Beca BINID UTN)
- Dos Becarios alumnos (Beca alumno UTN)
- Un Tesista de carrera de grado

Para lograr estos objetivos se desarrollan:
- Dictado de cursos, seminarios y conferencia para público especializado de la región.
- Promoción, coordinación y asistencia técnica de tesinas de grado para alumnos de ingeniería de sistemas de información e ingeniería electrónica de la frmza
- Promoción, coordinación, dirección y asistencia técnica a tesis doctorales, postgrado y/o maestría.
- Presentación de trabajos en congresos y reuniones técnicas/científicas.
- Publicación de trabajos en revistas con/sin referato.

5 BIBLIOGRAFÍA

[1] A. Diedrichs, C. Taffernaberry, G. Mercado, G. Grunwaldt, M. Pecchia, G. Tabacchi, M. González, N. Altamiranda, “RED SIPIA-LP Estudio de mecanismos de bajo consumo energético aplicados a Red de Sensores Inalámbricos en el ámbito de Agricultura de Precisión”, WICC 2016, ISBN: 978-950-698-377-2, Abril 2016.
[2] C. Taffernaberry, G. Mercado, “GW-CIAA-IoT: Gateway con CIAA para red inalámbrica de IoT”, WICC 2016, ISBN: 978-950-698-377-2, Abril 2016
[3] S. Chakrabarti, “IPv6 over Low-Power Wireless Personal Area Network (6LoWPAN)”, RFC 8066, ISSN: 2070-1721, IETF, February 2017
[4] S. Duquennoy, A. Elsts, B. A. Nahas and G. Oikonomo, "TSCH and 6TiSCH for Contiki: Challenges, Design and Evaluation," *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, Ottawa, ON, 2017, pp. 11-18.
[5] Contiki-NG: disponible en https://github.com/contiki-ng consultado el 29 de Marzo de 2020.
[6] “OpenWSN” disponible en http://www.openwsn.org/ consultado el 29 de Marzo de 2020.
[7] OpenMote: Open-Source Prototyping Platform for the Industrial IoT Conference Paper · September 2015 ISSN 1867-8211
[8] T. Winter, Ed. “RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks”, RFC 6550, ISSN: 2070-1721, IETF, March 2012
[9] Z. Shelby, K. Hartke, C. Bormann “The Constrained Application Protocol (CoAP)” RFC: 7252 ISSN: 2070-1721 June 2014
[10] “InfluxDB open-source time series database”. Disponible en: https://www.influxdata.com consultado el 29 de Marzo 2020.

Sistema de Entrenamiento para Navegación a Vela Ligera de Alto Rendimiento basado en Dispositivos Multisensoriales de Bajo Costo (categorías olímpicas, veleros de 1 o 2 tripulantes).

Secretaría de Investigación y Postgrado – Facultad de Ciencias Exactas, Químicas y Naturales - Universidad Nacional de Misiones
Félix de Azara 1552 – Posadas – Misiones – CP 3300

Ing. Rubén Luis María Castaño
rcastano@gmail.com

Ing. Selva Nieves Ivaniszyn
selvanieves@gmail.com

Lic. Claudio Omar Biale
claudio.biale@gmail.com

RESUMEN

El trabajo tiene por objeto el Desarrollo de un Sistema de Entrenamiento para Navegación a Vela Ligera de Alto Rendimiento basado en Dispositivos Multisensoriales de Bajo Costo (categorías olímpicas, veleros de 1 o 2 tripulantes) de la escuela de Vela Ligera de la Provincia de Misiones.

En una primera etapa se determinará con los entrenadores los datos a ser censados de cada embarcación, de cada competidor y de los puntos de control. Se evaluarán las mejores técnicas para la obtención de las mediciones como así también las técnicas para su almacenamiento, transmisión, análisis y presentación en un tablero de control que permita realizar el seguimiento de cada competidor o de la flota en tiempo real o hacer un análisis posterior para discutir con los competidores las técnicas y estrategias empleadas. Esta información de apoyo permitirá a los entrenadores realizar prácticas personalizadas en busca de un desempeño de alto rendimiento competitivo.

PALABRAS CLAVE: Navegación a Vela, Sailing, Yachting, Dispositivos Multisensoriales, IMU, GPS, Tracking, Low Cost.

CONTEXTO

Con la denominación "Desarrollo de un Sistema de Entrenamiento para Navegación a Vela Ligera de Alto Rendimiento basado en Dispositivos Multisensoriales de Bajo Costo (categorías olímpicas, veleros de 1 o 2 tripulantes)" el proyecto dio inicio el primero de enero de 2019, habiendo sido aprobado y dado de alta por Resolución 0150/2019.

La Provincia de Misiones se caracteriza por estar rodeada de importantes ríos de Sudamérica. En particular la ciudad de Posadas se halla ubicada a orillas del Río Paraná.

En cuanto a la actividad de navegación, Posadas, se ha posicionado como una de las canchas predilectas a nivel nacional desde que el Río Paraná aumentó su cota por los trabajos realizados en la Represa de Yacyretá. Las canchas se destacan por su amplitud, la poca corriente de sus aguas y la constante de sus vientos.

En los últimos 15 años se observa el desarrollo de una actividad en particular como una Escuela de Vela Ligera, que ha ido creciendo año tras años, dándose a conocer provincial, nacional e internacionalmente por la participación destacada de sus deportistas en competencias náuticas de vela ligera a nivel nacional e internacional.

Los Entrenadores locales de reconocida trayectoria han formado parte de los equipos nacionales de entrenadores para diversas competencias de las copas sudamericanas y mundialistas.

Como Profesor Regular Adjunto a cargo de las asignaturas de Comunicaciones y Redes,

Sistemas Operativos y Sistemas Distribuidos de la FCEQyN de la UNaM se busca un campo de aplicación de los contenidos abordados en las mismas. La problemática planteada puede encuadrarse en los temas de geoposicionamiento, triangulación, tratamiento de señales, digitalización, transmisión de datos, almacenamiento, administración eficiente de la energía, Sistemas embebidos, procesamiento distribuido, IoT entre otros, todos aspectos tratados en dichas asignaturas.

Actualmente muchas de las soluciones provienen del área de los sistemas embebidos y de la Internet de las Cosas (IoT), contemplando la integración de plataformas de Hardware y Software con énfasis en las comunicaciones, la ubicuidad y el procesamiento distribuido. Esto se puede apreciar en diversas publicaciones científicas y en las últimas recomendaciones y estándares de los organismos internacionales referentes como ser la IEEE y la ITU.

Algunas publicaciones relevantes vinculadas al objeto de estudio son:

P. Prempraneerach, K. Thamchaitas and P. Kulvanit, "Autonomous waypoint tracking of kayak boat using state- variable feedback control," 2012 Oceans - Yeosu, Yeosu, 2012, pp. 1-7. doi: 10.1109/OCEANS-Yeosu. 2012.6263616

N. D. Simões, J. L. Gonçalves, M. L. Caeiro, M. J. Boavida and F. D. Cardoso, "ZigBee/GPS tracking system for rowing races," 2011 IEEE EUROCON - International Conference on Computer as a Tool, Lisbon, 2011, pp. 1-4. doi: 10.1109/EUROCON.2011.5929248

I. Augustyniak, G. Hapel, P. Woś, J. Dziuban and P. Knapkiewicz, "“Mechatronic yacht” — A training unit,"

Perspective Technologies and Methods in MEMS Design, Polyana, 2011, pp. 127-129.

A. Maciuca, G. Stamatescu, D. Popescu and M. Struţu, "Integrating wireless body and ambient sensors into a hybrid femtocell network for home monitoring," 2nd International Conference on Systems and Computer Science, Villeneuve d'Ascq, 2013, pp. 32-37.

IEEE Colloquium on Motion Analysis and Tracking

1 INTRODUCCIÓN

La Provincia cuenta con una Ley de Promoción del Deporte en apoyo a los deportistas y clubes fundamentalmente en la representación de la provincia en competencias de orden nacional e internacional.

Existe interés por parte de los entrenadores de la escuela local de vela ligera en aplicar tecnología para la mejora de las estrategias de prácticas y entrenamientos en forma personalizada y grupal de los deportistas.

Se estima que la solución obtenida se podrá aplicar con pocas modificaciones a otros deportes de alto rendimiento como ser ciclismo, maratón, entre otros.

En relación de los recursos humanos representará una experiencia en cuanto a la aplicación de soluciones a problemáticas del medio e interés social en referencia a la contención durante la infancia, adolescencia y juventud a través del deporte como así también con impacto a personas adultas sin límites de edad en la práctica de esta actividad.

En cuanto al impacto tecnológico, acompaña a otras actividades de investigación y desarrollo que se vienen aplicando desde la Universidad, posicionando a la misma como referente en actividades de Investigación y Extensión y presencia en la Comunidad local y Regional.

El uso de redes de área amplia de baja potencia es una solución alternativa al problema de consumo de energía en los nodos y a la transmisión en grandes distancias. Diversos trabajos implementan soluciones de monitoreo mediante tecnología LoRa, en [18] se logra transmitir hasta una distancia de 2 km con una pérdida de paquetes menor al 5% en zonas planas. En [19] se utiliza una red para controlar botes de la clase Optimist. En [20] se logra seguir a 255 objetos logrando una transmisión de datos en buenas condiciones hasta 3 km.

2 LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación, desarrollo e innovación que se inició y se encuentra en proceso es el desarrollo de un prototipo a ser

usado en los veleros para determinar su ubicación mediante GPS, analizar ajustes mediante información provenientes de las unidades inerciales IMU y la transmisión de datos mediante tecnología LoRa.

Se realizará un análisis de protocolos de comunicación LoRa, el diseño de los paquetes de datos e impacto de la pérdida de paquetes.

En cuanto a equipamiento se dispone de cinco placas de desarrollo Arduino Uno, un Raspberry pi 3, módulos de WIFI, sensores de temperatura y humedad (AM2302) y compás digital (HMC 5883L).

Se planea la adquisición de módulos LoRa basados en el módulo SX 1276, unidades GPS basados en módulos L86-M33 y unidades inerciales MPU6050.

3 RESULTADOS OBTENIDOS / ESPERADOS

Los principales resultados del Proyecto se evidenciarán a través de los distintos hitos que permitirán evaluar el avance de las actividades.

Objetivos e Hitos de Evaluación

- Análisis, desarrollo e integración de componentes (Hard y Soft) para la medición y recolección de datos mediante sistemas embebidos aplicados a la problemática.
  - Obtención de prototipo para recolección y medición de datos.
- Análisis y selección de métodos para la determinación de la posición, velocidad y dirección de las embarcaciones ligeras (GPS, Sistemas de balizamiento, etc.)
  - Obtención de prototipo para la determinación de la posición, velocidad y dirección de la embarcación.
- Análisis y selección de métodos para el almacenamiento, transmisión de la información atendiendo a la preservación del consumo de energía.
  - Obtención de prototipo del enlace de comunicaciones para la transmisión de los datos censados.
- Gestión del conocimiento de las técnicas de entrenamiento y análisis del impacto de la tecnología aplicada.
  - Obtención de base de datos de los datos censados y análisis estadístico de los mismos.
- Desarrollo de un Sistema de Entrenamiento para Navegación a Vela Ligera de Alto Rendimiento basado en Dispositivos Multisensoriales de Bajo Costo (categorías olímpicas, veleros de 1 o 2 tripulantes).
  - Obtención de Tablero de Control con la información para el seguimiento de los deportistas, las técnicas de entrenamiento aplicadas y la evaluación de la performance de los mismos.

4 FORMACIÓN DE RECURSOS HUMANOS

Los integrantes del grupo de investigación son Investigadores y Auxiliares en Investigación, categorizados y no. Docentes con amplia trayectoria en las cátedras afines con el Proyecto y que han integrado grupos de trabajo sólidos en diversas actividades y proyectos.

Como aportes del grupo se pueden considerar los siguientes:

- Director y Vicedirector de Proyecto han tomado cursos de timonel de navegación a vela.
- Director y Vicedirector de Proyecto se hallan vinculados a Clubes de Vela participando en diferentes oportunidades en la organización y realización de eventos de competencias y acompañando a integrantes de la familia a otros eventos interactuando con entrenadores locales e internacionales.
- Docentes integrantes del proyecto son responsables de cátedras como Sistemas Operativos, Sistemas Distribuidos, Comunicación y Redes, Introducción a las Bases de Datos, Ingeniería de Software, Programación Orientada a Objetos.
- Director y Vicedirector de Proyecto integrantes de la Comisión de Creación de

la Carrera Ingeniería en Computación y actualmente integrante de la Comisión de Seguimiento Curricular y Directora de Carrera, respectivamente.

- Integrantes del Proyecto han tomado diferentes cursos de postgrado en tecnologías orientadas a la producción como así también en el desarrollo regional.
- Elaboración de RSL para trabajo final de tesis de maestría en el tema del proyecto en cuestión.
- Se ha aprobado por el Comité Académico de la Maestría en Tecnologías de la Información un Plan de Trabajo Final de Tesis de Maestría a la espera de la Resolución de CD de FCEQyN - UNaM.

5 BIBLIOGRAFÍA

[1] *W. San-Um, P. Lekbunyasin, M. Kodyoo, W. Wongsuwan, J. Makfak, y J. Kerdsri, «A long-range low-power wireless sensor network based on U-LoRa technology for tactical troops tracking systems», en 2017 Third Asian Conference on Defence Technology (ACDT), Phuket, Thailand, 2017, pp. 32-35.*

[2] *A. Bergeron y N. Baddour, «Design and Development of a Low-Cost Multisensor Inertial Data Acquisition System for Sailing», IEEE Transactions on Instrumentation and Measurement, vol. 63, n.o 2, pp. 441-449, feb. 2014.*

[3] *K. Bai, X. Ge, W. Zheng, y J. Han, «Design of Optimal Route Basing on VPP's Results», en 2008 Fourth International Conference on Natural Computation, Jinan, Shandong, China, 2008, pp. 357-362.*

[4] *«Bai et al. - 2008 - Design of Optimal Route Basing on VPP's Results.pdf».*

[5] *J. Neville, D. Rowlands, A. Wixted, y D. James, «Determining over ground running speed using inertial sensors», Procedia Engineering, vol. 13, pp. 487-492, 2011.*

[6] *S. Fan, Y. Liu, C. Xu, J. Guo, y Z. Wang, «GPS kinematic Precise Point Positioning based on sequential least squares estimation», en 2010 18th International Conference on Geoinformatics, Beijing, China, 2010, pp. 1-5.*

[7] *Y. Kim, T. Koh, H. Kim, S. Cho, H. Hong, y M. Chung, «Maritime safety system using Bluetooth low energy and global positioning system», en 2018 International Conference on Electronics, Information, and Communication (ICEIC), Honolulu, HI, 2018, pp. 1-5.*

[8] *A. G. S. Junior, A. P. D. Araujo, M. V. A. Silva, R. V. Aroca, y L. M. G. Goncalves, «N-BOAT: An Autonomous Robotic Sailboat», en 2013 Latin American Robotics Symposium and Competition, Arequipa, Peru, 2013, pp. 24-29.*

[9] *E. Rodskar, R. Volden, y E. Skjong, «Sailing into the Future: Industrial Internet of Things at Sea with X- Connect», IEEE Electrification Magazine, vol. 5, n.o 3, pp. 33-39, sep. 2017.*

[10] *A. Weintrit, «The need to develop new geodesic-based computational algorithms for marine navigation electronic devices and systems», en 2015 International Association of Institutes of Navigation World Congress (IAIN), Prague, Czech Republic, 2015, pp. 1-12.*

[11] *F. Plumet, C. Petres, M.-A. Romero-Ramirez, B. Gas, y S.-H. Ieng, «Toward an Autonomous Sailing Boat», IEEE Journal of Oceanic Engineering, vol. 40, n.o 2, pp. 397-407, abr. 2015.*

[12] *F. Daiber, M. Jones, F. Wiehr, K. Cheverst, F. Kosmalla, y J. Häkkilä, «UbiMount: 2nd workshop on ubiquitous computing in the mountains», en Proceedings of the 2017 ACM International Joint Conference on*

Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers on - UbiComp '17, Maui, Hawaii, 2017, pp. 1022-1026.

[13] *P. Prempraneerach, K. Thamchaitas and P. Kulvanit, "Autonomous waypoint tracking of kayak boat using state- variable feedback control," 2012 Oceans - Yeosu, Yeosu, 2012, pp. 1-7. doi: 10.1109/OCEANS-Yeosu. 2012.6263616*

[14] *N. D. Simões, J. L. Gonçalves, M. L. Caeiro, M. J. Boavida and F. D. Cardoso, "ZigBee/GPS tracking system for rowing races," 2011 IEEE EUROCON - International Conference on Computer as a Tool, Lisbon, 2011, pp. 1-4. doi: 10.1109/EUROCON.2011.5929248*

[15] *I. Augustyniak, G. Hapel, P. Woś, J. Dziuban and P. Knapkiewicz, "'Mechatronic yacht" — A training unit," Perspective Technologies and Methods in MEMS Design, Polyana, 2011, pp. 127-129.*

[16] *A. Maciuca, G. Stamatescu, D. Popescu and M. Struţu, "Integrating wireless body and ambient sensors into a hybrid femtocell network for home monitoring," 2nd International Conference on Systems and Computer Science, Villeneuve d'Ascq, 2013, pp. 32-37.*

[17] *IEEE Colloquium on Motion Analysis and Tracking.*

[18] *Li, L., Ren, J., & Zhu, Q. «On the application of LoRa LPWAN technology in Sailing Monitoring System» en 2017 13th Annual Conference on Wireless On-demand Network Systems and Services (WONS). IEEE, 2017, pp. 77-80.*

[19] *Sanchez-Iborra, R., G Liaño, I., Simoes, C., Couñago, E., & Skarmeta, A. F. «Tracking and monitoring system based on LoRa technology for lightweight boats», Electronics, 8(1), 15. 2019*

[20] *Gruber, J. M., & Brossi, B. «Position*

“Tecnologías de la información y las comunicaciones mediante IoT aplicadas a soluciones en el medio productivo y medioambiental”

Jorge R. Osio[1,2], Marcelo Cappelletti[1,2], Mauro Salina[1], Mauro Gomez[1], Leonel Navarro[1], Juan. E Salvatore[1], Daniel Alonso[1], Diego Encinas[1,3], Martín Morales[1,4]

[1] Programa Tecnologías de la Información y la Comunicación (TIC) en aplicaciones de interés social, Instituto de Ingeniería y Agronomía, UNAJ
[2] Línea CeTAD - GCA - Instituto Leici - Fac. de Ingeniería -UNLP
[3] Instituto LIDI - Fac. de Informática -UNLP
[4] Codiseño HW SW para Aplicaciones en Tiempo Real - UTN - FRLP

{Josio, mcappelletti, jsalvatore, dalonso, dencinas, martin.morales}@unaj.edu.ar

Resumen

El presente proyecto se basa en la utilización de internet de las cosas (IoT) como herramienta fundamental para proveer soluciones tecnológicas a problemáticas de interés social como son la crisis energética, la ausencia de soluciones tecnológicas a sectores productivos marginados, el cuidado del medioambiente y el diseño de interfaces para educación a distancia. Entre las líneas de investigación que se llevarán adelante para contribuir con las problemáticas planteadas, se propone en primer lugar, atender las necesidades de sectores tales como el frutícola, hortícola y florícola, que carecen de soporte y no disponen de herramientas para automatizar procesos y recibir información crítica en tiempo real. Por otro lado, se plantea el diseño de un algoritmo inteligente de eficiencia energética que permita minimizar el consumo en sistemas alimentados con energía convencional y energía alternativa (este último es muy utilizado por el sector productivo fruti-hortícola). Por último, mediante las interfaces IoT se colaborará en el diseño de laboratorios remotos; en el sensado de parámetros en arroyos y en el diseño de aplicaciones robóticas que provean soluciones de monitoreo remoto, detección automática de objetos y procesamiento de datos.

Palabras clave: *Internet de las cosas (IoT), Eficiencia Energética, Laboratorios Remotos, Robótica, telemetría, contaminación en arroyos, Software Embebido*.

Contexto

Las líneas de Investigación aquí desarrolladas son parte del Proyecto de Investigación Científico-Tecnológico “Tecnologías de la información y las comunicaciones mediante IoT para la solución de problemas en el medio socio productivo” que se lleva a cabo en la Universidad Nacional Arturo Jauretche (UNAJ), bajo la dirección del Mg. Jorge Osio.

A su vez, parte de las líneas de investigación desarrolladas se encuentran enmarcadas en los

convenios de colaboración en actividades de investigación con el área de Sociales para el sensado de parámetros en arroyos, el proyecto de la Dra. María Joselevich sobre Laboratorios Remotos en la enseñanza de la Física y el convenio firmado con la Asociación de Productores Horti-Florícolas de Florencio Varela y Berazategui (APHOVABE).

1. Introducción

El concepto de internet de las cosas ha adquirido gran relevancia en los últimos años, debido a la posibilidad que ofrece de interconectar objetos entre sí y la conectividad a internet que provee a las redes de dispositivos [1]. Aunque los sistemas IoT son más que conocidos, su implementación y desarrollo no ha sido completamente explotado en determinados sectores. IoT aplicado a la horticultura es un área en la que la evolución tecnológica nunca ha sido una prioridad, entre los factores que han postergado la investigación en esta área se tiene la falta de capacitación y falta de soporte e interacción con los productores para determinar cuáles son las principales problemáticas que se pueden resolver. Para este tipo de ambientes se dispone de un conjunto de protocolos de comunicación e interfaces como GSM/GPRS, Wifi y zigbee que se pueden implementar mediante herramientas basadas en Arquitecturas de Sistemas Embebidos como las desarrolladas en [2] y [3]. Las investigaciones realizadas mediante IoT se podrían aplicar en el monitoreo de cultivos, herramientas de soporte para la toma de decisiones, controlar automáticamente riego, protección de heladas, fertilización y monitoreo e iluminación artificial, entre otras .
Otra problemática común en el sector fruti-hortícola, es el acceso al suministro de energía, en cuyo caso requiere de sistemas costosos de energía alternativa. Actualmente, se considera una problemática a tratar de manera urgente el uso eficiente de la energía [4], debido a que se está investigando mucho sobre energías alternativas pero hay mucho por hacer respecto a la reducción en el consumo de energía en función del factor humano y del control automático de los sistemas eléctricos/electrónicos. En este sentido, se debe tener en cuenta la necesidad de minimizar el consumo en los sistemas de energías alternativas, principalmente energía fotovoltaica [4], debido a que estos sistemas almacenan la energía en baterías altamente contaminantes y al lograr optimizar el consumo eficiente de energía se extendería notablemente la vida útil de la misma. Estos sistemas son de gran interés, principalmente por ser de uso común en los sectores agrícolas, donde se aplicarán las investigaciones realizadas conjuntamente con el proyecto "Estudio de sistemas inteligentes para optimizar el aprovechamiento de la radiación solar en la actividad agroindustrial del territorio de la UNAJ" dirigido por el Dr. Marcelo Cappelletti. Para minimizar el consumo, se considera necesario implementar un algoritmo inteligente que tenga en cuenta todos los factores intervinientes, para esto es necesario diseñar un algoritmo basado en redes neuronales.
Respecto a la aplicación de IoT en robótica, se tiene como objetivo el diseño de un sistema que posibilite el monitoreo remoto y permita la detección de objetos mediante procesamiento de imágenes. Los resultados de esta línea de investigación se orientarán a la realización de tareas en ambientes nocivos para el ser humano y en la detección de objetos mediante procesamiento automático. Se prevé que el uso de sistemas robóticos se intensifique en los próximos años, es por eso que se considera necesario diseñar, desarrollar y construir nuevos sistemas que brinden soluciones a situaciones problemáticas de manera eficiente.
Respecto a la aplicación de IoT en Laboratorios remotos, desde el proyecto se ha diseñado la interfaz de Hardware y Software embebido

entre el servidor y los laboratorios desarrollados en el proyecto de la Dra. María joselevich [5]. Además, se está diseñando el hardware y software de una sonda multiparamétrica para el sensado de parámetros en arroyos, que permitan determinar el vertido de efluentes contaminantes por parte de la industria.

Metodología:

Para optimizar el uso racional de la energía, primero se deben tener en cuenta todos los factores de consumo, entre ellos el factor humano, el factor ambiental y el destino de cada uno de los recursos. Con un sistema diseñado a medida se puede lograr optimizar el consumo al máximo mediante la implementación de un algoritmo de decisión, basado en aprendizaje automático, que pueda interpretar los parámetros de entrada y generar una salida eficiente. Para tener en cuenta todos estos factores se debe desarrollar un sistema complejo que tenga la capacidad de combinar todos estos aspectos de manera óptima e inteligente.

En cuanto a las soluciones tecnológicas en las actividades frutihortícolas, en su mayoría requieren de la disponibilidad oportuna de datos, tanto del clima como del suelo, la iluminación y el agua. El desarrollo del sistema informático se realizará a partir de un algoritmo de software con soporte para diferentes aplicaciones, que en base a la información proporcionada por una red de sensores, emita mensajes de alertas a los usuarios y provea información de interés. Para que la propuesta sea aplicable en cualquier lugar, se plantea la utilización de una interfaz GSM/GPRS que permite transmitir información mediante las antenas celulares. El diseño del sistema local a utilizar en huertas a campo abierto y bajo cubierta (invernaderos), estará formado por un sistema de procesamiento (basado en tecnologías de microcontroladores, arduino, LPC1769 y raspberry Phy), al cual se conectarán los sensores y el dispositivo de comunicación GSM/GPRS.

Respecto al Laboratorio remoto, se ha implementado un sistema físico que controla el movimiento de una cuerda mediante PWM y posee una cámara incorporada a una raspberry. Para el acceso al sistema se configuró una extensión de Moodle y una plataforma que permite seleccionar la frecuencia y visualizar la cámara. En cuanto a la sonda multiparamétrica se ha diseñado el hardware y software para la implementación de una red de sensores para la obtención de parámetros y la transmisión mediante wifi a un servidor. Utiliza sistemas embebidos de última generación, los cuales poseen alto poder de cómputo y bajo costo. Respecto al sistema robótico, se utilizarán técnicas de procesamiento digital de imágenes para la detección de objetos y plataformas de control robótico como el sistema ROS.

Antecedentes del Grupo de Trabajo

Se pretende encauzar temas de estudio de posgrado actuales y futuros, de docentes y profesionales surgidos de la UNAJ, dentro de estas líneas de investigación. El Director del Proyecto participa en proyectos de investigación desde el año 2005 en la Facultad de Ingeniería de la UNLP como colaborador y a partir de 2010 como Investigador. Actualmente, el grupo de trabajo investiga en los temas: Transmisión de datos IoT (mediante la interfaz Ethernet y GSM/GPRS), Eficiencia Energética, Robótica y Arquitecturas de procesadores embebidos; pertenecientes a estudios de Especializaciones, Maestrías y Doctorados que se realizan en el marco de acuerdos de colaboración entre diferentes institutos de la UNAJ a través de proyectos de investigación. Como resultado del trabajo realizado recientemente en la UNAJ se han publicado artículos en diferentes congresos nacionales de Ingeniería Informática y Ciencias de la Computación [2-4] y [6].

2. Líneas de Investigación y Desarrollo

El grupo de investigación que se ha constituido en la UNAJ es multidisciplinario, y sus miembros cuentan con experiencia en sistemas multiprocesador, Arquitecturas de Procesadores Embebidos, IoT, sistemas operativos embebidos, aprendizaje automático, protocolos de comunicación TCP y UDP y robótica.
En cuanto a la eficiencia energética, en la actualidad se considera que un sistema de iluminación es eficiente energéticamente cuando permite adaptar el nivel de iluminación en función de la variación de la luz solar o en función de la detección o no de presencia. En esta propuesta se considera que las soluciones existentes no han sido suficientemente desarrolladas, ya que para un uso eficiente de la energía se deben tener en cuenta varios factores y combinar una amplia variedad de métodos para lograr un algoritmo inteligente que permita optimizar el consumo al máximo. Respecto a la investigación aplicada al sector fruti-hortícola, se plantea el diseño de una red de sensores que interactúan entre sí y se comuniquen con un servidor mediante una red IoT, en este sentido se propone usar el protocolo LoRa para determinar la eficiencia de dicho protocolo en aplicaciones que exigen el uso energía alternativa. Dentro de esta línea se enmarca el sensado de parámetros en arroyos mediante sondas que entre otras cosas deben funcionar a batería y transmitir datos de forma inalámbrica a un nodo central.
En cuanto a la robótica, esta línea está en pleno desarrollo y se buscan aplicaciones innovadoras como la detección de objetos para reciclado mediante procesamiento de imágenes y la capacidad de inspeccionar sectores de difícil acceso o en ambientes donde la contaminación impide la permanencia de seres humanos.

Para las aplicaciones de eficiencia se deberán desarrollar algoritmos de aprendizaje automático supervisado y no supervisado basados en regresión lineal múltiple y clustering respectivamente.

Temas de Estudio e Investigación

Programación de diferentes Arquitecturas de Procesadores (freescale, LPC1769, CIAA, Arduino, Raspberry Phy) embebidos para la implementación de nodos en las redes de sensores y para la implementación de algoritmos de control. Utilización y configuración de sistemas operativos de tiempo real y Linux embebido para la ejecución de los algoritmos, plataformas como ROS y la adquisición de datos de sensores. Implementación de una red de sensores inalámbrica mediante protocolos estándar de IoT (LoRA) e interfaces RF. Investigación relacionada con los protocolos TCP y UDP para el envío de paquetes, donde en condiciones de red conocidas [3], se deben determinar las tasas de transmisión mínimas para asegurar el funcionamiento óptimo del sistema de eficiencia energética ([7] y [8]), medición de parámetros en arroyos y el laboratorio remoto. Determinación de la eficiencia en la ejecución de algoritmos de control de un robot humanoide sobre un sistema multicore. Estudio y utilización de algoritmos de aprendizaje automático supervisado y no supervisado [9].

3. Resultados Obtenidos/Esperados

- Se implementó un sistema de telemetría mediante la interfaz GSM/GPRS mediante el módulo SIM800, que permite supervisar el estado del cultivo (temperatura, iluminación y humedad) mediante una aplicación web.
- Se implementó un sistema de control a través de una comunicación inalámbrica punto a punto Wifi de un robot comercial cuya arquitectura

está basada en arduino. El control se realiza por medio de una página web que puede ser accedida desde un teléfono celular como así también de una PC.

- Por último, se encuentra en la etapa final un sistema que permite optimizar el consumo energético en instituciones educativas mediante la ejecución de algoritmos que toman en cuenta diferentes factores. Este sistema utiliza la plataforma IoT desarrollada en [3] para la implementación y ejecución del algoritmo.

Los resultados esperados para los nuevos temas que se están desarrollando en el proyecto son:

- Se espera para el año en curso realizar las pruebas de funcionamiento en campo del sistema autónomo y escalable que se encarga de minimizar el consumo de energía conjuntamente con la implementación de un sistema informático que provea información precisa de magnitudes climatológicas, del estado del suelo y del nivel de iluminación para determinados cultivos. De la misma forma se buscarán resultados en la mediciòn de parámetros en arroyos. En función de los resultados que se obtengan se realizarán las correcciones pertinentes.
- Finalmente, la incorporación de esta nueva línea buscará proveer soluciones de software embebido para la implementación del Laboratorio remoto propuesto en el Proyecto de la Dra. María Joselevich.

4. Formación de Recursos Humanos

Dentro de la temática de la línea de I+D, todos los miembros del proyecto participan en el dictado de asignaturas de la carrera de Ingeniería Informática de la UNAJ.

En este proyecto existe cooperación a nivel Nacional. Hay dos Doctores en Ingeniería, un Magister, un Especialista y un integrante realizando Maestría en temas relacionados.

A su vez, se cuenta con un Becario de Entrenamiento CIC en temas de eficiencia energética y dos Becarios de Vinculación en temas de eficiencia energética y monitoreo de parámetros en invernaderos.

Adicionalmente, se cuenta con la colaboración de estudiantes avanzados.

5. Bibliografía

[1] L. Fatmasari Rahman, ”Choosing your IoT Programming Framework: Architectural Aspects”, 2016 IEEE 4th International Conference on Future Internet of Things and Cloud.

[2] J. Osio, M. Cappelletti , G. Suárez, L. Navarro, F. Ayala, J. Salvatore , D. Alonso ,D. Encinas, M. Morales, “Diseño de aplicaciones de IoT para la solución de problemas en el medio socio productivo”, UNSJ, San Juan, WICC 2019.

[3] Osio J., C. Acquarone, E. Hromek, J. Salvatore , “Plataforma de desarrollo para IoT”, IV conaiisi, 2017

[4] J. Osio, J. Salvatore , D. Alonso , V. Guarepi , M. Cappelletti , M. Joselevich , M. Morales, “Tecnologías de la información y las comunicaciones mediante IoT para la solución de problemas en el medio socio productivo”, UNNE, Ciudad de Corrientes, WICC 2018.

[5] M A. Revuelta, Laboratorio remoto en un entorno virtual de enseñanza aprendizaje,, FIUNLP, 2016

[6] J. Osio, J. Salvatore, E. Kunysz, D. Montezanti, D. Alonso, V. Guarepi, M. Morales, “Análisis de Eficiencia en Arquitecturas Multiprocesador para Aplicaciones de Transmisión y Procesamiento de Datos”, ITBA, CABA, WICC 2017

[7] M. Fernandez Salgado, Eficiencia energética en los edificios, AMV ediciones, 2011

[8] M. Aparicio, Energía Solar Fotovoltaica 2ª Edición, Edit. marcombo, 2009

[9] J. Hurwitz, D. Kirsch, “Machine Learning for Dummies”, Ed. Wiley, 2018.

Wireless Wine:
Estimación de rendimiento y ubicación de sensores para la predicción de heladas en los viñedos

Diego Dujovne[1], Thomas Watteyne[2], Gustavo Mercado[3],
Ana Diedrichs[3], Carlos Taffernaberry[3], Jorge Pérez Peña[4]

[1]Escuela de Informática y Telecomunicaciones, Facultad de Ingeniería,
Universidad Diego Portales, Santiago, Chile
diego.dujovne@mail.udp.cl

[2]INRIA - Institut National de Recherche en Informatique et en Automatique
Paris – France
thomas.watteyne@inria.fr

[3]GridTICs – Grupo en Tecnologías de la Información y las Comunicaciones
Departamento de Electrónica / Facultad Regional Mendoza / UTN
{gustavo.mercado, ana.diedrichs, carlos.taffernaberry}@gridtics.frm.utn.edu.ar

[4]INTA - Instituto Nacional de Tecnología Agropecuaria
EEA Mendoza INTA
perezpena.jorge@inta.gob.ar

Abstract

Los productores de viñedos han estado reuniendo información operacional significativa sobre sus campos, y las fincas más avanzadas pueden producir varios años de datos sobre cómo se ha plantado, regado, fertilizado y cómo han madurado las uvas. A pesar de esta riqueza de datos operacionales, hoy en día no existe ninguna herramienta que pueda complementar y gestionar eficientemente estos datos para (1) proporcionar una previsión precisa del rendimiento y (2) predecir los efectos destructivos de los eventos de heladas. La previsión del rendimiento es el proceso para estimar la cantidad de producción de uva para cada sección de un campo en términos de kilos por unidad de superficie.

El primer objetivo de WirelessWine es desarrollar una solución basada en el "aprendizaje de máquinas" lista para usar, que combine la riqueza de la información operativa de los productores, los datos meteorológicos regionales, las estaciones meteorológicas a nivel de campo y los dispositivos de IoT desplegados directamente en el campo para proporcionar una previsión precisa del rendimiento.

En la región de Mendoza, las heladas han hecho que la producción de uva de 2016 sea un 40% inferior a la de 2015. Los productores pueden combatir un evento de heladas calentando el campo, pero tienen que saber que el evento de heladas viene con un par de horas de anticipación. Los pronósticos meteorológicos a escala regional no proporcionan la precisión necesaria para predecir de manera fiable los eventos de heladas. *El segundo objetivo del WirelessWine es identificar, de forma iterativa, la ubicación de los sensores (en el follaje, en el racimo, a diferentes alturas, etc.) que permite la predicción más precisa del evento de helada.*

Keywords: Industrial Internet of Things, Smart Agriculture, Yield Estimation, Frost Prediction, Vineyards, Wireless Sensors Networks, Machine Learning.

Contexto

Este proyecto contituye un equipo interdisciplinario y perfectamente complementario de los mejores expertos en agricultura (INTA), en IoT inalámbrica de baja potencia (Inria, UDP y gridTICs), y en aprendizaje automático (UTN). Este equipo de expertos trabaja en estrecha colaboración con los

productores de vino de clase mundial que proporcionan los problemas del mundo real a los que se enfrentan.

Los socios del proyecto provienen de Argentina, Chile y Francia, tres países en los que la elaboración de vino es una parte central de la cultura y la economía. La estrecha colaboración con los productores de vino, el enfoque pragmático orientado a las soluciones y las fuertes conexiones con la industria son garantías de la transferencia de la tecnología a la industria, además del excelente resultado académico del proyecto.

El proyecto cuenta con financiamiento del "Programa Regional STIC-AmSud 2018 (Investigación - Innovación)" bajo el nombre de "Wireless Wine" y el ID "19STIC-01".

Las instituciones asociadas son:

- Escuela de Informática y Telecomunicaciones - Universidad Diego Portales Santiago, Chile
- gridTICs - Grupo de I&D en Tecnologías de la Información y Comunicaciones. Universidad Tecnológica Nacional - Facultad Regional Mendoza, Mendoza Argentina
- Equipo EVA - INRIA - Institut National de Recherche en Informatique et en Automatique. París, Francia
- INTA - Estación Experimental Agrícola Mendoza - Instituto Nacional de Tecnología Agropecuaria (EEA Mendoza INTA), Mendoza Argentina

1. INTRODUCTION

Motivación: Hoy en día, dentro de la industria del vino, hay dos problemas principales que afectan profundamente a su proceso de producción:

Previsión de rendimiento: Como se discutió en Chien [1], la estimación de la cosecha es fundamental. Si bien los agricultores esperan obtener la mejor fruta disponible para aumentar el precio del producto, también están interesados en proporcionar la máxima producción de sus campos. Dada la estimación, se puede calcular el momento de la cosecha, los costos y los gastos de transporte. Además, tanto la sobreexplotación como la subexplotación pueden generar retrasos en la maduración de los frutos y problemas de enfermedades. Además, en lo que respecta a las bodegas, el error de estimación de la cosecha afecta a la logística de transporte y al espacio de almacenamiento, que se dimensionan y planifican de antemano para reservar recursos.

El método típico para estimar el rendimiento de un campo de viñedos implica la participación de varios, que se calculan a partir de muestras tomadas del viñedo para contar bayas y flores, y medir el tamaño de los racimos. Se trata de una tarea manual, y todas las muestras deben tomarse durante un período de tiempo reducido, para que la muestra sea válida, [2] y [3].

Los métodos estándar utilizan el muestreo de frutas; están dirigidos a los resultados en lugar de a la causa. Las muestras representan instantáneas de la evolución de las vides durante el crecimiento de las bayas, lo que da lugar a un proceso largo y costoso.

El rendimiento de la producción es muy sensible a las condiciones particulares del campo para el tipo de plantas, su edad, los diferentes tratamientos que se dan durante el proceso de maduración: riego y raleo entre otros, y las condiciones meteorológicas hasta la maduración. La relevancia de las condiciones climáticas en los viñedos se plantea como “clasificadores” de las diferentes regiones vitícolas del mundo en el trabajo de Tonietto [4], teniendo en cuenta el índice de sequedad, el índice heliotérmico y el índice de noches frescas. Walsh [5] describe el estado actual de la tecnología, proponiendo el uso de mediciones a macroescala desde satélites y a microescala desde sensores específicos para viñedos. El autor también cubre brevemente las aplicaciones que proponemos en este trabajo: mejorar la detección de heladas y la previsión del rendimiento, a partir de datos medio-ambientales.

Sabbatini y otros [6] describen el método del peso del racimo de la cosecha, el método de la fase de retardo (que mide durante el período en que las uvas dejan de crecer en tamaño) y el método del grado-día de crecimiento, mediante la inclusión de uno de los parámetros climatológicos.

A partir de nuestra investigación de campo

previa a esta propuesta, encontramos que los productores de vino (independientemente del tamaño de sus campos) sufren una intensa variabilidad en la producción, y los métodos estadísticos no han ayudado a reducir los errores de previsión del rendimiento.

Por último, el cambio climático global ha mostrado su influencia en los viñedos, como lo demuestran Orduña y otros. [7]. La consecuencia es la necesidad de obtener un conocimiento más profundo de la variabilidad del clima y del microclima y de construir modelos en torno a estos parámetros, a fin de proporcionar las herramientas para tomar mejores decisiones sobre el terreno. Además, Jones [8] destaca la necesidad de medir las condiciones climáticas en todas las escalas: macroescala, mesoescala, topoescala (clima del lugar) y microescala (hasta el nivel de la copa), como proponemos en el presente proyecto

Predicción de heladas: Sólo durante el 2016, más de 35000 hectáreas de viñedos fueron afectadas en Mendoza, Argentina, [9] [10]. Además, en Chile, las heladas provocaron una pérdida de hasta el 50% de la producción de viñedos durante el año 2013 en el valle de Casablanca, y afectaron la productividad de más de tres mil hectáreas de viñedos [11]. Ambas representan las regiones de producción de vino más relevantes de América Latina.

Además, un estudio del Instituto Tecnológico de Karlsruhe [12] advierte que los viñedos de Mendoza y San Juan, en Argentina, representan las regiones de mayor riesgo del mundo en cuanto a clima extremo y peligros naturales.

Los eventos de heladas son difíciles de predecir, dado que se trata de un fenómeno localizado. Pero un evento de helada puede destruir toda la producción en cuestión de minutos: incluso si el daño no es visible justo después del evento, los efectos pueden aparecer al final de la temporada, reduciendo tanto la cantidad como la calidad de la cosecha.

Existen varias medidas para contrarrestar las heladas, que incluyen calentadores de aire por medio de la quema de gas, gasolina u otros combustibles, y la eliminación del aire utilizando grandes ventiladores distribuidos a lo largo del campo. Sin embargo, cada una de estas contramedidas es costosa. Por consiguiente, es fundamental predecir los episodios de heladas con la mayor precisión posible, a fin de iniciar la contramedida en el momento oportuno, reduciendo la posibilidad de que se produzcan falsos negativos o falsos positivos.

La experiencia adquirida en el proyecto STIC -AmSud PEACH [13], finalizado en 2017, que tenía el objetivo de detectar eventos de heladas en las parcelas de duraznos, proporcionará un punto de partida para construir una solución a la industria del vino. Aunque los viñedos tienen una estructura y un comportamiento diferentes a los de los durazneros, podemos aplicar los conceptos básicos de despliegue, combinando el IoT y algoritmos de aprendizaje automático como plataformas de desarrollo validadas.

Metodología

Proponemos utilizar la infraestructura de sensoramiento de IoT combinada con técnicas de aprendizaje automático para mejorar el proceso de previsión del rendimiento, incluyendo los datos meteorológicos recogidos en cada uno de las parcelas, complementados con los datos actuales e históricos captados por el productor.

El Internet de las Cosas ha evolucionado a partir de las antiguas redes de sensores inalámbricos como un concepto más amplio que abarca no sólo la conectividad de los dispositivos, sino todo el ámbito de la detección, la captura, el transporte, el procesamiento y la visualización. Además, la IoT es actualmente adoptada por la industria a nivel mundial como una de las tecnologías de más rápido crecimiento [14], lo que ha fomentado la necesidad de estandarización [15].

Actualmente, los estándares de comunicaciones de IoT incluyen las capas físicas de Time-Scheduled Channel Hopping y MAC [16], que proporcionan robustez contra las interferencias y eficiencia en el uso de la energía, 6LoWPAN [17] para soportar la conectividad desde cualquier dispositivo

conectado a Internet, el enrutamiento proporcionado por RPL (Protocolo de Enrutamiento para Redes de Baja Potencia y Pérdida) [18] que permite la escalabilidad y la disponibilidad, y CoAP [19] para soportar una capa de transporte transparente. Nuestros equipos tienen una amplia experiencia de trabajo en esta área obtenida de sus anteriores despliegues de campo, específicamente en escenarios agrícolas, y trabajos de investigación, incluyendo el desarrollo de aplicaciones y tareas de estandarización.

El uso del Aprendizaje Automático, una de las herramientas más utilizadas en el campo de la investigación de la Inteligencia Artificial, se ha aplicado a las más diversas áreas [20] en las que los problemas no podían ser resueltos por modelos analíticos estándar o enfoques estadísticos. Proponemos, a partir de la experiencia actual del proyecto PEACH, utilizar estas técnicas para analizar datos actuales e históricos y generar resultados como herramientas de estimación para mejorar la predicción de heladas y la previsión de rendimientos en viñedos.

2. OBJETIVOS DEL PROYECTO

El objetivo general del proyecto es utilizar la información microclimática para anotar conjuntos de datos con características para:

- Comprender el impacto de la temperatura, la humedad y otros factores en el proceso de maduración en los viñedos para mejorar la predicción del rendimiento
- Identificar las ubicaciones más significativas de los sensores en los viñedos para la predicción de heladas, utilizando los resultados de proyectos anteriores, para despliegues de IoT y técnicas de aprendizaje automático.

Los objetivos específicos son:

1. Especificar las características que se utilizan actualmente para la predicción de rendimientos, tanto cuantitativos como cualitativos
2. Identificar nuevas características del medio ambiente como candidatas para mejorar la predicción del rendimiento y caracterizar las mediciones requeridas
3. Crear un prototipo de despliegue de medidas basadas en IoT en un campo de viñedos de pruebas.
4. Reutilizar el Modelo de Aprendizaje Automático del proyecto PEACH para predecir los eventos de heladas en los viñedos, y probar la precisión de la solución
5. Identificar las ubicaciones de los sensores que dan la predicción más precisa de la helada
6. Diseñar un nuevo Modelo de Aprendizaje de Máquinas para predecir el rendimiento de la producción y probar la precisión de la solución
7. Complementar los datos del sensor del campo con datos históricos de muestreo de bayas y racimos de los productores.
8. Generar una primera campaña de medición con un diseño inicial predeterminado.
9. Generar una segunda campaña de medición con un diseño mejorado según los resultados de la localización.
10. Extraer los resultados de cada una de las campañas y publicar los resultados.

4. CONTRIBUCIONES

Los tres grupos que participan en este proyecto han desarrollado, implantado y desplegado sistemas de IoT para diferentes tipos de aplicaciones, incluidas las plataformas agrícolas, ambientales, de transporte y experimentales, lo que proporciona una sólida base para el trabajo en conjunto. Además, la sinergia entre los grupos puede demostrarse mediante la evolución del proyecto PEACH, que fue el resultado de la colaboración mutua en materia de detección e instrumentación para parcelas de duraznos.

3.1 FORMACIÓN DE RECURSOS HUMANOS

UDP: Universidad Diego Portales [Chile]

Estudiantes universitarios desarrollarán sus proyectos de programa final de acuerdo con los objetivos del proyecto Wireless Wine.

UTN: [Argentina]

El grupo gridTICS tiene seis estudiantes de licenciatura, un ingeniero electrónico y dos estudiantes de posgrado que colaborarán en el desarrollo de este proyecto. La ingeniera Ana Diedrichs es estudiante de doctorado y su tema de investigación es la predicción de heladas utilizando el aprendizaje automático y las redes de sensores inalámbricos.

Inria: Institut National de Recherche en

Informatique et en Automatique [Francia]
Los principales contribuyentes a este proyecto son un docente/investigador y dos post-doctorandos.

REFERENCIAS

[1] Chien, M. “Crop Estimating in Vineyards” Technical Booklet (2009), College of Agricultural Sciences, Penn State University.
[2] Dunn, G. M. "Yield forecasting." *Technical Booklet* (2010).
[3] B. Komm, and M. Moyer. "Vineyard yield estimation." (2015).
[4] J. Tonietto, and A. Carbonneau. "A multicriteria climatic classification system for grape-growing regions worldwide." *Agricultural and Forest Meteorology* 124, no. 1 (2004): 81-97.
[5] P. Walsh, “In Vino Veritas On Wine, Weather and Climate”, The Weather Channel, March 25, 2017. [https://weather.com/farming/news/weathers-effect-grapes-andwine-20140325]
[6] P Dami, I Sabbatin and G. Howell "Predicting Harvest Yield in Juice and Wine Grape Vineyards.", 2012
[7] De Orduna, R. Mira. "Climate change associated effects on grape and wine quality and production." *Food Research International* 43, no. 7 (2010): 1844-1855.
[8] G. Jones, R. Reid, and A. Vilks. "Climate, grapes, and wine: structure and suitability in a variable and changing climate." In *The Geography of Wine*, pp. 109-133. Springer Netherlands, 2012.
[9] L. Saieg, Los Andes. “Casi 35 mil hectáreas afectadas por heladas” November 20th, 2016, Mendoza, Argentina [In spanish] [http://www.losandes.com.ar/article/casi-35-mil-hectareas-de-vid-afectadas-por-heladas].
[10] A. Barnes, Decanter.com, “El Nino hampers Argentina’s 2016 wine harvest”, May 23rd, 2016, United Kingdom. [http://www.decanter.com/wine-news/el-nino-argentina-2016-wine-harvest-305057/].
[11] Bayer, “Viñas de casablanca estiman pérdidas por usd 14 millones,” October 2013, http://www.cropscience.bayer.cl/noticias/vernoticia.asp?id=3121
[12] M. Lehné, Karlsruhe Institute of Technology, “Winemakers lose every year millions of dollars due to natural disasters”, Germany, April 26th, 2017. [http://www.kit.edu/kit/english/pi_2017_051_winemakers-lose-billions-of-dollars-everyyear-due-to-natural-disasters.php].
[13] A. L. Diedrichs, F. Bromberg, D. Dujovne, K. Brun-Laguna and T. Watteyne, “Prediction of Frost Events Using Machine Learning and IoT Sensing Devices”, in IEEE Internet of Things Journal, vol. 5, no. 6, pp. 4589-45Co-Authors: Thomas97, Dec. 2018. DOI: 10.1109/JIOT.2018.2867333
[14] L. Columbus. “Roundup of Internet of Things Forecasts and Market Estimates, 2016” Forbes.com, November 27th, 2016. [https://www.forbes.com/sites/louiscolumbus/2016/11/27/roundup-of-internet-of-thingsforecasts- and-market-estimates-2016/#13a440a6292d]
[15] M. Palattella, N. Accettura, X. Vilajosana, T. Watteyne, L. Grieco, G. Boggia, and M. Dohler. "Standardized protocol stack for the internet of (important) things." *IEEE Communications Surveys & Tutorials* 15, no. 3 (2013): 1389-1406.
[16] IEEE Standard for Information technology: Local and metropolitan area networks Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (WPANs). IEEE St. 802.15.42006.
[17] G. Mulligan, "The 6LoWPAN architecture." In *Proceedings of the 4th*
34 *workshop on Embedded networked sensors*, pp. 78-82. ACM, 2007.
[18] T. Watteyne, A. Molinaro, M. Richichi, "From manet to ietf roll standardization: A paradigm shift in wsn routing protocols." *IEEE Communications Surveys & Tutorials* 13, no. 4 (2011)
[19] C. Bormann, A. P. Castellani, and Z. Shelby. "Coap: An application protocol for billions of tiny internet nodes." *IEEE Internet Computing* 16, no. 2 (2012): 62-67.
[20] M. Jordan, and T. Mitchell. "Machine learning: Trends, perspectives, and prospects." *Science* 349, no. 6245 (2015)

Monitoreo a Distancia de Centros de Transformación Eléctrica Smart Grid

Luis Álvarez[1], Gustavo Mercado[2], Evaristo Martinez[1], Roberto Martinez[1], Martin Ruiz[1], Claudio Gonzalez[1], Sebastián Tobar[2], Ariel Verdejo[2]

[1]IRESE - Instituto Regional de Estudios Sobre Energía
Departamento de Electromecánica / Facultad Regional Mendoza / UTN
alvarez.irese@frm.utn.edu.ar

[2] gridTICs – Grupo en Tecnologías de la Información y las Comunicaciones
Departamento de Electrónica / Facultad Regional Mendoza / UTN
{gustavo.mercado, sebastian.tobar, ariel.verdejo}@gridtics.frm.utn.edu.ar

Resumen

Los grandes aprovechamientos de generación requieren inversiones muy grandes y además afectan considerablemente el medio ambiente. Esto ha llevado a desarrollar la incorporación de energías renovables en los sistemas de distribución domiciliaria; las fuentes de energía más comunes son: Paneles Fotovoltaicos Domiciliarios o Pequeños Parques Fotovoltaicos, Mini-centrales de Generación Hidráulica y la Generación Eólica

Este novedoso régimen de generación produce grandes variaciones de voltaje en los alimentadores, que deben ser analizadas, mediante flujos de carga, para determinar el uso de métodos paliativos y correctivos de la anormalidad, tales como la inyección de potencia activa y reactiva. Cada alimentador está conformado por una cierta cantidad de transformadores de consumo domiciliario, cuyas potencias varían dependiendo de los usos y tipo de usuario. Por este motivo es necesario contar con un dispositivo electrónico económico que permita transmitir información a distancia de los diferentes parámetros eléctricos más relevantes de manera inalámbrica

Este trabajo se describe un sistema prototipo, de adquisición y tratamiento de datos característicos de los alimentadores y transformadores de una red de distribución eléctrica. El sistema estará conformado por elementos de bajo costo de sensado, transmisión y almacenamiento de información en la modalidad de "Internet de las Cosas".

También contará con un sistema de software de análisis de comportamiento de la red.

Palabras Clave: Smart Grid, Supervisory Control And Data Acquisition, Uso Racional de la Energía, Energías no Convencionales, Internet de las Cosas

Contexto

El presente trabajo está inserto en el proyecto de investigación acreditado por la Universidad Tecnológica Nacional código ENUTNME0005396 denominado "Monitoreo a Distancia de Centros de Transformación - Smart Grid".

1. INTRODUCCIÓN

Los grandes aprovechamientos de generación requieren inversiones muy grandes y además afectan considerablemente el medio ambiente. Esto ha llevado a desarrollar la inclusión de energías renovables en los sistemas de distribución domiciliaria, la tensión de distribución puede ser generalmente en 13,2 kV ó 33 kV y las fuentes de energía más comunes que nos encontramos en nuestra región son: Paneles fotovoltaicos domiciliarios o pequeños parques fotovoltaicos que no superan los 10 MW, mini-centrales de generación hidráulica y la generación eólica [1].

Todas ellas de potencias bajas acordes a los

límites de potencias a transmitir en esos niveles de voltaje. Debido a su intermitencia de generación es conveniente su monitoreo permanente y la manera que afecta a la red de distribución.

Durante su funcionamiento ayudan a la red domiciliaria a mejorar niveles de voltaje y liberar capacidad de transporte de los alimentadores debido a que ahora la energía no proviene de la red, por lo que es inyectada por estas fuentes de energía en el mismo alimentador; también estas fuentes de energía durante las cargas máximas del alimentador puede que no estén disponibles, ya que no se dispone de la energía solar de 18 a 9 hs y existe una probabilidad elevada de que la generación eólica no se encuentre disponible durante la noche quedando así solamente la energía de las minicentrales hidráulicas. Este régimen de generación produce grandes variaciones de voltaje en el alimentador que deben ser analizadas mediante flujos de carga para sensibilizarse con la inclusión de estos tipos de generación distribuida que son provocadas por la inyección de potencia activa y reactiva de las mismas, que actúan en forma positiva mejorando los niveles de voltaje y las pérdidas. Estas se modifican en forma negativa cuando la generación distribuida se encuentra fuera de servicios por los motivos indicados anteriormente. Cada alimentador está conformado por una cantidad de transformadores de consumo domiciliario de potencias que varían entre 80 kVA y 630 kVA, dependiendo si es en lugares más alejados (zonas rurales) o con alta densidad de población (zonas urbanas). Por este motivo es necesario contar con un dispositivo electrónico económico que permita transmitir información a distancia de los diferentes parámetros eléctricos más relevantes de manera inalámbrica desde transformadores de distribución hasta un nodo central que recopile y procese la información de los diferentes centros de distribución [2]. El proyecto inicial consiste en tomar un alimentador de referencia con todos sus transformadores y monitorear a distancia los parámetros eléctricos mar relevante como lo son: Voltaje, Corriente, coseno de fi, estado de carga y sensado de movimiento del transformador. El sistema planteado nos permitirá obtener información sobre el desequilibrio de fases (Desbalance de corrientes), sensado de movimiento por motivos de robo, sobrecarga de los mismos, determinación de la vida útil a partir de los valores de sobrecarga y tiempo de duración de la misma, determinación del estado de energización de los equipos y alimentadores, estadística de falla del sistema, que permite determinar parámetros que miden la calidad del servicio y determinación de pérdidas en la transmisión, las cuales se determinan mediante el uso de programas de flujo de carga. El IEC recomienda el uso de dispositivos estandarizados para estas tareas de gestión y control [3].Con esos datos obtenidos y trasmitidos a una central se puede llegar a conocer como re-configurar el alimentador dependiendo del estado de carga y del desequilibrio de las fases. Además conocer el lugar exacto de una falla, como los valores de los parámetros al momento de la misma también consecuentemente poder actuar rápidamente en caso de falla y/o robo de transformador. Con todos esos datos almacenados poder usar modelos matemáticos y poder analizar de manera real cómo se comporta la curva de vida de los diferentes transformadores.

Este proyecto propuesto también propone el desarrollo de ingeniería en electrónica y de comunicación a muy bajo costo en las áreas de electrónica y comunicación eficiente y a muy bajo costo, bajo el esquema de Smart Grid, como la Red Inteligente, considerada como la próxima generación red eléctrica, utiliza flujos bidireccionales de electricidad e información para crear una red automatizada de suministro de energía ampliamente distribuida [4].

Antecedentes

Este proyecto es continuación del proyecto FONARSEC 2013. Denominación: "Red inteligente ciudad Gral San Martín Mendoza", Proyecto FITS Energía 2013 - UREE. Proyecto 13, Resolución Nº 470/14 financiado por el Fondo sectorial de energía.

Participantes: CAPP (EDESTE, EMESA y UTN FRM), con vigencia desde el 1/1/2015 hasta el 31/12/2019 [5][6][7].

Al cabo de los tres primeros años del proyecto se cuentan con los resultados siguientes:

Parque de 5000 medidores inteligentes tanto monofásicos como trifásicos instalados en clientes finales residenciales, comerciales, industriales, de alumbrado público y de riego agrícola, integrados a una red de comunicación y reportando a un sistema de recolección y administración de datos que permita: Gestión de la telelectura de los mismos, recibir avisos y eventos, controlar la suspensión y reanudación del suministro, y aportar datos para avanzar con los desarrollos, estudios, sistemas, integraciones, optimizaciones.

Planta de generación fotovoltaica de 2MW instalada y funcionando interconectada a la red de distribución en Media Tensión de EDESTE S.A. telecontrolada, telemedida y telesupervisión, y con cuatro módulos de autogeneración de bajas potencias (5 a 15 KW) interconectados a la red de distribución de Baja Tensión instalados y funcionando en edificios públicos en la zona.

3. OBJETIVOS y AVANCES DEL PROYECTO

Debido a la actual crisis energética que atraviesa nuestro país es necesario desarrollar tecnología y sistemas de información de bajo costo que permitan operar y monitorear la red eléctrica de distribución, con el objetivo de alcanzar mejoras en la distribución de carga con alto factor de eficiencia, disminuyendo los ciclos de falla y mejorando sustancialmente la calidad de servicio brindada a los usuarios.

El proyecto de investigación tiene como objetivos principales lo siguiente:

1. Desarrollar un sistema de bajo costo que permita monitorear a distancia los parámetros eléctricos de los transformadores de distribución y alimentadores menores a 5 MVA, para poder operar de manera eficiente la red de distribución local.

2. Almacenar, en una base los datos, la información obtenida para realizar estudios numéricos de análisis de fallas, flujos de carga, análisis de cortocircuito y diagramas de estabilidad de la red.

3. Poder conocer de manera precisa el estado de la red y en caso de falla la ubicación exacta de la misma poder actuar de manera más rápida y eficiente

Avances del Proyecto

Para llevar adelante el proyecto de investigación, deberá dividirse en dos grandes etapas más significativas [8]:

1 Relevamiento de la red de distribución local:

A. Como primera instancia de debe acordar, mediante convenios, con una empresa distribuidora de energía eléctrica local, para que brinde información de sus instalaciones existentes y de sus características.

B. Realizar el relevamiento de factibilidad en un alimentador determinado en media tensión (13,2 KV) con los transformadores de distribución incluidos en las aguas abajo.

C. Tomar mediciones de los parámetros eléctricos más relevantes y volcarlos a una base de datos para luego analizarlos con diferentes modelos matemáticos y físicos.

2)- Diseño del sistema lógico para el monitoreo a distancia [9].

D. Diseño de equipos remotos de bajo costo para la trasmisión de datos a distancia, en la modalidad de Internet de las Cosas.

E. Determinación del medio y protocolo de comunicación a utilizar.

F. Diseño de un modelo matemático para estudios estadísticos de análisis de falla y vida de los transformadores de distribución.

G. Diseño de un software SCADA para recopilar esa información y poder visualizarla de manera adecuada desde una sala de control.

4. FORMACIÓN DE RECURSOS HUMANOS

Uno de los principales objetivos del proyecto es la capacitación de los recursos humanos.

La meta como investigadores es fortalecer la capacidad para realizar investigación científica, generar conocimientos y facilitar la transferencia de tecnología que permita el desarrollo humano.

Este proyecto de investigación posibilitará la colaboración inter-institucional y la ejecución de proyectos conjunto entre grupos I+D de diferentes disciplinas.

Para lograr estos objetivos se dispuso del siguiente personal:

Dos Investigadores formados
Tres Investigadores de apoyo
Dos Becarios graduado (Beca BINID UTN)
Dos Becarios alumnos (Beca alumno UTN)
Un Tesista de carrera de grado

Adicionalmente se realizarán tareas, como:

- Dictado de Cursos, Seminarios y Conferencia para público especializado.

- Promoción, coordinación y asistencia técnica de tesinas de grado para alumnos de la FRMza.

- Promoción, coordinación y asistencia técnica a pasantes alumnos, cursantes de carreras de grado y de pre-grado en el ámbito de la UTN FRMza.

- Presentación de Trabajos en Congresos y Reuniones Técnicas/Científicas.

- Publicación de Trabajos en revistas con/sin referato.

- Publicación de todo el código fuente desarrollado en el presente proyecto, en `plataformas abiertas como github.

5. BIBLIOGRAFÍA

[1] Qing-Chang Zhong, "Distributed Energy Resourse – How to Achieve Completely Autonomous Power in the Next Generation of Smart Grid", IEEE Smart Grid Compendium, 2015.

[2] F. Borges and M. Mühlhäuser, "EPPP4SMS: Efficient Privacy-Preserving Protocol for Smart Metering Systems and Its Simulation Using Real-World Data" TRANSACTIONS ON SMART GRID NOVEMBER 2014 VOLUME 5 NUMBER 6 ITSGBQ (ISSN 1949-3053)

[3] IEC, "Coping with the Energy Challenge The IEC's role from 2010 to 2030 Smart electrification - The key to energy efficiency" International Electrotechnical Commission ISBN 978-2-88912-890-7, 2010

[4] Xi Fang, Satyajayant Misra, Guoliang Xue, and Dejun Yang, "Smart Grid – The New and Improved Power Grid: A Survey" IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 14, NO. 4, FOURTH QUARTER 2012

[5] G. Mercado et al "SG-SM - Smart Grid San Martín. Red de Distribución y Generación de Energía Inteligente.en Ciudad Gral San Martin Mendoza" XVII Workshop de Investigadores en Ciencias de la Computación - WICC 2015, Salta 16-17, Abril 2015

[6] G. Mercado, J. Da Peña, G. Vivone, M. Ledda, R. Caceres, C. Taffernaberry, S. Pérez, L. Álvarez, A. Arena, G. Henderson, M. Funes, J. Catalá, "Diseño, Implementación y Análisis de Red de Distribución Inteligente y Generación Distribuida de Energía", Anales de CIIS 2016, May 2016, Universidad de Córdoba, Córdoba.

[7] A. P. Arena, L. Álvarez, J. Da Peña, R. Stasi, G. López, A. Burlot, L. Barnabo, M. Ledda, R. Caceres, C. Taffernaberry and G. Mercado, "Diseño, Implementación y Análisis de una Red Eléctrica Inteligente (Smart Grid)", Anales del ENIDI 2017, Mendoza 22 al 24 de Noviembre 2017

[8] A. Álvarez y E. Martinez "PID Monitoreo a Distancia de Centros de Transformación. SMART GRID" UTN FRM, SCTyP. 2018

[9] G. Mercado, S. Tobar, C. Taffernaberry y A. Verdejo "nod-e^{2}sg-sigfox: Nodo AMI IoT para Smart Grid en Sigfox", Memoria Técnica PID5396 - IRESE/gridTICs UTN FRM, 2019

Smart Micro Grid de Campus Universitario

Raúl Moralejo[1], Sebastián Tobar[1], Gustavo Mercado[1], Carlos Taffernaberry[1], Antonio Storni[1], Marcelo Ledda[1], Luis Alvarez[2], Andrés Benito[3], Ariel Verdejo[1], Nicolás Altamiranda[1].

[1]gridTICs – Grupo en Tecnologías de la Información y las Comunicaciones
Departamento de Electrónica / Facultad Regional Mendoza / UTN
{raul.moralejo, sebastian.tobar, gustavo.mercado}@gridtics.frm.utn.edu.ar

[2]IRESE - Instituto Regional de Estudios Sobre Energía
Departamento de Electromecánica / Facultad Regional Mendoza / UTN
alvarez.irese@frm.utn.edu.ar

[3]CLIOPE - Energía, Ambiente y Desarrollo Sustentable
Departamento de Electromecánica / Facultad Regional Mendoza / UTN
aobenito31@gmail.com

Resumen

Desde un contexto global, la red eléctrica inteligente (o REI; smart grid en inglés) se puede definir como la integración dinámica de los desarrollos en ingeniería eléctrica, almacenamiento energético y los avances de las tecnologías de la información y comunicación (o TIC), dentro del negocio de la energía eléctrica (generación, transmisión, distribución, almacenamiento y comercialización, incluyendo las energías alternativas)

El concepto de "micro smart grid" se diferencia de la "smart grid" en el concepto del alcance de la administración de la red, mientras una smart grid supone una sistema de distribución de energía con control administrativo público, una micro smart grid, es un sistema cerrado e interno a una entidad privada, que pretende administrar la energía de acuerdo a criterios y normas propias de la institución. En este proyecto se pretende diseñar e implementar una "micro smart grid" de alcance de campus universitario y que sirva para un uso racional y eficiente de la energía, haciendo uso de tecnologías de la información y las comunicaciones, diseñadas específicamente para este proyecto, que permitan una gestión inteligente del sistema.

Palabras Claves: Smart Grid, Uso Racional de Energía, Energías no Convencionales, Internet de las Cosas.

Contexto

El presente trabajo está inserto en el proyecto de investigación acreditado por la Universidad Tecnológica Nacional código ENUTIME0005424TC denominado "*Smart Micro Grid de Campus Universitario - Desarrollo, implementación y prueba.*".

El proyecto es llevado adelante por investigadores de los grupos, gridTICs, IRESE y Cliope. Todos pertenecientes a la Facultad Regional Mendoza de la Universidad Tecnológica Nacional.

1. INTRODUCCIÓN

Desde un contexto global, la red eléctrica inteligente (o REI; smart grid en inglés) se puede definir como la integración dinámica de los desarrollos en ingeniería eléctrica, almacenamiento energético y los avances de las tecnologías de la información y comunicación (o TIC), dentro del negocio de la energía eléctrica (generación, transmisión, distribución, almacenamiento y comercialización, incluyendo las energías alternativas); permitiendo que las áreas de coordinación de protecciones, control, instrumentación, medida, calidad y administración de energía, etc., sean concatenadas en un solo sistema de gestión con el objetivo primordial de realizar un uso eficiente y racional de la energía. Al anterior concepto también se podría dar la integración de otros actores en el área de la medición y

control, como lo son las fuentes de gas y el servicio de agua. Así, las redes eléctricas inteligentes entran a hacer parte de un macro-concepto de dominio territorial, como lo es el de las ciudades inteligentes (smart cities en inglés). La red eléctrica inteligente es una forma de gestión eficiente de la electricidad que utiliza la tecnología informática para optimizar la producción y la distribución de electricidad con el fin de equilibrar mejor la oferta y la demanda entre productores y consumidores. El término red inteligente se asocia a menudo con el concepto de medidores inteligentes capaces de ofrecer una facturación detallada por franjas horarias, lo que permitiría a los consumidores no solo el elegir las mejores tarifas de entre las diferentes empresas eléctricas, sino también discernir entre las horas de consumo, lo que a su vez permitiría un mejor uso de la red. Este sistema también permitiría mapear con más precisión el consumo y anticipar mejor las necesidades futuras a nivel más local. La irrupción de las energías renovables en el panorama energético ha cambiado notablemente los flujos de energía en la red eléctrica: ahora los usuarios no sólo consumen, sino que también producen electricidad a través de la misma red. Por tanto, el flujo de energía es ahora bidireccional. Una red inteligente envía electricidad desde los proveedores a los consumidores usando una tecnología digital bidireccional para controlar las necesidades del consumidor. Esto ayuda a ahorrar energía, reducir costos e incrementar la usabilidad y transparencia. El concepto de "micro smart grid" se diferencia de la "smart grid" en el concepto del alcance de la administración de la red, mientras una smart grid supone una sistema de distribución de energía con control administrativo público, una micro smart grid, es un sistema cerrado e interno a una entidad privada, que pretende administrar la energía de acuerdo a criterios y normas propias de la institución. En este proyecto se pretende diseñar e implementar una "micro smart grid" de alcance de campus universitario y que sirva para un uso racional y eficiente de la energía, haciendo uso de tecnologías de la información y las comunicaciones, diseñadas específicamente para este proyecto, que permitan una gestión inteligente del sistema.

Estado actual de conocimiento del tema

Smart micro grid

La interconexión entre las infraestructuras energéticas [1] inicialmente independientes ofrece flexibilidad y eficiencia adicionales del sistema. La integración a nivel de distribución simplifica la implementación de las funcionalidades integradas del sistema de energía. Se proponen conceptos y principios de diseño de una red de microenergía inteligente (MEG) para alojar microrredes, sistemas de poligeneración distribuida, instalaciones de almacenamiento de energía y las infraestructuras asociadas de distribución de energía. El sistema de administración de energía es responsable de la operación inteligente del MEG al tiempo que admite múltiples criterios, como seguridad, economía y protección del medio ambiente. Para realizar la visión del MEG inteligente, se investiga un sistema de administración de energía basado en la teoría de juegos de ingeniería con capacidad óptima de autoenvío.

Una microrred es un grupo localizado de fuentes de electricidad [2] y sumideros (cargas) que normalmente opera conectado y síncrono con la grilla centralizada tradicional. macrogrid), pero puede desconectar y mantener la operación de forma autónoma según lo dicten las condiciones físicas y / o económicas.

De acuerdo con el Grupo de Intercambio de Microrredes del Departamento de Energía de EE. UU., Los siguientes criterios definen una microrred:

Una microrred es un grupo de cargas interconectadas y recursos de energía distribuidos dentro de límites eléctricos claramente definidos que actúa como una única entidad controlable con respecto a la red. Una microrred puede conectarse y desconectarse de la red para permitir que funcione tanto en modo conectado a la red como en modo isla.

En cambio, las definiciones de microrred se centran principalmente en dos características: una microrred es un sistema controlado localmente una microrred puede funcionar tanto conectada a la red tradicional (megarejilla) como a una isla eléctrica.
Como ejemplos se pueden mencionar:

Mesa Del Sol

El desarrollo mixto comercial-residencial Mesa del Sol en Albuquerque [3], Nuevo México es el sitio de una micro red en pleno funcionamiento. El proyecto ha sido creado por la Organización de Desarrollo de Tecnología Industrial y Nueva Energía de Japón (NEDO), junto con el Estado de Nuevo México, la Mesa del Sol, la Empresa de Servicios Públicos de Nuevo México, el Laboratorio Nacional Sandia y varias empresas japonesas incluyendo Shimizu. Toshiba, Tokyo Gas, Mitsubishi, Fuji Electric, Furukawa Battery y otros.

El sistema comprende un sistema fotovoltaico solar de 50 kW y un gabinete de microrred que contiene una celda de combustible de 80 kW, un generador de gas natural de 240 kW, un banco de baterías de plomo ácido, almacenamiento térmico frío y caliente y un enfriador de adsorción. Bajo el acuerdo de NEDO, la microred fue entregada a la Universidad de Nuevo México (UNM) a fines de abril de 2014.

La isla de Kythnos

La isla de Kythnos se encuentra en el mar Egeo, cerca de Atenas [4]. El proyecto Kythnos Island Project fue financiado por el programa Europeo de Microrredes FP 5, cuyo objetivo fue probar estrategias de control centralizadas y descentralizadas para el islamiento.

Es una microrred autónoma de pequeña escala, integrada por una red trifásica de bajo voltaje, generación de energía solar fotovoltaica, almacenamiento de batería y un generador de respaldo. La red se compone de líneas eléctricas aéreas y un cable de comunicación que funciona en paralelo para cumplir con los requisitos de control y monitoreo.

Hay 10 kW de PV en dos ubicaciones, un banco de baterías nominal de 53 kWh y un grupo electrógeno diesel de 5 kW. Un segundo generador fotovoltaico de aproximadamente 2 kW conectado a un inversor SMA en el techo de los edificios del sistema de control proporciona energía para el monitoreo y la comunicación, respaldado por un banco de baterías de 32 kWh cercano.

Tres inversores SMA conectados en una configuración paralela maestra-esclava suministran energía a las 12 residencias solo de verano, cuyas cargas mínimas son principalmente de iluminación y bombeo de agua. Cuando los clientes exigen más potencia de la que los sistemas fotovoltaicos pueden proporcionar directamente, uno o más de los inversores de batería de 3,6 kW se activan. Los inversores de la batería pueden funcionar en modo isócrono o de caída. Operar en modo de caída de frecuencia permite pasar información a los controladores de carga de conmutación, que limitan las cargas si el estado de carga de la batería es bajo y también restringe la potencia de salida de los inversores fotovoltaicos si el banco de baterías está lleno.

Antecedentes

Este proyecto es continuación del proyecto FONARSEC 2013. Denominación: "Red inteligente ciudad Gral San Martín Mendoza", Proyecto FITS Energía 2013 - UREE. Proyecto 13, Resolución Nº 470/14 financiado por el Fondo sectorial de energía. Participantes: CAPP (EDESTE, EMESA y UTN FRM), con vigencia desde el 1/1/2015 hasta el 31/12/2019 [5][6][7].

Al cabo de los tres primeros años del proyecto se cuentan con los resultados siguientes:

Parque de 5000 medidores inteligentes tanto monofásicos como trifásicos instalados en clientes finales residenciales, comerciales, industriales, de alumbrado público y de riego agrícola, integrados a una red de comunicación y reportando a un sistema de recolección y administración de datos que permita: Gestión de la telelectura de los mismos, recibir avisos y eventos, controlar la suspensión y reanudación del suministro, y

aportar datos para avanzar con los desarrollos, estudios, sistemas, integraciones, optimizaciones.

Planta de generación fotovoltaica de 2MW instalada y funcionando interconectada a la red de distribución en Media Tensión de EDESTE S.A. telecontrolada, telemedida y telesupervisión, y con cuatro módulos de autogeneración de bajas potencias (5 a 15 KW) interconectados a la red de distribución de Baja Tensión instalados y funcionando en edificios públicos en la zona.

3. OBJETIVOS Y AVANCES DEL PROYECTO

Objetivo Principal

Desarrollar VÍNCULOS DE COMUNICACIÓN de alta confiabilidad y disponibilidad y SISTEMAS INFORMATIZADOS eficientes, que permitan GESTIONAR LA INFORMACIÓN obtenida a partir de la incorporación de tecnología apropiada en la red de baja tensión y con generación distribuida de energía no convencional en un campus universitario, buscando los siguientes objetivos específicos:

Objetivos específicos

1. Facilitar la normalización del servicio.

2. Disminuir las interrupciones de servicio y los tiempos de reposición.

3. Aumentar la confiabilidad, seguridad y disponibilidad de las redes de comunicación existentes.

4. Desarrollar metodologías para mejorar la eficiencia en las inversiones y los costos de Operación y Mantenimiento.

5. Desarrollar programa de mejoramiento de la calidad de producto.

6. Mejorar el control tendiente a disminuir el hurto de energía, fraude y vandalismo.

7. Evaluar la eficiencia de los sistemas fotovoltaicos

8 Investigar sobre las variables que afectan la eficiencia, monitoreando, simulando y modelando el sistema

9. Cuantificar los beneficios ambientales asociados con la implementación del sistema

Avances del Proyecto

El sistema está compuesto por tres partes constitutivas, a saber [8]:

- Sub Sistema AMI [9]

Medidor de energía (nod-e^{2}sg15-4) [10]: Consistente en nodo IoT de adquisición y transmisión, interoperabilidad semántica y sistema de comunicación.

WSN [11]: Consistente en gateway de comunicación de red interna

Plataforma de Smart Micro Grid [12]: Plataforma de almacenamiento y tratamiento de la información de la grid.

- Sub Energía Solar

Realiza tareas de simulación y monitoreo de los sistemas solares, cuantificando la energía incidente por medio de piranómetros de alta sensibilidad, temperatura, y producción eléctrica. Se modelará la relación entre las variables a fin de optimizar el funcionamiento

- Sub Eficiencia Energética

Realiza tareas de estudio y determinación de normalización del servicio, disminución de interrupciones de servicio y los tiempos de reposición, aumento la confiabilidad, seguridad y disponibilidad de las redes eléctricas, desarrollo de metodologías para mejorar la eficiencia en las inversiones y los costos de Operación y Mantenimiento

4. FORMACIÓN DE RECURSOS HUMANOS

Uno de los principales objetivos del proyecto es la capacitación de los recursos humanos.

La meta como investigadores es fortalecer la capacidad para realizar investigación científica, generar conocimientos y facilitar la transferencia de tecnología que permita el desarrollo humano.

Este proyecto de investigación posibilitará la colaboración inter-institucional y la ejecución de proyectos conjunto entre grupos I+D de diferentes disciplinas.

Para lograr estos objetivos se dispuso del siguiente personal:

Cinco Investigadores formados
Dos Investigadores de apoyo
Un Becario doctoral (beca UTN)
Tres Becarios graduados (Beca BINID UTN)
Tres Becarios alumnos (Beca alumno UTN)
Dos Tesistas de carrera de grado

Adicionalmente se realizarán:

- Dictado de Cursos, Seminarios y Conferencia para público especializado.

- Promoción, coordinación y asistencia técnica de tesinas de grado para alumnos de la FRMza.

- Promoción, coordinación y asistencia técnica a pasantes alumnos, cursantes de carreras de grado y de pre-grado en el ámbito de la UTN FRMza.

- Promoción, coordinación, dirección y asistencia técnica a Tesis doctorales, postgrado y/o maestría.

- Presentación de Trabajos en Congresos y Reuniones Técnicas/Científicas.

- Publicación de Trabajos en revistas con/sin referato.

- Publicación de todo el código fuente desarrollado en el presente proyecto, en github.com/GridTICs/smartmicrogrid.

5. BIBLIOGRAFÍA

[1] Shengwei Mei, Rui Li, Xiaodai Xue, Ying Chen, Qiang Lu, Xiaotao Chen, Carsten D. Ahrens, Ruomei Li, and Laijun Chen, "Paving the Way to Smart Micro Energy Grid: Concepts, Design Principles, and Engineering Practices" IEEE © 2017 CSEE

[2] About Microgrids,

https://building-microgrid.lbl.gov/about-microgrids

[3] Mesa del Sol,

https://building-microgrid.lbl.gov/mesa-del-sol

[4] Kythnos Island.

https://building-microgrid.lbl.gov/kythnos-island

[5] G. Mercado et al "SG-SM - Smart Grid San Martín. Red de Distribución y Generación de Energía Inteligente.en Ciudad Gral San Martin Mendoza" XVII Workshop de Investigadores en Ciencias de la Computación - WICC 2015, Salta 16-17, Abril 2015

[6] G. Mercado, J. Da Peña, G. Vivone, M. Ledda, R. Caceres, C. Taffernaberry, S. Pérez, L. Álvarez, A. Arena, G. Henderson, M. Funes, J. Catalá, "Diseño, Implementación y Análisis de Red de Distribución Inteligente y Generación Distribuida de Energía", Anales de CIIS 2016, May 2016, Universidad de Córdoba, Córdoba.

[7] A. P. Arena, L. Álvarez, J. Da Peña, R. Stasi, G. López, A. Burlot, L. Barnabo, M. Ledda, R. Caceres, C. Taffernaberry and G. Mercado, "Diseño, Implementación y Análisis de una Red Eléctrica Inteligente (Smart Grid)", Anales del ENIDI 2017, Mendoza 22 al 24 de Noviembre 2017

[8] R. Moralejo y S. Tobar "PID Smart Micro Grid de Campus Universitario - Desarrollo, implementación y prueba". UTN FRM, SCTyP. 2018

[9] A. K. Chakraborty and N. Sharma, "Advanced metering infrastructure: Technology and challenges," 2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), Dallas, TX, 2016.

[10] S. Tobar, C. Taffernaberry y A. Verdejo "nod-e^{2}sg15-4: Nodo AMI IoT para Smart Grid en IEEE802.15.4", Memoria Técnica PID5424 - gridTICs UTN FRM, 2019.

[11] C. Taffernaberry, G. Mercado, "GW-CIAA-IoT: Gateway con CIAA para red inalámbrica de IoT", WICC 2016, ISBN: 978-950-698-377-2, Abril 2016

[12] R. Moralejo y A. Storni. "Plataforma de gestión para MicroGrids", Sesión IoT y eCiencia‖, Anales del 2do Encuentro Latinoamericano de eCiencia, Octava Conferencia TICAL 2018. Cartagena de Indias, Colombia del 3 al 5 de septiembre de 2018.

Base de Datos

y

Minería de Datos

Tratamiento de Secuencias de ADN y Clustering de Pacientes con Cáncer Colorrectal.

Laura Avila*, Victoria Santa María**, Luis López*, Marcelo Soria***, Cristóbal R. Santa María*,
*DIIT-UNLaM, **Instituto Lanari-FMed-UBA, ***FAUBA
Florencio Varela 1903 San Justo Pcia. de Buenos Aires
54-011-44808952
laura_avila75@yahoo.com.ar
vctrsntmr@hotmail.com
llopez@ing.unlam.edu.ar
soria@agro.uba.ar
csantamaria@unlam.edu.ar

RESUMEN

Con este trabajo se continua la línea de investigación consistente en evaluar y desarrollar procedimientos computacionales adecuados para analizar la relación clínica entre el microbioma intestinal y la presencia del cáncer colorrectal. En esta oportunidad se ha trabajado con muestras propias obtenidas en el medio local. Corresponden a los microbiomas de 20 pacientes, 10 sanos y 10 enfermos, del Sector de Coloproctología del Hospital Italiano de Buenos Aires que fueron secuenciados a partir de materia fecal. La identificación bacteriana se realizó utilizando el gen marcador 16S rRNA para obtener la distribución de frecuencias a distintos niveles taxonómicos en cada paciente. El presente artículo describe el proceso que se ha realizado desde que las muestras salen del secuenciador hasta que son procesadas para su valoración clínica. Con tal objetivo los pacientes fueron agrupados por medio de algoritmos de aprendizaje no supervisado y se desarrolló el aspecto matemático de una distancia que trata de ajustar el clustering computacional a los objetivos clínicos. La metodología de trabajo empleada ha sido validada mediante la participación en la red Global Research Network to Investigate the CRC-associated Microbiome of non-Western Countries creada por la Universidad de Leeds, UK.

Palabras Clave: Secuencias, Microbioma, Clasificación, Cáncer Colorrectal

CONTEXTO

El Grupo de Investigación y Desarrollo en Data Mining del Departamento de Ingeniería e Investigaciones Tecnológicas de la UNLaM viene realizando evaluaciones y desarrollos de algoritmos con el fin de evidenciar los aspectos médicos de interés para diagnosticar y observar la evolución de patologías gastrointestinales tales como el cáncer colorrectal. Con tal finalidad desde 2015 ha desarrollado, dentro del programa de Incentivos, los proyectos de investigación C169 “Aplicaciones de Data Mining al Microbioma Humano” y C200 “Aplicación de Técnicas de Data Mining para Análisis del Microbioma Humano según Funcionalidades Metabólicas”. Actualmente lleva adelante, por primera vez a partir de muestras tomadas a pacientes autóctonos, el proyecto C220 del mismo Programa, “Explotación de Datos del Microbioma de Pacientes con Cáncer Colorrectal” en el marco de un convenio de colaboración con el Hospital Italiano de Buenos Aires firmado entre UNLAM e HIBA durante 2019.

1. INTRODUCCIÓN

Se estima que hasta el 90% de las condiciones de salud y enfermedad están asociadas de alguna manera al microbioma. Por ese motivo y por la posibilidad de intervenciones con prebióticos y antibióticos, los estudios metagenómicos basados en la Secuenciación de Nueva

Generación abren una nueva era en la prevención y tratamiento [1]. El cáncer colorrectal, que presenta características moleculares particulares y estrecha relación con la dieta "occidental" [2], es una patología de estudio de las más frecuentemente abordadas debido a su alta incidencia. La metagenómica orientada hacia el uso de genes marcadores como el 16S rRNA permite establecer el perfil taxonómico del microbioma de pacientes con cáncer colorrectal. En este camino es posible que en algún momento el análisis del microbioma alcance a transformarse en una herramienta auxiliar para el diagnóstico y evaluación de la enfermedad. Sin embargo, toda esta potencialidad depende en gran medida de que sea ajustada la interrelación entre lo bioinformático y lo médico. Cada algoritmo a utilizar, cada parámetro a ajustar, requieren de una evaluación acerca del grado en que colaboran a mejorar, en términos médicos, la herramienta de análisis. En tal sentido el vasto campo que constituye el dominio de las técnicas de proceso desde que se obtienen las secuencias de ADN del secuenciador hasta el desarrollo de métodos de aprendizaje automático que afinen la precisión en la clasificación médica, define la problemática a investigar por el grupo en el marco del convenio referido con el Hospital Italiano de Buenos Aires.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El trabajo pretende estudiar en detalle la aplicación de procedimientos computacionales supervisados y no supervisados sobre los microbiomas para clasificar y predecir patologías. Comprende tanto el enfoque a través del gen marcador, el caso de los resultados que aquí se presentan, como el enfoque a partir de la información de funcionalidad metabólica aportada por el metagenoma completo. Se intentan alcanzar varios objetivos:

-Dominar la tecnología de almacenamiento, comparación y distribución funcional según las secuencias obtenidas del microbioma intestinal de pacientes por videocolonoscopía o por materia fecal.

-Determinar los métodos computacionales más convenientes para los agrupamientos de microbiomas de pacientes de forma que revelen óptimamente sus características clínicas.

-Realizar lo propio respecto de algoritmos de predicción entrenados y testeados para la evaluación clínica.

-Dejar allanado el camino para la aplicación experimental de todos estos métodos a mayor cantidad de muestras de pacientes locales obtenidas por investigadores del grupo.

El primer trabajo que ha sido necesario realizar fue el de establecer la secuencia de procesos para poder hacer el análisis estadístico posterior. En la Figura 1 se muestra el diagrama de los procesos que se han efectuado.

Importación de los reads. Para su importación a QIIME2 [3] y [4], se ha creado un archivo delimitado por comas (.csv), en formato *fastq manifest*. Este archivo es el que ha permitido conectar los identificadores de las muestras con las rutas absolutas de los archivos fastq.gz que contiene las secuencias directas e inversas, indicando la dirección de la lectura para cada una.

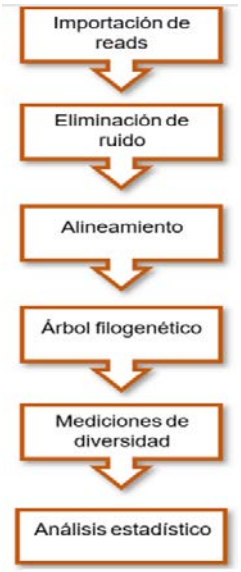


Figura 1. Pasos del proceso

Eliminación del ruido. En este paso del proceso se realiza el filtrado de las secuencias y se

eliminan las lecturas ambiguas o de baja calidad. También se realiza la eliminación de quimeras para evitar la falsa diversidad que se podría generar en análisis posteriores. Asimismo, se descartan las muestras cuyo recuento final sea inferior a 10000. Como resultado de dicha etapa se genera una tabla de frecuencias de las secuencias agrupadas en Unidades Taxonómicas Operacionales (OTU) como representantes de las lecturas.

Alineamiento. Se ha creado una tabla de secuencias alineadas, mediante algoritmos de alineamiento múltiple.

Árbol filogenético. Antes de realizar el árbol, ha sido necesario filtrar las secuencias, ya que el proceso de alineamiento añade ruido.

Mediciones de diversidad. Todo el trabajo realizado en los pasos previos ha permitido el estudio de la diversidad alfa y beta, mediante métricas filogenéticas y no filogenéticas [5]. Las medidas de alfa diversidad que se han utilizado son: Índice de diversidad de Shannon, OTUs observadas, análisis de correlación de Spearman. En cuanto a las medidas de diversidad beta, se han realizado análisis con diferentes distancias: Jaccard, Bray-Curtis, unweigthted Unifrac, weighted Unifrac y test de Adonis.

Se ha analizado la composición microbiana de las muestras según el grupo de pertenencia (sano o con presencia de cáncer de colon) y se ha aplicado test de Kruskal-Wallis a dichas matrices para determinar si existen diferencias estadísticamente significativas entre los grupos. Para ello se ha observado la relación entre la cantidad de OTUs y la diversidad de Shannon, y las variables categóricas de la metadata.

Para el análisis de diversidad beta se tuvieron en cuenta varias métricas y se realizó el análisis de permutaciones sobre las matrices de distancia (permutaciones) PERMANOVA, en la búsqueda de diferencias en la composición de los grupos de pertenencia con respecto al género. Por otro lado, se generaron gráficos a partir del análisis de componentes principales con las distintas métricas, que permitieron visualizar los resultados de la diversidad beta.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los análisis de diversidad alfa realizados con las métricas de Shannon y por OTUs observadas, confirman que no existen sesgos en cuanto a su condición clínica o a su género. Como ejemplo, se presenta la Figura 2 y la Tabla 1.

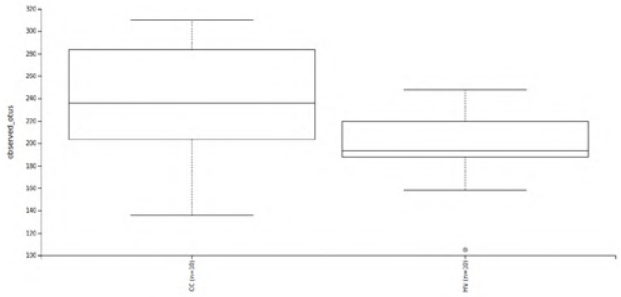


Figura 1. Gráfico de significación de diversidad de OTUs observadas dividida según el grupo (CC: cáncer de colon, HV: sano) al que pertenecen las muestras.

Resultados Kruskal-Wallis (todos los grupos)	
H	0,0701
p-valor	0,7911

Tabla 1. Test de Kruskal-Wallis para la diversidad de OTUs observadas según el grupo (CC: cáncer de colon, HV: sano) al que pertenecen las muestras.

Además de analizar la diversidad alfa, se analizó también la diversidad beta con diferentes métricas, puesto que este tipo de diversidad informa sobre el grado de diferenciación entre comunidades microbianas. Este análisis se ha realizado mediante componentes principales y usando el análisis de permutaciones PERMANOVA sobre las matrices de distancias. Ninguno de ellos reveló diferencias en la

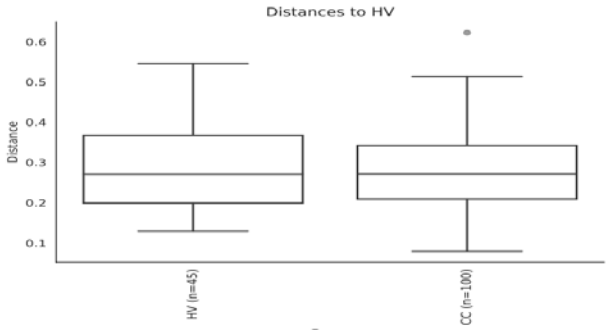


Figura 3. Gráfico que muestra la Unifrac entre cada grupo de interés respecto a los pacientes sanos.

composición entre los grupos, como se ve en la

Figura 3, que se muestra como ejemplo ya que los demás resultan similares.
El test estadístico aplicado para confirmar que las diferencias no son significativas ha sido el test PERMANOVA. Sus resultados se ven en la Tabla 2.

Resultados de PERMANOVA

Método	PERMANOVA
Nombre del test estadístico	Pseudo-F
Medida de la muestra	20
Número de grupos	2
Test estadístico	0,97616
p-valor	0,385
Número de permutaciones	999

Tabla 2. Test permanova a nivel taxonómico género.

Para abordar el análisis de la composición taxonómica de las muestras según los grupos de interés, en QIIME2 se ha asignado taxonomía a la tabla de secuencias representativas mediante el clasificador SILVA 132. Luego se colapsa la taxonomía a nivel 6 (nivel especie), en una tabla rarefaccionada, que permite exportarse para realizar otros tipos de análisis estadísticos.
La tabla de frecuencias taxonómicas rarefaccionada que se ha exportado del QIIME2 se cruzó con los metadatos de las muestras, es decir, la tabla obtenida incluye, además de las frecuencias absolutas de cada uno de los 239 Otus halladas, la clasificación en sano o enfermo, la edad y el sexo de cada paciente.
La información obtenida a la salida de QIIME2 se dispuso en tablas donde cada fila representa un microbioma, es decir un paciente, y en cada columna se ubican los taxones correspondientes al nivel taxonómico que se tiene en cuenta. Así hay tablas por género, familia, orden, clase y phylum que son los niveles a los cuales se realizan los estudios. La Tabla 3 muestra un ejemplo:

Columna1	OTU1	OTU2	OTU3	OTU4	OTU5	OTU6
GCRFNG_AR	0	0	5	0	42	0
GCRFNG_AR	0	0	0	0	22	0
GCRFNG_AR	160	0	2	0	213	0
GCRFNG_AR	0	0	0	0	359	1

Tabla 3. Ejemplo de tabulación a nivel género.

En la Tabla 3 la columna 1 corresponde a la identificación del paciente y las siguientes se asocian a las distintas Unidades Taxonómica Operacionales en las que se han agrupado las secuencias del gen marcador en cada microbioma. El número dentro de cada celda de la Tabla 3 es la cantidad de veces que la OTU respectiva se ha presentado en el correspondiente microbioma o, lo que es lo mismo, la cantidad de secuencias que han sido asignadas en ese microbioma a ese taxón. El total de OTUS halladas en todos los pacientes fue de 239. Es decir, se hallaron 239 géneros distintos en los que distribuir las secuencias de los genes marcadores aunque, claramente, no en todos los microbiomas se presentaron todos los géneros. La información de la tabla incluye en las tres últimas columnas, que no se ven, la clasificación en sano o enfermo, la edad y el sexo de cada paciente. A partir de ella se realizaron distintos procesos. El cálculo de la correlación lineal entre las variables y la clasificación de enfermedad o salud dio, como se esperaba, alta correlación lineal (-0.78) entre la edad y la enfermedad con un valor p del orden de 10^{-5} lo que autoriza a sostener tal correlación no solo en la muestra sino a nivel poblacional. A continuación, se realizó un clustering no jerárquico utilizando la distancia euclídea y encadenamiento promedio. Se tomaron en cuenta solo las variables que correspondían a cada taxón descartando la edad, el sexo y la clasificación médica respecto de la enfermedad. La Tabla 4 muestra los resultados.

Cluster	N° de Pacientes	Silueta
1	18	0,31
2	2	-0,14
Total	20	0,26

Tabla 4. Agrupamientos de pacientes

Es claro que el agrupamiento de bajo índice silueta total no revela nada sobre la condición clínica de los pacientes. La correlación entre la variable de clasificación de la enfermedad y la variable conglomerado asignado es exactamente

0 con un valor p de exactamente 1 lo que indica la imposibilidad de rechazar la hipótesis de no correlación a nivel poblacional. Sin embargo, se aprecian ciertas diferencias importantes de frecuencias promedio entre pacientes enfermos y sanos. Para evaluar la influencia real de cada diferencia en la disimilaridad de casos enfermos y sanos, las frecuencias medias pueden ser estandarizadas y luego calcular las diferencias para cada taxón (en el ejemplo el género u OTU). Tales diferencias se toman en valor absoluto mediante la cuenta $D_i = |fCC_i - fHV_i|$. Así se obtiene un perfil de diferencias de frecuencias medias estandarizadas entre pacientes sanos y enfermos como el que se muestra en la Figura 4 para el nivel taxonómico género.

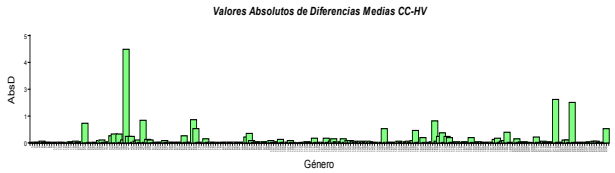


Figura 4. Diferencia de frecuencias medias sanos-enfermos

A continuación se realiza la cuenta $P_j = 1000 \frac{D_j}{\sum_{k=1}^{239} D_k}$ para todo j=1,2,…,239 obteniéndose un peso para cada diferencia. Con estos pesos se arma una distancia entre microbiomas i y k cuya fórmula es: $d = \sqrt{\sum_{j=1}^{239} (f_{ij} - f_{kj})^2 P_j}$. Esta distancia se propone para un nuevo clustering no jerárquico con encadenamiento promedio. Se observa entonces que los casos enfermos son todos bien clasificados, mientras que solo resultan bien clasificados la mitad de los pacientes sanos. La correlación lineal entre ambas variables arroja un coeficiente de 0.58 y el valor p fue 0.01 lo que indica que puede rechazarse la inexistencia de correlación con una probabilidad 0.01 de error. La nueva distancia pesada parece desempeñarse mejor para evaluar la similitud entre pacientes de acuerdo a su clasificación clínica. Los pesos obtenidos podrían muy bien utilizarse para medir distancias entre casos que no hayan integrado la muestra original. Resulta auspicioso que todos los casos enfermos hayan sido bien clasificados, pues la correlación de Spearman, que se utiliza también en variables cualitativas, dio relativamente alta y con muy baja probabilidad de error al extenderse a la población.

4. FORMACIÓN DE RECURSOS HUMANOS

En el equipo de trabajo participan un magister y un especialista en data mining, un doctor en biología, un médico, 2 ingenieros en sistemas y una matemática Está en curso una tesis de maestría.

5. BIBLIOGRAFÍA

[1] Di Bella, et al. 2013. High throughput sequencing methods and analysis for microbiome research. Journal of Microbiological Methods. Vol. 95, Issue 3, pp 401-414. https://doi.org/10.1016/j.mimet.2013.08.011

[2] Carbonetto, B., et al. 2016. Human Microbiota of the Argentine Population- A Pilot Study. Frontiers in microbiology, 7, 51. https://doi.org/10.3389/fmicb.2016.00051

[3] Bolyen E, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37: 852–857. https://doi.org/10.1038/s41587-019-0209-9

[4] D'Argenio V, et al. 2014. Comparative Metagenomic Analysis of Human Gut Microbiome Composition Using Two Different Bioinformatic Pipelines. BioMed Research International. Vol. 2014 Article ID 325340 https://doi.org/10.1155/2014/325340

[5] Xia, Y., & Sun, J. (2017). Hypothesis Testing and Statistical Analysis of Microbiome. Genes & diseases, 4(3), 138–148. https://doi.org/10.1016/j.gendis.2017.06.001

PREDICCIÓN DEL RESULTADO DE OSEOINTEGRACIÓN EN IMPLANTES DENTALES MEDIANTE MÚLTIPLES CLASIFICADORES

Nancy B. Ganz[1, 2, *], Alicia E. Ares[1, 2] y Horacio D. Kuna[2]
[1] Instituto de Materiales de Misiones (IMAM-CONICET). Félix de Azara 1552, N3300LQH, Posadas, Misiones, Argentina.
[2] Facultad de Ciencias Exactas, Químicas y Naturales (FCEQyN). Universidad Nacional de Misiones (UNaM). Félix de Azara 1552, N3300LQH, Posadas, Misiones, Argentina.
[*]E-mail: nancy.bea.ganz@gamil.com

RESUMEN

El campo de la ciencia de datos ha tenido muchos avances respecto a la aplicación y desarrollo de técnicas en el sector de la salud. Estos avances se ven reflejados en la predicción de enfermedades, clasificación de imágenes, identificación y reducción de riesgos, así como muchos otros. Sin embargo, no se han encontrado trabajos que apliquen procesos o técnicas de aprendizaje automático específicamente a la predicción de fracasos en implantes dentales. Este trabajo tiene por objetivo investigar el beneficio de la utilización de múltiples algoritmos de clasificación, para la predicción de fracasos en implantes dentales de la provincia de Misiones, Argentina. La experimentación es realizada con tres conjuntos de datos, un conjunto concerniente a historias clínicas de implantes dentales confeccionado para el estudio de caso, un conjunto generado artificialmente y un conjunto obtenido del repositorio de datos kaggle. En comparación con los clasificadores individuales, el enfoque propuesto obtiene el mayor porcentaje de acierto de los fracasos, logrando un 75% de casos correctamente identificados.

Palabras Clave: *combinación de clasificadores, clasificación, aprendizaje automático, implantes dentales, predicción.*

CONTEXTO

Esta línea de investigación se lleva a cabo dentro del Programa de Materiales y Fisicoquímica (PROMyF) en el Laboratorio de Ciencia de los Materiales del Instituto de Materiales de Misiones (IMAM), de la Facultad en Ciencias Exactas, Químicas y Naturales (FCEQyN), de la Universidad Nacional de Misiones (UNaM), en el marco de un plan de tesis doctoral, bajo el nombre de *"Predicción de fracasos en Implantes Dentales mediante la aplicación de Ciencia de Datos"*. Está financiado por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) a través de una "Beca Interna Doctoral" otorgada por Resolución D Nº 4869. Además, cuenta con el auspicio del Colegio de Odontólogos de la Provincia de Misiones.

1. INTRODUCCIÓN

Existen diversos antecedentes que combinan o integran clasificadores, para mejorar el porcentaje de acierto o incluso para no sesgar la decisión sobre los resultados de un solo clasificador [1]–[9]. Inspirados en las ideas de estos trabajos, proponemos un procedimiento de aprendizaje automático mediante la utilización de múltiples clasificadores para un dominio de poco conocimiento, como es el caso de los implantes dentales.

El conjunto de datos utilizado se basa en historias clínicas de pacientes que se han sometido a procesos quirúrgicos de colocación de implantes dentales en la Provincia de Misiones, Argentina.

Un paso importante en este trabajo de investigación fue la búsqueda de los mejores clasificadores individuales para el problema. Después de investigar los métodos existentes y teniendo en cuenta los enfoques de combinación, proponemos la utilización de cinco clasificadores: Random Forest (RF) [10], C-Support Vector (SVC) [11], K-Nearest Neighbors (KNN) [12], Multinomial Naive

Bayes (MNB) [13] y Multi-layer Perceptron (MLP) [14]. Estos clasificadores obtuvieron en evaluaciones exploratorias el mejor desempeño en comparación con otras combinaciones exploradas, que han incluido diferentes métodos como: Rpart, Ada, Gradient Boosting Machine (GBM) [15] y distintos clasificadores Naive Bayes [13].

El enfoque de integración propuesto para los cinco clasificadores consistió en combinar los resultados de sus predicciones para determinar el grado de pertenencia a la clase. Así, se pretende lograr una mayor precisión de la que cada uno de los clasificadores logra de manera individual para la etiqueta de clase objetivo (fracaso). La integración de los modelos se realiza con el método votación suave ponderada [16], [17]. Este método implica la definición de pesos asignados a cada clasificador, para ello se realiza una búsqueda en cuadricula para obtener los valores de los pesos que mejor se ajustan a los clasificadores en función de su desempeño.

Para la experimentación y validación del enfoque propuesto, se empela un conjunto de datos de casos reales de implantes dentales, un conjunto de datos generado artificialmente con la técnica de sobremuestreo de minorías sintéticas (SMOTE) [18], basado en los casos reales de implantes dentales, y un conjunto de datos obtenido del repositorio kaggle (Heart Disease).

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta línea de investigación propone estudiar, evaluar y aplicar distintos algoritmos de clasificación, para predecir los fracasos de un conjunto de datos de implantes dentales de la Provincia de Misiones. A través de la presente línea de investigación, se logró confeccionar el conjunto de datos necesario para proceder al diseño de un procedimiento, que a través de una metodología híbrida, permita identificar los factores de fracaso. Así mismo, se reconoce como objetivo de interés, el estudio de los tipos de tratamientos de superficie (arenado, ataque ácido, tratamiento térmico, etc.) de los implantes utilizados, para incorporar como característica a evaluar.

3. RESULTADOS OBTENIDOS

Entre los resultados obtenidos, destacamos la creación de un procedimiento que consistió en:

Paso 1. Leer el conjunto de datos.

Paso 2. Seleccionar la característica objetivo para la predicción.

Paso 3. Obtener el subconjunto de características más relevante, a través del método Chi Squared, el cual es un método muy utilizado para la selección de características categóricas [19]–[21]. Este evalúa el valor de una característica calculando el valor de la estadística de X^2 con respecto a la clase. Está dado por: $X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$.

Paso 4. Dividir el subconjunto resultante del paso 3 de forma aleatoria para preservar la distribución de ambas clases en: 70 % para entrenamiento y 30 % para evaluación [22]–[25] (ver tabla 1). Garantizando que todos los casos se encuentren representados en ambos conjuntos.

Paso 5. Para obtener un modelo robusto y optimizar los resultados de los clasificadores, se realizó una búsqueda en cuadrícula para ajustar los híper parámetros [22], [25], [26]. Esta búsqueda se efectuó con los datos de entrenamiento de cada uno de los conjuntos de datos. Para este proceso se especificó: un espacio de búsqueda, un algoritmo de optimización o ajuste (GridSearchCV) [27], un método de evaluación (validación cruzada de 10 iteraciones) y una medida de rendimiento (precisión de equilibrio).

Paso 6. Entrenar cada clasificador con los valores óptimos de los híper parámetros hallados en el paso 5.

Paso 7. Realizar la predicción de cada clasificar con los datos de prueba.

Paso 8. Integrar las predicciones mediante la aplicación del método votación suave ponderada. Este método recoge las probabilidades de clase predichas por cada clasificador, multiplica por un peso asignado al

mismo y los promedia. La etiqueta de clase final se deriva de la etiqueta de clase con la probabilidad promedio más alta. Está dado por: $\hat{y} = \arg max_i \sum_{j=1}^{m} w_j p_{ij}$. El peso fue determinado mediante una búsqueda en cuadricula utilizando un parámetro de prueba w con valores comprendidos entre 0 y 1. Esta búsqueda fue sometida a una validación cruzada de 10 iteraciones, donde se medió la precisión de cada clasificador para la clase en cuestión, seleccionando el valor de w que logró la mejor precisión [7], [9].

Paso 9. Validar los resultados hallados (ver Tabla 2). Los parámetros utilizados para evaluar el rendimiento de los clasificadores individuales y comparar con el enfoque propuesto, fueron: matriz de confusión, sensibilidad, especificidad y precisión [28], [29].

Tabla 1. Características de los conjuntos de datos utilizados para la evaluación experimental. De izquierda a derecha se presenta: nombres de los conjuntos de datos, número de muestras, número de atributos por tupla, cantidad de características seleccionadas por el método Chi-Square (X^2) y tamaño de los conjuntos de entrenamiento y de prueba.

Conjuntos de Datos	**Muestra**	**Características**	X^2	**Entrenamiento**	**Prueba**
Implantes Dentales	1165	33	17	815	350
Artificial	1748	33	21	1223	525
Problemas Cardíacos	303	13	10	212	91

Tabla 2. Eficiencia en el acierto de los clasificadores RF, SVC, KNN, MNB, MLP y el enfoque propuesto (Integrado) sobre los conjuntos de datos *Implantes Dentales*, *Artificial* y *Problemas Cardíacos*.

Conjuntos de Datos	**Clasificadores**	*Clase objetivo (Fracaso)* **Sensibilidad**	*Clase no-objetivo (Éxito)* **Especificidad**
Implantes Dentales	RF	59 %	98 %
	SVC	64 %	99 %
	KNN	64 %	99 %
	MNB	72 %	79 %
	MLP	66 %	97 %
	Integrado	75 %	96 %
Artificial	RF	81 %	97 %
	SVC	81 %	99 %
	KNN	81 %	99 %
	MNB	60 %	81 %
	MLP	82 %	97 %
	Integrado	89 %	97 %
Problemas Cardíacos	RF	81 %	71 %
	SVC	70 %	79 %
	KNN	70 %	76 %
	MNB	77 %	74 %
	MLP	72 %	68 %
	Integrado	94 %	58 %

En la tabla 1, se observa que los clasificadores SVC y KNN lograron el mejor rendimiento sobre la clase no objetivo para todos los conjuntos de datos en comparación con los demás clasificadores, incluso superan al enfoque propuesto en todos los casos. Para la clase objetivo, se aprecia que la integración de las predicciones de los cinco clasificadores permitió alcanzar el mayor porcentaje de acierto. Para esta clase, además se observa que el rendimiento de los clasificadores individuales fue variado. Si bien, el rendimiento de la integración de las predicciones no fue la mejor opción para la clase no objetivo, no quiere decir que haya sido la peor en comparación con las predicciones individuales. Mientras que la integración de las probabilidades para la clase objetivo fue la mejor opción, ya que permitió obtener el mayor porcentaje de acierto.

4. CONSLUSIONES

Se demostró que en este dominio es mejor integrar las predicciones de los clasificadores, para no sesgar la decisión sobre un solo resultado. Asimismo, utilizar predicciones integradas permite conocer diversos puntos de vistas o resultados para un mismo caso, ya que se utiliza más de un clasificador, permitiendo asegurar una asignación de etiqueta o clasificación más precisa.

Como resultado, cada clasificador logró hasta un 72% de acierto de la clase objetivo para el conjunto de implantes dentales, mientras que el enfoque propuesto permitió alcanzar el 75% de casos correctamente identificados como fracasos.

Finalmente, se plantea como trabajo futuro validar el enfoque propuesto con otros conjuntos de datos del área de la salud o la medicina. Además, se propone la inclusión o ampliación de los clasificadores utilizados, para evaluar la posibilidad de ajustar el porcentaje de acierto de ambas clases. Así como, realizar una validación con expertos humanos.

5. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto es parte de las líneas de investigación del "Programa de Materiales y Fisicoquímica" de la FCEQyN – UNaM, con cuatro integrantes: un investigador independiente del CONICET, un docente categoría I perteneciente al Depto. de Informática de la FCEQyN – UNaM, un doctorando y un tesista de grado de la carreara de Licenciatura en Sistemas de Información.

6. BIBLIOGRAFÍA

[1] J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho, "Multiple classifiers in biometrics. part 1: Fundamentals and review," *Inf. Fusion*, vol. 44, no. December 2017, pp. 57–64, 2018.

[2] Y. Miao, H. Jiang, H. Liu, and Y. dong Yao, "An Alzheimers disease related genes identification method based on multiple classifier integration," *Comput. Methods Programs Biomed.*, vol. 150, pp. 107–115, 2017.

[3] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, 2014.

[4] C. Catal and M. Nangir, "A sentiment classification model based on multiple classifiers," *Appl. Soft Comput. J.*, vol. 50, pp. 135–141, 2017.

[5] M. Pandey and S. Taruna, "Towards the integration of multiple classifier pertaining to the Student's performance prediction," *Perspect. Sci.*, vol. 8, pp. 364–366, 2016.

[6] J. Yan, D. B. Bracewell, F. Ren, and S. Kuroiwa, "Integration of Multiple Classifiers for Chinese Semantic Dependency Analysis," *Electron. Notes Theor. Comput. Sci.*, vol. 225, no. C, pp. 457–468, 2009.

[7] D. Ruano-Ordás, I. Yevseyeva, V. B. Fernandes, J. R. Méndez, and M. T. M. Emmerich, "Improving the drug discovery process by using multiple classifier systems," *Expert Syst. Appl.*, vol. 121, pp. 292–303, 2019.

[8] L. Oliveira, U. Nunes, and P. Peixoto, "On Exploration of Classifier Ensemble Synergism in Pedestrian Detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp.

16–27, 2010.

[9] H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-garadi, "Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions," *Inf. Fusion*, vol. 46, no. June 2018, pp. 147–170, 2019.

[10] L. Breiman, "Random Forest," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[11] C. Chang and C. Lin, "LIBSVM : A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–39, 2011.

[12] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *Am. Stat.*, vol. 46, no. 3, pp. 175–185, 1992.

[13] C. D. Manning, P. Raghavan, and H. Schutze, "Text classification and Naive Bayes," in *Introduction to Information Retrieval*, Cambridge University Press, 2009, pp. 253–287.

[14] B. Irie and Sei Miyake, "Capabilities of Three-layered Perceptrons," *IEEE nternational Conf. Neural Networks*, pp. 641–648, 1988.

[15] M. Mohandes, M. Deriche, and S. O. Aliyu, "Classifiers Combination Techniques: A Comprehensive Review," *IEEE Access*, vol. 6, pp. 19626–19639, 2018.

[16] L. N. Eeti and K. M. Buddhiraju, "A modified class-specific weighted soft voting for bagging ensemble," *Int. Geosci. Remote Sens. Symp.*, vol. November, pp. 2622–2625, 2016.

[17] X. Fan and H. Shin, "Road vanishing point detection using weber adaptive local filter and salient-block-wise weighted soft voting," *IET Comput. Vis.*, vol. 10, no. 6, pp. 503–512, 2016.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[19] X. Jin, A. Xu, R. Bie, and P. Guo, "Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles," *Data Min. Biomed. Appl.*, vol. 3916, pp. 106–115, 2006.

[20] M. Moran and G. Gordon, "Curious Feature Selection," *Inf. Sci. (Ny).*, vol. 485, pp. 42–54, 2019.

[21] I. Sumaiya Thaseen and C. Aswani Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 29, no. 4, pp. 462–472, 2017.

[22] B. T. Pham, M. D. Nguyen, K. T. T. Bui, I. Prakash, K. Chapi, and D. T. Bui, "A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil," *Catena*, vol. 173, no. September 2018, pp. 302–311, 2019.

[23] G. Isabelle, W. Maharani, and I. Asror, "Analysis on Opinion Mining Using Combining Lexicon-Based Method and Multinomial Naïve Bayes," *2018 Int. Conf. Ind. Enterp. Syst. Eng. (ICoIESE 2018)*, vol. 2, no. IcoIESE 2018, pp. 214–219, 2019.

[24] K. Bhattacharjee and M. Pant, "Hybrid Particle Swarm Optimization-Genetic Algorithm trained Multi-Layer Perceptron for Classification of Human Glioma from Molecular Brain Neoplasia Data," *Cogn. Syst. Res.*, vol. 58, pp. 173–194, 2019.

[25] D. Chong, N. Zhu, W. Luo, and X. Pan, "Human thermal risk prediction in indoor hyperthermal environments based on random forest," *Sustain. Cities Soc.*, vol. 49, no. April, p. 101595, 2019.

[26] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," *J. Inf. Sci.*, vol. 44, no. 1, pp. 48–59, 2018.

[27] sciki-learn, "Tuning the hyper-parameters of an estimator," 2019. [Online]. Available: https://scikit-learn.org/stable/modules/grid_search.html#grid-search. [Accessed: 04-Jul-2019].

[28] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.

[29] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.

BIOINGENIERÍA INFORMÁTICA APLICADA A LA PREDICCIÓN DE ENFERMEDADES CARDIOLÓGICAS Y SU IMPLEMENTACIÓN EN EL HOSPITAL DELICIA CONCEPCIÓN MASVERNAT DE LA CIUDAD DE CONCORDIA, PROVINCIA DE ENTRE RÍOS

[1] FALAPPA, Marcelo Alejandro, [2] BENEDETTO Marcelo Gabriel, [2] CARABIO, Ana Lía Ramona, [2] SILVA LAYES, María Elizabeth, [2] BENÍTEZ, Duval Horacio

[1] Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur Avenida Alem 1253 - Bahía Blanca (B8000CPB) - Tel.: +54(0291)4595135 mfalappa@cs.uns.edu.ar

[2] Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos Monseñor Tavella 1424 – Concordia, Entre Ríos (3200) - Tel.: +54(0345)4231433 {marben, anacar}@fcad.uner.edu.ar, elizabeth.silva@gmail.com, dhbenitez@gmail.com

RESÚMEN

El sector sanitario, sin lugar a dudas es uno de los ámbitos en el que se administran grandes volúmenes de datos; principalmente en el área clínica. Esto conduce a identificar una importante necesidad de encontrar maneras de administrar, integrar, analizar e interpretar ese gran conjunto de datos; procurando identificar patrones de comportamiento que sean de utilidad en la toma de decisiones médicas. En este sentido, la minería de datos es una herramienta que nos ayuda en el análisis de los mencionados datos, sirviendo de apoyo tanto en el área de cuidados de la salud como en la investigación médica; a través de su aplicación que abarca desde el diagnóstico médico hasta la capacidad para analizar datos desde una variedad de perspectivas tendientes a descubrir patrones ocultos [4].

Actualmente, existen diversas investigaciones que permiten encarar tratamientos de enfermedades específicas, aunque siguen siendo estudios puntuales que, si bien tienen muy buenos resultados, no son de aplicabilidad a la práctica cotidiana de consultas médicas. Teniendo en cuenta esta realidad, la investigación que se plantea persigue lograr el desarrollo de un componente de software que permita incorporar la predicción de probabilidad de riesgos de enfermedades cardíacas como sistema de soporte a las decisiones clínicas incorporada al acto asistencial; así como también en el proceso de gestión de medicina preventiva. Si bien existen diferentes proyectos aplicados que permiten conocer sobre la probabilidad de riesgos, los mismos no terminan siendo integrados a la Historia Clínica Electrónica, de manera de tener una aplicación inmediata en el acto asistencial [11].

En este proyecto se plantea como principal objetivo desarrollar un componente de software capaz de generar, con aprendizaje automatizado, un modelo con capacidades predictivas sobre enfermedades cardiológicas; que permitan un mejor soporte a decisiones de diagnóstico clínico y un avance significativo en la medicina preventiva para actos asistenciales de pacientes.

Palabras clave: Minería de Datos, Sistemas de Soporte a Decisiones Clínicas, Modelos Predictivos.

CONTEXTO

La línea de investigación descripta en este artículo se encuentra enmarcada en el contexto del proyecto PID 7060 denominado: "*Bioingeniería Informática aplicada a la predicción de enfermedades cardiológicas y su implementación en el Hospital Delicia Concepción Masvernat de la ciudad de Concordia, Provincia de Entre Ríos*" de la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos. El citado proyecto se enmarca dentro del Convenio Específico de Colaboración entre Programas de Investigación y Postgrado, celebrado en el año 2008 entre la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos y el Departamento de Ciencias e Ingeniería de la Computación de la

Universidad Nacional del Sur y en convenio con el Hospital Delicia Concepción Masvernat de la ciudad de Concordia, Provincia de Entre Ríos.

Uno de los principales objetivos de este proyecto es la formación de recursos humanos para investigación en la Facultad de Ciencias de la Administración de la UNER, especializados en la línea prioritaria de investigación denominada "Ingeniería de Software y Lenguajes de Programación" establecida por Res. 25/11 del C.D. Al ser también ésta una de las principales líneas de investigación del Departamento de Ciencias e Ingeniería de la Computación de la UNS, se justifica la creación de un equipo de investigación inter-universidades que sea contenedor del desarrollo de investigadores de la U.N.E.R. en el área.

INTRODUCCIÓN

Actualmente es inconmensurable la información con la que cuentan las instituciones de salud sobre sus pacientes, tanto a nivel administrativo como asistencial. Los datos referidos al área clínica son utilizados por el médico para obtener información del paciente, ya sea para efectuar diagnósticos, tratamientos, medicación; así como también por los propios sistemas informáticos para generar múltiples alertas (como pueden ser interacciones medicamentosas, control sobre medicamentos contraindicados para una patología específica, resultados anormales en estudios, entre otros).

El análisis de este gran cúmulo de datos nos brinda una oportunidad inmejorable para [1]:
- Realización de estudios epidemiológicos (análisis de rendimientos de campañas de información, prevención, sustitución de fármacos, entre otros),
- Cálculo de expectativas de vida,
- Identificación de terapias médicas satisfactorias para diferentes enfermedades,
- Asociación de síntomas y clasificación diferencial de patologías,
- Estudio de factores de riesgo para la salud en distintas patologías,
- Segmentación de pacientes para una atención más inteligente y
- Identificación de terapias médicas y tratamientos erróneos para determinadas enfermedades.

De esta manera, aparecen nuevas oportunidades para mejorar, no sólo la medicina personalizada sino también el rendimiento y resultado de los sistemas sanitarios. Esto conlleva a identificar necesidades referidas a nuevas maneras de administrar, integrar, analizar e interpretar tales conjuntos de datos; que posibiliten encontrar patrones de comportamiento de utilidad en la toma de decisiones médicas. En este aspecto, la minería de datos es una herramienta fundamental que resulta imprescindible en el análisis de dichos datos.

Si bien, en la actualidad existen investigaciones que conllevan a desarrollos informáticos que sirven de apoyo en tratamientos de distintas enfermedades, las implementaciones que permitan incorporar el apoyo a la toma de decisiones clínicas en el marco de la atención del paciente en tiempo real son escasas.

Teniendo en cuenta que en nuestro país una de las principales causas de muerte son las enfermedades cardiovasculares, se considera necesario el desarrollo de sistemas fiables que sirvan de apoyo a las decisiones clínicas en dos momentos fundamentales (la respuesta del sistema cuando el paciente concurre a una visita médica y una respuesta proactiva que permita brindar apoyo a través de la aplicación de medicina preventiva); logrando reducir el tiempo del diagnóstico y aumentar la precisión del mismo en este tipo de enfermedades [11].

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO.

La línea de trabajo principal del proyecto de investigación radica en el estudio, análisis y comparación de técnicas de minería de datos para ser aplicados en la generación de un

modelo predictivo sobre enfermedades cardiovasculares.

Se prevé desarrollar un componente de software que pueda integrarse a una historia clínica electrónica para brindar apoyo en la toma de decisiones en el diagnóstico clínico.

OBJETIVOS

Objetivo General:

Desarrollo de un componente de software capaz de generar, con aprendizaje automatizado, un modelo con capacidades predictivas sobre enfermedades cardiológicas; que permitan un mejor soporte a decisiones de diagnóstico clínico y un avance significativo en la medicina preventiva para actos asistenciales de pacientes.

Objetivos específicos:

- Evaluar diferentes técnicas de minería de datos para generar un modelo aplicado a la predicción de probabilidad de riesgo en enfermedades cardiovasculares.
- Generar un componente de software (*Web Service*) para ser integrado a la Historia Clínica Electrónica (HCE) de las instituciones sanitarias; con implementación de una prueba piloto para instancias asistenciales en el Hospital público "*Delicia Concepción Masvernat*" de la ciudad de Concordia, provincia de Entre Ríos.
- Desarrollar un módulo de soporte a la medicina preventiva, con capacidades de generar alertas tempranas sobre pacientes con riesgo cardiológico, con umbrales parametrizables basado en rango etario y sexo.
- Medir las alertas generadas por el sistema y realizar una comparativa con datos estadísticos de referencia de la Organización Mundial de la Salud (OMS); de manera de evaluar las capacidades del modelo generado determinando el sesgo propio por las características de la población de referencia.

RESULTADOS ESPERADOS

Siendo que el proyecto se encuentra en su fase inicial, durante la ejecución del mismo se pretenden alcanzar los siguientes resultados por actividades:

Relevamiento bibliográfico: Exploración de técnicas de minería de datos aplicables, detectando clasificadores relevantes para el dominio cardiológico. Selección de variables de uso frecuente a incorporar al estudio. Revisión de modelos algorítmicos aplicables.

Preprocesamiento de datos: Recopilación de datos, conformación del *dataset* de entrenamiento, depuración, extracción de ruido, validación de datos. Preparación de datos en general, codificación de variables. Categorización de datos, binarios, multivaluado, evaluación de datos no categóricos, establecimiento de rangos y acondicionamiento de tipos. Preprocesamiento de datos no estructurados.

Selección de clasificadores: Medición del rendimiento de clasificadores evaluados, descartando todos aquellos que muestran poca incidencia cardiológica, detección de nuevos clasificadores que no han sido considerados aún, medición de indicadores, clasificación de instancias. Conformación del *dataset*.

Generación de modelos predictivos: Análisis y comparación de distintos métodos de *machine learning* aplicables, medición de indicadores relevantes de evaluación del modelo, medición de sensibilidad, especificidad, tasa de falsos positivos~~,~~ y tasa de falsos negativos. Evaluación de incorporación de pesos relativos de detecciones erróneas que permitan corregir el modelo, establecimiento del rango de actuación del sistema y frecuencia de actualización óptima.

Análisis de factores extras necesarios: Detección de factores, nuevos datos e indicadores~~;~~ que permitan generar nuevas

estrategias de captación de datos, para una mejora del modelo, proyección de necesidades futuras, recomendación de políticas de captación de datos relevantes.

Desarrollo del componente de software. Desarrollo del W*eb Service* de interoperabilidad con historias clínicas de instituciones de salud, e implementación del piloto en una institución del medio. Recomendaciones para la adopción del modelo.

Evaluación empírica del componente generado: Validación de alertas generadas, revisión en conjunto con profesionales médicos, análisis de las capacidades del modelo, determinación de resultados logrados, recomendaciones futuras, determinación de aportes factibles a políticas sanitarias de la población en general.

Registro, exposición del conocimiento adquirido y transferencia al medio: Registro y publicación de logros de proyecto, publicación científica, transferencia al medio, implementación de prueba piloto en el Hospital público "*Delicia Concepción Masvernat*" de la ciudad de Concordia, provincia de Entre Ríos, si los resultados logrados muestran aporte significativo y validan la hipótesis inicial.

FORMACIÓN DE RECURSOS HUMANOS

Se brindará a los integrantes del proyecto, formación en lo que se refiere a técnicas avanzadas de Minería de Datos, Análisis de Datos y Data Science.

Los integrantes del proyecto, se desempeñan en cátedras relacionadas directamente con el tema central de la investigación, por lo que este trabajo tendrá impacto directo e inmediato en la docencia.

Se procederá a dirigir becarios de investigación, así como también tesinas finales de grado, dirigidos por el director del proyecto de investigación y/o por los docentes investigadores del mismo. Para estos casos, también se prevé la presentación a convocatorias de becas ante organismos provinciales y nacionales.

Los integrantes participarán de reuniones científicas y técnicas que permitan actualizar los conocimientos en el tema de interés. También se trabajará en la presentación de trabajos en congresos nacionales e internacionales relacionados con el área del proyecto. Estos trabajos servirán para divulgar los conocimientos obtenidos durante el trabajo de investigación.

Uno de los principales objetivos del proyecto es que el personal docente de la UNER dedicado al mismo avance y/o concluya con sus estudios de posgrado, así como también se incorporen becarios realizando investigaciones en temas afines a la temática del proyecto.

El Director del proyecto, Marcelo Alejandro FALAPPA, es Doctor en Ciencias de la Computación egresado de la Universidad Nacional del Sur (UNS) en el año 1999. Actualmente es Profesor Asociado con Dedicación Exclusiva en el Departamento de Ciencias e Ingeniería de la Computación de la UNS a cargo de la materia curricular Bases de Datos, es Investigador Adjunto del CONICET especializado en Actualización de Bases de Datos Deductivas y Teoría de Cambio, y tiene Categoría II en el Programa de Incentivos.

El codirector del proyecto, Marcelo Gabriel BENEDETTO, es Magíster en Sistemas de Información, se encuentra realizando los cursos del Doctorado en Ciencias de la Computación de la U.N.S. y posee categoría III en el Programa de Incentivos.

La integrante, Licenciada en Sistemas de Información Ana Lía Ramona CARABIO, posee categoría V en el Programa de Incentivos y se encuentra realizando la Maestría en Redes de la U.N.LP., restando la presentación de la Tesis; la integrante María Elizabeth SILVA LAYES es Magíster en Sistemas de Información de la UNER. Finalmente, el Licenciado en Sistemas Horacio Duval BENÍTEZ se encuentra realizando los

cursos de la Maestría en Sistemas de Información de la UNER y posee categoría V en el Programa de Incentivos.

BIBLIOGRAFÍA

[1] Ana Lía Carabio, Elizabeth Silva Layes, Fabian Frola, O. Bioingeniería aplicada en el diagnóstico de enfermedades. 45 JAIIO, CAIS 2016, Séptimo Congreso Argentino de Informática y Salud (2016), 240–248.

[2] Gillespie, G. There's gold in them thar' databases. Health Data Management 8(11) (2000), 40–52.

[3] Andrew Wilson, Lehana Thabane, A. H. Application of data mining techniques in pharmacovigilance. British Journal of Clinical Pharmacology 57(2) (2003), 127–134.

[4] Kuttiannan Thangavel, Palanichamy Iaganathari, P. E. Data mining approach to cervical cancer patients analysis using clustering technique. Asian Journal of Information Technology 5 (4) (2006), 413–417.

[5] Sellappan Palaniappan, R. A. Intelligent heart disease prediction system using data mining techniques. IEEE/ACS International Conference on Computer Systems and Applications (2008), 108–115.

[6] Techentin, B. Big data and graph analytics in a health care setting. http://www.graphanalysis.org/SC12/03_Techentin.pdf, noviembre 2012. Online; accedido en 2018.

[7] Ayad Ghany Ismaeel, D. Y. M. Effective data mining technique for classification cancers via mutations in gene using neural network. International Journal of Advanced Computer Science and Applications 7(7) (2016), 69–76.

[8] Shantakumar Patil, Y. K. Intelligent and effective heart attack prediction system using data mining and artificial neural network. European Journal of Scientific Research Vol.31 No.4 (2009), 642–656.

[9] Examples of Big Data Analytics In Healthcare That Can Save People. https://www.datapine.com/blog/big-data-examples-in-healthcare/. 2018

[10] Bernard Marr. Big Data In Healthcare: Paris Hospitals Predict Admission Rates Using Machine Learning. https://www.forbes.com/sites/bernardmarr/2016/12/13/big-data-in-healthcare-paris-hospitals-predict-admission-rates-using-machine-learning/#1532778779a2 (2016)

[11] Elizabeth Silva Layes, Marcelo Alejandro Falappa, O. Sistemas de soporte a las decisiones clínicas. XIII Workshop de Investigadores en Ciencias de la Computación (2011), 113-115

Predicción de la enfermedad de Parkinson utilizando redes neuronales convolucionales

Renata S. Guatelli, Verónica I. Aubin, Silvia N. Pérez

Departamento de Informática e Investigaciones Tecnológicas
Universidad Nacional de La Matanza
rguatelli@unlam.edu.ar; vaubin@unlam.edu.ar; sperez@unlam.edu.ar

RESUMEN

La enfermedad de Parkinson (EP) es un desorden neurodegenerativo del sistema nervioso, de causa desconocida y curso crónico, progresivo e irreversible. En la actualidad se asume que los cambios patofisiológicos que permiten apreciar los síntomas de la enfermedad, no son visibles hasta al menos cuatro años luego de su inicio. Por esta causa, se buscan métodos alternativos que permitan detectar la enfermedad en forma temprana. Dado que las deficiencias del habla es uno de los síntomas de la enfermedad, esto puede dar origen a un biomarcador para el diagnóstico temprano y el monitoreo de la enfermedad.

Este trabajo propone un estudio a partir del aprendizaje profundo de los espectrogramas obtenidos de señales de voz grabadas con celulares. Como objetivo se plantea realizar aportes al diagnóstico de EP, contribuyendo asimismo al conocimiento de las características de la voz afectadas por la enfermedad.

Para tal fin se creará una base de datos de espectrogramas de los segmentos de audio que mejor permitan caracterizar la voz de los EP. Se desarrollarán modelos de redes neuronales convolucionales con distintas arquitecturas para distinguir los EP de los pacientes sanos, utilizando la validación adecuada para las características de dichos datos.

Palabras clave: *espectrograma, red neuronal, enfermedad de Parkinson.*

CONTEXTO

Esta investigación se realiza como parte del Proyecto de investigación PROINCE-C224, de la Universidad Nacional de La Matanza. Se desarrolla conjuntamente entre el Departamento de Salud y el de Ingeniería e Investigaciones Tecnológicas de la UNLaM.

Propone el análisis estadístico de señales acústicas de la voz como método objetivo y no invasivo para el diagnóstico precoz de la enfermedad de Parkinson (EP).

El mismo es continuación del proyecto PROINCE-C199 de la Universidad Nacional de La Matanza, llevado a cabo durante los años 2017 y 2018.

1. INTRODUCCIÓN

La enfermedad de Parkinson idiopática (EPI) es un desorden neurodegenerativo del sistema nervioso, de causa desconocida y curso crónico, progresivo e irreversible. Las manifestaciones motoras marcadoras de la enfermedad están representadas por temblor de reposo, rigidez y bradicinesia [1]. En su evolución natural se agregan otros síntomas como: trastornos de la marcha, desequilibrio y alteraciones de la voz lo que implica un marcado impacto en la calidad vida limitando la movilidad con caídas frecuentes y aislamiento social como resultado de severa disartria [2].

Su proceso de diagnóstico y seguimiento, aún hoy en día, es lento y complicado al no existir ninguna prueba especıfica. En su defecto, el personal sanitario tiene que evaluar concienzudamente el historial clínico, síntomas y exámenes físicos y neurológicos de cada paciente, y ofrecer un diagnóstico a menudo poco preciso. Es por ello, que se estima que hasta un 25% de los diagnosticados como enfermos de Parkinson realmente padecen otra enfermedad, es decir, el proceso actual provoca numerosos falsos positivos.

La rigidez asociada afecta al sistema respiratorio encargado de facilitar el flujo y la presión de aire necesarios para la emisión

vocal lo que se traduce en una disminución de presión sub-glótica. Esto se refleja en un tono inadecuadamente bajo, resonancia gutural posterior, intensidad baja, voz monótona e hipo nasalidad [3]. Las deficiencias del habla son comúnmente uno de los síntomas tempranos de la enfermedad, por lo que puede ser un buen biomarcador para el apoyo diagnóstico y el monitoreo de la enfermedad.

Se estima que entre un 60-80% de paciente con EPI presentan alteraciones de la voz caracterizados por cambios en frecuencia, duración e intensidad [4], [5]. Estas alteraciones se confunden en muchas ocasiones con los cambios naturales de los adultos mayores en relación con la presbifonía [6], [7] o estados depresivos [8]. Los pacientes con EPI reflejan la disminución en la capacidad de los músculos laríngeos para mantener una posición fija en la pronunciación sostenida de las vocales [9]. También se observa una disminución en la capacidad para producir prosodia.

Entre los hallazgos más frecuentes en la valoración de la voz de pacientes con EPI se encuentra el incremento u otra variación de la frecuencia fundamental (F0), la reducción del tiempo de producción de vocales, el aumento del VOT (del inglés Voice Onset Time, tiempo de inicio de la sonoridad de la vocal que sigue a la consonante), la disminución de la intensidad en la fonación, así como el decrecimiento del MPT (tiempo máximo de fonación), perturbación del tono Jitter e intensidad Shimmer, la razón ruido/armónicos, también ha mostrado diferencias significativas [10]. Metter y Hanson pusieron de manifiesto que el incremento en la F0 es paralelo a la gravedad de los síntomas y al avance de la enfermedad [11].

La alteración en prosodia expresiva está documentada y justificada por la reducción [12], variabilidad [11] e intensidad [13], [14], de la F0 en tareas de lectura de párrafos en los que los sujetos debían imitar frases acentuando su contenido emocional. Estos resultados ayudan en la búsqueda de los efectos que provoca esta enfermedad en la respiración, fonación, articulación y prosodia [15].

En los últimos años se han propuesto varios sistemas computarizados para identificar los síntomas tempranos de la enfermedad de Parkinson, monitoreando los patrones de voz [16], [17], [18]y [19]. En la Universidad de Oxford, [20], [18], definieron un conjunto de medidas de disfonía y compararon los resultados de usar cuatro algoritmos de selección de características y clasificación binaria. Los integrantes del grupo de investigación Informática Aplicada al Procesado de Señal de la Universidad Politécnica de Madrid (UPM) en colaboración con el Centro de Tecnología Biomédica, han desarrollado un simulador biomecánico para modelar las particularidades del aparato fonador y han evaluado diferentes propuestas de extracción de parámetros para detectar la existencia de patologías utilizando la señal de voz de los pacientes [21].

A través del análisis articulatorio de la repetición rápida de sílabas como /pa-ta-ka/, calculando características temporales y espectrales extraídas en los segmentos de tiempo de inicio de voz [22], propone diseñar un sistema experto para la detección temprana de la EP.

Las redes neuronales convolucionales (CNN) han permitido mejoras significativas en aplicaciones en visión por computadora, tales como detección de objetos, reconocimiento de rostros, clasificación de imágenes, procesamiento del lenguaje natural, procesamiento de audio, entre otros [23]. Los algoritmos de aprendizaje profundo o deep learning buscan en el aprendizaje automático características de alto de alto nivel a partir de grandes volúmenes de datos [24].

Los avances en las TICs (Tecnologías de la Información y Comunicación), han abierto la posibilidad del seguimiento online de la enfermedad mediante el análisis de grabaciones de voz recogidas con teléfonos móviles. Esto permite disponer de un gran número de datos de pacientes con mínimo trabajo de recopilación. Estas tecnologías junto a las aplicaciones de aprendizaje profundo hacen viable la asistencia sanitaria móvil.

Una señal de audio se puede representar visualmente como un espectrograma que captura el contenido de frecuencia variable de la señal, ver Figura 1. Es una representación en tres dimensiones, tiempo, frecuencia y amplitud de la distribución de energía de una señal. En la imagen se visualiza la intensidad del sonido mediante las variaciones de color, a lo largo del tiempo (eje horizontal) en función de la frecuencia (eje vertical).

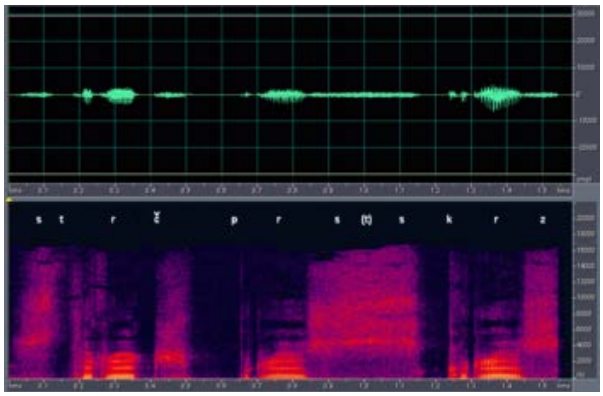


Figura 1. Espectrograma generado a partir de una señal de voz.

En este trabajo de investigación se propone discernir entre pacientes con EP y personas sanas a partir del uso de una CNN aplicada a los espectrogramas de señales de voz obtenidas con teléfonos móviles. Los sonidos que se utilizarán corresponden a los de la base de datos mPower, que contiene más de 64.000 grabaciones de voz, capturadas con teléfonos celulares discriminados entre enfermos de Parkinson y controles sanos. También se realizarán pruebas sobre la base de datos generada en el proyecto C-224, grabada en un ambiente controlado, siguiendo el protocolo de grabación acordado con profesionales del Hospital Nacional Alejandro Posadas, del Hospital Bernardino Rivadavia y con el asesoramiento del Dr. Jorge Gurlekian.

2. LINEA DE INVESTIGACION

La línea de investigación que se desarrolla pretende detectar EP utilizando un método no invasivo. En la literatura, existe un consenso sobre que el análisis acústico resulta útil para tal fin.

La propuesta aquí presentada consiste en obtener modelos que permitan colaborar con el diagnóstico temprano de la enfermedad, estudiando la aplicación del aprendizaje profundo de los espectrogramas, obtenidos a partir de registros de señales de voz. Dichas señales corresponden a grabaciones realizados a través de teléfonos celulares, ya que esta tecnología promueve la asistencia sanitaria móvil.

Asimismo, se explorarán distintas particularidades de la voz de los EP, pre-procesando los audios para extraer distintos fragmentos (por ej. comienzo, final, etc.) que permitan evaluar a través de las CNN con cuál de ellos se logra una mejor clasificación.

3. RESULTADOS

Como resultado de esta investigación se obtendrán modelos de detección de la enfermedad de Parkinson, utilizando espectrogramas obtenidos a partir de datos de voz recogidos a través de grabaciones de teléfonos móviles, implementando algoritmos de redes neuronales basadas en Deep Learning.

La metodología utilizada contempla tres etapas:

- Pre-procesado de los audios y creación de la base de datos
- Construcción de la red neuronal convolucional
- Procesamiento de la Base de Datos

En la primera etapa se procesarán los audios. Se seleccionarán y aplicarán distintos algoritmos para eliminar el ruido de fondo. Se seleccionarán los segmentos de audio significativos para el estudio del reconocimiento de EP y se crearán los espectrogramas de los audios seleccionados.

En la segunda etapa, una vez tratados los datos iniciales para introducirlos en la red neuronal convolucional, se desarrollarán los modelos correspondientes a distintas arquitecturas propuestas en la literatura.

Por último, en la **tercera etapa, s**e evaluarán y compararán los modelos generados. Para la evaluación de cada modelo se seleccionará el tipo de validación más adecuada teniendo en cuenta las características de los datos, asegurándose que el conjunto de entrenamiento y de test sean representativos por sí mismos. Para comparar los modelos

desarrollados se implementarán diferentes indicadores de desempeño como matriz de confusión y curva ROC, entre otros.
El proyecto de investigación se desarrolla conjuntamente entre el Departamento de Salud y el de Ingeniería e Investigaciones Tecnológicas de la UNLaM. Los resultados obtenidos impactarán sobre los recursos disponibles de ambos departamentos. Estos implican un avance del conocimiento en el campo de la detección de enfermedades neurodegenerativa, posibilitando la formación de recursos humanos capacitados en una nueva línea de investigación para esta casa de altos estudios.
Se propone además realizar vínculos con grupos de investigación de Argentina o el exterior, que se encuentren trabajando en temas afines a este proyecto.
Se divulgarán los resultados obtenidos y se hará transferencia de tecnología a potenciales usuarios del área de salud. Se prevé como primeros beneficiarios a la comunidad de La Matanza y luego a la comunidad en general, ya que se busca un método de bajo costo para el diagnóstico precoz del Parkinson.

4. FORMACIÓN DE RECURSOS HUMANOS

En el proyecto participan estudiantes de grado y jóvenes graduados de la UNLaM, quienes han obtenido becas de iniciación a la investigación.

En base a esta investigación está previsto el desarrollo de una Tesis de Maestría en Informática a cargo de Lic. Renata Silvia Guatelli.

5. BIBLIOGRAFÍA

[1] Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. Journal of neurology, neurosurgery & psychiatry, 79(4), 368-376.

[2] Köllensperger, M., Geser, F., Seppi, K., Stampfer-Kountchev, M., Sawires, M., Scherfler, C., & Pellecchia, M. T. (2008). Red flags for multiple system atrophy. Movement disorders: official journal of the Movement Disorder Society, 23(8), 1093-1099.

[3] Landázuri, E., Villamil, L., & Delgado, L. (2007). Parámetros acústicos de la voz en personas con enfermedad de Parkinson. Umbral Científico, (11), 90-103.

[4] De Letter, M., Santens, P., & Van Borsel, J. (2005). The effects of levodopa on word intelligibility in Parkinson's disease. *Journal of communication disorders*, *38*(3), 187-196.

[5] Törnqvist, A. L., Schalén, L., & Rehncrona, S. (2005). Effects of different electrical parameter settings on the intelligibility of speech in patients with Parkinson's disease treated with subthalamic deep brain stimulation. *Movement disorders*, *20*(4), 416-423.

[6] Fernández González, S., Ruba San Miguel, D., Marques Girbau, M., Sarrasqueta Sáenz, L.(2006). "Voz del anciano" Revista de Medicina de la Universidad de Navarra, vol. 50, Nº 3, 44-48.

[7] Morrison, M., & Rammage, L. (1996). Tratamiento de los trastornos de la voz. Elsevier España.

[8] Cummings, J. L., & Masterman, D. L. (1999). Depression in patients with Parkinson's disease. International journal of geriatric psychiatry, 14(9), 711-718.

[9] Gallena, S., Smith, P. J., Zeffiro, T., & Ludlow, C. L. (2001). Effects of levodopa on laryngeal muscle activity for voice onset and offset in Parkinson disease. Journal of Speech, Language, and Hearing Research.

[10] Jiménez-Jiménez, F. J., Gamboa, J., Nieto, A., Guerrero, J., Orti-Pareja, M., Molina, J. A., & Cobeta, I. (1997). Acoustic voice analysis in untreated patients with Parkinson's disease. Parkinsonism & Related Disorders, 3(2), 111-116.

[11] Metter, E. J., & Hanson, W. R. (1986). Clinical and acoustical variability in hypokinetic dysarthria. Journal of communication disorders, 19(5), 347-366.

[12] Flint, A. J., Black, S. E., Campbell-Taylor, I., Gailey, G. F., & Levinton, C. (1992). Acoustic analysis in the differentiation of Parkinson's

disease and major depression. Journal of Psycholinguistic Research, 21(5), 383-399.

[13] Hertrich, I., & Ackermann, H. (1995). Gender-specific vocal dysfunctions in Parkinson's disease: electroglottographic and acoustic analyses. Annals of Otology, Rhinology & Laryngology, 104(3), 197-202.

[14] J. Holmes, R., M. Oates, J., J. Phyland, D., & J. Hughes, A. (2000). Voice characteristics in the progression of Parkinson's disease. International Journal of Language & Communication Disorders, 35(3), 407-418.

[15] Aguilera Pacheco, O. R., Escobedo Beceiro, D. I., Sanabria Macias, F., & Nuñez Lahera, I. (2015). Alteración de parámetros acústicos de la voz y el habla en la enfermedad de Parkinson. In XIV Simposio Internacional de Comunicación Social. Comunicación Social: Retos y Perspectivas (Vol. 2).

[16] Asgari, M., & Shafran, I. (2010). Extracting cues from speech for predicting severity of parkinson's disease. In *2010 IEEE International Workshop on Machine Learning for Signal Processing* (pp. 462-467). IEEE.

[17] Sakar, C.O and Kursun, O. (2009) 'Telediagnosis of Parkinson's disease using measurements of dysphonia', Journal of Medical Systems, DOI: 10.1007/s10916-009-9272-y.

[18] Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE transactions on biomedical engineering, 59(5), 1264-1271.

[19] Alhussein, M., & Muhammad, G. (2018). Voice pathology detection using deep learning on mobile healthcare framework. *IEEE Access*, *6*, 41034-41041.

[20] Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2010). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. Journal of the royal society interface, 8(59), 842-855

[21] Díaz-Pérez, F., García-Nieto, E., Ros, A., & Claramunt, R. (2014). Best estimation of spectrum profiles for diagnosing femoral prostheses loosening. Medical engineering & physics, 36(2), 233-238.

[22] Montaña, D., Campos-Roca, Y., & Pérez, C. J. (2018). A diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease. Computer methods and programs in biomedicine, 154, 89-97.

[23] Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, *7*(3–4), 197-387.

[24] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018, October). A survey on deep transfer learning. In *International conference on artificial neural networks* (pp. 270-279). Springer, Cham.

Plataforma de Datos Abiertos Enlazados para la Gestión y Visualización de Datos Primarios de Ciencias del Mar

Carlos Buckle[1,2], Marcos Zarate[1,2,4], Renato Mazzanti[1,2,3], Claudio Delrieux[2,5], Mirtha Lewis[4]

[1] Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB) Sede Puerto Madryn, Chubut, Argentina,+54 280-4883585 Int. 117.

[2] Laboratorio de Investigación en Informática de la UNPSJB (LINVI-UNPSJB)

[3] Unidad de Gestión de la Información, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (UGI-CENPAT-CONICET)

[4] Centro para el Estudio de Sistemas Marinos, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (CESIMAR-CENPAT-CONICET)

[5] Depto. de Ingeniería Eléctrica y de Computadoras, Universidad Nacional del Sur (DIEC-UNS)

cbuckle@unpata.edu.ar, zarate@cenpat-conicet.gob.ar, renato@cenpat-conicet.gob.ar

Resumen

La gestión, preservación y disposición con acceso público de datos primarios producidos en investigaciones están cobrando cada vez mayor importancia. Nuestro país, define normativas de ciencia abierta enfocadas al uso democrático y eficiente del conocimiento científico, para potenciar el desarrollo y como herramienta para un mejor aprovechamiento de los recursos utilizados en la exploración y recopilación de datos científicos. Una necesidad común es integrar información de diferentes repositorios y unificarla bajo un mismo vocabulario con conceptos relacionados, para ser explorada por grupos de investigación, citada en producciones científicas, visualizada para su interpretación y sistematizada bajo estándares que permitan la interoperabilidad con otras aplicaciones o repositorios.

Este proyecto propone investigar posibles soluciones a estas necesidades aplicándolas al dominio de las ciencias del mar. Para ello, se dispone de información recolectada en campañas multidisciplinarias sobre el Atlántico Sur Occidental. Como aplicación se propone el desarrollo de una plataforma para la especificación, modelado, generación y publicación de los conjuntos de datos primarios como datos abiertos enlazados, con visualizaciones que faciliten su interpretación y comparación, como así también su explotación basada en semántica.

Palabras clave: Datos Primarios Científicos, Datos Enlazados, Visualización de Datos, Repositorios Digitales.

Contexto

El proyecto propone avanzar sobre resultados y líneas de investigación del proyecto precedente *UNPSJB-PI 1424-Infraestructura de Acceso a Datos Primarios con aporte de semántica en Repositorios Digitales*, en el cual se identificaron ventajas en los Datos Abiertos Enlazados (Linked Open Data) [1] como camino estándar hacia la integración de datos abiertos [2] heterogéneos de diferentes dominios, bajo vocabularios comunes y con posibilidades de razonamiento por partes de agentes de software. Este trabajo se focalizará específicamente en la

explotación de datos relacionados con las ciencias del mar y se desarrollarán visualizaciones adaptadas a las necesidades de los investigadores.

Esta propuesta tiene una concepción interdisciplinaria y por ello el grupo de trabajo incluye: investigadores de las ciencias de la computación de diferentes áreas (web semántica, bases de datos, visualización científica e inteligencia artificial) e investigadores de las ciencias biológicas, expertos en el dominio de aplicación.

El proyecto se inscribe dentro del Laboratorio de Investigaciones en Informática (LINVI-UNPSJB) y se integra con otros grupos de investigación de la Universidad Nacional del Sur (UNS), del Centro Nacional Patagónico (CENPAT-CONICET) y del CIT Golfo San Jorge (UNPSJB-UNPA-CONICET). Está avalado por el Consejo Directivo de la Facultad de Ingeniería de la UNPSJB y será financiado por la Secretaria de Ciencia y Técnica de la UNPSJB para llevar a cabo durante 2020-2022.

1. Introducción

La complejidad y el alcance de la recolección de datos de campañas oceanográficas exigen una aproximación interdisciplinaria y una proyección amplia en el uso de la información obtenida. Estas campañas utilizan el Sistemas Nacionales de Datos del Mar (SNDM) [3] como repositorio final de datos, y requieren de una gestión integrada, tanto para su tratamiento como para su explotación, de manera que sea posible generar productos de síntesis que faciliten la tarea de análisis y descubrimiento de conocimiento.

SNDM utiliza como Plataforma el software provisto por *International Oceanographic Data and Information Exchange* (IODE). Su implementación en Argentina no fue del todo exitosa y tiene algunas limitaciones:

- Brinda la información de los diferentes recursos en forma fragmentada y no hay recursos suficientemente detallados para el ingreso de datos biológicos.
- No disponen de un desarrollo para el manejo restringido de los datos, para que durante el periodo de embargo post campaña, los datos puedan ser compartidos entre los grupos de trabajo o proyectos en cooperación.
- No existe la posibilidad de generar un informe con los datos más relevantes de una campaña de investigación en el mar.
- No permite interpretar la información de manera visual, por ejemplo un mapa interactivo donde el usuario pueda realizar búsquedas personalizadas.

La ciencia de datos es considerada una disciplina fundamental para abordar la complejidad y el alcance de las temáticas que exigen una aproximación interdisciplinaria y una proyección amplia en el uso de la información. En el Atlántico Sur las mediciones de este tipo son escasas y no se dispone de un sistema de captura de información adecuado. Por lo tanto, es necesario desarrollar sistemas capaces de gestionar su integración y su comunicación, tanto para un aprovechamiento integral y secundario de los grupos e instituciones participantes como para usuarios externos que requieran de información.

La relevancia de la propuesta se identifica en diferentes niveles: a nivel tecnológico, la creación de una plataforma de Datos primarios Abiertos Enlazados permitirá resolver los problemas de integración de datos heterogéneos bajo un mismo vocabulario [4] y disponerlos para ser explotados mediante búsquedas semánticas [5] y consultas basadas en lenguaje natural. Complementariamente, el desarrollo de Visualizaciones de Datos del Mar (Data Visualización) [6] facilitará su interpretación y comparación.

A nivel de política científica para la región y el país, se logrará un prototipo de servicio estandarizado para repositorios de datos biológicos del Atlántico Sur Occidental, con información unificada provenientes de

diferentes plataformas de muestreo (buques oceanográficos, ROVs, submarinos tripulados, etc.), que podrá ser reusada, citada, explotada para la generación de nuevos conocimientos.

A nivel socio-económico, las posibilidades de reuso de datos permitirán reducir los altos costos de las campañas de investigación en el mar y aportará al uso democrático y eficiente de datos científicos.

A nivel de formación de recursos humanos del proyecto participan estudiantes de posgrado de la disciplina Ciencias de la Computación, investigadores en formación y estudiantes avanzados de la Licenciatura en Informática de la UNPSJB.

2. Motivación

Dada la importancia del acceso abierto a datos científicos recogidos en el Atlántico Sur Occidental, tanto para la investigación y el desarrollo (I+D) en el país, como para garantizar visibilidad en el contexto regional e internacional; y dado que los datos relacionados mediante Datos Abiertos Enlazados pueden proporcionar la forma de vincular campañas del mar con publicaciones científicas, conjuntos de datos biológicos y químicos a través de vocabularios controlados y ontologías, esta investigación considera relevante plantear como hipótesis que la plataforma de Datos Primarios Enlazados posibilita la realización de un modelo conceptual para la construcción de una plataforma web (prototipo) capaz de publicar y vincular las campaña sobre el mar argentino para que puedan ser compartidas y reutilizadas a nivel nacional e internacional. Esto conlleva a la necesidad de estudiar infraestructuras para la publicación de datos científicos [7] y a considerar que la aplicación de visualizaciones de datos colaborará con la tarea analítica, mediante herramientas gráficas que faciliten la tarea de interpretación y contrastación de datos provenientes de los repositorios.

3. Líneas de Investigación y Desarrollo

Este proyecto desarrolla como principal línea de investigación, el Modelado Conceptual en la Web Semántica [8] y la construcción de grafos de conocimiento oceanográfico [9, 10] mediante datos enlazados [11, 12] para la integración de datos científicos y su explotación [13, 14]. Pero además, será necesario abordar el tratamiento de grandes volúmenes de datos oceanográficos y meteorológicos [15], caracterizados por las 5V (Volumen, Velocidad, Variedad, Verosimilitud y Valor) de Big-data [16] y teorías, técnicas y herramientas para la visualización de datos científicos oceanográficos [17, 18].

4. Resultados esperados

El propósito de la investigación es desarrollar una plataforma de Datos Abiertos Enlazados para la gestión, explotación, vinculación y visualización de datos del mar recolectados en campañas sobre el Atlántico Sur Occidental.

Se busca proporcionar una gestión confiable de datos para garantizar la preservación y el acceso a nuestros activos nacionales de investigaciones sobre el mar Argentino. Con objetivos de democratización del conocimiento, incentivo a los grupos de investigación para sumar contribuciones y optimización de costos basada en reuso de datos. Se esperan como resultados:

1. Estudio del estado del arte respecto de las bases de datos del mar en general y en particular las que registran datos del mar Argentino, con posterior registro de metadatos a nivel de crucero oceanográfico (trayectorias, instrumentos utilizados, origen, destino, estaciones, etc.).

2. Especificación, Modelado, Generación y Publicación en una Plataforma de Datos Abiertos Enlazados, escalable, estándar y con incorporación de semántica del dominio de Ciencias del Mar.

3. Desarrollo de visualizaciones de datos (scientific data visualization) para la interpretación, contrastación y comparación, incorporando consultas en lenguaje natural.

4. Formación y Transferencia: Consolidación del grupo de investigación que trabaja en repositorios de datos primarios con semántica. Formación de estudiantes, doctorandos y jóvenes investigadores. Difusión de resultados y publicaciones. Transferencia a iniciativas similares.

5. Formación de recursos humanos

En este proyecto participa un docente de UNPSJB-Puerto Madryn, doctor en Ciencias de la Computación y actual becario posdoctoral CONICET focalizado en el desarrollo de Datos Enlazados para la gestión integrada de datos científicos multidisciplinares de ciencias oceanográficas, de biodiversidad y ambientales. También participa un becario doctoral CONICET perteneciente al CIT Golfo San Jorge que estudia y propone modelos de visualización de datos científicos. Sus directores son investigadores de la Universidad Nacional del Sur y del CESIMAR-CENPAT-CONICET, quienes también integran el equipo de trabajo en calidad de asesores y expertos de dominio.

Además, el equipo de trabajo incluye a seis docentes del Departamento de Informática de la Facultad de Ingeniería de la UNPSJB-Puerto Madryn. Uno de ellos iniciando una especialización en Gestión de Información Científica y Tecnológica y un joven investigador que estudiará explotación de grafos de conocimiento mediante visualizaciones.

También forman parte del proyecto 2 (dos) alumnos del ciclo superior de la carrera Licenciatura en Informática que podrán desarrollar su trabajo de tesina y realizar Instancias Supervisadas de Formación en la Práctica Profesional en el marco de este proyecto.

Referencias

[1] Tom Heath and Christian Bizer. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1):1–136, 2011.

[2] Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6, 06/2011 2011.

[3] Sistema Nacional de Datos del Mar, Ministerio de Ciencia Tecnología e Innovación Productiva Presidencia de la Nación Argentina. *http://www.datosdelmar.mincyt.gob.ar*. Accedido: 15/2/2020.

[4] CL Chandler, RC Groman, A Shepherd, MD Allison, D Kinkade, S Rauch, PH Wiebe, and DM Glover. Using controlled vocabularies and semantics to improve ocean data discovery. In *AGU Fall Meeting Abstracts*, 2013.

[5] Judith A Blake and Carol J Bult. Beyond the data deluge: data integration and bioontologies. *Journal of biomedical informatics*, 39(3):314–320, 2006.

[6] Reiner Schlitzer. Interactive analysis and visualization of geoscience data with ocean data view. *Computers & geosciences*, 28(10):1211–1218, 2002.

[7] Giuseppe Andronico, Valeria Ardizzone, Roberto Barbera, Bruce Becker, Riccardo Bruno, Antonio Calanducci, Diego Carvalho, Leandro Ciuffo, Marco Fargetta, Emidio Giorgio, Giuseppe Rocca, Alberto Masoni, Marco Paganoni, Federico Ruggieri, and Diego Scardaci. e-infrastructures for e-science: A global view. *Journal of Grid Computing*, 9:155– 184, 06 2011.

[8] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 284(5):34–43, May 2001.

[9] Marcos Zárate, Pablo Rosales, Pablo Fillottrani, Claudio Delrieux, and Mirtha Lewis. Oceanographic data management: Towards the publishing of pampa azul oceanographic campaigns as linked data. *In Proceedings of the 12th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW 2018)*, 2018.

[10] Marcos Zárate, Pablo Rosales, Germán Braun, Mirtha Lewis, Pablo Rubén Fillottrani, and Claudio Delrieux. Oceangraph: Some initial steps toward a oceanographic knowledge graph. *In Iberoamerican Knowledge Graphs and Semantic Web Conference*, pages 33–40. Springer, 2019.

[11] Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, II Vardeman, et al. Five stars of linked data vocabulary use. *Semantic Web*, 5(3):173–176, 2014.

[12] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*, pages 205–227. IGI Global, 2011.

[13] Tony Hey, Stewart Tansley, and Kristin Tolle. *The Fourth Paradigm: DataIntensive Scientific Discovery*. Microsoft Research, October 2009.

[14] Adam Leadbetter, Robert Arko, Cynthia Chandler, Adam Shepherd, and Roy Lowry. Linked Data An Oceanographic Perspective. *The Journal of ocean Technology*, 8(3), 2013.

[15] Tanu Malik and Ian Foster. Addressing data access needs of the long-tail distribution of geoscientists. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 5348–5351. IEEE, 2012.

[16] Mark A Beyer and Douglas Laney. The importance of 'big data': a definition. *Stamford, CT: Gartner*, pages 2014–2018, 2012.

[17] Peter Fox and James Hendler. Changing the equation on scientific data visualization. *Science*, 331(6018):705–708, 2011.

[18] Nikos Bikakis and Timos Sellis. Exploration and visualization in the web of big linked data: A survey of the state of the art. *arXiv preprint arXiv*:1601.08059, 2016.

Búsquedas Selectivas sobre Flujos de Documentos

Santiago Ricci[1], Pablo Lavallén[1], Gabriel H. Tolosa[1,2]
{sricci, plavallen, tolosoft}@unlu.edu.ar
[1]Departamento de Ciencias Básicas, Universidad Nacional de Luján
[2]CIDETIC, Universidad Nacional de Luján

Resumen

La cantidad de información digital que se genera día a día impone restricciones a los usuarios en cuanto a la facilidad de acceso. Considerando la necesidad de acceder a información relevante, la alta tasa de aparición de nuevo contenido genera la necesidad de contar con herramientas de búsqueda que puedan manejar el tamaño, complejidad y dinamismo de las fuentes de información digital actuales. Este problema no puede ser resuelto en el ámbito de un solo equipo de cómputo por lo que requiere de una arquitectura que involucra procesamiento paralelo y distribuido, la cual incluye diseñar y optimizar estructuras de datos y algoritmos eficientes que las gestionen.

Esta arquitectura es desafiada cuando los documentos aparecen en flujos en tiempo real como, por ejemplo, las publicaciones en las redes sociales. Un caso paradigmático son las publicaciones en Twitter, en la cual millones de usuarios[1] alrededor del mundo publican "documentos cortos"(*tweets*) desde diferentes tipos de dispositivos (generalmente, móviles), los cuales deben estar disponibles casi de inmediato (segundos) por lo que las estructuras de datos deben soportar un alto dinamismo. Esto contrasta con la búsquedas web clásicas, donde el índice invertido se actualiza en modo batch ya que existe un tiempo entre actualizaciones debido a la necesidad de recolectar los nuevos documentos a indexar.

Un abordaje actual a este problema es la partición de la colección en porciones (*shards*) de acuerdo a algún criterio (por ejemplo, temático) de manera tal de enviar las consultas solamente a un número reducido n de nodos ($n \ll P$) que contengan particiones de la colección que potencialmente pueden satisfacer la consulta. Este problema se lo conoce como "búsquedas selectivas"(*selective search*) e incluye métodos que permiten seleccionar los recursos adecuados, algoritmos de fusión de resultados parciales y estrategias adaptadas de caching.

Este trabajo presenta las líneas de investigación en el contexto de las búsquedas en tiempo real utilizando una arquitectura basada en búsquedas selectivas. Las propuestas abarcan el estudio, diseño y evaluación de los criterios de actualización del índice invertidos por partición, las estrategias de cache a implementar y el algoritmo de búsqueda final y cómo estos impactan en la performance que se pretende optimizar (eficiencia y/o efectividad).

Palabras clave: algoritmos de búsqueda, estructuras de datos, búsquedas selectivas.

Contexto

Esta presentación se encuentra enmarcada en el proyecto de investigación "Búsquedas Selectivas sobre Flujos de Documentos", de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Luján. Convocatoria: Proyectos de Investigación Articulados con Centros de Investigación, Docencia y Extensión y/o Centros Regionales, Delegaciones y Sede CABA (RESREC:213-19).

Introducción

Las búsquedas de escala web son llevadas a cabo por motores de búsqueda que se encuentran preparados para gestionar eficientemente aspectos de escalabilidad, eficiencia y efectividad [2]. Esto se puede plantear como el problema de diseñar sistemas distribuidos de gran escala que permitan satisfacer las expectativas de los usuarios, utilizando eficientemente los recursos disponibles [5]. Hoy en día, este problema es considerado uno de los más atrayentes tanto en aspectos teóricos como prácticos en ciencias de la computación.

[1]De acuerdo a https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/, más de 300 millones de usuarios. Última consulta: 10/03/2020.

La arquitectura típica de un motor de búsqueda está compuesta por un nodo front-end (broker) que proporciona la interfaz de usuario y un conjunto (grande) de nodos *workers* (P) que almacenan los datos y soportan las búsquedas. Para alcanzar una alta eficiencia, se han desarrollado estrategias sofisticadas para recorrer las estructuras de datos (índice invertido) en los *workers* como los algoritmos de poda dinámica (*dynamic pruning*) Maxscore, WAND y Block-Max-Wand [8]. A partir de una consulta de usuario (*query*), éstos permiten "navegar"por las estructuras de datos sin evaluar todos los documentos de modo de acelerar el procesamiento sin disminuir la calidad de las respuestas. Complementariamente, se implementan varios niveles de cache: a nivel del *broker* se mantiene generalmente un cache de resultados[4], el cual almacena la lista final de los resultados correspondientes a las consultas más frecuentes o recientes. Además, un cache de listas de posteo (*posting list cache*) [14] se implementa en cada nodo de búsqueda. Este cache almacena en memoria las listas de posteo de términos populares o valiosos, para evitar el acceso al disco.

Cuando esta arquitectura es aplicada sobre flujos de documentos de escasa extensión generados en tiempo real, como el caso de Twitter, aparecen nuevos problemas, siendo necesario revisar esta arquitectura y diseñar y aplicar nuevas estrategias, políticas, estructuras de datos y algoritmos.

Diversos trabajos se enfocan en este tema, ya sea, basándose en indexación adaptativa o en control de la evolución del tamaño de las estructuras de datos. En el primero de los casos, Chen et al. [6] implementa un sistema de indexación orientado incorporar al índice solamente aquellos tweets que tengan mayor probabilidad de formar parte de los resultados de búsqueda (llamados "distinguidos"), retrasando la indexación de los restantes (llamados "ruidosos"). En el otro caso, Rissola et al. [12] propone estrategias de poda dinámica del índice invertido (invalidación) que controlan su tamaño y permiten aumentar la eficiencia de la búsqueda al recorrer estructuras más compactas.

Por otro lado, se han propuesto estrategias alternativas para acelerar el proceso de búsqueda en grandes colecciones de documentos. Una de ellas son las "búsquedas selectivas"(*selective search*), la cual implica particionar la colección en subcolecciones que luego son asignadas a diferentes nodos de acuerdo a algún criterio con el objetivo de consultar un número reducido de ellos para satisfacer una consulta. Como es de suponer, un aspecto clave aquí es el criterio empleado para particionar y distribuir la colección entre los nodos. Dicho criterio es llamado política de asignación. En relación a esto, Kulkarni et al. [11] estudiaron tres tipos de asignación (aleatoria, basada en origen y basada en el tema) e investigaron cómo realizar la búsqueda sólo en un subconjunto de *shards* para cada consulta sin sacrificar calidad. Sus resultados muestran que los costos de búsqueda se reducen al menos un orden de magnitud utilizando la estrategia de dividir por tema, esto es, mediante una técnica de clustering para la asignación de documentos a conjuntos disjuntos temáticos. En cuanto a la precisión, al menos el 86 % de las consultas obtuvo una performance más alta con esta estrategia versus la aleatoria y la basada en origen.

Por su parte, Hafizoglu et al. [9] plantean la utilización de un índice denominado *cluster-skipping* por cada *shard* que permite mejorar la latencia en la resolución de las consultas, llegando a mejoras del orden del 55 % respecto de la búsqueda selectiva basada en tópicos. Además, muestran que la arquitectura mejora el imbalance producido al utilizar índices basados en tópicos, permitiendo un balanceo uniforme a través de los diferentes *shards*.

Sin embargo, ninguno de los trabajos previos considera la posibilidad que los documentos arriben en flujos, como es el caso de las publicaciones en redes sociales.

Finalmente, entre los trabajos sobre caching y búsquedas selectivas solamente se cuenta con el aporte de Dai y Callan [7]. En el mismo, se estudia si los *shards* basados en temas reducen la efectividad del *caching* basado en *QtfDf* [3] y propone una serie de variantes que tienen en cuenta esta arquitectura. Además, explora si el uso de un *log* de consultas relacionadas a los *shards* puede mejorar los resultados del *caching* de *posting lists*. Los resultados muestran que la distribución sesgada de las *posting lists*s pueden causar una disminución en la tasa de aciertos. Sin embargo, este efecto puede ser eliminado utilizando los logs de consultas en la política de asignación de los shards, distribuyendo los más populares en diferentes núcleos de los procesadores, lo cual a su vez, mejora la performance.

Líneas de I+D

Las líneas de I+D del grupo se basan, principalmente, en mejorar la eficiencia en la recuperación de información de gran escala sobre flujos de documentos en tiempo real, con énfasis en los siguientes temas:

a. Diseño de políticas de asignación

Como se mencionó anteriormente, en este trabajo se parte de la hipótesis de que una arquitectura basada en búsquedas selectivas pueden mejorar la búsqueda sobre flujos de documentos cortos en tiempo real sobre todo en eficiencia. Así, dentro de este tipo de arquitecturas, los criterios empleados para la asignación de los documentos en los diferentes *shards* juega un rol fundamental, dado que la misma impacta fuertemente en el logro de una distribución de carga lo más equitativa posible entre los nodos que componen el *cluster*. Aquí no hay que perder de vista que dicha política será empleada sobre un flujo de documentos en tiempo real, lo cual implica una indexación incremental lo que supone desafíos desde el punto de vista del rendimiento a la hora de asignar documentos a las particiones. Los objetivos aquí perseguidos entonces son el diseño y evaluación, respecto al estado del arte, de políticas y técnicas de asignación de documentos a las diferentes particiones utilizando, entre otras, técnicas de clasificación de aprendizaje automático. En particular, se propone tomar como base los algoritmos/métodos, considerados estado del arte, utilizados para seleccionar recursos, concretamente Rank-S [10] y Taily [1]. A partir de éstos, se propondrán variantes que consideren que los documentos son cortos y que aparecen en flujos. Lo cual, además, lleva a que consideren que las estadísticas[2] no son tan robustas y que deben recalcularse de forma on line.

b. Métodos de selección de los nodos

Luego de haber distribuido los documentos entre las particiones, otra cuestión de suma importancia en el contexto de la búsqueda selectiva, es cuáles y cuántos *shards* consultar a la hora de resolver una *query*. Aquí se pueden distinguir dos grandes clases de algoritmos de selección que modelan a los *shards* de diferentes formas: los basados en vocabulario y los basados en muestreo [1]. Los primeros, representan las particiones mediante una serie de estadísticas de los términos del vocabulario manejado por el motor de búsqueda. En cambio, los basados en muestreo son aquellos que utilizan un índice central que contiene muestras representativas de cada *shard*. En este proyecto, específicamente, en base a escenarios diferentes, con requerimientos diversos planteados para el diseño de las políticas de asignación, se propone estudiar el *tradeoff* entre la cantidad de particiones y número seleccionado para resolver la consulta.

c. Técnicas de caching en los nodos y en el broker

Para el estudio de las políticas de caching, se propone partir del trabajo de Dai et al. [7], quienes trabajan con caches estáticos, luego extenderlo a dinámicos y estudiar el impacto de diferentes políticas de reemplazo, con la posibilidad de combinarlas entre *posting lists* y resultados. En general, estas pruebas de evaluación se realizan habitualmente con datos reales provenientes de archivos de log de motores de búsqueda comerciales. Existen datasets disponibles para investigación que son de dominio público y – además – se cuenta con vínculos académicos con grupos en la temática con quienes se comparten datos de prueba..

Finalmente, para poder estimar la performance de diferentes arquitecturas y configuraciones se propone complementar la experimentación real con modelos de simulación. Éstos posibilitan evaluar un sistema sobre escenarios para los cuales no se disponen recursos (por ejemplo, un cluster de 2000 o más nodos) y determinar costos, performance, escalabilidad y otras métricas.

Resultados y objetivos

Dentro del escenario en el cual se cuenta con flujos de documentos que ingesta el sistema y se deben incorporar a una o varias particiones para soportar la búsqueda, la hipótesis de trabajo que aparece es que es posible aumentar la performance del sistema de búsqueda a partir de la utilización de métodos ad-hoc para la asignación de documentos,

[2]Con el término "estadísticas"se pretende indicar el conjunto de indicadores que se utilizan para la descripción de una colección (o partición) y que son necesarios también para la selección de recursos (por ejemplo, distribución de a frecuencia de los términos en los documentos y en la colección, tamaño de la misma, tasa de actualización, etc.)

junto con estrategias de selección de particiones que consideren este mismo criterio.

De aquí, se define como objetivo general de esta propuesta el desarrollo de técnicas y algoritmos para la asignación de flujos de documentos a diferentes particiones de un índice para luego soportar búsquedas sobre un subconjunto de éstas, maximizando la performance (tanto eficiencia como efectividad).

En cuanto a los objetivos específicos, se consideran:

- Desarrollar un método online para rutear los documentos a los nodos (o particionar una colección) para la asignación a las diferentes particiones (*shards*) basado en parámetros de la arquitectura (por ejemplo, número de procesadores, núcleos, etc.).
- Proponer y evaluar nuevos métodos de selección de los nodos a consultar para responder una consulta a partir de considerar la dinámica de ingreso de los documentos y la estrategia de ruteo.
- Estudiar el *tradeoff* entre cantidad de *shards* y número de nodos a consultar (k) para obtener la respuesta a una consulta dada para las métricas de interés (eficiencia/efectividad).
- Incorporar y evaluar técnicas de *caching* en los nodos y en el *broker* para incrementar el *throughput* del sistema (*queries/segundo*).
- Desarrollar un método que permita especificar una métrica a maximizar para un *query* dado (por ejemplo, para soportar aplicaciones que requieren una aproximación cercana a una búsqueda exhaustiva).
- Desarrollar un modelo de estimación de la performance de la arquitectura en base a las variables más determinantes para un hardware dado (por ejemplo, un *cluster* particular).

Los resultados de este proyecto poseen aplicación práctica directa en la industria. Hoy en día la mayoría de los sistemas que gestionan información implementan estrategias de búsquedas, incluso sobre documentos de texto o flujos de tweets. En este último caso, principalmente debido a que Twitter se utiliza como una plataforma multipropósito, tanto de diseminación de información social como para negocios (marketing), publicidad, entre otras. Cualquiera de estas aplicaciones que requieran soportar búsquedas en tiempo real deben contar con estrategias eficientes que optimicen el uso de recursos ofreciendo parámetros de efectividad aceptables (que dependen del problema). Además, teniendo en cuenta que se propone realizar la experimentación sobre plataformas distribuidas (por ejemplo, Spark [13]), se genera conocimiento para implementar los modelos/servicios a diferentes escalas, permitiendo, por ejemplo, que una pequeña/mediana organización acceda a desplegar un cluster con hardware commodity.

Formación de Recursos Humanos

En cuanto a formación de recursos humanos se propone que los integrantes del proyecto se incorporen a tareas de investigación de forma regular. Además, se espera que al menos uno de los integrantes inicie estudios de postgrado. Complementariamente, durante los dos años del proyecto se espera incorporar un pasante (estudiante de la Lic. en Sistemas de Información) para que inicie tareas de investigación y ofrecerle la posibilidad de seguir esta línea para su trabajo final (tesina de grado).

Referencias

[1] R. Aly, D. Hiemstra, and T. Demeester. Taily: shard selection using the tail of score distributions. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 673–682, 2013.

[2] R. Baeza-Yates, C. Castillo, F. Junqueira, V. Plachouras, and F. Silvestri. Challenges on distributed web retrieval. In *2007 IEEE 23rd international conference on data engineering*, pages 6–20. IEEE, 2007.

[3] R. Baeza-Yates, A. Gionis, F. Junqueira, V. Murdock, V. Plachouras, and F. Silvestri. The impact of caching on search engines. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 183–190, 2007.

[4] R. Blanco, E. Bortnikov, F. Junqueira, R. Lempel, L. Telloli, and H. Zaragoza. Caching search

engine results over incremental indices. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 82–89, 2010.

[5] B. B. Cambazoglu and R. Baeza-Yates. Scalability challenges in web search engines. In *Advanced topics in information retrieval*, pages 27–50. Springer, 2011.

[6] C. Chen, F. Li, B. C. Ooi, and S. Wu. Ti: an efficient indexing mechanism for real-time search on tweets. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 649–660, 2011.

[7] Z. Dai and J. Callan. Inverted list caching for topical index shards. In *European Conference on Information Retrieval*, pages 577–583. Springer, 2018.

[8] S. Ding and T. Suel. Faster top-k document retrieval using block-max indexes. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 993–1002, 2011.

[9] F. Hafizoglu, E. C. Kucukoglu, and I. S. Altingovde. On the efficiency of selective search. In *European Conference on Information Retrieval*, pages 705–712. Springer, 2017.

[10] A. Kulkarni. *Efficient and effective large-scale search*. PhD thesis, Carnegie Mellon University, 2013.

[11] A. Kulkarni and J. Callan. Selective search: Efficient and effective search of large textual collections. *ACM Transactions on Information Systems (TOIS)*, 33(4):1–33, 2015.

[12] E. A. Ríssola and G. H. Tolosa. Improving real time search performance using inverted index entries invalidation strategies. *Journal of Computer Science & Technology*, 16, 2016.

[13] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, I. Stoica, et al. Spark: Cluster computing with working sets. *HotCloud*, 10(10-10):95, 2010.

[14] J. Zhang, X. Long, and T. Suel. Performance of compressed inverted list caching in search engines. In *Proceedings of the 17th international conference on World Wide Web*, pages 387–396, 2008.

Guía de recomendaciones para el tratamiento del Big Data como evidencia digital.

Lilia Palomo, Laura Sánchez Piccardi y Sergio Mario Guillet
Facultad de Ciencias Políticas, Sociales y Jurídicas - Universidad Católica de Santiago del Estero
lilia.palomo@ucse.edu.ar – lsanchezpiccardi@gmail.com – smguillet@yahoo.com.ar

RESUMEN

La evidencia digital es un tema que está avanzando en diversas latitudes dentro nuestro territorio nacional, respetando la fisonomía, características y tiempos de cada administración de justicia.

En la provincia de Santiago del Estero en particular, existen diversos avances en proceso de concreción,

En el desarrollo de esta investigación se trabajó en referenciar aspectos poco conocidos o ciertamente ignorados sobre la gestión de datos basada en macrodatos (Big Data) dado que esta no contiene por sí misma un tamiz ético y se encuentra escasamente regulada por normas procesales respecto a su aporte probatorio en la faz judicial.

De ahí nuestra pretensión de esbozar normas de buenas prácticas orientadas a un uso sustentable del Big Data y referenciar cómo deben funcionar los medios de prueba clásicos en un contexto informático y cual resulta la potencialidad acreditativa de los nuevos tipos de datos surgidos a partir de la utilización de las nuevas tecnologías.

En este trabajo se presentan los avances realizados durante el proyecto de investigación para desarrollar una guía de referencia con las características diferenciales del Big Data como fuente de datos en el contexto de la evidencia digital y su incidencia en el proceso de la investigación judicial, siguiendo de cerca las políticas y legislación vigentes.

Palabras clave: Big Data, evidencia digital, prueba digital, prueba electrónica, forensia, investigación judicial.

CONTEXTO

Este artículo forma parte del proyecto de investigación de cátedra: "Implicancias jurídicas del Big Data en la evidencia digital y su incidencia en el proceso judicial (período 2019)", de la facultad de Ciencias Políticas, Sociales y Jurídicas perteneciente a la Universidad Católica de Santiago, en el que se presentan los principales resultados de una línea de investigación que surge de ese proyecto, ante la necesidad de analizar las condiciones y estrategias del uso del Big Data desde la perspectiva de sus implicaciones jurídicas como evidencia digital, siguiendo de cerca las políticas y legislación vigentes.

1. INTRODUCCIÓN

Parece natural que los conflictos requieran cada vez en mayor medida de análisis y certificaciones digitales de los datos que se aportan a un procedimiento de resolución de disputas, sea judicial o arbitral. La base de las demandas y de la acreditación de prueba se basa, cada vez en mayor medida, en datos en formato electrónico; se está volviendo cada vez más relevante la capacidad de analizar y explicar las acciones que llevan a cabo individuos, empresas e instituciones por medio de sus sistemas de información automatizados.

La tradicional "prueba electrónica", se ha transformado con el advenimiento del fenómeno del "Big Data" por cuanto surge la necesidad de un análisis holístico de las múltiples interacciones entre todos aquellos sistemas de información que contienen datos relevantes para el esclarecimiento de los hechos.

La evidencia digital es un tema que está avanzando en diversas latitudes dentro nuestro territorio nacional, respetando la

fisonomía, características y tiempos de cada administración de justicia.

En la provincia de Santiago del Estero en particular, existen diversos avances que se pueden mencionar:

- "Aproximación teórica de las estrategias de delivery de datos unificados del ámbito organizacional. [1]
- "Protocolo de actuación para la extracción de evidencia digital y su vinculación con los códigos de procedimientos, de Santiago del Estero, en materia de prueba científica, en un gabinete de investigación forense".[2]
- "Articulación entre la Evidencia Digital y el Código de Procedimiento, en Materia de Prueba Científica". CIIDDI. 2018. [3]
- "Big Data: Modelado y visualización de grandes volúmenes de datos jurídicos". [4]
- "Creación del Gabinete de Informática Forense (GIF), del Poder Judicial de Santiago del Estero" [5].

Entonces, no solo es preciso, que se estudie el fenómeno informático del Big Data, en lo que hace a la gestión de la información y al cúmulo de datos que se producen; sino es necesario comprender cabalmente cómo juegan los principios probatorios en materia de recolección de prueba informática, determinar cómo funcionarán los medios de prueba clásicos en un contexto informático y analizar la potencialidad acreditativa de los nuevos tipos de datos surgidos a partir de la utilización de las nuevas tecnologías.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Se mencionan aquí las principales líneas de investigación y desarrollo abordadas en el marco del proyecto:

- Enfoques técnicos y estratégicos del Big Data disponibles en la actualidad.
- Panorama actual de los delitos informáticos basados en Big Data.
- Condiciones del Big Data como fuente del delito o como producto del análisis.
- Formas, que desde un Procedimiento, mejoren el tratamiento de la evidencia digital desde su hallazgo hasta su presentación.

3. RESULTADOS OBTENIDOS

En relación al cumplimiento de los objetivos propuestos en la investigación, se elaboró una Guía de referencia, de las características diferenciales del Big Data como fuente de datos en el contexto de la evidencia digital y su incidencia en el proceso de investigación judicial, siguiendo de cerca las políticas y legislación vigentes.

En este periodo, que nuestra civilización cuenta con vehículos autónomos dirigidos por Inteligencia Artificial, de Smart cities o plantas de producción autogestionadas, almacenes totalmente controlados por robots, ropa inteligente, electrodomésticos, controles de fuga de agua, semáforos inteligentes, etc., sumado a este concepto de IoT que se basa en la interconexión de dispositivos cotidianos a través de internet, está el Big Data producto de la gestión de datos basada en macrodatos.

Siguiendo con la distinción realizada por Igual Molina (2016) [6], dentro del ámbito operativo y de acción del Big Data, se distinguen 3 niveles: 1º nivel *consciente*, constituido por la información y datos puestos en la red por personas de existencia ideal o personas jurídicas, empresas, a través de sitios Webs, Blogs, o Redes sociales: Twitter, LinkedIn, Facebook, artículos o publicaciones científicas como este trabajo, o Bases de datos; 2º nivel *inconsciente* de la generación de datos, resultado del uso de dispositivos de comunicación, que proveen información sobre lugares, gustos, actividades, ente otras, provocando una pérdida de privacidad y donde ganan las grandes empresas que aprovechan el nuevo oro que extraen con la práctica del Data Mining; y estaría un 3º nivel *impersonal* de los datos, que corresponde a la producción e intercambio de datos entre máquinas, como la inteligencia artificial, entre otras.

Sobre la base de estas aseveraciones, ahora ya, desde el punto de vista jurídico se puede afirmar que los datos por los cuales se preocupa el derecho y los protege son los

Datos personales, estos son al decir de la Ley 25.326 de Protección de los Datos Personales: *"toda Información de cualquier tipo referida a personas físicas o de existencia ideal determinadas o determinables"*. Y más específicamente *gozan de* "*mayor protección los Datos sensibles*". Definidos como Datos personales, aquellos que revelan origen racial y étnico, opiniones políticas, convicciones religiosas, filosóficas o morales, afiliación sindical e información referente a la salud o a la vida sexual[1].

Entonces, en el desarrollo del presente trabajo, procuramos esbozar simples sugerencias a fin de armonizar y profundizar el componente de legislación y reglamentación del fenómeno del Big Data, en nuestro país. La ley de Datos Personales 25326 y el Convenio de Estrasburgo para la Protección de las Personas con respecto al Tratamiento Automatizado de Datos de carácter Personal[2], es un muy buen marco para a partir de allí, ampliar los alcances, en línea con el Reglamento General de Protección de Datos (RGPD)[3]de la Unión Europea y demás instrumentos legales internacionales citados en el punto IV de este trabajo. En una tarea responsable y siempre propendiendo a la unificación normativa a nivel nacional e internacional para garantizar la seguridad jurídica.

Asimismo formulamos una estimación respecto a lo que se considera como Buenas Prácticas al momento de incorporar Evidencia Digital o Electrónica en un proceso judicial cuya fuente fuera el análisis y la recopilación de información proveniente de la aplicación de técnicas para procesar grandes volúmenes de datos. En este aspecto sugerimos:

- Respecto a la Evidencia Digital, en todos los fueros su presentación, como aporte probatorio sea efectuada en formato digital y papel poniendo a disposición o indicando donde se encuentran los dispositivos de almacenamiento de probanzas, además conjuntamente se ofrezcan informes técnicos o pericias informáticas necesarias, con indicación respecto de su extracción, su conservación y su custodia, observando las pautas de guías forenses[4] normalmente aceptadas por la comunidad jurídica.
- La medida cautelar del congelamiento y conservación de datos, -quickfreeze- puede ser solicitada en toda clase de procesos a fin de asegurar la evidencia electrónica. Como así también las clonaciones de discos.
- Resulta posible peticionar informes sobre la identidad y domicilio de quien utilizó determinada dirección de IP, ya que esta información no está alcanzada por la garantía del art. 19 de la Constitución Nacional, ni encuadra en los parámetros de la ley de protección de datos personales (25.326) o en las previsiones de las leyes n° 19798 (telecomunicaciones), y n° 25520 (inteligencia nacional). Por el contrario la facultad de requerir datos de tráfico, y datos de contenido, es decir -la interceptación o captación de comunicaciones-, es una atribución específica de los magistrados, salvo los supuestos excepcionalmente previstos por la ley y por mínimo tiempo[5]. Los magistrados de todas las instancias y fueros se encuentran facultados para requerir los datos de tráfico o tránsito.
- Se debe requerir autorización específica del Juez para acceder a servicios de mensajería privada en general, recomendando a los operadores judiciales que en las órdenes emitidas, se prevea

[1] Vease http://servicios.infoleg.gob.ar/infolegInternet/anexos/60000-64999/64790/texact.htm

[2] Ver https://rm.coe.int/16806c1abd

[3] https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_es

[4] Por ejemplo: "Guidelines for Evidence Collection and Archiving" disponible en: https://tools.ietf.org/html/rfc3227 o "Guía de obtención, preservación y tratamiento de Evidencia Digital" disponible en: www.mpf.gob.ar/resoluciones/PGN/2016/PGN-0756-2016-001.pdf

[5] Ley 27319 Delitos Complejos Artículo 18.

puntualmente este supuesto para el trabajo de peritos forenses.

- Al redactar las guías, protocolos o estándares reglamentarios de temas vinculados con evidencia digital o procedimientos forenses sobre dispositivos electrónicos, recomendamos respetar el "principio de neutralidad tecnológica"[6]
- Para incorporar datos alojados en servicios de hosting fuera del país, antes de iniciar la gestión del exhorto, y atendiendo a los preceptos del Convenio de Budapest solicitar como buenos oficios al sitio de almacenamiento información del suscriptor (datos de abonado), y metadatos o datos de tráfico.

Como actividad de divulgación, en el mes de octubre 2019, se desarrolló el curso "El Big Data y sus Implicancias en el Contexto Jurídico y Empresarial", en el cual se presentó el tema del Big Data integrado con evidencia digital y las tecnologías disruptivas en los mercados. Siendo los expositores, investigadores y becarios cuyos temas y contendidos estaban relacionados, de manera de promover así la interacción vertical y horizontal de la temática entre diferentes facultades pertenecientes a la U.C.S.E. (Ciencias Políticas, Sociales y Jurídicas, Ciencias para la Innovación y el Desarrollo y Ciencias Económicas), y las carreras Abogacía, Ing. en Informática y Contador Público.

Además, estos resultados se incorporaran a los contenidos de la cátedra Informática de la carrera de Abogacía, correspondientes a la planificación 2020.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo fue conformado por dos docentes de las carreras Ing. en Informática y Abogacía. Además, integró el equipo de investigación un asesor legal externo, perteneciente a la Secretaría de Información Jurídica del Poder Judicial de Santiago del Estero.

También, se seleccionaron 2 estudiantes en calidad de becarios, uno por la carrera de Abogacía de la F.C.P.S.yJ., y otro de la carrera de Ingeniería en Informática perteneciente a la F.C.I.D. (Disp. 46/19).

5. BIBLIOGRAFIA

[1] Palomo, Lilia; Sánchez Piccardi, M. Laura y Lesca, Norma. (2017-2018). "Aproximación teórica de las estrategias de delivery de datos unificados del ámbito organizacional. Investigación de la Fac. Ciencias para la Innovación y el Desarrollo de la U.C.S.E.

[2] Palomo, Lilia; Guillet, Sergio y Mues, Silvina. (2018). "Protocolo de actuación para la extracción de evidencia digital y su vinculación con los códigos de procedimientos, de Santiago del Estero, en materia de prueba científica, en un gabinete de investigación forense" Investigación de la Fac. de Ciencias para la Innovación y el Desarrollo de la U.C.S.E.

[3] Palomo, Lilia; Guillet, Sergio y Mues, Silvina. (2018). "Articulación entre la Evidencia Digital y el Código de Procedimiento, en Materia de Prueba Científica". Paper presentado en el Congreso Iberoamericano de Investigadores y Docentes de Derecho e Informática CIIDDI. 2018.

[4] Lesca, Norma; Palomo, Lilia y Mulki, Juan. (2017-2018). "Big Data: Modelado y visualización de grandes volúmenes de datos jurídicos". Investigación de la Fac. Ciencias para la Innovación y el Desarrollo de la U.C.S.E.

[5] Excmo. Superior Tribunal de Justicia de la provincia de Santiago del Estero. (2017). "Anexo 1, Gabinete De Informática Forense". Recuperado de: www.jussantiago.gov.ar/jusnueva/Novedades/2017-Anexo1GIF-26-10-2017.pdf

[6] Molina, David Igual. (2016). "Fintech: Lo

[6] Sobre su incidencia véase: https://fernando-acero.livejournal.com/54049.html

que la tecnología hace por las finanzas". Barcelona, España, Editorial I, S.L. Recuperado de: https://books.google.com.ar/books?id=_NeADQAAQBAJ&pg=PT128&lpg=PT128&dq=nivel+consciente+del+big+data&source=bl&ots=DqsEagsrXS&sig=ACfU3U3TA_W9p7AkfwcB-T2R5J3Wp-4KmQ&hl=es&sa=X&ved=2ahUKEwiA3djVna7oAhVcCrkGHcL3DwYQ6AEwBXoECAoQAQ#v=onepage&q=nivel%20consciente%20del%20big%20data&f=false

[7] Cotino Hueso, Lorenzo. *(*2017). *"*Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales". Recuperado de: https://dialnet.unirioja.es/servlet/articulo?codigo=6066829

[8] Rivolta**,** Mercedes**. (**2007**)**. "Medios de prueba electrónicos: estado de avance en la legislación argentina". Recuperado de: www.saij.gob.ar/doctrina/dacc070049-rivolta-medios_prueba_electronicos_estado.htm

[9] Sánchez Piccardi, María Laura. (2017). Tesis de maestría: "Estado del arte de la industria 4.0". Universidad del Norte Santo Tomás de Aquino, Tucumán, Argentina.

Modelos, Algoritmos y Aplicaciones en Búsquedas a Gran Escala

Gabriel H. Tolosa[1,2], Agustín Marrone[1], Andrés Giordano[1], Agustín Gonzalez[1], Tomás Jurán[1,2] , Esteban A. Ríssola[1,3]
{tolosoft, eamarrone, agiordano, agonzalez, tjuran}@unlu.edu.ar
esteban.andres.rissola@usi.ch

[1]Departamento de Ciencias Básicas, Universidad Nacional de Luján
[2]CIDETIC, Universidad Nacional de Luján
[3]Faculty of Informatics, Università della Svizzera italiana

Resumen

La publicación de información digital crece día a día a tasas exponenciales. Esto exige mayores capacidades de hardware a los proveedores de servicios, e impone restricciones a los usuarios en cuanto a la facilidad de acceso. Además, teniendo en cuenta que los usuarios requieren información relevante lo más rápido posible, la alta tasa de aparición de contenido desafía a las herramientas de búsqueda, las cuales deben considerar y manejar eficientemente el tamaño, la complejidad y el dinamismo de las fuentes actuales de información digital.

En el caso del procesamiento de colecciones masivas de documentos, uno de los desafíos en cuanto a la eficiencia está dado por analizar la menor cantidad de documentos posible para satisfacer una consulta. Por otro lado, si los documentos ocurren en tiempo real (flujos) se requieren estrategias eficientes de ruteo hacia los nodos de búsquedas y de indexación incremental.

Estos problemas requieren, en general, procesamiento distribuido, paralelo y algoritmos altamente eficientes. En la mayoría de los casos, la partición del problema y la distribución de la carga de trabajo son aspectos de las estrategias que requieren ser optimizados de acuerdo al problema.

Este trabajo presenta las líneas de investigación en el contexto de los problemas de búsquedas aplicados a datos masivos en colecciones o flujos de documentos. Las propuestas abarcan el estudio, diseño y evaluación de estructuras de datos y algoritmos con énfasis en la eficiencia y el uso racional de los recursos de hardware.

Palabras clave: algoritmos eficientes, motores de búsqueda, estructuras de datos, grandes datos.

Contexto

Esta presentación se encuentra enmarcada en el proyecto de investigación "Estrategias y Algoritmos para Problemas de Búsquedas a Gran Escala" (Disposición CD-CB Nº 350/19) del Departamento de Ciencias Básicas (UNLu).

Introducción

La generación y publicación de información digital crece día a día a tasas exponenciales, lo cual impone restricciones a los usuarios en cuanto a la facilidad de acceso. Considerando la necesidad de acceder a información relevante, esta alta tasa de aparición de contenido (de diversas formas) genera la necesidad de contar con herramientas de búsqueda que puedan manejar el tamaño, complejidad y dinamismo de las fuentes de información digital actuales [25].

El área de Recuperación de Información (RI) es una disciplina dentro de la Ciencias de la Computación que trata con la representación, el almacenamiento y la recuperación de información relevante para los usuarios [2]. Desde las últimas décadas, RI ha experimentado un crecimiento importante en cuanto a sus alcances debido a la utilización de muchas de sus técnicas en los motores de búsqueda para la web (*"web search engines"* o WSE), los cuales han impuesto múltiples desafíos a la comunidad de investigación en el tema.

De forma sintética, un WSE es una aplicación distribuida, que se ejecuta en un *cluster* de computadoras y que debe responder a las consultas de los usuarios (*queries*) con estrictas restricciones de tiempo, en general, en unos pocos milise-

gundos. Pero, además, debe: (1) procesar el repositorio de documentos más grande que se conoce, la web, (2) mantener estructuras de datos específicas e (3) implementar sofisticados algoritmos para lograr satisfacer eficaz y eficientemente los requerimientos. Se puede considerar a los WSE como una de las primeras aplicaciones de procesamiento de datos masivos, variados y dinámicos, características de las hoy llamadas aplicaciones de Big Data [13].

Garantizar la efectividad y la eficiencia es primordial para el despliegue práctico de la web y otras aplicaciones que requieren búsquedas (procesamiento) a gran escala. En los últimos años, esta tarea se ha complejizado por el constante crecimiento del tamaño del problema (cantidad de documentos) y requerimientos de los sistemas (y usuarios). Se han incorporado a la disciplina nuevas estrategias y nuevos escenarios. Por ejemplo, en el primer caso, se utilizan técnicas basadas en aprendizaje automático para aprender a clasificar o rankear [22] y, en el segundo, aparecen requerimientos de búsqueda en redes sociales, las cuales poseen como estructura subyacente un grafo masivo [1]. Este ecosistema ofrece oportunidades para que la comunidad de RI investigue nuevas arquitecturas, se revisen los supuestos acerca de los modelos utilizados, se consideren los costos asociados a las soluciones y, en general, se mejore el *tradeoff* entre la eficacia y la eficiencia y puedan escalar a más datos y usuarios. Además, surgen nuevos problemas que requieren soluciones prácticas.

En el caso del procesamiento de colecciones masivas de documentos, uno de los desafíos en cuanto a la eficiencia está dado por analizar la menor cantidad de documentos posible para satisfacer una consulta. La arquitectura interna de un WSE está compuesta, básicamente, por un nodo *Broker* que recibe las consultas de los usuarios y las dirige a *Search Nodes* (*Back-End*) los cuales poseen las estructuras de datos (índices) necesarias para responder eficientemente. En el *Broker* se pueden implementar algoritmos de selección de recursos (para determinar a cuáles *Search Nodes* enviar la consulta), estrategias de ranking (para ofrecer una mejor respuesta a los usuarios) y caching (para disminuir el procesamiento en el *Back-End*), principalmente. La estructura de datos comúnmente utilizada para soportar la recuperación eficiente es el índice invertido. De forma simple, está compuesta por un vocabulario (V) con todos los términos extraídos de los documentos y, por cada uno de éstos, una lista de los documentos donde aparece dicho término junto con información de frecuencia (*posting list*).

El procesamiento de consultas es uno de los desafíos más difíciles de manejar en un sistema de búsquedas debido al constante crecimiento tanto en datos como en usuarios [15]. Además, el tema de la eficiencia es contínuamente identificado como uno de los más importantes en RI [8]. Por lo tanto, en este proyecto se abordan, principalmente, problemas de eficiencia relacionados con búsquedas a gran escala en dos ámbitos puntuales: colecciones y flujos de documentos.

En sentido amplio, la idea general de aumentar la eficiencia en las búsquedas permite procesar mayor cantidad de datos con menos recursos. Por ejemplo, en un datacenter con 25.000 equipos (nodos), una reducción de sólo un 2 % en el tiempo de ejecución significa que se requieren utilizar 500 nodos menos, lo que impacta positivamente en el mantenimiento de éstos, el consumo energético y la generación de calor. Esto, además, permite disponer de nodos ociosos para asignar a otras tareas o, simplemente, mantenerlos para mejorar la tolerancia a fallas del sistema completo.

Estos problemas requieren procesamiento distribuido, paralelo y algoritmos altamente eficientes [5]. Para su abordaje, se utilizan plataformas de almacenamiento y procesamiento no tradicionales [13], tales como Hadoop [27] y Spark [29], sobre las cuales se situan capas de software que implementan algoritmos de búsquedas, aprendizaje automático y optimización. En la mayoría de los casos, la partición del problema y la distribución de la carga de trabajo son aspectos de las estrategias que requieren ser optimizados de acuerdo al problema [26].

Líneas de I+D

Las líneas de I+D del grupo se basan en mejorar la eficiencia en la recuperación de información de gran escala y el procesamiento de grandes volúmenes de datos, con énfasis en:

a. Algoritmos para Poda Dinámica

Una de las áreas importantes cuyas soluciones impactan en la eficiencia son los algoritmos para *top-k*. Como se mencionó, se utiliza un índice invertido como estructura de datos el cual puede (o no) residir completamente en memoria principal y sobre el cual se utilizan dos enfoques principales para

el recorrido de las *posting lists* de los términos que forman la consulta: DAAT (Document-at-a-Time) y TAAT (Term-at-a-Time). A partir de un *query* formado por n términos ($q = \{t_1, t_2 \dots t_n\}$) DAAT recorre las n listas en paralelo intentando determinar en qué momento detener la evaluación sin llegar al final de todas (*early termination* o *dynamic pruning*) garantizando que el conjunto final de respuesta es el definitivo y equivalente a una búsqueda exhaustiva, pero con mucho menos costo computacional. En el caso de TAAT, las listas de los términos se procesan una a la vez, siguiendo la misma idea.

En la actualidad, las dos estrategias predominantes siguen el criterio DAAT y son Maxscore [23] y WAND [3]. En ambos casos, la idea subyacente es contar con un valor umbral (*upper bound*) que permite determinar en qué momento finalizar la evaluación. Sobre esta base se han realizado varias extensiones y propuestas, en algunos casos combinando la estrategia con una estructura de índice particular, como Block-Max [9]. En este sentdio, un avance relevante es el trabajo de Mallia et al. [14] en el cual se propone utilizar bloques de tamaño variable en la estructura de datos. El particionado de las listas en bloques se modela como un problema de optimización, introduciendo una estructura comprimida para representar la información de los bloques. Sus resultados muestran una mejora de la performance de aproximadamente $2x$ respecto del mejor algoritmo hasta el momento (Block-Max-WAND). En el mismo sentido, un muy reciente trabajo [15] muestra que una estrategia de saltos de bloques múltiples (*long skipping*) permite obtener mejoras para casos particulares como *queries* cortos de hasta un 37 %. Por el lado de Maxscore, Jian et al. [10] proponen realizar saltos (*skipping*) también sobre las listas de términos *esenciales* y muestran que el algoritmo procesa significativamente menos documentos.

En esta línea se investigan los algoritmos DAAT que forman parte del estado del arte y, adicionalmente, se propone uno nuevo como extensión de MaxScore. En éste, en vez de almacenar sólo un *upperbound* global, se almacena un conjunto de ellos en una estructura similar a una *skip list* [7], dotando al algoritmo de más información para mejorar la eficiencia del procesamiento. Resultados preliminares muestran que la evaluación de documentos en postings se puede reducir hasta un 50 % favoreciendo a consultas con términos muy populares. No obstante, también se ha advertido que se requiere una implementación altamente eficiente para compensar el overhead del procesamiento de los *upper-bounds*.

b. Procesamiento Mixto de Posting Lists

Las técnicas de procesamiento de consultas TAAT forman parte de una línea de investigación opacada por la eficiencia de DAAT en grandes bases de datos textuales como la Web, donde éstas últimas permiten evitar analizar un gran número de documentos. Además, TAAT requiere de una cantidad de *acumuladores* de *scores* (puntaje que determina su relevancia) parciales de cada documento hasta poder calcular el *score* final que indica su orden en el ranking.

Sin embargo, como bajo ciertas circunstancias no se presentan estos problemas descritos, en la búsqueda de estrategias más eficientes de procesar consultas, se propone combinar estrategias TAAT con Maxscore, una de las técnicas DAAT previamente mencionadas.

Se utilizan como base la técnica TAAT descrita por Persin [20] (Filtros de Persin) y la versión de Maxscore propuesta por Jonassen & Bratsberg [11], derivada del trabajo de Turtle & Flood [23].

El desarrollo principal propone crear una estructura de datos donde la *posting list* completa está divida en dos secciones: una primera sección con documentos ordenados por la frecuencia del término en el documento y una segunda sección ordenada por identificador de documento.

Luego, generar un algoritmo que recorra esta estructura, procesando la primer sección utilizando los Filtros de Persin y finalmente, si fuera necesario, realizar MaxScore con los documentos restantes, manteniendo los *scores* parciales calculados. De esta manera, se aprovecha la eficiencia de TAAT en colecciones pequeñas y la ventaja de DAAT en colecciones grandes.

Para esto se requiere mantener la sección de documentos ordenados por frecuencia lo suficientemente extensa para resolver la consulta sin utilizar MaxScore, pero lo suficientemente corta para no perder la eficiencia. Parte de esto involucra definir un umbral que divida la *posting list* entre sus dos secciones y cuyo *overhead* no presente un deterioro significativo en la eficiencia del procesamiento. Dado que la generación de la estructura mixta y sus parámetros auxiliares puede realizarse *offline*, el trabajo se enfoca en optimizar el tiempo de procesamiento de la consulta con los algoritmos descriptos.

c. Búsquedas sobre Flujos en Tiempo Real

Complementando los algoritmos de búsqueda mencionados en la línea anterior, otra estrategia para acelerar el proceso está basada en la partición de la colección de documentos en porciones, P, (*shards*) de acuerdo a algún criterio (por ejemplo, temático) de manera tal de enviar las consultas solamente a un número reducido n de nodos ($n << P$) que contengan particiones de la colección que potencialmente pueden satisfacer la consulta. Dichas estrategias son desafiadas cuando los documentos aparecen en flujos en tiempo real como, por ejemplo, las publicaciones en las redes sociales [21]. Un caso paradigmático son las publicaciones en Twitter, en la cual millones de usuarios alrededor del mundo publican "documentos cortos"(*tweets*) desde diferentes tipos de dispositivos (generalmente, móviles), los cuales deben estar disponibles casi de inmediato (segundos) por lo que las estructuras de datos deben soportar un alto dinamismo [4].

Los problemas principales a abordar aquí abarcan tanto la asignación de los documentos *nuevos* a las particiones como la indexación incremental. El primero de los casos puede ser abordado con técnicas de clustering combinadas con ruteo de documentos. Aquí, es necesario analizar (y redefinir) estrategias efectivas para documentos cortos.

En esta linea de investigación se propone combinar la problemática de las búsquedas en tiempo real utilizando una arquitectura basada en búsquedas selectivas donde los criterios de actualización del índice invertidos por partición, las estrategias de caché a implementar y el algoritmo de búsqueda final impactan en la performance que se pretende optimizar (eficiencia y/o efectividad). Los objetivos generales perseguidos son el desarrollo de técnicas y algoritmos para la asignación de flujos de documentos a diferentes particiones de un índice para luego soportar búsquedas sobre un subconjunto de éstas, maximizando la performance (tanto eficiencia como efectividad).

d. Compresión del índice

El tamaño de una colección de documentos influye directamente en el espacio que su representación lógica ocupa en disco condicionando la posibilidad de ubicarlo en memoria principal. Por ello, a gran escala es necesario usar técnicas de compresión sin pérdida. Sin embargo, el hecho de utilizar una estructura comprimida añade un tiempo de descompresión durante la resolución de una consulta, el cual puede ser compensado al transferir menor cantidad de datos [16]. Además, una compresión eficiente permite que más datos se almacenen en memoria principal [31, 16, 6, 12].

En sentido estricto, las listas invertidas que conforman el índice son secuencias de enteros que pueden comprimirse utilizando métodos específicos.

Entre los diversos esquemas de compresión, se pueden mencionar a los libres de parámetro, que comprimen (y descomprimen) cada entero de forma individual (*Unario* y *Gamma* [16] , o *Variable Byte* [6, 12, 30]); y a los adaptativos a lista, que comprimen una secuencia o bloque de enteros. Además, si se tiene en cuenta la relación costo-beneficio entre tamaño resultante y tiempo de descompresión, se destacan *PFor* (y sus variantes) [30, 12, 28] y *Elias-Fano* [24], en sus versiones particionadas [19]. A su vez, otra codificación interesante es *Interpolative* [17], que a pesar de ofrecer el mejor *compression ratio*, posee una velocidad de descompresión ineficiente para tratar el problema [12, 15].

En general, las listas que lo conforman se suelen dividir en particiones (o *chunks*) de 128 elementos cada una [12, 30, 19]. Asimismo, debido a que también pueden producirse ineficiencias de compresión cuando la cantidad de elementos a comprimir es menor al tamaño de bloque [6], suelen utilizarse métodos alternativos tales como *Variable Byte* (eficiente en tiempo) o *Interpolative* (eficiente en espacio).

Con todo este ecosistema de métodos disponibles, sus variantes y sus ventajas y desventajas en esta línea de investigación se propone modelar un esquema multicompresión que combina diversas codificaciones de acuerdo a las propiedades particulares de cada una de las particiones de las listas invertidas, contemplando también una solución de compromiso entre tasa de compresión y velocidad de resolución de consultas. Siguiendo algunos trabajos prelimimares [18] se trabaja en el uso de codificaciones adicionales y de otros tamaños de partición, abordando también el problema del *overhead* requerido.

Resultados y objetivos

La eficiencia sigue siendo un tema de significativa importancia en la comunidad de RI y se lo reco-

noce como uno de los que siguen ofreciendo desafíos importantes [8]. Principalmente, la necesidad de escalar en datos y hardware ofrece múltiples oportunidades para desarrollos científico/tecnológicos.

Por lo tanto, el objetivo principal del proyecto es estudiar, desarrollar, aplicar, validar y transferir modelos, algoritmos y técnicas que aborden el problema de las búsquedas sobre datos masivos, tanto en documentos como en grafos. Se pretende estudiar los problemas mencionados relacionados con técnicas de optimización para aplicaciones de búsqueda a partir de propiedades de los datos. En particular:

- Proponer, diseñar y evaluar variaciones de los algoritmos para poda dinámica para recuperación *top-k* basados en múltiples *upper-bounds*, minimizando la cantidad de postings procesadas.
- Desarrollar una estructura de datos y un algoritmo de recorrido que considere un ordenamiento mixto por identificadores de documentos o por frecuencias de acuerdo a diferentes criterios.
- Desarrollar un esquema de multicompresión de las posting lists que consideren diferentes *codecs* para cada bloque, de acuerdo a propiedades de los datos en éstos.
- Diseñar y evaluar estrategias de indexación distribuida y resolución de consultas para flujos de documentos que ocurren en tiempo real (por ejemplo, publicaciones en redes sociales).

Formación de Recursos Humanos

En el marco de estas líneas de investigación se están dirigiendo tres tesis de Licenciatura en Sistemas de Información (UNLu). Además, asociados al proyecto de investigación hay una estancia de investigación de la Secretaría de CyT (UNLu), una Beca Estímulo a las Vocaciones Científicas (CIN) y una pasantía interna UNLu.

Referencias

[1] N. K. Ahmed, N. Duffield, T. L. Willke, and R. A. Rossi. On sampling from massive graph streams. *Proc. VLDB Endow.*, 10(11):1430–1441, Aug. 2017.

[2] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval: The Concepts and Technology Behind Search*. Addison-Wesley Publishing Company, 2nd edition, 2011.

[3] A. Z. Broder, D. Carmel, M. Herscovici, A. Soffer, and J. Zien. Efficient query evaluation using a two-level retrieval process. In *Proc. of the Twelfth Int. Conf. on Information and Knowledge Management*, CIKM '03, pages 426–434. ACM, 2003.

[4] M. Busch, K. Gade, B. Larson, P. Lok, S. Luckenbill, and J. Lin. Earlybird: Real-time search at twitter. In *Proc. of the 28th Int. Conf. on Data Engineering*, ICDE '12. IEEE, 2012.

[5] B. B. Cambazoglu and R. A. Baeza-Yates. Scalability and efficiency challenges in large-scale web search engines. In *Proc. of the Eighth ACM Int. Conf. on Web Search and Data Mining, WSDM*, 2015.

[6] M. Catena, C. Macdonald, and I. Ounis. On inverted index compression for search engine efficiency. In *Proc. of the 36th European Conf. on IR Research on Advances in Information Retrieval - Volume 8416*, page 359–371. ACM, 2014.

[7] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. MIT press, 2009.

[8] J. S. Culpepper, F. Diaz, and M. D. Smucker. Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in lorne (swirl 2018). *SIGIR Forum*, 52(1):34–90, Aug. 2018.

[9] S. Ding and T. Suel. Faster top-k document retrieval using block-max indexes. In *Proc. of the 34th Int. Conf. on Research and Development in Information Retrieval*, pages 993–1002. ACM, 2011.

[10] K. Jiang and Y. Yang. Faster maxscore query processing with essential list skipping. In X. Luo, J. X. Yu, and Z. Li, editors, *Advanced Data Mining and Applications*, pages 549–559. Int. Publishing, 2014.

[11] S. Jonassen and S. E. Bratsberg. Efficient compressed inverted index skipping for disjunctive text-queries. In *Proc. of the 33rd European*

Conf. on Advances in Information Retrieval - Volume 6611, page 530–542. ACM, 2011.

[12] D. Lemire and L. Boytsov. Decoding billions of integers per second through vectorization. *Softw. Pract. Exper.*, 45(1):1–29, Jan. 2015.

[13] S. Madden. From databases to big data. *IEEE Internet Computing*, 16(3):4–6, 2012.

[14] A. Mallia, G. Ottaviano, E. Porciani, N. Tonellotto, and R. Venturini. Faster blockmax wand with variable-sized blocks. In *Proc. of the 40th Int. Conf. on Research and Development in Information Retrieval*, pages 625–634. ACM, 2017.

[15] A. Mallia and E. Porciani. Faster blockmax wand with longer skipping. In L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff, and D. Hiemstra, editors, *Advances in Information Retrieval*, pages 771–778. Int. Publishing, 2019.

[16] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[17] A. Moffat and L. Stuiver. Binary interpolative coding for effective index compression. *Inf. Retr.*, 3(1):25–47, July 2000.

[18] G. Ottaviano, N. Tonellotto, and R. Venturini. Optimal space-time tradeoffs for inverted indexes. In *Proc. of the Eighth ACM Int. Conf. on Web Search and Data Mining*, WSDM '15, pages 47–56. ACM, 2015.

[19] G. Ottaviano and R. Venturini. Partitioned elias-fano indexes. In *Proc. of the 37th Int. Conf. on Research & Development in Information Retrieval*, page 273–282. ACM, 2014.

[20] M. Persin. Document filtering for fast ranking. In *Proc. of the 17th Annual Int. Conf. on Research and Development in Information Retrieval*, page 339–348. ACM, 1994.

[21] E. Rissola and G. Tolosa. Improving real time search performance using inverted index entries invalidation strategies. *Journal of Computer Science & Technology*, 16(1), 2016.

[22] N. Tax, S. Bockting, and D. Hiemstra. A cross-benchmark comparison of 87 learning to rank methods. *Information Processing & Management*, 51, 11 2015.

[23] H. Turtle and J. Flood. Query evaluation: Strategies and optimizations. *Inf. Process. Manage.*, 31(6):831–850, Nov. 1995.

[24] S. Vigna. Quasi-succinct indices. In *Proc. of the Sixth ACM Int. Conf. on Web Search and Data Mining*, page 83–92. ACM, 2013.

[25] Y. Wang, L. Wu, L. Luo, Y. Zhang, and G. Dong. Short-term internet search using makes people rely on search engines when facing unknown issues. *PLOS ONE*, 12(4):1–9, 2017.

[26] Y. Wang, L. Wu, L. Luo, Y. Zhang, and G. Dong. Short-term internet search using makes people rely on search engines when facing unknown issues. *PloS one*, 12(4), 2017.

[27] T. White. *Hadoop: The Definitive Guide*. O'Reilly Media, Inc., 1st edition, 2009.

[28] H. Yan, S. Ding, and T. Suel. Inverted index compression and query processing with optimized document ordering. In *Proc. of the 18th Int. Conf. on World Wide Web*, WWW '09, page 401–410. ACM, 2009.

[29] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica. Spark : Cluster Computing with Working Sets. *Proc. of the 2nd USENIX Conf. on Hot topics in cloud computing*, page 10, 2010.

[30] J. Zhang, X. Long, and T. Suel. Performance of compressed inverted list caching in search engines. In *Proc. of the 17th Int. Conf. on World Wide Web*, WWW '08. ACM, 2008.

[31] M. Zukowski, S. Heman, N. Nes, and P. Boncz. Super-scalar ram-cpu cache compression. In *Proc. of the 22nd Int. Conf. on Data Engineering*, ICDE '06, page 59. IEEE, 2006.

Avances en el proyecto de Análisis y elaboración de datos para el desarrollo de un sistema de indicadores de ayuda social

Gustavo Illescas, María Rosa Dos Reis, Moisés Bueno, Gustavo Tripodi, Daniel Xodo

Instituto de Investigación en Tecnología Informática Avanzada, Grupo en Informática de Gestión - Centro asociado CIC. Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires.
Paraje Arroyo Seco, Tandil (7000), Argentina. +54 249 4385680
illescas@exa.unicen.edu.ar, bueno@econ.unicen.edu.ar, dosreis@econ.unicen.edu.ar, gtripodi@exa.unicen.edu.ar, dxodo@exa.unicen.edu.ar

RESUMEN

El objetivo del proyecto es desarrollar un sistema de indicadores para gestionar, el sistema de ayuda social KOINONIA[1], a partir de la aplicación de algoritmos y técnicas de Inteligencia Artificial y de Gestión del Conocimiento sobre los datos disponibles y detectables en la red de asistencia solidaria y en la población asistida, mediante tecnologías WEB.

A través del proyecto y sus líneas la transferencia que realiza la universidad se fundamenta en ser partícipe necesario en la generación de sistemas y herramientas informáticas con muy alto conocimiento aplicado a fin de lograr la mayor eficiencia y eficacia en todos los casos. Los servicios desarrollados y en desarrollo de KOINONIA son expuestos a distintos niveles y experiencias en las ONGs que los utilizan, con realidades que presentan características comunes (atienden necesidades sociales y la demanda de recursos suele ser superior a la oferta); pero también grandes diferencias en las formas y tratamiento de acciones o gestión en la toma de decisiones (atención de alimentación, problemas de adicciones, violencia de género, etc.).

Palabras clave: Indicadores, Ayuda Social, Multicriterio, ANP, AHP, PLN

CONTEXTO

La propuesta emerge del proyecto de incentivos actual (03/C282, Análisis y Elaboración de Datos para el Desarrollo de un Sistema de Indicadores de Ayuda Social. UNCPBA (2017-2020). El mismo se desarrolla dentro del Instituto de Tecnología Informática Avanzada (INTIA) de la Facultad de Ciencias Exactas (EXA), Universidad Nacional del Centro de la provincia de Buenos Aires (UNCPBA).
Colaboran con el proyecto miembros de Universidades extranjeras, entre otros:
• Laboratorio en Ingeniería de Software de la Universidad Carlos III (UC3M), Madrid, España.
• Centro de Investigación en Matemáticas (CIMAT), A.C. Unidad Zacatecas, México.
• Universidad Autónoma de Coahuila, México.

1. INTRODUCCIÓN

El presente trabajo muestra los avances en el proyecto presentado en el XIX Workshop de Investigadores en Ciencias de la Computación (2017) que a su vez es el

[1] PROYECTO KOINONIA. Página web oficial: www.proyectokoinonia.org.ar

proyecto actual de incentivos del cual participan los autores del mismo.
Además de lo expuesto en el resumen, también hacen a la problemática del proyecto el alcance geográfico de las soluciones, la potencialidad de uso global de las mismas, los costos y requerimientos técnicos acorde a las soluciones y el conocimiento necesario para su implementación y uso.

En todos estos aspectos la capacitación es un flujo constante de transferencia. No se puede diseñar e implementar una solución a ser aplicada por las ONGs si no se conoce y se capacita en cómo trabaja y opera en el medio social. Es decir, quienes creen las soluciones deben estar capacitados en los 'quehaceres' de las ONGs que utilizarán los servicios. Del mismo modo, la aplicación eficientemente realizada y con los conocimientos académicos puestos al servicio de la gestión de ONGs, por si solos no serán utilizados. Es sumamente necesario capacitar a las ONGs en su utilización y garantizar canales de diálogo a los efectos de mantener actualizaciones y/o incorporación de nuevos requerimientos.

Además de las capacitaciones en las herramientas informáticas, las herramientas desarrolladas son transferencia específica a la Organización Civil Proyecto Koinonia, que por su parte serán utilizadas por otras organizaciones sociales en forma global y gratuita. Los servicios actuales en diversas etapas del desarrollo son:
-Ayud@rg es una solución integral al trabajo en redes sociales de las instituciones que vinculan la oferta y demanda de recursos. Este sistema es el puente entre la necesidad y la ayuda que brindan las instituciones con vocación social.
-Encontr@rg es un espacio que promociona Organizaciones y sus servicios Solidarios centralizando su información y potenciando la comunicación y creación de relaciones.
-El Sistema de Asignación Global (SAG) permite que las instituciones puedan asignar en forma eficiente y eficaz los recursos con los que cuentan. Originalmente es un sistema para asignación de bolsones de alimentos, pero con el transcurso del tiempo deriva en un sistema de asignación general de recursos escasos en demanda mayor.
-El sistema Registro Global de Beneficiarios (RGB) permite centralizar la información referida a las personas asistidas por diferentes instituciones, en confidencia y permitiendo relevar la información específica que hace al deber ser de cada institución.
-El sistema de Formularios Dinámicos es una herramienta personalizable que se puede utilizar para relevar información a través de formularios de web. Mediante diferentes metodologías se puede procesar los datos ingresados en los formularios e incorporarlos a diversos proyectos que ejecutan las instituciones.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

El proyecto actual se encuentra articulado con dos proyectos de extensión:
• "TICs en la Gestión Solidaria". Convocatoria: Universidad, Cultura y Sociedad SPU 2016. Aprobado por Res SPU. Entidad Beneficiaria: Asociación Civil Proyecto Koinonía. Unidad Ejecutora: UNCPBA a través del INTIA (Facultad de Cs. Exactas). Estado: Finalizado el 31/10/2018. Monto financiado: $ 90.000,00. Directora: Dos Reis, María Rosa
• "Creación de Capital Social a través de las Redes de Conocimiento Solidarias". Convocatoria: Programas y Proyectos de Extensión 2017. Aprobado por RR Nº 1978/17. Unidad Ejecutora: Facultad de Ciencias Exactas. Monto financiado: $15.800. Directora: Dos Reis, María Rosa.
Adicionalmente se ha participado de las siguientes actividades:
Actividades de Extensión
• Seminario de "INNOVACIÓN en ADMINISTRACIÓN Y DECISIÓN en ORGANIZACIONES" con foco en la mejora de Decisiones en la Empresa y en el Estado. (SIADO 1-2019) Universidad Nacional de Lujan. Gustavo Tripodi (Asistente). 08/2019
• Seminario: "PROCESOS ANALÍTICOS PARTICIPATIVOS INTELIGENTES SISTÉMICOS (PAPIS-2019)". Universidad

Nacional de Lujan. Gustavo Tripodi (Asistente). 12/2019

Participación en eventos CyT

• Evento: "#MITAGDS Made in the Americas 2019 Global Digital Services Summit". Argentina, CABA. Organizada por: BANCO INTERAMERICANO DE DESARROLLO (BID). Gustavo Tripodi (Asistente). 07/2019

• Evento: "PRENDETE". Argentina, Tandil. Organizada por: FACULTAD DE CIENCIAS ECONOMICAS UNCPBA; Modo de participación: Gustavo Tripodi (Asistente, Mentor y Coordinador de uno de los equipos participantes). 08/2019

• Evento: "AGROBIT". Argentina, Tandil. Organizada por: FACULTAD DE CIENCIAS ECONOMICAS, UNCPBA; CEPIT. Gustavo Tripodi (Asistente). 11/2019

• Evento: Punta Tech Meetup (PTM) 2020. Gustavo Tripodi (Asistente). 01/2020

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados alcanzados a la fecha han sido publicados en numerosos artículos entre los que se encuentran (últimos tres años):

• Xodo, Daniel; Illescas, Gustavo; Bueno, Moisés; Dos Reis, María Rosa: "Análisis y elaboración de datos para el desarrollo de un sistema de indicadores de ayuda social". XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires), abril, 2017.

En este trabajo se presenta el proyecto sobre el que se exponen los avances.

• Dos Reis, María R.; Bueno, Moisés; Illescas, Gustavo; Weimann, Mariano: "Propuesta de Generalización y Diseño Web de Herramienta de Soporte para la Toma de Decisiones en Organizaciones de Ayuda Social -PROYECTO KOINONÍA". Anales del XXX ENDIO - XXVIII EPIO. Córdoba, mayo de 2017.

El trabajo plantea la migración de un prototipo para toma de decisiones en organizaciones de ayuda social que brindan servicios sociales alimenticios del tipo distribución de bolsones, a una Herramienta Web con diversas mejoras. Incluye un redimensionamiento para que pueda ser utilizada en diversos dominios organizacionales. Las técnicas que implementa la Solución Web incluyen Aprendizaje Automático e Inteligencia Artificial, Minería de Datos, Descubrimiento de Conocimiento en Base de Datos, Teoría de Utilidades, Programación Dinámica y Algoritmo de Greedy.

• Dos Reis, María Rosa; Bueno, Moisés: "TICs en la Gestión Solidaria". VI Jornadas de Extensión del Mercosur. Ponente en mesa Educación, Comunicación y Cultura. 25, 26 y 27 de abril de 2018. Tandil.

Este trabajo tiene por objeto presentar el Proyecto de Extensión "TICs en la Gestión Solidaria"; una continuación de "Proyecto Koinonía en la Universidad", avalado por SPU, donde se presentan los resultados obtenidos hasta el momento, las fortalezas y oportunidades del Proyecto, así como las debilidades y amenazas.

• Dos Reis, Maria R. – Barreiro Sandoval, Camila R. – Bueno Moisés E: "Organización de las prioridades de la Herramienta AYUD@RG por medio de AHP". Anales del XXXI ENDIO - XXIX EPIO, 2018.

En base a la identificación del conjunto de tareas que requieren atención, en este trabajo se propone la aplicación de un modelo de apoyo a la toma de decisiones, específicamente Analytic Hierarchy Process (AHP) y Analytic Network Process (ANP), para determinar un orden de prioridades de alternativas a implementar, de acuerdo a un conjunto de criterios de evaluación de software, que permita atender las problemáticas con mayor urgencia, desde el análisis del decisor (Saaty, Vargas, 2012; Favret, Rodríguez, Labat, 2015).

• Illescas Gustavo, Figini Iris, Martinez Mariano: "Data Sources Integrator for Management by Indicators". XXIV Congreso Argentino de Ciencias de la Computación (CACIC 2018). Facultad de Ciencias Exactas UNCPBA, 8 al 12 de octubre de 2018. Tandil

El trabajo expone la investigación y desarrollo de una herramienta para la integración de fuentes de datos en el contexto de la gestión por indicadores.

• Varona, Braian R. – Aciti, Claudio – Dos Reis, María R. – Bueno, Moisés E: "An Approach to Automatic Information Extraction in Social Aid Texts". Anales del Anales del CoNaIISI 2018. 6to Congreso Nacional de Ingeniería Informática y Sistemas de Información, Mar del Plata, noviembre de 2018.

El artículo presenta un método híbrido que permite extraer automáticamente información relevante en textos de ayuda social, lo cual permite que los procesos sean optimizados y se reduzcan los esfuerzos de búsqueda.

Se utilizan técnicas de análisis lingüístico y Procesamiento de lenguaje natural (PNL), que se centran principalmente en un área principal de estudio. La solución propuesta se implementa en la herramienta AYUD@RG.

• Dos Reis, Maria R. – Xodo, Daniel H. – Bueno Moisés E: "Propuesta de definición de métricas del capital social en Redes Sociales de Organizaciones No Gubernamentales (ONGs)". Anales del XXXII ENDIO - XXX EPIO, 2019, mayo de 2019.

Es de esperarse que, a partir de esta propuesta, con sustento en el análisis de la Teoría de Redes, se cree valor al medir la sinergia lograda en las acciones conjuntas, se profundice el conocimiento generado al estudiar las acciones sociales en redes de organizaciones, y se cuantifique para su mejor operación y gestión, optimizando así las acciones futuras para todos los actores intervinientes. (Fukuyama, 2001; Nakagawa, 2004).

• Dos Reis, Maria R. – Bueno Moisés E: "Optimización de la Gestión de Proyectos Sociales con Métricas de Aplicación de CPM. Caso PROYECTO KOINONIA". Anales del XXXII ENDIO - XXX EPIO, 2019. Tucumán, mayo de 2019.

Aplicar CPM permitió atender problemas derivados de la existencia de actividades y caminos críticos en el tratamiento y gestión de proyectos sociales, logrando una mayor eficiencia en la planificación y ejecución de los proyectos acompañados desde PROYECTO KOINONIA. Adicionalmente, analizando y aprehendiendo de los proyectos ejecutados, se podría compartir el conocimiento intra e inter-organizacional en la planificación o realización de "campañas" con objetivos similares.

• Santillán Cooper, M. - Serrano, F. - Armentano, M. - Schiaffino, S. - Dos Reis, M. R. – Bueno, M.: "Aplicación de técnicas de PLN en sistema de donaciones". ASAI 2019, Simposio Argentino en Inteligencia Artificial, 48º Jornadas Argentinas de Informática Salta septiembre 2019.

En este artículo se describe el enfoque basado en Procesamiento de Lenguaje Natural adoptado en el sitio AYUD@RG para poder detectar de manera satisfactoria las donaciones que las personas quieren realizar a partir de texto libre. Los principales desafíos los constituyen las características propias el lenguaje natural, como la ambigüedad en los términos, los sinónimos, los errores ortográficos (Varona, 2018; Gonzalez, 2013)

4. FORMACIÓN DE RECURSOS HUMANOS

La estructura del equipo de trabajo se muestra en la siguiente tabla:

Apellido y nombre	Título	Cargo	Funciones
Illescas, Gustavo	Dr.	Prof. UNCPBA	Director
Bueno Moisés	MBA.	Prof. UNCPBA	Integrante
Dos Reis Rosa	MBA.	Prof. UNCPBA	Integrante
Tripodi Gustavo	MBA.	Prof. UNCPBA	Integrante
Martinez, Mariano	Ing.	Aux. UNCPBA	Integrante
Montero, Norma	Ing.	Aux. UNCPBA	Integrante
Xodo, Daniel	MBA.	Prof. UNCPBA	Colaborador (Director hasta 09/2019)
Etchepare, Federico	Tesis de grado finalizada 2019	Alumno de grado	Colaborador
Servat, Agustín	Tesis de grado finalizada 2019	Graduado	Colaborador
Figini, Iris	Tesis de	Graduado	Colaborador

	grado finalizada 2018		
Weimann, Mariano	Tesis de grado finalizada 2017	Graduado	Colaborador

El proyecto contempla la participación docente de Facultades de Ciencias Exactas y Cs Económicas; y alumnos de ambas. Por otro lado, se encuentran en desarrollo y finalizadas las siguientes tesis.

• Tesis de Doctorado en Administración. La Mag. Dos Reis, es doctorando del Programa Doctoral en Administración de la Facultad de Ciencias Económicas-UNCPBA.

• Tesis de Doctorado en Administración: El Mg. Ing. Tripodi, se encuentra en fase de redacción de la propuesta de tesis doctoral para el Programa Doctoral en Administración de la Facultad de Ciencias Económicas-UNCPBA.

En relación al presente proyecto se han finalizado las siguientes tesis:

• Tesis de Maestría en Administración de Negocios: UTN-FRTL. Maestrando: Lic. Marcelo Matassa. Dir. Mag. Daniel Xodo. Finalizado 2017.

Tesis de grado en ingeniería de sistemas:

• Generalización y diseño web de herramienta de soporte para la toma de decisiones en organizaciones de ayuda social. Integrantes: Weimann, Mariano Ezequiel. Directora: Mag. María Rosa Dos Reis. Codirector: Mag. Moisés Bueno. Finalizado 2017.

• Integrador de fuentes de datos para la gestión por indicadores. Integrante: Iris Figini. Director: Dr. Illescas Gustavo. Codirector: Ing. Mariano Martinez. Finalizado 2018.

• Un Enfoque híbrido para la detección automática de recursos en textos cortos. Integrante: Varona, Braian Raúl. Directores: Mag. María Rosa Dos Reis, Mag. Claudio Aciti. Finalizado 2018.

• Herramienta para la gestión por indicadores. Integrantes: Federico Etchepare, Agustín Servat. Directores: Dr. Ing. Illescas Gustavo, Dr. Arturo Mora Soto. CIMAT, Zacatecas, México. Finalizado 2019.

5. BIBLIOGRAFÍA

Favret F., Rodríguez F, Labat M. (2015): "Regresión con SVM para reducir la inconsistencia de la matriz AHP". ASAI 2015, 16º Simposio Argentino de Inteligencia Artificial.

Fukuyama, F. (2001). "Social capital, civil society and development". Third World Quarterly, Vol 22, No 1, pp. 7– 20.

Gonzalez-Aguirre, A., Rigau, G.: Construcción de una base de conocimiento léxico multilingüe de amplia cobertura: Multilingual central repository. Linguamática 5(1), 13–28 (2013).

Nakagawa, Y. and Shaw, R. (2004). "Social Capital: A Missing Link to Disaster Recovery". International Journal of Mass Emergencies and Disasters. March 2004, Vol. 22, No. 1, pp. 5–34.

PROYECTO KOINONIA. Página web oficial: www.proyectokoinonia.org.ar.

Saaty T., Vargas L. (2012): *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*. Second Edition. International Series in Operations Research & Management Science - © Springer Science+Business Media New York.

Varona, B.: Un enfoque híbrido para la detección automática de recursos en textos cortos. Trabajo final de grado, Facultad de Ciencias Exactas, UNCPBA (2018).

Xodo D., Illescas G., Bueno M., Dos Reis M. (2017): "Análisis y elaboración de datos para el desarrollo de un sistema de indicadores de ayuda social". Anales del XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires).

Análisis de relaciones intra-institucionales e interdisciplinares de una universidad a partir de la producción registrada en *Microsoft Academic*: el caso de la Universidad Nacional de La Plata

José Federico Medrano[1], Sandra Miguel [2],

jfmedrano@fi.unju.edu.ar, smiguel@fahce.unlp.edu.ar

[1]VRAIn / Visualización y Recuperación Avanzada de Información / Facultad de Ingeniería Universidad Nacional de Jujuy (UNJu) - Ítalo Palanca 10, +54 (388) 4221587

[2] Instituto de Investigaciones en Humanidades y Ciencias Sociales (IdIHCS) (UNLP-CONICET) / Facultad de Humanidades y Ciencias de la Educación / Universidad Nacional de La Plata, Calle 51 e/ 124 y 125, (1925) Ensenada, Argentina

RESUMEN

El objetivo del trabajo es conocer las relaciones de colaboración entre investigadores y/o grupos de investigación de una misma institución, separando por áreas del conocimiento, empleando metodologías e indicadores de análisis de redes sociales. El estudio se realizará aplicado al caso de la producción científico-académica de la Universidad Nacional de La Plata indexada en el buscador académico *Microsoft Academic*. Se tomará una muestra del corpus total que corresponde a los registros con año de creación entre 2010-2019. Se analizará la colaboración entre científicos a partir de un estudio descriptivo de las redes de coautoría que existen entre ellos y se realizará un análisis textual de las relaciones entre los temas y ámbitos de investigación. Los resultados permitirán reflejar los vínculos entre investigadores de las mismas o distintas áreas y conocer los niveles de interdisciplinariedad y relaciones entre unidades de investigación en la producción científico-académica de la institución. Con el trabajo no solo se espera contribuir al conocimiento de las relaciones de colaboración intra-institucionales e interdisciplinares, sino también mostrar el uso de este tipo de base de datos bibliográfica como fuente de datos para este tipo de estudios métricos y de análisis de redes sociales.

Palabras clave: *Recuperación de Información; Microsoft Academic; Redes de Coautoría; Análisis de Redes Sociales*

CONTEXTO

La línea de investigación aquí presentada se encuentra enmarcada dentro del Proyecto tetra anual (2017-2020) con Código 11/H827, que

lleva por nombre "*La investigación sobre temas locales. Análisis de la producción científica sobre Argentina desde la perspectiva bibliométrica y altmétrica*", aprobado y financiado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de La Plata (UNLP). Este proyecto es dirigido por la Dra. Sandra Miguel, directora del Departamento de Bibliotecología de la Facultad de Humanidades y Ciencias de la Educación de la UNLP.

1. INTRODUCCIÓN

La colaboración científica es un aspecto de gran relevancia pues se sitúa en la base del desarrollo científico de cualquier disciplina. La colaboración científica es entendida como el desarrollo de actividades conjuntas de investigación entre investigadores, instituciones y países, en las que se comparten recursos intelectuales, económicos, físicos, etc. para la generación y transferencia de conocimientos, productos o servicios (Miguel & Arias, 2012).

En la UNLP la colaboración constituye uno de los ejes estratégicos de sus políticas científicas y tecnológicas. En este sentido, y a través de diferentes programas, promueve la vinculación científica, académica y tecnológica y la participación en redes nacionales e internacionales en las diferentes áreas del conocimiento. En la UNLP más del 80% de la producción científica con visibilidad regional e internacional se publica en colaboración entre autores de la misma institución, o de otras instituciones del país y/o del extranjero. Esto significa una capacidad instalada de cooperación científica de la Universidad, que si bien puede variar entre un campo temático y otro, se va consolidando como un nuevo modo de producción de conocimientos en la mayoría de las disciplinas (Miguel, 2013).

Según (Moreno, 1997), la colaboración entre grupos tiene un efecto positivo sobre la producción científica, tanto cuantitativa como cualitativa, en especial cuando ésta se lleva a cabo por comunidades científicas y grupos del más alto nivel. Se ha demostrado que las investigaciones que se realizan en equipo son más productivas y adquieren mayor número de citaciones que cuando se trabaja en forma aislada o con escasa colaboración. Esta dinámica de generación de conocimiento se puede evaluar a través de redes de colaboración, que llegan a ser un reflejo de las relaciones dadas entre grupos, quienes pueden unir instituciones a través de sus vínculos laborales y académicos (Gutiérrez & Velasco, 2017).

Los estudios previos sobre la colaboración institucional han intentado medir la interacción de investigadores e instituciones a partir de las redes de coautoría reflejadas en la producción científico-académica recolectada en distintas bases de datos, al respecto las más empleadas han sido *Scopus*, la *Web of Science*, los Sistemas de Información de Ciencia y Tecnología nacionales, la producción registrada en el CV, en los portales de acceso abierto SciELO y RedALyC, entre otros. Sin embargo, dichos análisis no han sido tenidos en cuenta empleando bases de datos bibliográficas de libre acceso como el motor académico *Microsoft Academic (MA)*[1].

Microsoft Academic Graph (MAG) es un grafo de conocimiento de publicaciones académicas que brinda soporte a MA y está estructurado en torno a los siguientes tipos de entidades: publicación, autor, afiliación del autor

[1] https://academic.microsoft.com/

(institución), lugar de publicación (revistas y conferencias), campo de estudio (tema). Contiene fechas de publicación, así como pares de citas y, por supuesto, datos de coautoría. Debido a que MAG es un grafo de conocimiento, facilita poderosos análisis cienciométricos del resultado de la publicación, el impacto, la colaboración y más. Como lo indica (Hug & Brändle, 2017), MAG fue lanzado en Junio de 2015 y modela "*las actividades de comunicación académica de la vida real como un grafo heterogéneo*" (Sinha, y otros, 2015). Los datos para MAG se recogen principalmente de los *feeds* de metadatos de editores y páginas web indexadas por el buscador Bing[2] (el buscador de propósito general de Microsoft). Este enorme "grafo" puede ser accedido por dos vías, la primera de ellas a partir del motor de búsqueda académico *Microsoft Academic* y la otra por medio de la Academic Knowledge API (AK API)[3].

Existen estudios que se encargaron de remarcar la utilidad y capacidad de emplear a MA como herramienta de análisis bibliométrico (Hug & Brändle, 2017; Hug, Ochsner, & Brändle, 2017; Thelwall, 2018; Kousha & Thelwall, 2018). Es en este sentido que MA resulta atractivo a investigadores por la estructuración de los datos que ofrece, la AK API actuando de puente entre el usuario y MAG, facilita enormemente las tareas de extracción y recolección de publicaciones, tarea nada sencilla para su gran rival *Google Scholar*.

Por ello, este trabajo plantea un enfoque novedoso empleando a MA como origen de datos para recolectar toda la producción científico-académica de la UNLP indexada en esta enorme base de datos bibliográfica de libre acceso. Este conjunto de datos permitirá analizar las redes de coautoría que existen entre los investigadores, o grupos de investigación, del mismo modo se plantea el estudio de las relaciones entre los temas y ámbitos de investigación a partir de un análisis textual de dichas publicaciones. Esto será posible gracias a las entidades que se encuentran representadas en MAG y que permiten ser recolectadas a partir de distintas consultas por medio de la AK API.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La presente investigación se enfocará en dos aspectos claves:

- Análisis de relaciones intra-institucionales e interdisciplinares de una universidad.
- Empleo de bases de datos bibliográficas de libre acceso.

Quizás el primero de ellos no es del todo novedoso, puesto que ya ha sido estudiado desde distintas ópticas y con diferentes enfoques, y en diferentes ámbitos (Muñoz & Delgado, 2016; Miguel & González, 2010; Alcaide, y otros, 2006), sin embargo es el segundo aspecto el que le aporta novedad y relevancia. Este tipo de bases de datos están siendo objeto de estudio por investigadores del campo de la Bibliometría y Ciencias de la Información, por representar una clara alternativa a las bases de datos tradicionales en esta materia, *Scopus* y *Web of Science* (Harzing, 2019; Haunschild, Hug, Brändle, & Bornmann, 2018; Thelwall, 2017).

[2] https://www.bing.com/

[3] https://www.microsoft.com/en-us/research/project/academic-knowledge/

Por último, el análisis textual que se propone llevar a cabo incluye el empleo de técnicas de Procesamiento del Lenguaje Natural y Aprendizaje Automático para hallar patrones, tendencias y relaciones entre el contenido almacenado.

3. RESULTADOS OBTENIDOS/ESPERADOS

El objetivo de este trabajo es identificar los indicadores de colaboración científica y las redes de coautoría de los investigadores de la UNLP en la última década a partir de un estudio descriptivo de las redes de coautoría que existen entre ellos, empleando el motor académico *Microsoft Academic* como fuente de datos.

Realizando un análisis textual de las publicaciones (campos textuales como título, resumen y campo de estudio) se intentará establecer las relaciones entre los temas y ámbitos de investigación más comunes. Esto permitirá caracterizar desde una visión cuantitativa la producción por campo de estudio o área del conocimiento.

4. FORMACIÓN DE RECURSOS HUMANOS

El Equipo de Trabajo está conformado por docentes investigadores de la UNJu y UNLP. Los mismos llevan adelante esta línea de investigación desde hace años. Al ser un proyecto que involucra a profesionales de distintas instituciones, es un claro ejemplo de colaboración interinstitucional. Además, cada año se incorporan al proyecto alumnos avanzados de distintas carreras, quienes trabajan en temas relacionados. Del mismo modo, los integrantes del equipo participan en el dictado de asignaturas/cursos de grado y postgrado de la UNLP, UNJu y UCSEDASS.

5. BIBLIOGRAFÍA

Alcaide, G., Zurián, J., Benavent, R., Arroyo, A., Orive, J., & Serrano, S. (2006). Redes de coautoría y colaboración de las instituciones españolas en la producción científica sobre drogodependencias en biomedicina 1999-2004. *Trastornos Adictivos, 8*(2), 78-114. Retrieved from https://academic.microsoft.com/paper/2011648008

Gutiérrez, J., & Velasco, N. (2017). Redes de coautoría como herramienta de evaluación de la producción científica de los grupos de investigación. *Revista General de Información y Documentación, 27*(2), 279-297. Retrieved from https://academic.microsoft.com/paper/2771760668

Harzing, A.-W. (2019). Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics, 120*(1), 341-349. Retrieved from https://academic.microsoft.com/paper/2944642116

Haunschild, R., Hug, S., Brändle, M., & Bornmann, L. (2018). The number of linked references of publications in Microsoft Academic in comparison with the Web of Science. *Scientometrics, 114*(1), 367-370. Retrieved from https://academic.microsoft.com/paper/2763175323

Hug, S., & Brändle, M. (2017). The coverage of Microsoft Academic: analyzing the publication output of a university. *Scientometrics, 113*(3), 1551-1571. Retrieved from https://academic.microsoft.com/paper/2605124925

Hug, S., Ochsner, M., & Brändle, M. (2017). Citation analysis with microsoft academic. *Scientometrics, 111*(1), 371-378. Retrieved from https://academic.microsoft.com/paper/2522876520

Kousha, K., & Thelwall, M. (2018). Can Microsoft Academic help to assess the citation impact of academic books. *Journal of Informetrics, 12*(3), 972-984. Retrieved from https://academic.microsoft.com/paper/2886862840

Miguel, S. (2013). *Serie Indicadores Bibliométricos UNLP: Informe 2013.* La Plata: Universidad Nacional de La Plata.

Miguel, S., & Arias, R. (2012). La colaboración científica en la UNLP. Retrieved from https://academic.microsoft.com/paper/798563433

Miguel, S., & González, C. (2010). Composición y dinámica de los grupos de investigación del Departamento de Bibliotecología de la Universidad Nacional de La Plata en el periodo 2000-2009. *I Jornada de Intercambio y Reflexión acerca de la Investigación en Bibliotecología*. Retrieved from https://academic.microsoft.com/paper/409268175

Moreno, C. (1997). Técnicas bibliométricas aplicadas a las estudios de usuarios. *Revista General de Información y Documentación, 7*(2), 41-68. Retrieved from https://academic.microsoft.com/paper/1597984441

Muñoz, E., & Delgado, Y. (2016). Detección de comunidades a partir de redes de coautoría en grafos RDF. *ACIMED, 27*(1), 90-99. Retrieved from https://academic.microsoft.com/paper/2256194420

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J., & Wang, K. (2015). An Overview of Microsoft Academic Service (MAS) and Applications. *Proceedings of the 24th International Conference on World Wide Web*, (pp. 243-246). Retrieved from https://academic.microsoft.com/paper/1932742904

Thelwall, M. (2017). Microsoft Academic: A multidisciplinary comparison of citation counts with Scopus and Mendeley for 29 journals. *Journal of Informetrics, 11*(4), 1201-1212. Retrieved from https://academic.microsoft.com/paper/2766265189

Thelwall, M. (2018). Microsoft Academic automatic document searches: Accuracy for journal articles and suitability for citation analysis. *Journal of Informetrics, 12*(1), 1-9. Retrieved from https://academic.microsoft.com/paper/2768506257

Ciencia de datos aplicada al mejoramiento genético de la raza Aberdeen Angus.

Osvaldo Spositto, Gabriel Blanco, Lorena Matteo, Marcelo Levi, Julio Bossero.
Departamento de Ingeniería e Investigaciones Tecnológicas. Universidad Nacional de La Matanza. Florencio Varela 1902, San Justo, Prov. Buenos Aires, Argentina
{sposito, g2blanco, lmatteo,mlevi,jbossero}@unlam.edu.ar

RESUMEN

La Diferencia Esperada entre Progenies (DEP) es un indicador numérico que predice la calidad genética de las futuras crías de un toro o una vaca respecto de una base de comparación. Este es un valor genético que proporciona la mejor manera de comparar reproductores por la producción esperada en sus descendencias. En este trabajo estudiamos el uso de técnicas de Minería de Datos, del tipo Supervisadas y No Supervisadas, para identificar patrones o grupos de características en los valores genéticos de los animales reproductores que determinen el peso de los terneros al nacer de la cría de la raza Aberdeen Angus. El objetivo es brindar una herramienta complementaria para que un criador ganadero pueda seleccionar mejor los reproductores que al ser apareados con sus vientres, produzcan progenies superiores. En estos primeros estudios emplearon datos provenientes de 360 animales proporcionados por un establecimiento ganadero de la provincia de Buenos Aires. Se espera que a partir de los modelos aprendidos por los algoritmos se pueda extraer información preliminar sobre el valor genético de un animal, que pueda resultar de gran utilidad en el sector ganadero en la toma de decisiones en un programa de mejoramiento genético.

Palabras clave: Diferencia Esperada entre Progenies, Minería de Datos, Algoritmos Supervisados y No Supervisados, Mejoramiento Genético.

CONTEXTO

La línea de investigación aquí presentada está enmarcada dentro del Programa de Incentivos para Docentes Investigadores de la Secretaría de Políticas Universitarias (PROINCE) 2019-2020. El mismo lleva el título: *Uso de Minería de Datos para Mejoramiento Genético en la raza Aberdeen Angus.* Este proyecto es financiado por la Universidad Nacional de La Matanza (UNLaM). Los datos utilizados para este estudio provienen de los rodeos Aberdeen Angus de la estancia El Doce y de Cabaña Las Lilas[1] ambas ubicadas en la localidad de Chascomús, en la provincia de Buenos Aires.

Esta Cabaña perteneciente a la Asociación Argentina de Angus[2], es uno de los caudales genéticos más importantes de la Argentina y del Mercosur.

1. INTRODUCCIÓN

El propósito de un programa de mejoramiento genético de una raza de carne es conocer y promover los mejores animales basados en registros de comportamiento y evaluación de sus progenitores [1]. Los productores ganaderos se basan en ellos para identificar y procurar aquellos animales que mejor se adapten a las condiciones de producción existentes y que al mismo tiempo conduzcan a un incremento del beneficio económico de la actividad. Para esto es necesario valerse de información objetiva y precisa sobre los reproductores, que permita a los criadores, tomar decisiones de selección y hacer un uso diferencial de los mismos. La ganadería consta de distintos factores, que se deben considerar para que la misma sea exitosa y rentable, como ser la alimentación, la reproducción, la sanidad y la genética, entre otros.

La reproducción de bovinos mediante la Inseminación Artificial (IA) [2,3] es bastante sencilla y tiene muchas ventajas. Esta técnica se está aplicando desde hace bastante tiempo en el país. Hoy la Inseminación Artificial a Tiempo Fijo (IATF) es una técnica que, mediante la utilización de hormonas, permite sincronizar los celos y ovulaciones. Gracias a esto, es posible, inseminar una gran cantidad de animales en un corto período de tiempo [4].

[1] http://laslilas.com
[2] https://www.angus.org.ar/

Una de las herramientas utilizadas, por los ganaderos, para realizar las evaluaciones genéticas de los toros reproductores es la *Diferencia Esperada entre Progenie* (DEP), en base a estos valores, los productores pueden tomar decisiones de selección en base a información objetiva [5]. Los DEP anticipan cómo será el comportamiento promedio de las futuras crías de un toro en comparación con las que producirán el resto de los padres. Por el lado de las madres, a partir de esta información, la selección de los reproductores a utilizar como padres pasa a ser una de las más importantes decisiones de manejo que tiene el productor, permitiéndole seleccionar aquellos animales acordes a sus propios objetivos, su medio ambiente, su sistema de producción, e ir logrando avances genéticos que son acumulativos dentro del rodeo.

Seguidamente se detallan brevemente las descripciones de las siglas que componen los DEP's. Estas fueron extraídas del Anuario Las Lilas 2017 [6]:

- PN (Peso al nacer): Expresado en kilos, indica las diferencias genéticas para el PN de las crías de un padre determinado.
- PD (Peso al destete): Expresado en kilos. Indica el mérito genético de un reproductor en transmitir potencial de crecimiento directo a sus crías hasta el momento del destete.
- CM (Combinado materno): Esta variable combina el peso al destete y la aptitud materna en un solo valor.
- CE (Circunferencia escrotal): Expresada en centímetros y ajustada por edad de vida, es un indicador indirecto de la fertilidad de los rodeos.
- PF (Peso final): Expresado en kilos, indica la aptitud que tiene un reproductor en transmitir a su progenie capacidad de crecimiento post-destete.
- AM (Aptitud materna): Predictor de la producción lechera y aptitud materna que transmite un toro a sus hijas.
- AOB (Área del Ojo de Bife): Es la altura de la 12a costilla. Es un indicador del peso total y rendimiento de cortes despostados de la res.
- GD (Grasa dorsal): Expresada en milímetros, el espesor de grasa dorsal a la altura de la 12a costilla es un predictor genético de la precocidad y facilidad de terminación de las reses.
- MAR (Grado de Marmoreo): Es un indicador del porcentaje de grasa intramuscular del músculo dorsal largo.

Este trabajo se realiza bajo la hipótesis que si se aplica una metodología para realizar Minería de Datos (MD) a partir de los datos del material genético de los progenitores machos, más ciertos datos de las hembras, como la edad, el historial de partos, etc., se puede construir un modelo predictivo que mejor determine la etiqueta Peso al Nacer (Ver Tabla 1). Por otro lado, mediante los algoritmos No Supervisados, se intenta demostrar las relaciones existentes entre las variables, que también tengan mayor injerencia en el PN de los terneros.

Tabla 1. Descripción de las variables del conjunto de datos.

Nomenclatura	Tipo de dato	Descripción
Peso Adulto	*Numérico*	Peso real del padre
PAN	*Numérico*	Peso al nacer
PAD	*Numérico*	Peso al destete
PAF	*Numérico*	Peso final
Circ. Escrotal	*Numérico*	Circunferencia escrotal
FRAME	*Numérico*	Altura del animal
Certificado	*Numérico*	Edad promedio de los vientres primerizos
PN	*Numérico*	DEP del Toro Progenitor
PD	*Numérico*	
AM	*Numérico*	
CM	*Numérico*	
PF	*Numérico*	
CE	*Numérico*	
AOB	*Numérico*	
GD	*Numérico*	
MAR	*Numérico*	
Peso al nacer	*Numérico*	Datos de la vaca
Peso al destete	*Numérico*	
Cantidad nacimientos	*Numérico*	
Cantidad abortos	*Numérico*	
Baja antes del destete	*Numérico*	
Edad en meses	*Numérico*	
Certificado	*Numérico*	DEP del Toro Progenitor de la vaca
CE del padre	*Numérico*	
PN	*Numérico*	
PD	*Numérico*	
AM	*Numérico*	
CM	*Numérico*	
PF	*Numérico*	
CE	*Numérico*	
AOB	*Numérico*	
GD	*Numérico*	
MAR	*Numérico*	
Peso_Nac	*Texto*	Atributo Clase

En el primer año de la investigación se seleccionaron diferentes técnicas del tipo Supervisadas y No Supervisadas [7,8].

Las primeras son aquellas orientadas a predecir el valor de un atributo (etiqueta o clase) de un conjunto de datos, conocidos otros

atributos (atributos descriptivos). A partir de datos, cuya etiqueta se conoce, se induce una relación entre dicha etiqueta y otra serie de atributos. Esas relaciones sirven para realizar la predicción en datos cuya etiqueta es desconocida. Las segundas, también conocidas como técnicas de Clustering, agrupan datos dentro de un número de clases preestablecidas o no, partiendo de criterios de distancia o similitud. De esta forma se agrupan las clases que sean similares entre sí y distintas con las otras clases.

Existen en la actualidad distintas metodologías para llevar a cabo un proceso de MD, en este trabajo se optó, siguiendo la literatura consultada, por Cross Industry Standard Process for Data Mining (CRISP-DM) [8,9]. Esta tecnología interrelaciona diferentes fases del proceso entre sí, de tal manera que se consolida un proceso iterativo y recíproco.

Para la ejecución del modelo construido y las pruebas para realizar una comparación entre las diferentes técnicas, se utilizó el software WEKA[3] (acrónimo de Waikato Environment for Knowledge Analysis, en español «entorno para análisis del conocimiento de la Universidad de Waikato»). Para realizar las mismas, fue necesario la construcción de un modelo con los datos históricos de anteriores rodeos y sus resultados respecto al peso al nacer de sus crías. Para optimizar el modelo fue necesario normalizar las variables de entrada. Normalizar significa, en este caso, comprimir o extender los valores de la variable para que estén en un rango definido. Se empleó la fórmula de Normalización mínimo-máximo, la cual transforma linealmente los datos a un intervalo, para este caso, entre 0 y 1, donde el valor mínimo se escala a 0 y el máximo a 1 [10], que se define como:

$$X_{normalizada} = X - X_{min} / X_{max} - X_{min} \quad \textbf{(1)}$$

Para la evaluación de los clasificadores se utilizó una Matriz de Confusión (MC) y el análisis o curvas ROC (acrónimo de Receiver Operating Characteristic) [8], estos se encuentran dentro del software WEKA para cada clasificador. Con estas herramientas se evalúa entre otras cosas:

- La sensibilidad indica la capacidad de nuestro clasificador para dar como casos positivos los casos que realmente lo son; proporción de PN altos correctamente identificados.
- La especificidad indica la capacidad de nuestro estimador para dar como casos negativos los casos que realmente lo sean; por ejemplo: proporción de PN bajos correctamente identificados como tal.
- Este trabajo tiene como objetivo principal generar conocimiento especializado en el área de Minería de Datos en lo referente a un programa de mejoramiento genético. Específicamente, esta línea se centra principalmente en el estudio de dos ejes: los algoritmos Supervisados y No Supervisados.
- Además, como ya se mencionó, los datos de entrada, para realizar las pruebas contaron con los valores genéticos de sus padres/abuelos. El mayor obstáculo que se presentó es la cantidad y calidad de los mismos, si bien es posible encontrar tendencias y verificar ciertos resultados basados en el conocimiento del dominio, aún es necesario trabajar en tales datos, siendo lo deseado obtener nuevas cifras de establecimientos ganaderos similares.
- En algunos casos, incrementar el número de variables, mejora el rendimiento de los clasificadores, se pretende agregar variables tales como: el tipo de alimentación de los animales, el factor climático, etc.
- Se estudiarán nuevos algoritmos y con varias configuraciones para comprobar cual tiene la mayor capacidad de exactitud en su predicción.
- Por último, utilizar diferente herramienta computacional, por ejemplo probar con software más sofisticados como Matlab[1], SPSS[2], etc.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Este trabajo tiene como objetivo principal generar conocimiento especializado en el área de Minería de Datos en lo referente a un

[3] www.cs.waikato.ac.nz/~ml/weka/

programa de mejoramiento genético. Específicamente, esta línea se centra principalmente en el estudio de dos ejes: los algoritmos Supervisados y No Supervisados.

Además, como ya se mencionó, los datos de entrada, para realizar las pruebas contaron con los valores genéticos de sus padres/abuelos. El mayor obstáculo que se presentó es la cantidad y calidad de los mismos, si bien es posible encontrar tendencias y verificar ciertos resultados basados en el conocimiento del dominio, aún es necesario trabajar en tales datos, siendo lo deseado obtener nuevas cifras de establecimientos ganaderos similares.

En algunos casos, incrementar el número de variables, mejora el rendimiento de los clasificadores, se pretende agregar variables tales como: el tipo de alimentación de los animales, el factor climático, etc.

Se estudiarán nuevos algoritmos y con varias configuraciones para comprobar cual tiene la mayor capacidad de exactitud en su predicción.

Por último, utilizar diferente herramienta computacional, por ejemplo probar con software más sofisticados como Matlab[1], SPSS[2], etc.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Este grupo de investigación viene trabajando en proyectos PROINCE en años anteriores, también asociados a la misma temática:

- PROINCE C176 (2015-2016). "*Análisis Comparativo de Modelos de Clasificación de Minería de Datos (Data Mining). Su Aplicación en la Predicción de Perfiles de Alumnos en Riesgo de Deserción"*.
- PROINCE C199 (2017-2018). "*Modelos de Minería de Datos para el Diagnóstico Pre_ coz de Enfermedades Neurodegenerativas"*.
- PROINCE C205 (2017-2018). *"Uso de Minería de Datos para Acelerar la Recuperación de Documentos"*.

Los resultados de estas investigaciones dieron lugar a varias publicaciones [11-15].

Tal como quedó expuesto, se utilizaron dos tipos distintos de técnicas de MD. Se presentaron 2 trabajos en el año 2019. En uno se realizó una comparación en cuanto al desempeño de tres algoritmos del tipo Supervisado [16]:

- Árboles de Decisión (AD).
- Redes Neuronales Artificiales (RNA).
- Máquinas de Soporte Vectorial (MSV o SVM, del inglés Support Vector Machine)

En el otro trabajo presentado, se compararon cuatro algoritmos No Supervisados [17]:

- Expectation Maximization (EM).
- FarthestFirst.
- Simple K-Means.
- Mapas Auto Organizados (Redes SOM).

En el primero, se encontró que el modelo propuesto conserva una precisión (proporción de instancias clasificadas correctamente) aceptable en el caso del algoritmo AD, con un porcentaje levemente superior al 70%.

Para los algoritmos No supervisados, fue necesario un subconjunto de atributos menor del conjunto total inicial, que incluya aquellos relevantes para la tarea de agrupamiento. Se llegó a la conclusión que los datos relevantes para agrupar a los terneros según su PN involucra principalmente las características de su Madre y del Abuelo Materno.

En el segundo año se espera que, los criterios mencionados anteriormente, logrados de forma automática por los algoritmos, se puedan comparar con la estimación de los expertos en la realidad de los terneros a nacer este año. Es importante notar que la clasificación puede diferir. Sobre todo, teniendo en cuenta, la dificultad de calcular el peso del recién nacido. Por último, se están realizando con la Sociedad Rural de Chascomús y la Asociación Argentina de Angus, esto posibilitaría la obtención de datos de mejor calidad y de nuevos establecimientos ganaderos.

4. FORMACIÓN DE RECURSOS HUMANOS

Parte del grupo de desarrollo del proyecto trabaja desde el año 2015 en diversas áreas relacionadas con la Minería de Datos.

Actualmente forman parte del equipo, además de docentes de la UNLaM, una Ingeniera Agrónoma que trabaja en el SENASA y dos alumnos becarios de investigación.

Por otra parte, los docentes-investigadores

que integran el proyecto dictan clases en la cátedra de Inteligencia de Negocio de la carrera Licenciatura en Gestión de Tecnología y en la cátedra Base de Datos y Data Mining y Data Warehouse de la carrera de Ingeniería Informática. Se prevé, además, la capacitación y formación de recursos humanos, a través de cursos de actualización y posgrado en el área de estudio; la transferencia de conocimiento y resultados; y la posibilidad de brindar charlas informativas del desarrollo e implementación del proyecto a distintas instituciones del sectores ganadero, como la Sociedad Rural de Chascomús y la Asociación Argentina de Angus.

5. BIBLIOGRAFÍA

1. Firpo Brenta, L. y otros. (2012). Selección genética y mejoramiento animal. Disponible en: http://www.produccion-animal.com.ar /genética_seleccion_cruzamientos/bovinos_en_general/24-Seleccion_genetica.pdf. Último acceso: 06/09/2019.
2. Agrocor. (2011). Inseminación artificial en bovinos Curso Teórico Práctico de Inseminación Artificial en Bovinos. Disponible en: https://www.engormix.com /ganaderia-carne/articulos/inseminacion-artificial-en-bovinos-t26957.htm. Último acceso: 06/09/2019.
3. Díaz, P. Fonseca, V. Martínez P. y Rey A. (2003). Inseminación Artificial en bovinos. Biblioteca Digita, U. de Chile. Disponible en: www.biblioteca.org.ar/libros/8913.pdf. Último acceso: 06/02/2020.
4. Sommantico, S. (2018). Inseminación Arti_ ficial a Tiempo Fijo: la tecnología de la que se habla mucho y se usa poco. Disponible en: https://www.infocampo.com.ar/inseminacion-artificial-a-tiempo-fijo-la-tecnologia-de-la-que-se-habla-mucho-y-se-usa-poco/ Último acceso: 26/02/2020.
5. Guitou H. y Monti A. (1998). Interpretación y uso correcto de los DEPs como herramienta de selección. INTA Castelar. Disponible en: https://es.scribd.com/document/337947981/20-Interpretacion-Deps. Último acceso: 26/02/2020.
6. Cómo interpretar la evaluación genética. Anuario Las Lilas 2018-2019. Cabaña Las Lilas. Centro de Genética. Pág. 107. Disponible en: http://laslilas.com/pdf/Anuario-Genetica-2017.pdf. Último acceso: 26/02/2020.
7. Perez Lopez, C. y otros. (2007). Minería de datos. Técnicas y herramientas ISBN: 9788497324922. Ed. Paraninfo Cengage L. Madrid. España.
8. Hernández Orallo, J. y otros. (2004). Introdu_ cción a la minería de datos". Pearson. Edición: I.
9. Comparación de diferentes algoritmos de clustering en la estimación de coste en el desarrollo de software. (2005). U. de Alcalá, Madrid Disponible en: http://www.sc.ehu.es /jiwdocoj/remis/docs/GarreAdis05.pdf. Ultimo acceso: 26/02/2020.
10. Han Jiawei. Data Mining: Concepts and Techniques. 3ra. Edición. (2011). ISBN 978-0-12-381479-1.
11. Predicción del riesgo de abandono universitario utilizando métodos supervisados. (2016) IPECyT 2016. F. R. B. Blanca.
12. "Comparación de Algoritmos de Aprendizaje Supervisado para la obtención de perfiles de alumnos desertores". (2016). el CONAIISI 2016. Salta. Argentina.
13. "Modelos de minería de datos para el diagnóstico de enfermedad de Parkinson mediante el análisis de voz". Presentado CONAIISI 2017. Santa Fe. Argentina.
14. "Aceleración en la Recuperación de Información utilizando Algoritmos de Minería de Datos de R". Presentado en CACIC 2018. U. N. del Centro, Tandil.
15. "Selection of voice parameters for Parkinson´s disease prediction from collected mobile data". (2019) XXII Symposium on Image, Signal Processing and Art. Vision. Bucaramanga, Colombia.
16 *"Clasificación del Peso al Nacer de Terneros Aberdeen Angus mediante Algoritmos Supervisados"*. Trabajo presentado y no publicado en JAIIO 2019. UNSa.
17 *"Peso al Nacer de Terneros Aberdeen Angus mediante Algoritmos No Supervisados"*. Trabajo presentado en CONAIISI 2019. UNLaM. Buenos Aires. Argentina.

Búsqueda y Recopilación de Información sobre Legislación referida a Residuos Informáticos

Eduardo Rodríguez[1], Luciana Burzacca[1,2], Claudia Deco[1,2], Cristina Bender[1,2], Santiago Costa[1]

[1] Facultad de Química e Ingeniería del Rosario, Universidad Católica Argentina
[2] Universidad Nacional de Rosario
{ejrodriguez, lburzacca, cdeco, cbender, santiagocosta}@uca.edu.ar

Resumen

En la Argentina todavía existe poca conciencia sobre la importancia que posee el tema de los residuos informáticos. Este tipo de residuos tiene un crecimiento exponencial y constante. Produce un gravísimo impacto ambiental y pérdida económica ocasionada a partir de su no reutilización y su no reciclado. Dentro de este proyecto se propone tomar conocimiento sobre el marco legal en el ámbito internacional, nacional y local acerca de la gestión de Residuos de Aparatos Eléctricos y Electrónicos (RAEE); y para ello relevar y organizar esos datos recopilados en una matriz legal. En el proceso de la elaboración de esta matriz encontramos una gran diversidad de normas y hallamos un vacío legal nacional. Es por esto que estamos trabajando en el problema de buscar, diseñar una base de datos para organizar y almacenar la información correspondiente a la matriz legal, en pos de brindar un acceso eficiente a la misma.

Palabras Claves: RAEE, Recuperación de Información, Búsqueda de Información, Legislación.

Contexto

Esta línea de I+D se está llevando a cabo a través de un proyecto del Departamento de Investigación Institucional de la Facultad de Química e Ingeniería del Rosario de la Universidad Católica Argentina. El proyecto involucrado es: PID UCA "Propuesta de un Sistema de Gestión Integral de Residuos Electrónicos" (2017 - 2021).

Introducción

El exponencial desarrollo tecnológico de las últimas décadas tiene, como contrapartida, la acumulación de basura electrónica: el tipo de residuo de mayor crecimiento a escala mundial. Se definen como Aparatos Eléctricos y Electrónicos (AEE) a todo aparato que funciona con corriente eléctrica o campos electromagnéticos y que se utilizan con una tensión nominal menor a 1.000V en corriente alterna y 1.500V en corriente continua. Entre ellos están incluidos: grandes y pequeños electrodomésticos, equipos de informática y telecomunicaciones, aparatos de iluminación, herramientas eléctricas, equipos médicos. Un AEE se convierte en un Residuo de Aparato Eléctrico y Electrónico (RAEE) cuando el usuario lo descarta sin intención de que vuelva a utilizarse.

La velocidad con la que avanzan las

tecnologías informáticas hace que los componentes tecnológicos tengan un alto grado de obsolescencia tecnológica en un corto período. Actualmente se estima que las empresas, para mantenerse competitivas y actualizadas, con un entorno seguro y con usuarios con un nivel adecuado de productividad, reemplazan sus equipos de computación personal en un plazo de entre tres y cuatro años [1,2]. Se estima que el 50% de estos residuos están arrumbados en oficinas, hogares, entes públicos o depósitos; más del 40% se entierra o se descarta en basurales y rellenos y cerca del 10% ingresa en esquemas informales o formales de gestión de residuos [3, 4]. Cuando los RAEE no son gestionados de forma adecuada, resultan contaminantes para el medio ambiente.

En este proyecto se propone:

- La colaboración interinstitucional e interdisciplinaria entre especialistas de cada campo.
- Conocer la situación actual de RAEE en la región.
- Estudiar e investigar soluciones al problema del RAEE a nivel mundial aplicables a nuestra realidad regional.
- Analizar y testear distribuciones de Sistemas Operativos que extiendan la vida útil de equipos permitiendo su reutilización.
- Analizar el problema de construcción y mantenimiento de una matriz legal referida a los RAEE.
- Proponer un Sistema Integral de Gestión de RAEE.

La realización de este proyecto permite traducir el conocimiento generado en investigación y en un importante desarrollo tecnológico para un sector de alta relevancia como es recuperar equipos y realizar una reinserción responsable de los mismos en la sociedad. Un aporte fundamental es tener conocimiento disponible y eficaz sobre las disposiciones legales que afectan el tratamiento.

Líneas de Investigación, Desarrollo e Innovación

Debe tenerse en cuenta que los AEE son aparatos complejos que incluyen numerosas partes y componentes: piezas y carcasas de diversos materiales, plaquetas de circuitos impresos, tubos de rayos catódicos, pantallas de cristal líquido, cables, componentes eléctricos y electrónicos, diversos fluidos, contrapesos de hormigón, cartuchos de impresión, etc. Estos componentes pueden ser de materiales muy diversos, algunos con alto valor de mercado, como chatarra ferrosa y plástica, aluminio, oro, plata o estaño, y otros que pueden ser riesgosos para los ecosistemas y las personas, como plomo, compuestos fluorocarbonados, mercurio, entre otros.

Esto hace que, por un lado, cuando un AEE se descarta tenga sentido recuperar los materiales y componentes valiosos para reinsertarlos en el ciclo productivo; algunas estimaciones indican que el 97% de los componentes y materiales contenidos en un AEE son recuperables o reciclables. Al mismo tiempo, resulta imprescindible hacer una gestión adecuada que minimice los riesgos ambientales. Esto implica que todas las etapas de la gestión de los RAEE deben realizarse en condiciones seguras, sin mezclarse con otros flujos de residuos y evitando manipulaciones o roturas que puedan exponer al ambiente o a los trabajadores.

La gestión de los RAEE representa un problema creciente en un mundo donde el recambio tecnológico es cada vez más acelerado y en el que apenas una fracción mínima de los mismos se recicla o refuncionaliza.

En la Argentina no es una problemática menor. Según los datos del Observatorio Mundial de Residuos Electrónicos [5], en Argentina se generan anualmente alrededor de 8,4 kg de RAEE por persona. Si se multiplica por los 42 millones de habitantes, se obtiene unas 360 mil toneladas de RAEE generados anualmente. El desafío es, entonces, darle una disposición final a los equipos de acuerdo a los marcos legales que actualmente se encuentran en vigencia.

Con la intención de poder proponer soluciones al problema, en este proyecto se estudian varias líneas de acción: Análisis de legislación específica, Identificación de aspectos e impactos asociados a la actividad de reciclado/recupero de los residuos informáticos, y Diseño de sistemas de logística de recolección que resulten sustentables y viables económicamente e inspirados en las mejores prácticas de gestión ambientales existentes a nivel mundial.

Argentina posee regulaciones dispersas y no homogéneas. El marco regulatorio con el que cuenta debe ser fortalecido. A nivel nacional, con respecto a Residuos de Aparatos Eléctricos y Electrónicos nos encontramos con que no hay ninguna ley vigente. Actualmente en nuestro país se encuentran registrados dos proyectos de ley de presupuestos mínimos de gestión de residuos de aparatos eléctricos y electrónicos en la Cámara de Diputados: el Proyecto 0072-D-2018 presentado por el Diputado Juan Carlos Villalonga que establece presupuestos mínimos de protección ambiental, para la gestión de Aparatos Eléctricos y Electrónicos (AEE) y de los Residuos de Aparatos Eléctricos y Electrónicos (RAEE); y el Proyecto 5563- D-2018 presentado por el Diputado Daniel Filmus “Gestión de Residuos de Aparatos Eléctricos y Electrónicos”. Ambos proyectos incorporan el principio político de Responsabilidad Extendida del Productor (REP) y ponen foco en incorporar el análisis de Ciclo de Vida de un producto

En la etapa actual de este proyecto se pretende recopilar información sobre legislación referida a Residuos Informáticos. Para organizar la legislación vinculada a sus actividades, las empresas e instituciones mantienen una matriz legal cuyo contenido corresponde a estas actividades. Por esto, nos pareció conveniente organizar los datos recopilados vinculados a los RAEEs, en una matriz legal que consta de los siguientes campos:

Normativa (por ej. Convenio de Basilea), *Temas de implicancia* (por ej. Proteger el medio ambiente y la salud humana de los efectos nocivos provocados por la generación, manejo, movimientos transfronterizos y eliminación de desechos peligroso), *Jurisdicción* (por ej. Internacional-ONU), *Órgano Emisor* (por ej. AMUMA-Acuerdo Multilateral sobre Medio Ambiente), *Autoridad de Aplicación, Fecha de sanción/Aprobación, Fecha de Publicación/Promulgación, Descripción, Observaciones, Obligaciones que Establece y Definiciones.* Esta matriz debe estar actualizada con las leyes, decretos, resoluciones, etc., vigentes. Esto involucra un esfuerzo considerable para buscar, recopilar y mantener esta información.

Todas las empresas e industrias tienen la obligación de poseer una matriz legal según su actividad; por ejemplo, una aceitera posee una matriz con la legislación correspondiente a desechos peligrosos, entre otras. La matriz construida en nuestro proyecto puede adaptarse o insertarse en las matrices o submatrices legales que tiene cada empresa.

Resultados y Objetivos

El objetivo general de este proyecto es la elaboración de propuestas de acción para la puesta en marcha de un Sistema Integral de Gestión de RAEE para la región. Para esto se tienen los siguientes objetivos específicos:

- Establecimiento de línea de base e indicadores de gestión.
- Analizar el problema de construcción y mantenimiento de una matriz legal.
- Propuesta de marco legal determinando las responsabilidades de los actores involucrados.
- Elaboración de Proyectos de Unidades de Gestión de RAEE para reutilizado y reciclado.
- Elaboración de una propuesta de recolección, logística y acopio de RAEE.
- Análisis económico-financiero del sistema integral.
- Modelización y optimización del sistema general.
- Análisis de SO orientados a la reutilización de equipos.

En una primera instancia se estuvo trabajando en la selección de Sistemas Operativos (SOs) apropiados para extender la vida útil de residuos electrónicos. En este sentido, en la actualidad existe una gran cantidad de equipos informáticos dispuestos por grandes empresas y gobierno, que luego de un extensivo período de uso no se encuentran en condiciones de ser reinsertados en ambientes productivos, dado que pasado su ciclo de vida útil la performance de estos se torna prácticamente nula. Esto genera el análisis y búsqueda de SOs capaces de maximizar los recursos disponibles en equipos de computación personal de hasta catorce años de antigüedad. Como parte de la solución se analizó el mercado actual de SOs considerando aquellos con características enfocadas en un mínimo consumo de recursos y un máximo aprovechamiento de las capacidades de hardware; y se determinó, como resultado experimental que Linux Mint 19 Tara - Xfce (32-bit) fue el más óptimo [6, 7].

Luego se colaboró en la puesta en marcha de una Planta de Reciclado y Reacondicionado de Material Informático en Desuso (ver descripción en https://tau.org.ar/raee/). Para esto, integrantes del equipo de esta investigación han participado junto con la Asociación Civil Nodo TAU y el Grupo Obispo Angelelli, y en convenio con el Programa Nueva Oportunidad del gobierno de la provincia de Santa Fe. Además un grupo de alumnas de la Carrera Ingeniería Ambiental está realizando el estudio de Impacto Ambiental de dicha Planta. Además, la Planta cuenta con un Laboratorio que está destinado al testeo de placas madres y eventual recambio de componentes discretos en las mismas y en fuentes de alimentación y monitores planos. Los procedimientos que se llevan a cabo no difieren de los que se pueden realizar en un laboratorio de electrónica, es decir desoldado y soldado de componentes discretos.

Actualmente, se está trabajando en el problema de buscar y recopilar información legal relacionada con los RAEEs a los fines de poblar una matriz legal. Al momento se cuenta con 90 registros. En el proceso de la elaboración de esta matriz hallamos un vacío legal en cuanto al tratamiento de RAEEs, lo que dificulta esta actividad ya que según la legislación vigente muchos de los RAEEs caen dentro de la categoría de residuos peligrosos. El paso siguiente será la definición de la forma de almacenarla en un repositorio local a los fines de hacerla disponible para empresas e industrias de la región, así como al ámbito educativo, y

de esta forma contribuir a la matriz legal de cada industria o empresa.

Formación de Recursos Humanos

El equipo de trabajo está integrado por los magisters Eduardo Rodríguez y Cristina Bender, la doctora Claudia Deco, la especialista Luciana Burzacca, y el licenciado Santiago Costa, todos investigadores de la Universidad Católica Argentina.

El Lic. Costa está cursando la Maestría en Redes de Datos en la Universidad de La Plata y la Lic. Burzacca está realizando una Maestría en la Universidad Tecnológica Facultad Rosario.

Además, se han incorporado 4 alumnos de las carreras Ingeniería Industrial e Ingeniería Ambiental con los cuales se trabaja a partir de subproblemas identificados durante el trabajo de investigación. Los mismos se proponen para la realización de proyectos finales de grado.

Los conocimientos adquiridos se vuelcan en la comunidad académica a través de las actividades docentes que los investigadores de este proyecto realizan en las carreras Ingeniería Industrial e Ingeniería Ambiental de nuestra universidad.

La conformación del grupo de investigación por docentes de la carrera de Ingeniería Industrial y la carrera de Ingeniería Ambiental permite un enfoque multidisciplinario y adecuado de esta problemática. La viabilidad del proyecto está sustentada en los avances obtenidos mediante trabajos previos de los integrantes en el área y su interacción con otros grupos de investigadores tanto latinoamericanos como europeos.

Referencias

[1] Fernández Protomastro, G. (2014). Buenas Prácticas para la gestión de los residuos de aparatos eléctricos y electrónicos - RAEE. 1a ed. Ciudad Autónoma de Buenos Aires: Grupo Uno. 178 p. ISBN 978-987-29862-3-0

[2] Fernández Protomastro, G. (2013). Minería Urbana y la Gestión de los Residuos Electrónicos. - 1a ed. Ciudad Autónoma de Buenos Aires: Grupo Uno. 317 p. ISBN 978-987-29862-1-6

[3] Greenpeace, (2011). Basura Electrónica, la otra cara de la tecnología. Disponible en: http://www.greenpeace.org/argentina/Global/argentina/report/2011/contaminacion/basura_electronica_otra_cara_tecnologia.pdf. Consultado 07/04/2018.

[4] Greenpeace, (2012). Minería y Basura Electrónica, el manejo irracional de los recursos. Disponible en: http://www.greenpeace.org/argentina/Global/argentina/report/2012/contaminacion/inform-raee-V-1.pdf Consultado 07/04/2018.

[5] Baldé, C. P., Forti, V., Gray, V., Kuehr, R., & Stegmann, P. (2017). Observatorio Mundial de los Residuos electrónicos. 2017. Ginebra: ONU.

[6] Rodríguez, Eduardo, Deco, Claudia; Burzacca, Luciana; del Giglio, Cristian; Bender, Cristina; Costa, Santiago. (2019). Tratamiento de residuos informáticos seleccionando Sistemas Operativos apropiados para extender su vida útil.. 7º Congreso de Ciencias Ambientales COPIME 2019. Buenos Aires, Argentina, octubre de 2019.

[7] Rodríguez, Eduardo, Deco, Claudia; Burzacca, Luciana; del Giglio, Cristian; Bender, Cristina; Costa, Santiago. (2019). Selección de Sistemas Operativos apropiados para extender la vida útil de residuos electrónicos (RAEE), WICC 2019, Área "Arquitectura, Redes y Sistemas Operativos", Universidad Nacional de San Juan. Argentina ISBN: 978-987-3619-27-4. 5 p.

Sistemas Inteligentes. Aplicaciones en Minería de Datos y Big Data

L. Lanzarini[1], W. Hasperué[1], C. Estrebou[1], A. Villa Monte[1,3], P. Jimbo Santana[4], G. Reyes Zambrano[5], G. Camele[1,3], P. López[2], J. Corvi[2], A. Fernandez Bariviera[6], J. A. Olivas[7]

[1] Instituto de Investigación en Informática LIDI*, Facultad de Informática, UNLP, La Plata, Argentina
[2] Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina
[3] Becario postgrado UNLP
[4] Facultad de Ciencias Administrativas, Universidad Central del Ecuador, Quito, Ecuador
[5] Facultad de Ciencias Físicas y Matemáticas, Universidad de Guayaquil, Guayaquil, Ecuador
[6] Dpto de Economía, Universitat Rovira i Virgili, Reus, España
[7] Dpto. Tecnología y Sistemas de la Información, Universidad de Castilla-La Mancha, Ciudad Real, España
*** Centro asociado de la Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)**

{laural, whasperue, cesarest, avillamonte, gcamele, pdlopez}@lidi.info.unlp.edu.ar
prjimbo@uce.edu.ec, gary.reyesz@ug.edu.ec, julieta.corvi@gmail.com , aurelio.fernandez@urv.net, joseangel.olivas@uclm.es

CONTEXTO

Esta presentación corresponde a las tareas de investigación que se llevan a cabo en el III LIDI en el marco del proyecto "Sistemas inteligentes. Aplicaciones en reconocimiento de patrones, minería de datos y big data" perteneciente al Programa de Incentivos (2018-2021) y del proyecto PITAP-BA "Computación de Alto Desempeño, Minería de Datos y Aplicaciones de Interés Social en la Provincia de Bs.As." evaluado y subsidiado por la Comisión de Investigaciones Científicas de la Provincia de Bs.As. (2017-2019).

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes para la resolución de problemas de Minería de Datos y Big Data utilizando técnicas de Aprendizaje Automático. Los sistemas desarrollados se aplican particularmente al procesamiento de grandes volúmenes de información y al procesamiento de flujo de datos.

En el área de la Minería de Datos se está trabajando, por un lado, en la construcción de conjuntos de reglas de clasificación difusas que faciliten y permitan justificar la toma de decisiones y, por otro lado, en el diseño de técnicas de agrupamiento para flujos de datos con aplicación al análisis de trayectorias vehiculares para predecir congestión de tránsito.

Con respecto al área de Big Data se está trabajando en el diseño y desarrollo de algoritmos de selección de características en grandes bases de datos de muchas columnas. Las implementaciones que se están llevando a cabo serán utilizadas en problemas de genómica molecular con el objetivo de determinar *gene signatures*. En esta misma línea se está trabajando sobre algoritmos de selección de características para el tratamiento de flujos de datos.

Por otro lado y como transferencia tecnológica concreta, se efectuó un análisis sobre la producción de leche en ganado bovino a partir de la base de datos de la Asociación de la Región Pampeana de Entidades de Control Lechero (ARPECOL).

En el área de la Minería de Textos se han desarrollado estrategias para resumir documentos a través de la extracción de los párrafos más representativos utilizando métricas de selección y técnicas de optimización.

Palabras clave: Minería de Datos, Minería de Textos, Big Data, Redes Neuronales, Resúmenes Extractivos, Stream Processing. Selección de Características.

1. INTRODUCCION

El Instituto de Investigación en Informática LIDI tiene una larga trayectoria en el estudio, investigación y desarrollo de Sistemas Inteligentes basados en distintos tipos de estrategias adaptativas. Los resultados obtenidos han sido medidos en la solución de problemas pertenecientes a distintas áreas. A continuación se detallan los resultados obtenidos durante el último año.

1.1. MINERÍA DE DATOS

Extracción de Reglas de Clasificación

Las técnicas de minería de datos son de gran interés para las organizaciones, ya que facilitan la adopción de decisiones tácticas y estratégicas, generando una ventaja competitiva. En el caso especial de las organizaciones que conceden créditos, es importante definir claramente los criterios de rechazo/aprobación. En este sentido, las reglas de clasificación son un instrumento adecuado, siempre que el conjunto de reglas sea fácil de comprender y de aplicar y además posea una tasa de acierto razonable. La simplicidad de la regla se logra generando antecedentes cortos, formados por pocas condiciones.

En el III LIDI se analizan y diseñan soluciones basadas en técnicas de optimización para construir conjuntos de reglas de clasificación con las características antes mencionadas. Los modelos construidos son capaces de operar con atributos nominales y numéricos y emplean redes neuronales competitivas como punto de partida para reducir el tiempo de búsqueda. Además, para facilitar la comprensión del modelo, se incorpora la lógica difusa en la construcción del antecedente. En esta dirección, el énfasis está puesto en la importancia del uso de conjuntos difusos al expresar las condiciones que implican atributos numéricos. Por su intermedio, no sólo se facilita la comprensión y la aplicación de las reglas sino que también se consigue una precisión significativamente mayor respecto de las condiciones *crisp* o no difusas.

En el caso particular de las reglas aplicables a riesgo crediticio se utiliza la información del prestatario y el entorno macroeconómico en el momento de conceder el préstamo.

Los resultados obtenidos de aplicar las técnicas propuestas a distintas bases de datos de instituciones financieras en el Ecuador han sido publicados en [1-3]. Se trata de instituciones que se especializan en la colocación masiva de créditos así como en créditos de consumo y líneas de crédito empresarial. En todos los casos se ha observado que la incorporación de la lógica difusa genera conjuntos de reglas con mayor precisión.

Agrupamiento de flujos de datos

Las técnicas de agrupamiento han sido ampliamente utilizadas a la hora de construir modelos capaces de resolver tareas descriptivas, es decir, modelos que buscan establecer similitudes y diferencias entre las distintas situaciones que pueden ocurrir a fin de identificar cuáles son las características más importantes que deben ser tenidas en cuenta a la hora de explicar el comportamiento de un proceso.

Si se analiza la evolución de este tipo de técnicas puede verse que, en un principio, los modelos han operado buscando establecer similitudes en un conjunto de datos conocido a priori. Utilizando el enfoque tradicional, dicho conjunto puede ser analizado todas las veces que sea necesario hasta determinar las características más representativas del problema. Esta forma de trabajo, si bien sigue siendo válida en numerosas situaciones, presenta problemas cuando deben procesarse volúmenes de información sumamente grandes. En estos casos puede ocurrir que no se disponga de espacio suficiente para su almacenamiento o incluso que no resulte conveniente guardarlos por la velocidad con la que la información se actualiza.

Por lo antes expuesto, se está trabajando con técnicas de agrupamiento para flujos de datos capaces de establecer similitudes analizándolos una única vez [4]. Por lo tanto,

la información no será almacenada sino que será procesada inmediatamente al momento de arribar. Procesar un flujo de datos implica realizar modificaciones sobre el modelo de manera automática porque, como se dijo previamente, ya no se tendrán los datos originales para efectuar modificaciones sino que deben ir efectuándose registros más genéricos a medida que se ingresa la información. Este tipo de entrada se ajusta perfectamente al procesamiento de información temporal.

En esta línea de investigación interesa especialmente considerar la modalidad de recolección de los datos, la cantidad de veces que los mismos pueden ser utilizados y los parámetros de configuración que debe indicar el usuario.

Los resultados obtenidos hasta el momento se están aplicando en el análisis de trayectorias GPS. Los avances tecnológicos facilitan el registro de información sobre las trayectorias GPS de los vehículos en las carreteras públicas. El análisis inteligente de estos datos permite identificar patrones extremadamente útiles para la toma de decisiones en situaciones relacionadas con el urbanismo, el tráfico y la congestión de las carreteras, entre otros. En [5] se presentó un nuevo método de agrupamiento de trayectorias GPS que utiliza información angular para segmentar los recorridos y una función de similitud guiada por un pivote. El proceso de adaptación inicia distribuyendo los centroides de manera uniforme en la región a analizar formando un reticulado. Los resultados obtenidos luego de aplicar el método propuesto sobre una base de datos de trayectorias reales fueron satisfactorios y muestran una mejoría significativa en comparación con los métodos publicados en la bibliografía.

1.2. BIG DATA

Aplicaciones en Big Data

En esta línea se trabaja sobre el procesamiento, en streaming y en batch, de grandes volúmenes de datos. Para el procesamiento en streaming se están desarrollando estrategias basadas en técnicas de Aprendizaje Automático que permitan la selección de los atributos más relevantes de un flujo de datos, brindando resultados en tiempos de respuestas cortos los cuales se adaptan de manera dinámica a la llegada de nuevos datos [6].

Estas técnicas dinámicas se están implementando en el framework Spark Streaming, adecuado para procesamiento paralelo, distribuido y online.

Los algoritmos de selección de atributos sobre flujos de datos representan un desafío interesante ya que se deben tratar con cuidado el problema de *concept drift* para no perder información relevante.

Por otro lado y como una transferencia tecnológica concreta se ha trabajado en el tratamiento de la información proveniente de ARPECOL, una asociación que nuclea entidades de control lechero de la provincia de Buenos Aires. Las entidades de control lechero son organizaciones que brindan el servicio de medición de la producción de leche individual a los productores. Estas entidades toman muestras de la leche de las vacas para realizar análisis de laboratorio de la calidad de la leche producida (porcentaje de grasa, de proteínas, de solidos totales, conteo de células somáticas). En esta línea de investigación se colaboró con un proyecto del INIRA de la Facultad de Veterinaria de la UNLP que tiene por objetivo determinar factores genéticos para la identificación de las principales enfermedades reproductivas, mastitis y cojeras que afectan la lactancia de las vacas de tambo.

1.3. MINERIA DE TEXTOS

Hoy en día, la información que nos rodea lo hace en su gran mayoría en forma de texto. El volumen de información no estructurada crece continuamente de tal manera que resulta necesario separar por medio de técnicas de procesamiento de texto lo esencial de lo que no lo es. Los instrumentos de resumen

automático de textos tienen un gran impacto en muchos campos, como la medicina, el derecho y la investigación científica en general. A medida que aumenta la sobrecarga de información, los resúmenes automáticos permiten manejar el creciente volumen de documentos, generalmente asignando pesos a las frases extraídas en función de su importancia en el resumen previsto. La obtención del contenido principal de un documento determinado en menos tiempo del que llevaría hacerlo manualmente sigue siendo una cuestión de interés. En [7] se presentó un nuevo método capaz de generar automáticamente resúmenes extractivos de documentos mediante la ponderación adecuada de las características de puntuación de las frases utilizando optimización por cúmulo de partículas. El método propuesto es capaz de identificar las características que más se aproximan al criterio utilizado por el individuo al hacer el resumen. Para ello, combina una representación binaria y otra continua, utilizando una variación original de la técnica desarrollada por los autores de este documento. Las experimentaciones realizadas confirman que el uso de la información etiquetada por el usuario en el conjunto de entrenamiento ayuda a encontrar mejores métricas y pesos. Los resultados empíricos difundidos en [7] reflejan una mayor precisión en comparación con los métodos anteriores utilizados en este campo.

Por otro lado, en [8] se efectuó un estudio similar a partir de un modelo basado en una red neuronal que, a partir de la puntuación calculada con distintas métricas, es capaz de predecir la importancia de la sentencia dentro del documento. Una vez entrenada la red, este criterio puede aplicarse a otros documentos para obtener, como resultado, un resumen similar al que el usuario habría hecho manualmente.

2. TEMAS DE INVESTIGACIÓN Y DESARROLLO

- Estudio de técnicas de optimización poblacionales y redes neuronales artificiales para la obtención de reglas difusas de tipo IF-THEN.
- Modelización de trayectorias espacio-temporales con capacidad para establecer características comunes y detectar situaciones anómalas.
- Métodos estructurados y no estructurados aplicables a la representación de documentos.
- Representación de documentos de texto utilizando métricas.
- Obtención de resúmenes automáticos de texto.
- Implementación de técnicas inteligentes en el framework Spark Streaming
- Implementación de un algoritmo de selección de atributos en batch y en streaming.
- Uso de algoritmos de selección de atributos para la detección de gene signatures.
- Análisis de la base de datos de ARPECOL para la identificación de características genéticas que mejoren la producción de leche de las vacas de tambo.

3. RESULTADOS OBTENIDOS

- Desarrollo de un método de obtención de reglas de clasificación difusas con énfasis en la reducción de la complejidad del modelo aplicable a riesgo crediticio.
- Diseño e implementación de técnicas de agrupamiento para flujos de datos considerando especialmente los parámetros de configuración que debe indicar el usuario.
- Diseño e implementación de un nuevo método de agrupamiento de trayectorias GPS aplicable a la predicción de congestiones vehiculares.
- Desarrollo de un algoritmo de clustering que selecciona el número de clusters de manera dinámica implementado en el framework Spark streaming aplicado a

flujos de textos cortos.

- Identificación de las partes relevantes de un documento. Propuesta de distintas métricas y una representación vectorial de oraciones de diferentes longitudes.
- Análisis y comparación de resúmenes extractivos de documentos.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por: 2 profesores doctores con dedicación exclusiva, 4 tesistas de Doctorado en Cs. Informáticas (2 con beca de postgrado de la UNLP), 2 tesistas de grado y 2 profesores extranjeros.

Dentro de los temas involucrados en esta línea de investigación, en los últimos 2 años se han finalizado 1 tesis de doctorado, 1 tesis de especialista y 5 tesinas de grado de Licenciatura.

Actualmente se están desarrollando 3 tesis de doctorado, 2 tesis de especialista y 3 tesinas de grado de Licenciatura. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

[1] Jimbo P., Lanzarini L., Fernandez-Bariviera A. (2019) Variations of Particle Swarm Optimization for Obtaining Classification Rules Applied to Credit Risk in Financial Institutions of Ecuador. Journal Risks,8,1,1-14, MDPI, Open Access Journal

[2] Jimbo P., Lanzarini L., Fernandez-Bariviera A. (2019). Fuzzy Classification Rules with FRvarPSO Using Various Methods for Obtaining Fuzzy Sets. International Journal of Machine Learning and Computing.:International Association of Computer Science & Information Technology. issn 2010-3700.

[3] Jimbo P., Lanzarini L., Fernandez-Bariviera A. (2019). Particle Swarm Optimization for Obtaining Classification Rules Applied to Credit Risk in Financial Institutions of Ecuador. Risks 2020, Vol.8, Issue 1 doi:10.3390/risks8010002.

[4] Barbosa N., Travé-Massuyès L., Grisales-Palacio V. (2019) DyClee: Dynamic clustering for tracking evolving environments, Pattern Recognition. Volume 94, Pages 162-186, ISSN 0031-3203,

[5] Reyes-Zambrano G., Lanzarini L., Hasperué W., Fernández-Bariviera A (2020) GPS trajectory clustering method for decision making on intelligent transportation systems. Journal of Intelligent & Fuzzy Systems, vol. Pre-press, pp. 1-6. ISSN 1064-1246. DOI: 10.3233/JIFS-179644

[6] Molina, R, Hasperué, W. Villa Monte, A. *D3CAS: Distributed Clustering Algorithm Applied to Short-Text Stream Processing. Communications in Computer and Information Science. Springer. 2019. pp. 211-220.*

[7] Villa-Monte A., Lanzarini L., Fernández-Bariviera A., Olivas-Varela J.A. (2019) User-Oriented Summaries Using a PSO Based Scoring Optimization Method. Journal Entropy. 21 (6), 617. ISSN 1099-4300. https://doi.org/10.3390/e21060617

[8] Villa-Monte A., Lanzarini L., Corvi J., Fernández-Bariviera A.(2020). Document summarization using a structural metrics based representation. Journal of Intelligent & Fuzzy Systems. 1-10. 10.3233/JIFS-179648.

Laura Lanzarini:
0000-0001-7027-7564

Waldo Hasperué:
0000-0002-9950-1563

César Estrebou:
0000-0001-5926-8827

Augusto Villa Monte:
0000-0002-9854-3083

Aurelio Fernandez Bariviera:
0000-0003-1014-1010

José Ángel Olivas Varela:
0000-0003-4172-4729

Almacenamiento y procesado automático de imágenes satelitales para proyectos de monitoreo de bosque nativo a escala regional y local

Viva Mayer, Francisco M.[1,2] vivamayer.francisco@uader.edu.ar

Walter F. Sione[1,3] wsione@gmail.com

Francisco D. Maldonado[1] francisco.dario.maldonado@gmail.com

[1] CEREGEO - Centro Regional de Geomática
Facultad de Ciencia y Tecnología. Universidad Autónoma de Entre Ríos.
Ruta 11, km 10,5, Oro Verde. Entre Ríos.

[2] CICyTTP - Centro de Investigaciones Científicas y Transferencia de Tecnología a la Producción
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Matteri y España S/N° (CP 3105). Diamante. Entre Ríos.

[3] PRODITEL- Departamento de Ciencias Básicas.
Universidad Nacional De Luján – UNLU
Ruta 5 (CP 6700), Lujan. Buenos Aires. Argentina.

Resumen

La necesidad de información sobre el estado de conservación y la dinámica de cambios en coberturas de bosque nativo de nuestro país implica el procesamiento de grandes volúmenes de datos, de diferentes fuentes y de manera oportuna, por lo que es indispensable tender a la automatización de procesos.

En este trabajo se presentan los objetivos, la metodología, se discuten las características y el impacto a escala local y regional que tienen las líneas de investigación sobre automatización de procesos de monitoreos de estado actual y cambios en la cobertura de suelo. La formación de recursos humanos en esta línea y el trabajo multidisciplinario se plantean como necesidades para abordar estas líneas a futuro.

Palabras claves: Imágenes satelitales, Teledetección, Modelo de datos.

1.- Contexto

Esta línea de investigación y desarrollo se enmarca en el Proyecto de Tesis Doctoral "Desarrollo y Validación de un Sistema Informatizado de Monitoreo Automático de los desmontes para un área del Espinal del Centro-Norte de Entre Ríos, en base a Grandes Volúmenes de Datos de Teleobservación" del autor, y en líneas de investigación del Centro Regional de Geomática de la Facultad de Ciencia y Tecnología de la Universidad Autónoma de Entre Ríos.

En este trabajo se avanza con el desarrollo de herramientas aplicables en el marco de la Ley de Presupuestos Mínimos de Protección Ambiental de los Bosques Nativos de la Ley Nacional de Bosques 26331 dependientes de los órganos de competencia de las provincias.

2.- Introducción

Los ecosistemas se encuentran en un estado de flujo permanente en una variedad de escalas espaciales y temporales en todo el mundo. Las causas de estos flujos pueden ser tanto naturales como antropogénicas, o pueden ser una combinación de ambas. En este sentido, los datos proporcionados por la tele-detección (*remote* sensing) desde plataformas satelitales activas (radar) y pasivas (ópticas) se han presentado durante mucho tiempo como un recurso crítico para monitorear, medir y explicar dichos fenómenos naturales y físicos (Baghdadi & Zribi, 2016; Campbell & Wynne, 2011; Chuvieco, 2016). De hecho, la tele-detección podría caracterizarse justamente como una de las primeras ciencias de "*big data*" (Goodchild et al., 2012).

Sin embargo, es en los últimos años se observa un aumento significativo en la provisión y uso de datos de imágenes satelitales del espectro óptico para el control y seguimiento de las dinámica de las cambios de la superficie terrestre (natural y artificial) y de los cambios en el uso del suelo. Este aumento puede atribuirse a dos importantes hechos:

1. la necesidad de información precisa y oportuna de una sociedad que apunta hacia un desarrollo sustentable, que le permita mantener y/o mejorar el estilo de vida actual (Chuvieco, 2015; Hansen & Loveland, 2012);
2. los cambios en las políticas de acceso a la información de los organismos productores de datos de tele-detección desde plataformas satelitales (Soille et al., 2016; Woodcock et al., 2008), que ha resultado en un incremento de la disponibilidad (*data availability*) de la misma.

Estos avances han llevado a que la tele-detección produzca más información que la que resulta viable organizar y analizar mediante procedimientos tradicionales, que requieren de la intervención humana. Generando problemas de redundancia de información y/o desconocimiento de su existencia; o la adopción de técnicas y métodos de análisis que filtran, recortan o eliminan información a fin de volverlas humanamente manejables. Por ejemplo, para realizar un análisis activo de la cubierto vegetal de la provincia de Entre Ríos mediante imágenes de resolución media del espectro óptico de la plataforma Landsat8/OLI son necesarios ~ 6GB de datos cada 16 días. El procesado de esos datos $((7581 * 7301)_{pixel} * 4_{bandas} * 6_{tiles})$ incluyen billones de operaciones con números flotantes que requieren varios minutos o incluso hasta una hora para completar usando una sola estación de trabajo.

Este contexto de alta demanda de información y alta disponibilidad de datos ha llevado al auge de sistemas de computación de alto rendimiento (≪HPC≫ o *High Performance Computing*) y de plataformas de computación en La Nube. En Wang *et al.* (2016) encontramos un ejemplo donde se implementan múltiples paradigmas, técnicas y tecnologías de computación en la nube y HPC para un sistema de procesamiento de datos de tele-detección a escala global.

Otro ejemplo, lo encontramos en la plataforma de acceso gratuito Google Earth Engine (Gorelick et al., 2017). La cual ofrece a los investigadores y emprendedores una solución a los costos de infraestructura informática necesaria para el procesamiento y análisis de los datos de tele-detección; proporcionando acceso a un gran almacén de imágenes satelitales y la potencia

computacional necesaria para analizar esas imágenes.

Una característica de estas plataformas es que están orientadas a procesamientos a demanda (*on-demand*) de los usuarios. Por otro lado, no ofrecen estructuras sobre las cuales asentar las áreas de estudio u organizar los resultados de los procesos, quedando estas actividades en manos de los usuarios. Ambas características, son una ventaja para estudios de casos puntuales, pero se convierten en una desventaja al afrontar un estudio activo de resultados constantes para la toma de decisiones.

De los dos ejemplos presentados, solo Google Earth Engine se ofrece como una plataforma al público. Por otro lado, ambas plataformas no dan acceso al código fuente (*open source*) por consiguiente presenta dificultades para la implementación local en proyectos de pequeña y mediana escala.

En vista a las limitaciones actuales, en el marco de este proyecto se trabaja en la implementación de un sistema de gestión de información espacial, aplicable a escalas locales y regionales que organicen el almacenamiento, procesan de manera automática o semiautomática, y distribución de la información generada en base a datos de tele-observación y aplicables al monitoreo y gestión del bosque nativo.

Para el desarrollo de este proyecto, se optó por la utilización de soluciones software de código abierto (*open source*) o de libre distribución (*free software*). La utilización de herramientas open-source beneficia principalmente en la reducción de los costos de adquisición de licencias de software directas o indirectas previas y posteriores al proyecto; y también permite la libre distribución (adecuándose a las licencias existentes) de lo producido

Así, se decidió por el uso de PostgreSQL (PostgreSQL, 2018) como Sistema de Gestión de Base de Datos Relacional Orienta a Objetos (Object-Relational Database Management System, ORDBMS) y PostGIS (PostGIS, 2018) como extensión espacial para el mismo. PostgreSQL es un sistema de base de datos relacional de código abierto altamente escalable y compatible con SQL. Se ejecuta en todos los principales sistemas operativos, incluidos Linux, UNIX y Windows. Tiene una arquitectura comprobada que le ha valido una sólida reputación de fiabilidad, integridad de datos y corrección (Chen & Xie, 2008).

Por su parte, PostGIS, como una extensión espacial de PostgreSQL, cumple con los estándares OGC e ISO SQL/MM, que hacen posible el uso de PostgreSQL para el desarrollo moderno de aplicaciones espaciales. Así mismo, provee soporte para la gestión de información espacial en modelo ráster a partir de PostGIS 2.0+ (Dang, 2012).

La informatización de estas rutinas de procesamiento y transformación de datos se efectuara con Python (Python, 2020), haciendo uso de las librerías Geospatial Data Abstraction Library (GDAL, 2020), SciPy (SciPy, 2020) y de la rutina de extracción de nubes Fmask (Qiu et al., 2019; Zhu et al., 2015).

El resultado esperado de este trabajo de investigación y desarrollo es una estructura organizada permita establecer consultas elaboradas sobre múltiples capas de información; donde se reduzca la redundancia, se mejore la disponibilidad, la consistencia, y la integridad de los datos facilitando la interoperabilidad de los mismos con otros generadores de información y usuarios finales.

3.- Líneas de investigación y desarrollo

Las líneas son ejecutadas por un equipo multidisciplinario con capacidad para el procesamiento de imágenes, la adaptación y aplicación de técnicas de levantamiento de campo y comprender la dinámica de las coberturas vegetales mediante el uso de imágenes satelitales.

Una de las líneas en desarrollo es la estructura principal de la base de datos sobre la cual se construirá el sistema. La misma debe ser flexible y modular a fin de permitir el almacenamiento de una miríada de fuentes de información espacial, tanto de plataformas satelitales como de levantamiento de datos campo.

Otra línea se relaciona con el sistema que hace uso del modelo de base de datos y proporcione funcionalidades para la ingesta y consulta de los datos, así como una plataforma sobre la cual puedan construirse modelos automáticos o semiautomáticos de producción de información como por ejemplo para la detección de desmontes en tiempo casi-real y monitoreo del cambio de uso de suelo, entre otras.

4.- Resultados esperados

El principal resultado es una plataforma sobre la cual pueda implementarse una metodología automática operacional para la detección y monitoreo en tiempo casi-real de los desmontes en el bosque nativo del centro-norte de la provincia de Entre Ríos. Y a su vez permita producir periódicamente: alertas de desmontes, mapas de intensidad de los cambios del bosque natural ,mapas de cambios de uso producidos en áreas de bosque natural y tablas cuantitativas de los desmontes en el centro-norte de Entre Ríos.

Se espera que la misma sirva como trampolín para futuros desarrollos y avances en el libre acceso a la información.

5.- Formación de Recursos Humanos

El Magíster Lic. Francisco M. Viva Mayer lleva adelante su tesis de doctorado en Ciencias Aplicadas, con el título "*Desarrollo y Validación de un Sistema Informatizado de Monitoreo Automático de los desmontes para un área del Espinal del Centro-Norte de Entre Ríos, en base a Grandes Volúmenes de Datos de Teleobservación*". En el marco de los proyectos de equipo de trabajo se insertan también Licenciados en Biología, Licenciados en Sistemas, y alumnos de carrera de ambas carreras quienes se inician en actividades de investigación en estas líneas.

6.- Bibliografía

Baghdadi, N., & Zribi, M. (2016). *Optical Remote Sensing of Land Surface. Techniques and Methods* (1st ed.). ISTE Press - Elsevier.

Campbell, J. B., & Wynne, R. H. (2011). *Introduction to Remote Sensing* (5th ed.). Guilford Press.

Chen, R., & Xie, J. (2008). Open Source Databases and Their Spatial Extensions. In G. B. Hall & M. G. Leahy (Eds.), *Open Source Approaches in Spatial Data Handling* (pp. 105–129). Springer. https://doi.org/10.1007/978-3-540-74831-1_6

Chuvieco, E. (2015). *Teledetección Ambiental: la observación de la Tierra desde el espacio* (2nd ed.). Ariel Ciencia.

Chuvieco, E. (2016). *Fundamentals of satellite remote sensing: an environmental approach* (Second edition). Taylor & Francis.

Coppin, P. R., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, *25*(9), 1565–1596.

https://doi.org/10.1080/01431160310001016751

Dang, T. N. U. (2012). *PostGIS Raster: Extending PostgreSQL for The Support of Continuous Fields* [Master Thesis]. Université Libre de Bruxelles.

GDAL. (2020). *Geospatial Data Abstraction Library* (Version 3.0.4) [Computer software]. Open Source Geospatial Foundation. https://gdal.org

Goodchild, M. F., Guo, H., Annoni, A., Bian, L., Bie, K. de, Campbell, F., Craglia, M., Ehlers, M., Genderen, J. van, Jackson, D., Lewis, A. J., Pesaresi, M., Remetey-Fülöpp, G., Simpson, R., Skidmore, A., Wang, C., & Woodgate, P. (2012). Next-generation Digital Earth. *Proceedings of the National Academy of Sciences*, *109*(28), 11088–11094. https://doi.org/10.1073/pnas.1202383109

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, *122*, 66–74. https://doi.org/10.1016/j.rse.2011.08.024

PostGIS. (2018). *PostGIS: Homepage*. https://postgis.net/

PostgreSQL. (2018). *PostgreSQL: Homepage*. https://www.postgresql.org/about/

Python. (2020). *Python* (Version 3.8.2) [C, C++]. The Python Software Foundation. https://www.python.org/

Qiu, S., Zhu, Z., & He, B. (2019). Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sensing of Environment*, *231*, 111205. https://doi.org/10.1016/j.rse.2019.05.024

SciPy. (2020). *SciPy Ecosystem* [Python]. https://www.scipy.org/

Soille, P., Burger, A., Rodriguez, D., Syrris, V., & Vasilev, V. (2016). Towards a JRC Earth Observation Data and Processing Platform. *Proceedings of the 2016 Conference on Big Data from Space (BiDS'16)*, 4. https://doi.org/10.2788/854791

Wang, L., Ma, Y., Yan, J., Chang, V., & Zomaya, A. Y. (2016). pipsCloud: High performance cloud computing for remote sensing big data management and processing. *Future Generation Computer Systems*. https://doi.org/10.1016/j.future.2016.06.009

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S. N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P. S., Vermote, E. F., Vogelmann, J., Wulder, M. A., & Wynne, R. (2008). Free Access to Landsat Imagery. *Science*, *320*(5879), 1011–1011. https://doi.org/10.1126/science.320.5879.1011a

Zhu, Z., Wang, S., & Woodcock, C. E. (2015). Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment*, *159*, 269–277. https://doi.org/10.1016/j.rse.2014.12.014

Modelos para Aprendizaje Automático en Tiempo Real sobre Entornos de Big Data

Banchero Santiago[1], Fernandez Juan M.[1], Tonin Monzón Francisco[1], Giordano Luis A.[1]
Marrone Agustín H.[1], Lulic Maximiliano[1], Tolosa Gabriel H.[1,2]
{sbanchero, jmfernandez, ftonin, agiordano, amarrone, mlulic, tolosoft}@unlu.edu.ar
[1]Departamento de Ciencias Básicas, Universidad Nacional de Luján
[2]CIDETIC, Universidad Nacional de Luján

Resumen

En la actualidad existen incontables fuentes de información en tiempo real que provienen de redes de sensores, plataformas de observación del tiempo, mediciones de gases, observación de la tierra desde plataformas satelitales, ciudades inteligentes, entre un sin número de instrumentos que censan y transmiten datos. A su vez, hay una creciente demanda por el desarrollo de herramientas que permitan extraer conocimiento a partir de esos grandes repositorios de datos. El aprendizaje automático es un área de la inteligencia artificial, donde sus métodos contribuyen en el proceso de descubrimiento de conocimiento para la toma de decisiones inteligentes. Las demandas para la extracción de conocimiento en entornos de Big Data han acrecentado el interés por la utilización de técnicas tradicionales de aprendizaje automático en distintos problemas de repositorios masivos y entornos de flujos (o *streaming*) de datos donde muchas veces no es posible su almacenamiento, pero se requiere tomar decisiones en tiempo real.

Contexto

Este proyecto es el comienzo de una nueva línea de investigación del Departamento de Ciencias Básicas (UNLu) que tiene como principal aspiración profundizar los conocimientos sobre métodos actuales de aprendizaje de máquina como herramienta para el descubrimiento de conocimiento en problemas de Big Data sobre *streaming* de datos de diversas naturaleza.

Introducción

En la actualidad existen muchas aplicaciones que hacen un uso intensivo de datos, haciendo que el volumen y la complejidad de estos crezcan rápidamente. Los motores de búsqueda, redes sociales, e-ciencia (por ejemplo: genómica, meteorología y salud) y financieras (por ejemplo: banca y megatiendas entre otros) son algunas de sus aplicaciones [1, 15]. Esta problemática se conoce como el problema de Big Data [19].

Big Data se caracteriza principalmente por tres aspectos: (a) los datos son numerosos, (b) los datos no pueden ser categorizados en bases de datos relacionales regulares, y (c) los datos son generados, capturados y procesados muy rápidamente [13]. Si bien el volumen es y será un desafío significativo del Big Data, se debe prestar mucha atención a todas las dimensiones del problema: Volumen, Variedad y Velocidad (conocidos como las 3Vs) [6].

El concepto de *streaming* en Big Data tiene algunas características distintivas, en estos sistemas los datos se reciben como una secuencia continua, infinita, rápida, en ráfagas, impredecible y que varía en el tiempo [8]. El monitoreo (por ejemplo: tráfi-

co de red, redes de sensores, cuidado de la salud, etc.), seguimientos de *clics* en la web, transacciones financieras, detección de fraudes e intrusiones son algunas aplicaciones de *streaming* de Big Data [5]. Todos estos productores de datos que generan el *streaming* a menudo se encuentran distribuidos y con capacidades de procesamiento y memoria limitados.

En tareas de extracción de conocimiento dos pasos muy importantes son la selección de *features* (o atributos) y las tareas de mining de datos [17]. Los entornos de *streaming* de datos proponen nuevos retos con respecto a estas etapas. Una característica importante en selección de *features* es la habilidad para manejar grandes volúmenes de datos [23, 26]. Gran parte de las publicaciones existentes en la Web arriban como *streaming* (documentos, imágenes, contenidos multimedia, etc.), detectar un subconjunto de *features* útiles en estos flujos de datos en una tarea compleja debido a limitaciones de memoria, tiempos de respuesta, etc. [12, 29, 28].

Además del problema de selección de *features*, hay cuestiones enmarcadas dentro de las tareas de *stream mining* para extracción de conocimiento. La problemática aquí radica en que los patrones de datos evolucionan continuamente y se torna necesario diseñar algoritmos de minería para tener en cuenta los cambios en la estructura subyacente del *streaming* de datos [3, 2, 27]. Incluso la distribución subyacente puede cambiar en el tiempo, lo que genera que algunos modelos ya no sigan siendo válidos. Estos aspectos hacen que las soluciones de los problemas sean aún más difíciles desde un punto de vista algorítmico y computacional [7].

En este trabajo se proponen diversas líneas de investigación sobre los temas mencionados, con aplicaciones en flujos de datos y problemas reales. Se abordan tanto problemas de mejoras de rendimiento ante distintos niveles de exigencia de precisión como también la escalabilidad de las diferentes aplicaciones a datos reales.

Líneas de I+D

En este proyecto se llevan adelante lineas de I+D relacionadas principalmente con el análisis, adaptación y prueba de algoritmos de aprendizaje automático en entornos de *Streaming* de datos. Las principales líneas de trabajo en la actualidad proyecto son:

- Gestión de cache de consultas. Se busca determinar qué algoritmos de aprendizaje - en entornos de streaming - permiten lograr una mejor performance de clasificación en vías de optimizar la gestión de un cache de consultas.
- Algoritmos de clasificación multilabel. El objetivo de esta línea de trabajo es explorar los diferentes abordajes en un proceso de aprendizaje multi-etiquetas en ambientes de *streaming* de datos.
- Data Stream Clustering. Se busca desarrollar una metodología de detección de novedades o valores atípicos a través de algoritmos de agrupamiento. Analizando flujos de datos sintéticos y provenientes de redes sociales.

A continuación se hace una descripción somera de estas líneas de I+D.

a. Gestión de cache de consultas

Los árboles de decisión corresponden al aprendizaje supervisado y son ampliamente utilizados en problemas de clasificación. Estos algoritmos intentan, a partir de las instancias vistas, generar hipótesis con las cuales hacer predicciones de futuras instancias [18].

En aprendizaje sobre flujo de datos (*stream learning*), en general no es necesario computar estadísticas sobre todo el pasado, siendo suficiente con hacerlo sobre el pasado reciente [9]. Una de las formas mas clásicas y simples de mantener los ejemplos correspondientes a ese pasado es almacenar solo una ventana de instancias.

Los árboles de decisión adaptativos, o *Hoeffding Adaptive Tree* (HAT) [3], son una variante de *Hoeffding Tree* que utilizan ventanas deslizantes para

mantener ajustado el árbol, sin embargo no requiere que el usuario le especifique el tamaño de ventana a utilizar. Esto se debe a que el tamaño de ventana óptimo se calcula individualmente para cada nodo, utilizando detectores de cambios y estimadores llamados ADWIN [4].

Los resultados preliminares de aplicar HAT en el dominio de gestión de caché de consultas en motores de búsqueda han sido muy alentadores [25]. Como objetivo principal, se propuso evaluar la performance de un árbol de decisión adaptativo (HAT) para aplicar al diseño de una política de admisión para un motor de búsqueda web que recibe (y procesa) consultas en modo *streaming*. Se ha trabajado modelando la admisión como un problema de clasificación binario, intentando capturar los cambios de concepto en el tiempo, manteniendo un modelo siempre ajustado. Una vez ajustado el modelo, se integró el mecanismo de decisión como política de admisión y se evaluó la performance del caché de resultados, comparando el modelo resultante de utilizar árboles de decisión adaptativos con algoritmos de clasificación tradicionales. Siendo, de acuerdo a nuestro conocimiento, la primera vez que se propone un árbol de decisión adaptativo para la detección de términos de búsqueda frecuentes en motores de búsqueda, en los experimentos realizados, se ha observado un incremento del rendimiento del 18 % en comparación con la utilización de técnicas de clasificación tradicionales.

Los próximos pasos para esta línea de investigación están orientados a evaluar alternativas de clasificación a HAT. En este sentido se planea abordar dos alternativas, una de menor complejidad que tiene que ver con Naïve Bayes [8] y una de mayor complejidad asociada a un algoritmo de ensamble como *Random Forest* [11]. El objetivo de este trabajo es identificar posibles cotas de exactitud que puedan existir a la hora de escoger un algoritmo de aprendizaje automático sobre *stream* data que decida si una consulta es única o se repetirá en el corto plazo.

b. Algoritmos de clasificación multi-etiquetas

La clasificación multi-etiquetas es un nuevo paradigma de aprendizaje supervisado que generaliza las técnicas clásicas de clasificación para abordar problemas en donde cada instancia de una colección se encuentra asociada a múltiples etiquetas [10].

La mayor parte de los trabajos de investigación en este campo han sido realizados en contextos de aprendizaje por *batch* [8]; sin embargo, los ambientes de flujo continuo de datos (o *streaming*) presentan nuevos desafíos debido a las limitaciones de tiempo de respuesta y almacenamiento que acarrean. A esto se agrega la naturaleza evolutiva de este tipo de escenarios, que obligan a los algoritmos a adaptarse a cambios de concepto [22].

Una propuesta para esta linea de investigación sugiere aplicar algoritmos de clasificación multi-etiquetas a colecciones estructuradas y no estructuradas, combinando estos algoritmos con técnicas de procesamiento de lenguaje natural sobre la colección no estructurada. A su vez, por último, se proponen abordar estrategias de ensambles de algoritmos en búsqueda de una mejora en la calidad de la tarea de predicción de objetos no observados por el modelo.

c. Data Stream Clustering

Data Stream Clustering puede ser planteado como un problema tradicional de agrupamiento donde existen un conjunto finito (pero desconocido) de categorías - o *clusters* - que permiten describir la estructura de un conjunto de datos. El fundamento detrás de los algoritmos de *clustering* es que los objetos dentro de un *cluster* son más parecidos entre sí que los objetos que pertenecen a un *cluster* diferente [21, 16]. En este contexto, varios algoritmos de *clustering* sobre data *stream* han sido propuestos para realizar aprendizaje no supervisado.

El agrupamiento sobre un *stream* de datos requiere que un proceso sea capaz de agrupar objetos continuamente dentro de las restricciones de memoria y tiempo [8] que son las dos limitantes del problema. Los algoritmos de agrupamiento en flu-

jos de datos deben ser capaces de cumplir con los siguientes requerimientos [24, 3]:

1. Proveer oportunamente resultados mediante el procesamiento rápido e incremental de objetos;

2. adaptarse con rapidez a los cambios en la dinámica de los datos, es decir, un algoritmo debe detectar nuevos clusters que van apareciendo y otros que desaparecen;

3. escalar a la cantidad de objetos que constantemente arriban;

4. proveer un modelo de representación que no solo es compacto sino que además no crece con el número de objetos procesados;

5. capacidad para detectar *outliers*;

6. tratar con diferentes tipos de datos.

En el mundo real, pueden aparecer muchos datos atípicos o que son ruido debido a situaciones atribuidas a fallas de sensores, problemas con las conexiones, etc. Los algoritmos de agrupamiento deben ser resistentes a los valores atípicos. En un entorno de *streaming* de datos, esto es un gran desafío ya que tanto los *clusters* como los valores atípicos evolucionan con el tiempo. En otras palabras, es difícil saber ante el arribo de un nuevo dato si este se trata de un valor atípico o es el primer miembro de un nuevo grupo [30, 20, 14].

En esta línea de investigación se propone aplicar algoritmos de agrupamiento sobre flujos de datos para la detección de valores atípicos en diferentes niveles de dificultad. Inicialmente trabajará en la simulación de varios escenarios de flujos construidos de manera sintética con el objetivo de evaluar rendimiento de los algoritmos ante situaciones con variaciones en la distribución de los datos, cambios de conceptos e incorporación de ruido. Otra linea de trabajo será evaluar la detección de valores atípicos en flujos del mundo real, como por ejemplo Twitter.

Resultados y Objetivos

El objetivo principal de la propuesta es estudiar, desarrollar, aplicar, validar y transferir modelos, algoritmos y técnicas que permitan construir herramientas y/o arquitecturas para abordar algunas de las problemáticas relacionadas con el tratamiento de información masiva utilizando algoritmos de aprendizaje automático de Big Data para dar respuestas en tiempo real. Se propone profundizar sobre el estado del arte y definir, analizar y evaluar nuevos enfoques sobre aprendizaje automático a partir de *streaming* de datos. En particular se estudiarán las siguientes líneas principales:

1. Estrategias de gestión *streaming* de datos masivos para determinar las mejores herramientas para extracción de features y resolución de los problemas clásicos de ETL en el contexto del real-time.

2. Evaluar la escalabilidad de los algoritmos tradicionales del área de aprendizaje automático a problemas de respuestas en tiempo real sobre *streaming* de datos masivos en diferentes dominios.

3. Elaborar metodologías para el desarrollo de modelos en linea para toma de decisiones a partir de fuentes de información heterogénea.

Formación de Recursos Humanos

Este proyecto brinda un marco para que algunos docentes auxiliares y estudiantes lleven a cabo tareas de investigación y se desarrollen en el ámbito académico. Hasta la fecha se ha realizado un trabajo final correspondiente a la Lic. en Sistemas de Información (UNLu), se está dirigiendo uno más. Se espera dirigir al menos dos por año hasta la finalización del proyecto.

Referencias

[1] Aggarwal, C. C., Ashish, N., and Sheth, A. The internet of things: A survey from the data-centric perspective. In *Managing and mining sensor data*. Springer, 2013, pp. 383–428.

[2] Baldominos, A., Albacete, E., Saez, Y., and Isasi, P. A scalable machine learning online service for big data real-time analysis. In *Computational Intelligence in Big Data (CIBD), 2014 IEEE Symposium on* (2014), IEEE, pp. 1–8. 00017.

[3] Bifet, A. Adaptive stream mining: Pattern learning and mining from evolving data streams. In *Proceedings of the 2010 conference on adaptive stream mining: Pattern learning and mining from evolving data streams* (2010), Ios Press, pp. 1–212.

[4] Bifet, A., and Gavalda, R. Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 SIAM international conference on data mining* (2007), SIAM, pp. 443–448.

[5] Bifet, A., and Morales, G. D. F. Big data stream learning with samoa. In *Data Mining Workshop (ICDMW), 2014 IEEE International Conference on* (2014), IEEE, pp. 1199–1202.

[6] Chen, M., Mao, S., and Liu, Y. Big data: a survey. *Mobile Networks and Applications 19*, 2 (2014), 171–209. 00324.

[7] Gaber, M. M., Zaslavsky, A., and Krishnaswamy, S. Mining data streams: a review. *ACM Sigmod Record 34*, 2 (2005), 18–26.

[8] Gama, J. *Knowledge discovery from data streams*. CRC Press, 2010.

[9] Gama, J., and Gaber, M. M. *Learning from data streams: processing techniques in sensor networks*. Springer, 2007.

[10] Gibaja, E., and Ventura, S. A tutorial on multilabel learning. *ACM Computing Surveys (CSUR) 47*, 3 (2015), 52.

[11] Gomes, H. M., Bifet, A., Read, J., Bardal, J. P., Enembreck, F., Pfharinger, B., Holmes, G., and Abdessalem, T. Adaptive random forests for evolving data stream classification. *Machine Learning 106*, 9-10 (2017), 1469–1495.

[12] Huang, H., Yoo, S., and Kasiviswanathan, S. P. Unsupervised feature selection on data streams. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (2015), pp. 1031–1040.

[13] Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., Shiraz, M., and Gani, A. Big data: survey, technologies, opportunities, and challenges. *The Scientific World Journal 2014* (2014). 00028.

[14] Mansalis, S., Ntoutsi, E., Pelekis, N., and Theodoridis, Y. An evaluation of data stream clustering algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal 11*, 4 (2018), 167–187.

[15] Marz, N., and Warren, J. *Big Data: Principles and best practices of scalable realtime data systems*. Manning Publications Co., 2015.

[16] Nguyen, H.-L., Woon, Y.-K., and Ng, W.-K. A survey on data stream clustering and classification. *Knowledge and information systems 45*, 3 (2015), 535–569.

[17] Pruengkarn, R., Wong, K., and Fung, C. A review of data mining techniques and applications. *Journal of Advanced Computational Intelligence and Intelligent Informatics 21*, 1 (2017), 31–48.

[18] Quinlan, J. R. Induction of decision trees. *Machine learning 1*, 1 (1986), 81–106.

[19] Safaei, A. A. Real-time processing of streaming big data. *Real-Time Systems 53*, 1 (2017), 1–44. 00004.

[20] Shao, X., Zhang, M., and Meng, J. Data stream clustering and outlier detection algorithm based on shared nearest neighbor density. In *2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)* (2018), IEEE, pp. 279–282.

[21] Silva, J. A., Faria, E. R., Barros, R. C., Hruschka, E. R., Carvalho, A. C. d., and Gama, J. Data stream clustering: A survey. *ACM Computing Surveys (CSUR) 46*, 1 (2013), 1–31.

[22] Sousa, R., and Gama, J. Multi-label classification from high-speed data streams with adaptive model rules and random rules. *Progress in Artificial Intelligence 7*, 3 (2018), 177–187.

[23] Tang, J., Alelyani, S., and Liu, H. Feature selection for classification: A review. *Data Classification: Algorithms and Applications* (2014), 37.

[24] Tasoulis, D. K., Adams, N. M., and Hand, D. J. Unsupervised clustering in streaming data. In *Sixth IEEE International Conference on Data Mining-Workshops (ICDMW'06)* (2006), IEEE, pp. 638–642.

[25] Tonin Monzón, F., Banchero, S., and Tolosa, G. H. Árboles de decisión adaptativos en políticas de admisión a caché. In *IV Simposio Argentino de GRANdes DAtos (AGRANDA 2018)-JAIIO 47 (CABA, 2018)* (2018).

[26] Wang, J., Zhao, P., Hoi, S. C., and Jin, R. Online feature selection and its applications. *IEEE Transactions on Knowledge and Data Engineering 26*, 3 (2014), 698–710.

[27] Wang, L. Machine learning in big data. *International Journal of Advances in Applied Sciences 4*, 4 (2016), 117–123. 00048.

[28] Wu, X., Yu, K., Wang, H., and Ding, W. Online streaming feature selection. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (2010), Citeseer, pp. 1159–1166.

[29] Yin, C., Feng, L., Ma, L., Wang, J., Yin, Z., and Kim, J.-U. A feature selection algorithm of dynamic data-stream based on hoeffding inequality. In *2015 4th International Conference on Advanced Information Technology and Sensor Application (AITS)* (2015), IEEE, pp. 92–95.

[30] Yin, C., Zhang, S., Yin, Z., and Wang, J. Anomaly detection model based on data stream clustering. *Cluster Computing* (2017), 1–10.

CLUSTER PARA APRENDIZAJE Y PRÁCTICA DE *BIGDATA* Y SERVICIOS DE *LEARNING ANALYTICS*

Analia N. Herrera Cognetta, Nilda M. Pérez Otero, Francisco N. Colarich, Gustavo D. Castillo, Dalila J. Mamani, Mauro R. Patagua, Natalia E. Rodriguez, Roque E. Talavera y Diego M. Verrastro
Facultad de Ingeniería – Universidad Nacional de Jujuy – Argentina
{aniherrera012, nilperez, fcolarich, castilloguty, daly.jaquelin, mpatagua, nataliarod314, ema.tala015} @gmail.com, diegoxtr@hotmail.com

RESUMEN

Big Data es la nueva generación de almacenamiento, análisis de datos y de negocios. En el fenómeno del *Big Data* se considera, como fuentes de generación de datos a personas, *smartphones*, equipos que generan data de proyectos y experimentos, y, principalmente, a Internet. La tecnología *Big Data* permite recolectar, almacenar y preparar grandes volúmenes de datos, para analizar o visualizar la relación entre ellos, inclusive a partir de datos que se generan en tiempo real y provienen de redes sociales, sensores, dispositivos de diversa índole o fuentes de audio y video. En esta temática, el presente proyecto pretende, en una primera etapa, investigar tecnologías y arquitecturas para implementar un clúster en una plataforma web, donde estudiantes de la Facultad de Ingeniería de la UNJu, puedan aprender y practicar *Big Data*; en una segunda etapa se comenzará con una investigación en Analítica de Datos para evolucionar a *Learning Analytics*, utilizando algoritmos para medición, recopilación, análisis e informe de datos, sobre estudiantes universitarios y sus contextos, a fin de comprender y optimizar el aprendizaje y entornos en que se produce.

Palabras clave: Arquitectura escalable, *Big Data*, Plataforma web, Análisis predictivo de datos, Analítica de aprendizaje, *Learning Analytics*.

CONTEXTO

La línea de investigación aquí presentada se encuentra inserta en el proyecto **Cluster para aprendizaje y práctica de *Big Data* y servicios de *Learning Analytics***, ejecutado a partir de enero del corriente año por el grupo de la Facultad de Ingeniería de la Universidad Nacional de Jujuy. El proyecto, acreditado y financiado por la Secretaria de Ciencia y Técnica y Estudios Regionales de la Universidad Nacional de Jujuy, se encuentra bajo el Programa de Incentivos.

1. INTRODUCCIÓN

El *Big Data* es un concepto que hace referencia al tratamiento y análisis de enormes repositorios de datos, tan grandes, que resulta muy difícil y hasta imposible manejarlos con las herramientas de bases de datos y analíticas convencionales [1].

Esta tecnología se ocupa de las actividades relacionadas con los sistemas que manipulan grandes conjuntos de datos. Las dificultades vinculadas a la gestión de datos se centran en la recolección, almacenamiento, búsqueda, compartición, análisis y visualización de la información. El principal objetivo de manipular enormes cantidades de datos es, poder emplear dicha información en la creación de informes estadísticos y modelos predictivos, que pueden ser utilizados en muchas áreas del quehacer humano [2].

En el ámbito educativo, esta tecnología ha comenzado a brindar grandes beneficios. Por ejemplo, en la mejora de la gestión educativa, al desarrollo de nuevos métodos para la enseñanza y el aprendizaje, la creación de nuevas carreras y opciones profesionales, así como en la generación y almacenamiento de acervos digitales, que constituyen el producto de años de actividad académica, docente y de investigación.

Actualmente, se están gestando nuevos métodos sustentados en la tecnología para poder hacer el seguimiento de alumnos, mejorar su rendimiento, obtener datos objetivos de sus evaluaciones, predecir los riesgos académicos, o simplemente comprender el comportamiento de los grupos escolares. En este continuo cambio los profesores están más conscientes de la necesidad de actualizarse tecnológicamente, para ofrecer una educación más efectiva y adecuada a las necesidades de la población escolar actual. Es por ello que la analítica del *Big Data* se ha convertido en el recurso clave actual para entender y mejorar el proceso educativo [2].

Learning Analytics

Con el fin de comprender y optimizar el aprendizaje y los entornos en los que se produce el proceso educativo, no solo es importante conocer a los estudiantes, también es esencial analizar los datos sobre sus acciones y contextos de aprendizaje [3].

La analítica del aprendizaje o *Learning Analytics* (LA) es un término relativamente nuevo que ha ido creciendo en los últimos años por cuatro razones principales [4]:

1. **Aumento sustancial de datos**: la cantidad de datos disponibles en los contextos educativos, se obtienen a través de dispositivos digitales y tecnologías en línea, que permiten la captura y posterior análisis de la información.
2. **Formatos de datos mejorados**: en la actualidad hay formatos estandarizados para el registro de datos educativos.
3. **Avances en la computación**: la analítica es impulsada también por los avances de la computación, que permiten el análisis de grandes cantidades de datos.**Incremento de herramientas más sofisticadas para el análisis**: como es el caso de *Google Analytics* que desarrolló *MapReduce*, base importante para el proyecto *Hadoop*.

La LA puede ser de varios tipos [5]:

- **Descriptiva**: destinada a la agregación de datos y generación de información y visualizaciones relacionadas con los eventos y la interacción de los estudiantes con los cursos.
- **Predictiva**: destinada al desarrollo de modelos estadísticos y aprendizaje automático para, en función de la información disponible (cruda o tras la analítica descriptiva), obtener una visión lo más precisa posible de las potenciales situaciones a tratar en el futuro. Entre sus aplicaciones, se destaca la detección de estudiantes en riesgo de abandono y la predicción de resultados.
- **Prescriptiva**: nutrida a partir de la información proporcionada por los anteriores procesos de análisis y destinada a dar respuesta a sus resultados a través de funcionalidades como orientar a los estudiantes que están en riesgo de abandono para evitar que eso suceda.

Aspectos técnicos para el uso de *Big Data*

Una de las principales limitaciones para la adopción de la tecnología *Big Data* en cualquier institución son los recursos económicos, de infraestructura y de capital humano experimentado, que se requieren para su instrumentación. Sin embargo, están comenzando a aparecer alternativas relativamente económicas para poder realizar analítica con *Big Data,* y herramientas especializadas diseñadas para facilitar diferentes aspectos de esta tecnología. Por ejemplo, *Hadoop* es un *framework* de código abierto, diseñado para el almacenamiento y procesamiento a gran escala de conjuntos de datos en un gran número de máquinas, que permite la creación de aplicaciones para procesar grandes volúmenes de información, haciendo uso de un modelo sencillo de programación; es escalable y ofrece un buen nivel de tolerancia a fallos [6]

En este sentido, *Apache Hadoop* permite resolver varios aspectos importantes como [2]:

- Establecer el punto de contacto con la ciencia de los datos (data science) que provee
- Las técnicas necesarias para manipular y tratar la información desde un punto de vista estadístico/matemático.

- Hacer posible el procesamiento distribuido de grandes conjuntos de datos en arreglos de computadoras (clusters), utilizando modelos sencillos de programación.
- Escalar la infraestructura de cómputo, es decir, utilizar desde unos pocos servidores (nodos) y sistemas de almacenamiento de información, hasta miles de equipos, todos ellos ofreciendo una calidad idéntica de servicio.
- Permitir a las aplicaciones desarrolladas trabajar con miles de nodos y petabytes de datos, tantos como necesidades de análisis de información se requieran.
- Disponer de facilidades para realizar consultas avanzadas sobre los conjuntos de datos.
- Usar extensiones que facilitan el trabajo, manipulación y seguimiento de toda la información que maneja.
- Permite ejecutar procesos en paralelo en todo momento.

Se trata, en definitiva, de un *framework* de software de código abierto que se utiliza para almacenar, procesar y analizar grandes volúmenes de datos. Éste se orienta hacia la computación distribuida, en la cual la escalabilidad y la fiabilidad son los dos atributos más importantes. En otras palabras, *Hadoop* es el complemento perfecto de *Big Data*.

2. LINEAS DE INVESTIGACION Y DESARROLLO

El proyecto presenta dos etapas bien diferenciadas: en una primera etapa (aprendizaje) se investigará tecnologías software y arquitecturas hardware para el desarrollo y la implementación de un clúster, en una plataforma web que servirá de laboratorio a estudiantes de la Facultad de Ingeniería de la UNJu, para la práctica y aprendizaje de *Big Data*. La segunda etapa, capacitación, investigación y desarrollo en analítica de datos, se iniciará con investigación en Analítica de Datos para evolucionar a *Learning Analytics*, y analítica masiva de datos personalizados, utilizando algoritmos con herramientas orientadas por teorías del aprendizaje, técnicas pedagógicas y algoritmos para medición, recopilación, análisis e informe de datos sobre estudiantes universitarios y sus contextos, a fin de comprender y optimizar el aprendizaje y entornos en que se produce.

3. RESULTADOS OBTENIDOS/ESPERADOS

Es una realidad hoy la necesidad de que los estudiantes de las carreras vinculadas con las TICs, adquieran los conocimientos y capacidades sobre las últimas tecnologías, en la Institución Universitaria donde cursan. El contar con una herramienta que pudiera reforzar el conocimiento y apoyar el desarrollo de las habilidades prácticas en *Big Data*, será de gran utilidad tanto para estudiantes como para docentes.

Este proyecto ambiciona obtener un cluster que sirva como laboratorio para estudiantes de la Facultad de Ingeniería de la UNJu, y que a través de un proceso de madurez sea posible implementar estudios de analítica de datos para contribuir con la formación a través de *Learning Analytics*.

Por ello, el objetivo del proyecto consiste en lograr, a través de la investigación de nuevas tecnologías, diseñar, desarrollar e implementar un *cluster* para estudio y práctica de *Big Data*, para estudiantes de grado, alcanzando un nivel de madurez suficiente para el estudio y aplicación de *Learning Analytics*, siendo los objetivos particulares:

- Investigar y evaluar las herramientas del ecosistema Hadoop, para diseño y desarrollo.
- Implementar el entorno de trabajo que cumpla con los requerimientos del diseño.
- Diseñar y realizar las pruebas de usabilidad del clúster.
- Capacitación del equipo en Analítica de Datos.
- Brindar servicios de *Learning Analytics* para la toma de decisión.

Al tratarse de un proyecto que recién inicia (enero de 2020 a diciembre de 2021), no se cuenta aún con resultados obtenidos.

4. FORMACION DE RECURSOS HUMANOS

El grupo de investigación está formado por dos docentes, dos egresados y cinco estudiantes de las carreras de Licenciatura en Sistemas e Ingeniería Informática, todos de la Facultad de Ingeniería de la Universidad Nacional de Jujuy. Los integrantes del grupo fueron elegidos según sus especialidades y perfiles, un egresado tiene un Mg en Innovación Educativa por Competencias, quien aportará su conocimiento en esta temática para Learning Analytics, otro egresado trabaja actualmente en una empresa de desarrollo de software y cada estudiante se destaca por su compromiso y cumplimiento con respecto a la planificación propuesta, y en la profundidad y nivel académico demostrado en la investigación preliminar. Si bien las docentes cuentan con trayectoria en investigación, para los egresados y los estudiantes, es su primera experiencia, lo que representa un desafío y una oportunidad, desafío para las docentes de dirigir y organizar las actividades y el aprendizaje en investigación, para que resulte una vivencia enriquecedora, y que a la vez incentive a los jóvenes a seguir esta línea de trabajo, Una oportunidad para ellos de adquirir hábitos y metodologías para la investigación aplicada, que puede llegar a convertirse en su quehacer cotidiano, o la expertiz lograda, colabore en su desempeño profesional. Las carreras que estos estudiantes eligieron, tienen como característica fundamental, la evolución continua de métodos, herramientas, equipos y desarrollos, sobre los que deben tener conocimiento y manejo para mantenerse activos en el mercado laboral.

5. REFERENCIAS

[1] DANS, Enrique, "Big Data: una pequeña introducción", 2011. En línea: < https://www.enriquedans.com/2011/10/big-data-una-pequena-introduccion.html >, consultado en marzo de 2020.

[2] SALAZAR ARGONZA, Javier, "Big Data en la educación", Revista Digital Universitaria, 1 de enero de 2016, Vol. 17, Núm. 1, ISSN: 1607-6079 Disponible en Internet: <http://www.revista.unam.mx/vol.17/num1/art06/index.html>

[3] LANG, Charles, *et al.* (ed.). *Handbook of learning analytics*. SOLAR, Society for Learning Analytics and Research, 2017.

[4] SIEMENS, George; BAKER, Ryan SJ d. Learning analytics and educational data mining: towards communication and collaboration. En *Proceedings of the 2nd international conference on learning analytics and knowledge*. 2012. p. 252-254.

[5] ROLDÁN, Xavier Alamán, *et al*. GHIA: Modelado de Estudiantes, Analítica de Aprendizaje, Atención a la Diversidad y e-Learning. *IE Comunicaciones: Revista Iberoamericana de Informática Educativa*, 2019, n° 30, p. 78-89.

[6] ZENTENO, José Antonio Carrillo. Big Data-Analítica del aprendizaje y minería de datos aplicados en la Universidad. *Pro Sciences*, 2018, vol. 2, no 8, p. 39-54.

Contribuciones a las Bases de Datos Métricas

J. Arroyuelo, Maria E. Di Genaro, A. Grosso, V. Ludueña, C. Martínez, N. Reyes
Dpto. de Informática, Fac. de Cs. Físico-Matemáticas y Naturales, Universidad Nacional de San Luis
{*bjarroyu, digeme, agrosso, vlud, nreyes*}@unsl.edu.ar, *cintiavmartinez*@hotmail.com
Edgar Chávez
Centro de Investigación Científica y de Educación Superior de Ensenada, México
elchavez@cicese.mx
Karina Figueroa
Fac. de Cs. Físico-Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo, México
karina@fismat.umich.mx
Rodrigo Paredes
Dpto. de Cs. de la Computación, Fac. de Ingeniería, Universidad de Talca, Chile
raparede@utalca.cl

Resumen

Claramente, los nuevos modelos de bases de datos, capaces de contener y manejar todo tipo de datos no estructurados: imágenes, videos, música, secuencias biológicas, etc., no tienen la madurez y versatilidad que presentan las bases de datos convencionales. Estas nuevas bases de datos deben ser capaces de adaptarse al gran volumen de datos digitales, que son generados constantemente por fuentes muy disímiles; al igual que al tipo de requerimientos al que son sometidas, que pueden ser tan dispares como el tipo de datos administrados, debido que éstos pertenecen a campos muy diferentes.

Por esto, se hace necesario optimizar estos depósitos especializados, o desarrollar nuevos, y utilizar formas más sofisticadas de búsqueda sobre los mismos, que permitan enfrentar tales requerimientos. La administración del espacio disponible también se vuelve crucial debido a la gran cantidad de datos que se debe manipular para lograr respuestas adecuadas y eficientes. Esto obliga a los índices utilizados para acceder a este tipo de base de datos, a ser *conscientes de la jerarquía de memoria*.

Esta investigación pretende contribuir a la madurez de este nuevo modelo de bases de datos considerando distintas perspectivas. Para ello utiliza un modelo en el cual se puede utilizar métodos de acceso que contemplen estos aspectos, y que se adapta a tales requerimientos: las *Bases de Datos Métricas*.

Palabras Claves: bases de datos métricas, índices, búsquedas por proximidad.

Contexto

El presente trabajo se realizó en el marco de la línea *Bases de Datos no Convencionales*, del Proyecto Consolidado *Tecnologías Avanzadas de Bases de Datos*, (Cód. 03-2218 y en Programa de Incentivos 22-F814) de la Universidad Nacional de San Luis. En colaboración con investigadores de otros grupos de: Universidad de Talca (Chile), Universidad Michoacana de San Nicolás de Hidalgo y Centro de Investigación Científica y de Educación Superior de Ensenada (México).

La investigación que se realiza en este ámbito, está enfocada en lograr la consolidación de las Bases de Datos Métricas. Se espera contribuir a estos sistemas obteniendo índices que resulten más eficientes para memorias jerárquicas, dinámicos, con E/S eficiente y escalables (capaces de manejar grandes volúmenes de datos). Esto incluye además, plantear nuevas arquitecturas del procesador que mejoren a muy bajo nivel los administradores de estas bases de datos. Se espera contribuir en diferentes campos de aplicación: sistemas de información geográfica, robótica, visión artificial, diseño asistido por computadora, computación móvil, entre otros.

Introducción

El uso masivo de internet y la disponibilidad de dispositivos electrónicos en diversos ámbitos, como el productivo, artístico, laboral, recreativo, científico, de la salud, etc., ha generado una significativa aceleración tanto en el crecimiento del volumen de datos generados y almacenados, como la variedad de tipos de datos que aparecen. Este escenario ha exigi-

do que las bases de datos sean capaces de adaptarse tanto a los diferentes entornos, como a la gran variedad de usuarios de las mismas. Para ello deben administrar eficientemente todo tipo de datos (no estructurados) y responder consultas sobre los mismos de una manera totalmente diferente a la tradicional. Si se necesita encontrar las huellas digitales más similares a una dada, las búsquedas tradicionales (exactas) carecen de sentido. En la mayoría de estos casos, sobre estos tipos de datos, las *búsquedas por similitud* resultan más adecuadas, que las tradicionales.

El modelo habitual para las búsquedas por similitud es el de *espacios métricos*, y a pesar de la variedad de estas aplicaciones, todas comparten ciertas características que permiten la utilización de este modelo como su marco formal. Se define un espacio métrico como un universo de objetos $\mathbb{U}$ y una función de distancia definida entre ellos, $d : \mathbb{U} \times \mathbb{U} \longmapsto \mathbb{R}^+$, que mide cuán diferentes son estos objetos. Resolver este tipo de búsquedas puede ser tan sencillo como realizar una examinación secuencial del conjunto de datos, pero hacerlo eficientemente hace necesario el uso de los llamados *Métodos de Acceso Métricos* (MAMs). Sin embargo, debido a la diversidad de ámbitos en los que se aplica el modelo, es esencial la actualización y optimización de los MAMs, permitiendo su mejor adaptación a cada caso, además de la solución de problemas como la posibilidad de admitir actualizaciones (inserciones/eliminaciones), el soporte de conjuntos masivos de datos y la resolución de búsquedas complejas. Estos avances se reflejan en áreas como: reconocimiento de voz, reconocimiento facial, bases de datos médicas, minería de datos, biología computacional, entre otros.

Otro enfoque analizado es el desempeño de los administradores de bases de datos (DBMS) a bajo nivel. En ese sentido se está intentando caracterizar nuevas arquitecturas que permitan reducir el flujo de bits entre el procesador y la memoria, en relación a la cantidad de datos utilizados por cada programa, para mejorar el desempeño de los mismos.

Líneas de Investigación y Desarrollo

Arquitecturas de Procesadores Orientadas a Bases de Datos

Según algunos autores, se puede distinguir entre arquitectura, implementación y realización. Conforme a esta distinción, el conjunto mínimo de propiedades que determinan qué programas correrán y qué resultados producirán sobre el procesador, es lo que se denomina la arquitectura de una computadora. Es decir, es la interfaz entre el software y el hardware. La implementación está conformada por organización básica del flujo de datos y el control. Por último, la estructura física que comprende la implementación, conforma la realización [1].

Actualmente, la investigación sobre la implementación de procesadores ha reemplazado la investigación sobre arquitecturas de procesadores. La mayoría de los trabajos se ha enfocado en mejorar técnicas de sincronización y comunicación de procesadores (núcleos) a través de mensajes y/o memoria compartida, al igual que técnicas de predicción (tanto de control como de datos). Muchas de estas técnicas, surgidas en los años 60, que se han incorporado a los diseños de nuevos microprocesadores, se pueden aplicar a todo tipo de arquitecturas; tanto a una arquitectura RISC[1] (que intenta acercar el lenguaje de máquina al hardware del procesador), como a una arquitectura que se aleje del hardware e intente disminuir el tráfico de bits entre procesador y memoria. Si bien las arquitecturas RISC compitieron en desempeño con las arquitecturas CISC [2], las mismas poseen un alto tráfico de bits entre el procesador y la memoria para una determinada traza de ejecución. Esto finalmente favoreció a las CISC sobre las RISC, una vez que las CISC mejoraron sus técnicas de implementación.

Con el objetivo el plantear nuevas arquitecturas, que minimicen el tráfico de bits entre el procesador y la memoria, se está construyendo un simulador del set de instrucción AMD-64 o x86-64. Esto permitirá evaluar el tráfico de bits para benchmarks como Specint y Specfp para la arquitectura x86. A continuación, se evaluará el tráfico de bits para la arquitectura propuesta sobre los mismos benchmarcks, lo que implica construir no sólo el simulador de la arquitectura sino también el compilador C para la misma. Finalmente, se pretende aprovechar el conocimiento adquirido para, desde bajo nivel, mejorar el desempeño de los DBMSs.

Bases de Datos Métricas

Como se mencionó, se utilizarán los espacios métricos para modelizar aquellas bases de datos ca-

[1] acrónimo del inglés “Reduced Instruction Set Computer”.
[2] acrónimo del inglés “Complex Instruction Set Computer”.

paces de gestionar imágenes, videos, texto libre, secuencias de ADN o de proteínas, audio, etc. Por lo que se hará uso de MAMs a fin de responder eficientemente consultas por similitud sobre las mismas. Debido a lo costoso que resultan los cálculos de distancia, el número de cálculos realizados al crear el índice o al realizar búsquedas, es usado como medida de complejidad. Por ello, el objetivo aquí es optimizar los MAMs, analizando aquellos que han mostrado buen desempeño en las búsquedas para reducir su complejidad considerando, cuando sea necesario, la jerarquía de memorias. En general, dada una base de datos $X \subseteq \mathbb{U}$ y una consulta $q \in \mathbb{U}$, las consultas por similitud son de dos tipos: por rango o de *k-vecinos más cercanos* (k-NN).

Grafo de los k Vecinos

Entre las consultas por similitud en espacios métricos, una que resulta muy útil es la que obtiene los k-vecinos más cercanos de *todos* los elementos de la base de datos (*All-k-*NN). Esta consulta relaciona cada elemento $u \in X$, con los k objetos en X- $\{u\}$ que tengan la menor distancia a él. La forma ingenua de resolverlo es comparar cada objeto en la base de datos con todos los demás y devolver los k más cercanos a él. Esta solución tiene una complejidad de n^2 cálculos de distancia ($|X| = n$). Una solución más eficiente es preprocesar la base de datos construyendo un índice y luego buscando en el mismo los k-NN de cada elemento del conjunto.

Sin embargo, existen situaciones en las cuales el costo de la construcción del índice, para luego realizar n consultas del tipo k-NN, puede resultar excesivo. Este es el caso de una base de datos masiva, o cuando la función de distancia es demasiado costosa de calcular, o si se está trabajando con espacios métricos de alta dimensión. Estos casos pueden requerir revisar la base de datos completa, a pesar de la estrategia utilizada. Otro factor a considerar son los requerimientos de algunas aplicaciones particulares, que priorizan la velocidad de respuesta sobre la precisión de la misma [13, 7, 14, 8]. Para hacer frente a éstas circunstancias es que se han considerado las llamadas *búsquedas por similitud aproximadas*. Este tipo de consultas mejoran su complejidad aceptando algunos "errores" en la respuesta.

Sabemos que resolver el problema *All-k-*NN permite construir el *Grafo de los k-vecinos más cercanos* (kNNG)[12]. Dada una colección de objetos de un espacio métrico, el grafo de k vecinos más cercanos asocia cada nodo a sus k vecinos más cercanos. El kNNG resulta ser un índice eficiente, que admite mejoras y permite resolver búsquedas por similitud. Por ello hemos propuesto nuevas técnicas para resolver el problema de *All-k-*NN, que *no utiliza ningún índice* para buscar en él, y que permiten computar una aproximación del kNNG. Éstas conectan cada objeto u de la base de datos con k vecinos *cercanos*, relajando la condición que exige que no haya, en toda la base de datos, algún objeto más cercano a u que los k vecinos devueltos. Esto puede ocasionar que se pierda algún objeto muy cercano y en su lugar se devuelva otro un poco más lejano, pero a cambio la respuesta será más rápida. A este grafo se lo denominó *Grafo de vecinos cercanos* (knNG) [5].

Una primera aproximación aprovecha el profundo conocimiento que se tiene del *DiSAT* para plantear un enfoque novedoso. Aquí se consideró un caso particular del problema (k = 1) obteniendo el 1nNG. Esta propuesta utiliza la información obtenida durante la *construcción* del *DiSAT* para construir el 1nNG, conectando a cada elemento con un elemento cercano de la base de datos, que puede ser, o no, su vecino más cercano [5]. Esta propuesta permite recuperar el 1nNG con bajo costo, muy buena precisión y un error bajo, logrando un buen compromiso calidad/tiempo, y *sin realizar ninguna búsqueda.*

Las otras propuestas abordadas se enfocan en responder a los *All-k-*nN y computar el knNG. Estos planteos no utilizan el apoyo de ningún índice, no sólo no buscan en ellos, sino que ni siquiera recurren a la información provista por su construcción. El propósito de estos desarrollos es aprovechar de manera ingeniosa las propiedades de la *función de distancia*. En ellos se sugieren distintas maneras de seleccionar muestras de la base de datos, a partir de las cuales se obtiene un conjunto de distancias que serán el punto de partida de este proceso; analizando diferentes maneras de utilizar la información. En algunos casos se calculan los vecinos exactos [4] y en otros los aproximados, para todos los objetos de la base de datos, utilizando propiedades como la simetría o la desigualdad triangular. Los resultados de estas propuesta se muestran muy prometedores.

Métodos de Acceso Métricos

Muchas veces, debido al tamaño de los objetos almacenados en una base de datos, o su gran cantidad, los índices no caben en memoria principal. Entonces surge la necesidad de diseñar índices que

se almacenan en memoria secundaria. Muchos de estos índices se basan en "agrupar elementos". Teniendo esto en consideración, se han diseñado dos nuevos índices basados en la *Lista de Clusters*(*LC*) [7] que son totalmente dinámicos, es decir, admiten inserciones y eliminaciones de objetos y están especialmente diseñados para trabajar sobre grandes volúmenes de datos [11]. La *Lista de Clusters Dinámica* (*DLC*), tiene buen desempeño en espacios de alta dimensión, una buena ocupación de página y operaciones eficientes tanto en cálculos de distancia como en operaciones de I/O. Sin embargo, durante las búsquedas se debe recorrer completamente la lista de centros de los clusters, elevando los costos. El *Conjunto Dinámico de Clusters* (*DSC*), también mantiene los clusters en memoria secundaria, pero organiza los centros de clusters en un *DSAT* en memoria principal, permitiendo que las búsquedas realicen menos cálculos de distancia y accedan a menos páginas/clusters. Durante las inserciones, también se aprovecha la información de ese *DSAT* también, mejorando los costos en cálculos de distancia y manteniendo los costos de acceso a disco bajos. Ambos, *DLC* y *DSC*, han demostrado tener una razonable utilización de páginas de disco y son competitivas respecto a las alternativas representativas del estado del arte.

La calidad de los clusters generados es otro aspecto a tener en cuenta. El mismo se considera en una variante de la *DSC*, que en lugar de insertar los elementos en el índice a medida que van llegando, demora la incorporación de cada nuevo elemento a un cluster hasta tener varios elementos y poder determinar así un mejor agrupamiento de los objetos. Esto permite reducir el costo de construcción del índice, porque se realiza una escritura de un cluster en disco luego de varias inserciones y además implícitamente puede mejorar los costos de búsqueda al lograr clusters más compactos y que aseguran una total ocupación de la página del disco, achicando el tamaño del archivo y reduciendo los tiempos de acceso.

El dinamismo es otra característica necesaria en los MAM's. Esta necesidad provocó el desarrollo del *árbol de Aproximación Espacial Dinámico (DSAT)* [11] que permite realizar inserciones y eliminaciones. Está basado en el *árbol de Aproximación Espacial* (*SAT*) que, a pesar de ser uno de los índices de mejor desempeño en espacios de mediana a alta dimensión, es totalmente estático. El *DSAT* conserva el muy buen desempeño en las búsquedas, pero agrega un parámetro a sintonizar. Otra variante del *SAT*, el *árbol de Aproximación Espacial Distal* (*DiSAT*) [6], a pesar de ser estática, logra optimizar las búsquedas respecto de ambos (SAT y DSAT) y no necesita parámetros. Por ello, se ha propuesto la *Foresta de Aproximación Espacial Distal* (*DiSAF*) [3], basada en él pero que es dinámica. Es para memoria principal y aplica la técnica de dinamización de Bentley y Saxe al *DiSAT*, aprovechando el profundo conocimiento que se tiene sobre la aproximación espacial para mejorar al máximo su desempeño.

Otra faceta que hay que tener en cuenta, son los requerimientos de algunas aplicaciones que priorizan la rapidez en las respuestas aunque sea a costa de perder algunos elementos: se intercambia precisión (devolviendo sólo algunos objetos relevantes) por velocidad en la respuesta. Este tipo de búsquedas se denominan *aproximadas*. Para conjuntos de datos masivos, las búsquedas por similitud aproximadas permiten obtener un buen balance entre el costo de las búsquedas y la calidad de la respuesta obtenida. El *Algoritmo Basado en Permutaciones* (*PBA*) [2]. es uno de los mejores representantes de este tipo de consultas, logrando una respuesta de alta calidad a un bajo costo. Por esta razón, se ha utilizado como base del diseño de la *Lista Dinámica de Permutaciones Agrupadas* (*DLCP*) [9], que es dinámica y para memoria secundaria. Este índice, que combina *LC* con *PBA*, agrupa por distancia entre las permutaciones de los objetos, en lugar de por distancia entre objetos y se le puede indicar cuántos cálculos de distancia y/o operaciones de I/O utilizar, para obtener una respuesta rápida, aunque menos precisa. Además, se están considerando nuevas variantes para obtener mejores resultados.

Resultados y Objetivos

Las investigaciones realizadas sobre el modelo de espacios métricos, han permitido mejorar el desempeño de los MAMs estudiados, y los resultados obtenidos conducen a intentar aplicarlos a otros métodos de acceso [4, 5, 10, 6, 11, 3].

Se espera brindar nuevas herramientas eficientes de administración para bases de datos métricas, que logren acercar su desarrollo al de los modelos tradicionales de base de datos. Para ello, se buscará profundizar en el estudio de nuevos diseños de estructuras de datos, buscando incrementar su eficien-

cia en espacio y en tiempo: que se adapten mejor al nivel de la jerarquía de memorias donde se almacenarán y a las características de los datos a ser indexados. modelos tradicionales de base de datos. Se continuará estudiando diferentes técnicas que sin utilizar de índices, permitan resolver consultas eficientemente. Además, se espera mejorar el desempeño de las operaciones de bajo nivel en los DBMS, mediante una nueva arquitectura del procesador.

Actividades de Formación

Dentro de esta línea de investigación se forman alumnos y docentes-investigadores en:
Doctorado en Cs. de la Computación: una tesis sobre expresividad de lenguajes lógicos de consulta.
Maestría en Cs. de la Computación: una tesis sobre búsqueda por similitud aproximada (concluida) y otra sobre un índice dinámico eficiente.
Maestría en Informática: una tesis, de la Universidad Nacional de San Juan, sobre un índice dinámico para búsquedas aproximadas en disco (concluida).

Referencias

[1] G. Blaauw and F. Brooks, Jr. *Computer Architecture: Concepts and Evolution*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1997.

[2] E. Chávez, K. Figueroa, and G. Navarro. Effective proximity retrieval by ordering permutations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(9):1647–1658, Sept 2008.

[3] E. Chávez, M. Di Genaro, N. Reyes, and P. Roggero. Decomposability of disat for index dynamization. *Computer Science & Technology*, pages 110–116, 2017.

[4] E. Chávez, V. Ludueña, and N. Reyes. Solving all-k-nearest neighbor problem without an index. In *Procs. del XXV Congreso Argentino de Ciencias de la Computación CACIC 2019)*, pages 567–576. UniRío editora, 2019.

[5] E. Chávez, V. Ludueña, N. Reyes, and F. Kasián. All near neighbor graph without searching. *Computer Science & Technology*, 18:61–67, 2018.

[6] E. Chávez, V. Ludueña, N. Reyes, and P. Roggero. Faster proximity searching with the distal $\{$SAT$\}$. *Information Systems*, 59:15 – 47, 2016.

[7] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.

[8] P. Ciaccia and M. Patella. Approximate and probabilistic methods. *SIGSPATIAL Special*, 2(2):16–19, 2010.

[9] K. Figueroa, C. Martínez, R. Paredes, N. Reyes, and P. Roggero. Dynamic list of clustered permutations on disk. In *Computer Science and Technology Series: XXI Argentine Congress of Computer Science Selected Papers*, pages 201–211. EDULP, 2016.

[10] A. Camarena-Ibarrola L. Valero-Elizondo K. Figueroa, N. Reyes. Improving the list of clustered permutation on metric spaces for similarity searching on secondary memory. In *10th Mexican Conference on Pattern Recognition (MCPR2018)*, volume 10880, pages 82–92, 2018.

[11] G. Navarro and N. Reyes. New dynamic metric indices for secondary memory. *Information Systems*, 59:48 – 78, 2016.

[12] R. Paredes, E. Chávez, K. Figueroa, and G. Navarro. Practical construction of k-nearest neighbor graphs in metric spaces. In *Proc. 5th Workshop on Efficient and Experimental Algorithms (WEA)*, LNCS 4007, pages 85–97, 2006.

[13] H. Samet. *Foundations of Multidimensional and Metric Data Structures (The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2006.

[14] P. Zezula, G. Amato, V. Dohnal, and M. Batko. *Similarity Search: The Metric Space Approach (Advances in Database Systems)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. XVIII, 220 p., Hardcover ISBN: 0-387-29146-6.

Bases de Datos Espaciales y Espacio Temporales

Edilma Olinda Gagliardi, Maria Gisela Dorzán, Maria Teresa Taranilla,

Pablo Rafael Palmero y Carlos Andrés Casanova

Departamento de Informática
Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis, Argentina
{oli, mgdorzan, tarani, prpalmero, cacasanova}@unsl.edu.ar

RESUMEN

La línea de investigación *Bases de Datos Espaciales y Espacio Temporales*, se encuentra enmarcada dentro del Proyecto Consolidado *Tecnologías Avanzadas de Bases de Datos* (PROICO 03-2218) de la Universidad Nacional de San Luis, y está incluido en el Programa de Incentivos (Código 22/F814).

Esta línea de trabajo se orienta a vincular las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas. Como objetivo general, se plantea utilizar métodos, herramientas y técnicas para investigación de base, acompañando el desarrollo de aplicaciones para contribuir en la resolución de problemas provenientes de diferentes dominios de aplicación.

Palabras clave: Bases de Datos, Geometría Computacional, Metaheurísticas, Bases de Datos Espaciales y Espacio Temporales, AgroTIC.

CONTEXTO

El proyecto Tecnologías Avanzadas de Bases de Datos abarca el tratamiento de objetos, estructurados y no estructurados, de utilidad en diversos campos de aplicación: sistemas de información geográfica, motores de búsqueda en internet, computación gráfica, robótica, computación móvil, visión artificial, diseño asistido por computadora, entre otras.

El proyecto se conforma con tres líneas de investigación, orientadas al desarrollo de nuevos modelos para administrar y recuperar información almacenada en bases de datos no estructuradas, que requieren modelos no tradicionales tales como las bases de datos de texto, bases de datos espaciales, espacio temporales, bases de datos de imágenes, bases de datos de sonidos, espacios métricos, entre otros.

Los modelos de bases de datos espaciales y espacio temporales se utilizan en aplicaciones en las cuales se necesita guardar y consultar información actual e histórica de posiciones referenciadas espacialmente y cambios de forma que tuvieron los objetos de estudio en diferentes escenarios a lo largo del tiempo. En diferentes contextos, surge la necesidad de la incorporar métodos científicos, procesos y sistemas de descubrimiento de información de datos estructurados y no estructurados, para obtener información para asistir en la toma de decisiones en un sistema.

En este contexto, se vinculan las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas con el fin de utilizar métodos y herramientas que aporten en la resolución de problemas de diversos dominios de aplicación y para la resolución de problemas orientados a optimización.

El trabajo de investigación de esta línea se realiza en colaboración con investigadores de

grupos de la Universidad Nacional de San Luis, convenios entre organizaciones nacionales y provinciales con presencia en San Luis, y de universidades extranjeras mediante convenios de cooperación interinstitucional.

1. INTRODUCCIÓN

Las Tecnologías de la Información y la Comunicación (TIC) introducen mejoras en diversas actividades humanas y así producen cambios sociales, económicos y culturales, convirtiéndose en procesos clave para accionar sobre el presente y proyectar el futuro. La utilidad de las mismas en diferentes situaciones del mundo real conlleva a una mejora en temáticas referidas a calidad, gestión, economía, entre otras. En este sentido, los integrantes de la línea integraron el proyecto *Campo Conectado*, resultado de un convenio entre el Instituto Nacional de Tecnología Agropecuaria (INTA-Estación Experimental Agropecuaria San Luis) y la Universidad Nacional de San Luis con el objetivo de estudiar la potencialidad y alcance real de las TIC en el ámbito de la producción agropecuaria [1].

Otro eje de trabajo, que define objetivos de investigación está relacionado con problemas de optimización combinatoria, para los cuales no se puede garantizar encontrar la mejor solución en un tiempo razonable, para todas las instancias del problema. La optimización procura encontrar la mejor solución posible a un problema de este tipo dentro de un período de tiempo limitado. En Geometría Computacional, la optimización de configuraciones geométricas respecto de ciertos criterios de calidad, pertenecen a esta clase de problemas y pueden resolverse utilizando métodos de aproximación, tales como las técnicas metaheurísticas [2].

Una metaheurística es un proceso de generación iterativo que guía la búsqueda de soluciones combinando inteligentemente diferentes conceptos de campos diversos como inteligencia artificial, evolución biológica, inteligencia colectiva, sistemas inmunes, entre otros [10].

Algunos de los problemas de optimización estudiados son la Triangulación de Peso Mínimo (Minimum Weight Triangulation, MWT) y la Pseudo-Triangulación de Peso Mínimo (Minimum Weight Pseudo-Triangulation, MWPT), problemas de carácter NP-duro [8, 11]. La Triangulación de Dilación Mínima (Minimum Dilation Triangulation, MDT) donde la dilación mide la calidad de conexión entre puntos de la triangulación es otro problema estudiado. Para este problema no se conoce un algoritmo que lo resuelva en tiempo polinomial y tampoco se ha demostrado que sea NP-duro [9].

La utilización de configuraciones geométricas optimizadas respecto de algún criterio de calidad, resulta como soporte de estrategias en la resolución de problemas vinculados con bases de datos espaciales y espacio temporales.

2. LÍNEA DE INVESTIGACIÓN

En la línea de investigación se trata con varios dominios de aplicación de bases de datos espaciales y espacio temporales, utilizando técnicas y herramientas de apoyo en la resolución de problemas.

Entre los tópicos de estudio, se destacan:

- Diseño y aplicación de índices espacio temporales, en diversos escenarios de movimiento.
- Optimización de estructuras geométricas aplicando técnicas metaheurísticas en la optimización de problemas de complejidad NP.
- Aplicación de la Geometría Computacional y su marco disciplinar en aspectos propios de los problemas estudiados.
- Desarrollo de herramientas para la visualización de información y aplicaciones vinculadas con bases de datos espacio temporales.

Como objetivos específicos se propone:

- Desarrollar aplicaciones de Bases de Datos Espaciales y Espacio Temporales, aplicando herramientas provistas por la Geometría Computacional.
- Estudio de la indexación espacio temporal sobre objetos en movimiento para diversos escenarios. Desarrollo de las estructuras de almacenamiento, algoritmos de consulta y evaluación experimental.
- Estudio de configuraciones geométricas de puntos en el plano considerando medidas de calidad, mediante la aplicación de técnicas metaheurísticas y diversas estrategias algorítmicas.
- Desarrollar herramientas para la visualización vinculadas a las bases de datos mencionadas.

3. RESULTADOS OBTENIDOS

En relación al desarrollo de herramientas para la visualización, vinculadas a las bases de datos espaciales, se finalizó una aplicación para la difusión de material fotográfico de valor histórico preservado en la Fototeca de la Universidad Nacional de San Luis. La aplicación enlaza el pasado con el presente, divulgando el legado del fotógrafo puntano José La Vía, permitiendo redescubrir la ciudad de San Luis, a través de su historia fotográfica. Este desarrollo para dispositivos móviles con sistema operativo Android utiliza Realidad Aumentada para lograr una experiencia atractiva al transitar por la ciudad, poniendo en valor al fondo fotográfico custodiado por la Fototeca.

Con respecto a problemas de optimización abordados con técnicas metaheurísticas, se trabaja en el problema MDT para el cual, se utiliza el modelo de Programación Evolutiva introducido por L. J. Fogel [4], para desarrollar nuevos operadores y obtener resultados comparables con los obtenidos en trabajos previos. Este modelo se basa en el comportamiento adaptativo de las nuevas generaciones más que en la modificación de cromosomas de cada individuo. La Programación Evolutiva utiliza sólo dos operadores: mutación y selección, es decir que no se realiza recombinación. Los resultados serán contrastados con los obtenidos con los algoritmos: Greedy, Local Search (LS), Iterated Local Search (ILS), Simulated Annealing (SA), y Random Local Search (RLS) presentados en [3].

En el marco del proyecto Campo Conectado, se trabaja en desarrollos tecnológicos aplicados en la gestión de la producción agropecuaria para sistemas reales de producción.

En este contexto, se desarrolla una plataforma integral, soporte para diversos eventos y sistemas de información con dominio de aplicación en el sector agropecuario, mediante el uso de las TIC. Sus características principales están orientadas a la recolección de datos de diferentes fuentes con almacenamiento compartido, integración progresiva de diversas funcionalidades, explotación y visualización de la información [5].

En el marco de esta plataforma integral se está implementando una herramienta para el seguimiento espacio temporal de rodeos en establecimientos agropecuarios, mediante el uso de las TIC y de las Bases de Datos Espacio Temporales, apoyadas con herramientas de Geometría Computacional. La herramienta realiza la comunicación con las bases de datos y las consultas espaciales y espacio temporales de los objetos de tratamiento, así como la visualización del geoseguimiento disponible por medio de tecnología móvil [7].

Adicionalmente, se trabaja en el desarrollo de una herramienta de apoyo en la gestión de rodeos de cría, que permite la recolección de datos desde diferentes fuentes para la inferencia y el seguimiento de la condición corporal de los individuos del rodeo. La herramienta provee funcionalidades para el análisis de datos, generación de indicadores, obtención de conocimiento de apoyo para la toma decisiones y visualización de la información disponible por medio de tecnología móvil [6].

Estas herramientas estarán incluidas en la plataforma integral para permitir realizar un

seguimiento del rodeo y generar indicadores que aporten información relevante para el proceso de toma de decisiones. Dicha plataforma estará disponible en la web, con accesibilidad mediante tecnología móvil (I+D+i).

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo en la Universidad Nacional de San Luis, consolida su formación con actividades de cooperación mutua e intercambio recíproco de información científica, tecnología y desarrollo de nuevos conocimientos con investigadores locales y de otras universidades.

Entre las actividades más destacadas, se mencionan:

i) Formación de recursos humanos reflejada en tesis doctorales, tesis de maestría y Licenciados en Ciencias de la Computación.

ii) Actividades de formación académica, a través del dictado y realización de cursos de posgrado y de especialización.

iii) Realización de jornadas de investigación con docentes de otras universidades.

iv) Actividades de divulgación científica, conferencias y publicaciones en congresos.

La línea tiene como un objetivo continuar con las actividades relacionadas al presente proyecto, proponiendo nuevas actividades de formación académica, de formación de recursos humanos, investigación, desarrollo, y otras actividades académico-científicas vinculantes.

5. BIBLIOGRAFÍA

[1] *Campo Conectado*, proyecto interinstitucional. http://inta.gob.ar/noticias/campo-conectado-un-nuevo-proyecto-interinstitucional (2017).

[2] de Berg, M., Cheong O., van Kreveld, M., Overmars, M.,. *Computational Geometry: Algorithms and Applications.*, Springer-Verlag, Heidelberg (2008).

[3] Dorzán M. G. et al. *Approximated algorithms for the minimum dilation triangulation problem*. Journal Heuristics 189–209 (2014).

[4] Fogel, L. J. *Autonomous automata.* Industrial Research 4, 14-19 (1962).

[5] Gagliardi, E., Dorzán, M. G., Taranilla, M. T., Palmero, P., Casanova, C.: *Propuesta de plataforma para la integración de TIC orientadas al Agro* Anales de CAI 2017 Congreso de AgroInformática, 46JAIIO (2017)

[6] Gagliardi, E..; Dorzán, M.G.; Taranilla, M.T.; Palmero, P.; Casanova, C.: *Diseño de una herramienta de apoyo en la gestión de rodeos de cría.* Anales de CAI 2018 Congreso de AgroInformática, 47JAIIO (2018).

[7] Gagliardi, E.; Dorzán, M.G.; Taranilla, M.T.; Palmero, P.; Casanova, C.: *GeoSeguimiento de Rodeos, hacia una plataforma integral para el Agro*. Anales de CAI 2018 Congreso de AgroInformática, 47JAIIO (2018).

[8] Gudmundsson J., Levcopoulos C.; *Minimum weight pseudo-triangulations*. Computational Geometry. Theory and applications. Elsevier Vol. 38- Pages 139-153 (2007).

[9] Knauer, C., Mulzer, W.: *Minimum Dilation Tiangulations.* Tech. rep., FreieUniversitt Berlin, Fachbereich Mathematik und Informatik (2005).

[10] Michalewicz Z., Fogel D., *How to Solve It: Modern Heuristics*, Springer, (2004).

[11] Mulzer W., Rote G. *Minimum weight triangulation is NP-hard.*. Proceedings of the 22 Annual ACM Symposium on Computational Geometry (2006).

ESTRATEGIAS DE DESAMBIGUACION DE PERFILES Y SIMILITUD TEMÁTICA PARA UN METABUSCADOR DE LAS CIENCIAS DE LA COMPUTACIÓN

A. Canteros, U. Ramirez, E. Zamudio, M. Rey, A. Cantero, E. Martini, G. Pautsch, C. Biale, S. Krujoski, F. Rauber, A. Rambo, H. Kuna

Instituto de Investigación Desarrollo e Innovación en Informática (IIDII)
Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones.

hdkuna@gmail.com

RESUMEN

Un metabuscador académico para el área de las Ciencias de la Computación requiere gestionar los resultados de sus búsquedas y brindar sus servicios adecuadamente. En particular, los servicios de recomendación y gestión de resultados requieren el abordaje de problemas de desambiguación de las entidades que recupera, así como recomendación de autores. En este trabajo se presentan las líneas de investigación relacionadas con la evaluación de estrategias para la desambiguación de autores, junto con una línea relacionada con la recomendación de autores en base a los contenidos temáticos de sus perfiles.

Los resultados obtenidos en la evaluación de una estrategia de desambiguación demuestran que se pueden obtener un desempeño equivalente a la referencia. Asimismo, se describe un conjunto de datos en desarrollo para la evaluación de la recomendación de perfiles de autores en base a contenidos temáticos para el área de las Ciencias de la Computación.

Palabras clave: desambiguación, perfil, similitud, academico, tópico

CONTEXTO

Esta línea de investigación se desarrolla en el ámbito Programa de Investigación en Computación (PICom), perteneciente al Instituto de Investigación Desarrollo e Innovación en Informática de la Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones (IIDII, FCEQyN, UNaM).

El PICom desarrolla líneas de investigación relacionadas con la explotación de información y la robótica. Previamente se ha desarrollado un prototipo de metabuscador académico para las Ciencias de la Computación en el que se ha contribuido principalmente en áreas como la generación de estrategias de detección de datos anómalos (outliers e inliers), expansión de consultas, algoritmos de rankings, generación automática de perfiles de entidades, desambiguación de entidades, recomendación de expertos, y selección de grupos de expertos.

1. INTRODUCCIÓN

El metabuscador académico para las Ciencias de la Computación que desarrolla y

mantiene este grupo de investigación, actualmente en estado de prototipo, incorpora las contribuciones que se van desarrollando de acuerdo a las líneas de investigación propuestas en el marco del proyecto de investigación y sus actividades asociadas. En el último año, se han abordado problemas relacionados con el mejoramiento de procesos de recomendación, a partir de los resultados arrojados en los procesos de búsqueda, y como consecuencia se ha iniciado el tratamiento del problema de la desambiguación de entidades [1], incluyendo instituciones, autores, lugares de publicación.

La desambiguación de entidades es una tarea específica que involucra la generación de perfiles de dichas entidades, y luego un conjunto de procesos que contribuyan a determinar en qué medida dos o más perfiles con un mismo identificador (nombre) son similares entre sí [1].

Adicionalmente, la similitud de perfiles de de entidades puede tener otras aplicaciones, como la recomendación de contenidos adecuados para los perfiles, por ejemplo para determinar preferencias, o como la recomendación de perfiles, a partir de contenidos específicos. En este último caso, un metabuscador para las Ciencias de la Computación puede ofrecer un servicio de recomendación de perfiles en base a los contenidos temáticos de un documento.

En consonancia con la línea general de trabajo, se decide profundizar en el tratamiento de desambiguación de entidades, para intentar determinar las características adecuadas de una estrategia que permita adaptarse a los servicios brindados por un metabuscador académico para las Ciencias de la Computación.

Adicionalmente, y como consecuencia de los avances en el área de recomendación para el servicio del metabuscador, se identifica la necesidad de ampliar las estrategias de recomendación. En particular, se decide abordar la problemática de la generación de perfiles de autores de las producciones científico-tecnológicas y la recomendación de éstos en base a la similitud de los temas asociados a éste.

1.1 DESAMBIGUACION, SIMILITUD TEMÁTICA Y SU IMPACTO EN EL ÁMBITO DEL SISTEMA CIENTÍFICO TECNOLÓGICO NACIONAL

La importancia de la aplicación de estrategias de desambiguación y de similitud temática puede ser apreciada en contextos como el del Sistema Científico Tecnológico Nacional Argentino (SCTNA). En particular, el área de las Ciencias de la Computación incluye un conglomerado de organizaciones que tienen incidencia en la producción científico-tecnológica, tales como eventos de carácter científico, en la que se reciben trabajos que serán presentados durante la duración de los mismos. La evaluación que determina la aceptación o no del trabajo para ser presentado en un evento en particular se lleva a cabo por un conjunto de expertos que pertenecen a la nómina de varias organizaciones dentro del SCTNA.

La generación de perfiles de entidades, en particular de autores pertenecientes a un

sistema científico tecnológico, resulta determinante para permitir identificar en qué temas, áreas o tópicos, un autor define su expertise, y en qué medida lo logra. Asimismo, para propósitos como la recomendación, resulta necesario proporcionar herramientas que permitan desambiguar los perfiles de los expertos, así como el de proveer mecanismos para recomendar perfiles en base a la similitud temática.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En este trabajo, presentamos dos derivaciones de la línea de investigación principal referida al metabuscador para el ámbito académico de las Ciencias de la Computación. Estas derivaciones corresponden en primer lugar, a la evaluación de estrategias de desambiguación de autores de publicaciones científicas; y en segundo lugar, al desarrollo de estrategias para determinar similitudes temáticas entre perfiles de autores y documentos asociados a la producción científico tecnológica.

2.1 DESAMBIGUACIÓN DE AUTORES

La primer línea de trabajo define como objetivo, la evaluación de un enfoque de desambiguación de entidades, principalmente autores de producciones científico-tecnológicas. En particular, se evalúa la aplicación de estrategias de desambiguación a nombres de autores en el contexto de los resultados ofrecidos por el metabuscador de las Ciencias de la Computación.

Teniendo en cuenta el esquema del metabuscador, las entidades que lo componen y los datos con los que opera, se realizó un relevamiento y análisis de distintos enfoques de desambiguación de nombres de autores [2], [3].

Como resultado del relevamiento y análisis de las alternativas de desambiguación, se optó por la evaluación de AMiner en los resultados del metabuscador. AMiner presenta un framework de desambiguación de nombres de autores que se enfoca en resolver problemas tales como: determinar el número preciso de personas que comparten el mismo nombre; integrar datos continuamente, mejorando el proceso de desambiguación y la inclusión del usuario en el proceso [4].

2.2 SIMILITUD TEMÁTICA

La segunda línea de trabajo define como objetivo, la generación de procedimientos que permitan determinar el grado de similitud temática entre perfiles de autores y documentos asociados a producciones científico-tecnológicas. En particular, se pretende aplicar las estrategias de similitud temática [5] para asistir en la búsqueda de expertos adecuados para la evaluación de propuestas de trabajo y de artículos enviados a su publicación en eventos y revistas.

Esta línea de investigación abarca la identificación de estrategias del Procesamiento de Lenguaje Natural y del Aprendizaje Automático, destinadas al desarrollo de procesos que permitan elaborar perfiles de expertos a partir de artículos

académicos y otros contenidos textuales. Asimismo, se pretende desarrollar alternativas que permitan la evaluación de la similitud entre los perfiles de expertos y trabajos académicos del área de las Ciencias de la Computación.

3. RESULTADOS Y OBJETIVOS

La evaluación de la estrategia de desambiguación involucró la definición de un conjunto de datos generados a partir de consultas realizadas al metabuscador académico de Ciencias de la Computación. El conjunto de datos utilizados para la evaluación se constituyó a partir de 30 nombres de autores, de los cuales 18 (60%) elementos de la lista fueron tomados para entrenamiento y 12 (40%) para validación. Cada elemento de la lista contiene uno o más autores con el mismo nombre. Además, el conjunto de datos contiene los documentos de cada uno de esos autores, sumando en total 4.317 artículos.

Se utilizaron las métricas Precision, Recall y F1-score para evaluar los resultados del proceso de desambiguación. El experimento arrojó los siguientes resultados promedio. Precision: 0,7359; Recall: 0,60407; F1-score: 0,6635.

A modo de referencia, la ejecución del experimento sobre un conjunto de datos generados por AMiner [1] arrojó los siguientes resultados. Precision: 0,7685; Recall: 0,61661; y F1-score: 0,68423.

De acuerdo a los resultados obtenidos en la evaluación de la estrategia de desambiguación, se puede observar que la aplicación de dicha estrategia a distintos conjuntos de datos genera resultados similares. Por lo tanto, el método de desambiguación podría ser adecuado para su aplicación en el metabuscador.

En forma paralela a la evaluación de la estrategia de desambiguación de entidades, se comenzó a trabajar con la elaboración de un conjunto de datos que permita la evaluación exploratoria de estrategias para la generación de perfiles de expertos a partir del contenido temático de sus producciones científico-tecnológicas. Asimismo, se comenzó con el relevamiento de estrategias para determinar similitudes en base a dichas representaciones de los perfiles.
Se logró elaborar un conjunto de datos con 27.812 registros de producciones científico-tecnológicas en idioma español publicadas en el repositorio de la Universidad Nacional de La Plata (SEDICI). Estas producciones se corresponden principalmente con publicaciones en eventos, tesis, trabajos finales de grado y posgrado y artículos en revistas del ámbito académico de las Ciencias de la Computación de la República Argentina.

La evaluación exploratoria del enfoque para la generación de perfiles se realizó sobre un conjunto de 22.380 autores a partir de los registros contenidos en el conjunto de datos. Esta evaluación incluyó la generación de perfiles de expertos utilizando la técnica de word embeddings [6] a partir de atributos de

[1] https://github.com/neozhangthe1/disambiguation

los registros contenidos en el conjunto de datos, incluyendo: título y resumen [7]. Asimismo, se generaron conjuntos de tópicos mediante Latent Dirichlet Allocation (LDA) [8].

Sobre los resultados de la evaluación experimental, se han evaluado métricas de similitud (ej: Jensen-Shannon), los cuales aún deben ser contrastados con la opinión de expertos, debido a la falta de un conjunto de datos que permita contrastar los resultados en forma automática.

4. FORMACIÓN DE RECURSOS HUMANOS

Las líneas de investigación presentadas cuentan con doce integrantes relacionados con las carreras de Ciencias de la Computación de la UNaM. El grupo de investigación desarrolla dos tesis de grado articulando sus trabajos con becas de Estímulo a las Vocaciones Científicas del Consejo InterUniversitario Nacional (CIN) y becas UNaM; dos tesis de maestría en curso y una finalizada, de las cuales dos de ellas enmarcadas en becas del Programa Estratégico de Formación de Recursos Humanos en Investigación y Desarrollo (PERHID) del CIN. Asimismo, las líneas de investigación y sus integrantes se vinculan con grupos de la Universidad de Castilla-La Mancha, España y la Universidad de Sonora, México.

5. BIBLIOGRAFÍA

[1] I. Bhattacharya and L. Getoor, "Collective Entity Resolution in Relational Data," *ACM Trans Knowl Discov Data*, vol. 1, no. 1, Mar. 2007.

[2] A. Canteros, E. Zamudio, and H. Kuna, "Desambiguación de autores para un sistema de recuperación de expertos en un contexto académico," Simposio Argentino de Inteligencia Artificial, Sep. 2018, pp. 54–57.

[3] N. R. Smalheiser and V. I. Torvik, "Author name disambiguation," *Annu. Rev. Inf. Sci. Technol.*, vol. 43, no. 1, pp. 1–43, Jan. 2009.

[4] Y. Zhang, F. Zhang, P. Yao, and J. Tang, "Name Disambiguation in AMiner: Clustering, Maintenance, and Human in the Loop.," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Jul. 2018, pp. 1002–1011.

[5] O. Hanif, Z. Donghua, W. Xuefeng, and M. S. Nawaz, "Refining the Measurement of Topic Similarities Through Bibliographic Coupling and LDA," *IEEE Access*, vol. 7, pp. 179997–180011, 2019.

[6] M. Liu, B. Lang, Z. Gu, and A. Zeeshan, "Measuring similarity of academic articles with semantic profile and joint word embedding," *Tsinghua Sci. Technol.*, vol. 22, no. 6, pp. 619–632, 2017.

[7] B. Zhang and M. Al Hasan, "Name Disambiguation in Anonymized Graphs using Network Embedding," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Nov. 2017, pp. 1239–1248.

[8] M. Amami, G. Pasi, F. Stella, and R. Faiz, "An lda-based approach to scientific paper recommendation," in *International conference on applications of natural language to information systems*, 2016, pp. 200–210.

Recuperación de Información en Grandes Volúmenes de Datos

Fernando Kasián, Verónica Ludueña, Franco Merenda,
Marcela Printista, Nora Reyes, Patricia Roggero
LIDIC, Dpto. de Informática, Fac. de Cs. Físico Matemáticas y Naturales, Universidad Nacional de San Luis
{*fkasian, vlud, mprinti, nreyes,proggero*}@unsl.edu.ar, merenda.franco83@gmail.com
Karina Figueroa
Universidad Michoacana de San Nicolás de Hidalgo, México
karina@fismat.umich.mx
Claudia Deco
Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario
deco@fceia.unr.edu.ar

Resumen

Actualmente han surgido una cantidad de nuevos repositorios de información, en los cuales los datos son no estructurados y no se adaptan fácilmente al modelo relacional. Esto se debe tanto a la evolución de las tecnologías de información y comunicación, como a la gran cantidad y variedad de información disponible en formato digital. Estos diferentes tipos de datos tales como texto libre, imágenes, audio, video, secuencias biológicas de ADN o proteínas, entre otros; o bien no pueden ser fácilmente estucturados en claves y registros, o bien tal estructuración carece de sentido práctico, restringiendo de antemano los diversos tipos de consultas que se pueden requerir sobre ellos. Todo esto deja en evidencia la necesidad de procesar grandes conjuntos de datos, para obtener información útil a partir de ellos.

El objetivo de cualquier sistema de recuperación de información es obtener, desde una base de datos, lo que podría ser útil o relevante para el usuario a partir de una consulta. Para ello se utiliza alguna estructura de almacenamiento sobre dichos datos (índice), diseñadas especialmente para ese propósito, que permita responder a la consulta de manera eficiente.

Palabras Claves: *bases de datos masivas, computación de alto desempeño, recuperación de información.*

1. Contexto

Esta línea de investigación se encuentra enmarcada dentro del Proyecto Consolidado 3-03-2018 de la Universidad Nacional de San Luis (UNSL) y en el Programa de Incentivos (Código 22/F834): "Tecnologías Avanzadas Aplicadas al Procesamiento de Datos Masivos", dentro de la línea "Recuperación de Datos e Información", desarrollada en el Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC) de la UNSL. Este proyecto ha sido aprobado en la UNSL por Resolución del Consejo Superior 126/18 y finaliza en 2021.

El objetivo de esta línea de investigación es desarrollar herramientas eficientes para sistemas de recuperación de información sobre bases de datos masivas, que almacenan datos no estructurados. El análisis de nuevas técnicas que provean una buena interacción con el usuario, al igual que el desarrollo de nuevas estructuras de datos capaces de manipular eficientemente un gran volumen de datos no estructurados, están orientados en ese sentido.

Por lo tanto, lo que se pretende en esta línea es el diseño y desarrollo de índices que sirvan de apoyo a sistemas de recuperación dedicados a conjuntos de datos no estructurados masivos tales como: datos multimedia, texto, secuencias de ADN, huellas digitales, etc., proveyendo a los mismos de estructuras de datos eficientes y escalables, para memorias jerárquicas, que además aprovechen, cuando sea necesario, la aplicación de técnicas de computación de alto desempeño (HPC).

2. Introducción y Motivación

La evolución de las tecnologías de información y comunicación, el uso masivo de internet y la disponibilidad de dispositivos electrónicos en diversos ámbitos de la vida cotidiana, ha generado la aparición de un gran volumen de datos, que crece rápidamente y que provienen de fuentes tan disimiles como el tipo de datos producido. Este escenario obliga a redefinir las técnicas de procesamiento, análisis y obtención de información útil, y a formular nuevas metodologías para lograr una mayor aplicabilidad.

Aún en la actualidad, los sistemas tradicionales de computación utilizan principalmente información

que puede organizarse en claves y registros (estructurada), sobre la cual las búsquedas tradicionales tienen sentido y donde la estructura misma de los datos puede interpretarse y utilizarse en programas casi directamente. Sin embargo, dos de las características que aparecen en los datos en el contexto de problemas de "big data", el volumen y variedad de los mismos, hace imposible restringir las búsquedas sobre datos estructurados a las búsquedas tradicionales, ya que obligaría a representar una visión parcial del problema, dejando fuera información que podría ser relevante para la resolución efectiva del mismo.

Por lo tanto, en la era de "big data" es necesario administrar eficientemente información no estructurada y considerar tipos de búsqueda mucho más generales y complejas, que puedan servir de apoyo, por ejemplo, en la toma de decisiones. Las búsquedas por similitud son un tipo de búsqueda más general, que se sustentan, para lograr eficiencia en las respuestas, sobre *métodos de acceso* o *índices métricos* [5]. Un enfoque útil para sistemas de recuperación usando búsqueda por similitud es *la búsqueda basada en contenidos*. Dicho tipo de búsqueda usa el dato no estructurado en sí mismo para describir lo que se está buscando.

Entonces, si se consideran grandes cantidades de datos no estructurados, se pueden utilizar estos índices para lograr eficiencia en la respuesta, cuando se presentan al sistema consultas de recuperación de información. Dichos índices pueden tener distintas características que los hacen adecuados para aplicaciones reales: eficientes, dinámicos, escalables, resistentes a la *maldición de la dimensión*, entre otras.

El modelo habitual para las búsquedas por similitud es el de espacios métricos. Este modelo, además de brindar un marco formal, es independiente del dominio de la aplicación. Un espacio métrico está compuesto por un *universo* $\mathcal{U}$ de objetos y una *función de distancia* $d : \mathcal{U} \times \mathcal{U} \longrightarrow \mathbb{R}^+$, la cual cumple con las propiedades de una métrica. Sobre una *base de datos* $\mathcal{S} \subseteq \mathcal{U}$, se pueden considerar dos tipos básicos de búsqueda por similitud: la *búsqueda por rango* y la *búsqueda de los* k *vecinos más cercanos*. La función de distancia permite medir la disimilitud entre dos objetos. Sin embargo, para algunos de los tipos de datos no estructurados, el cálculo de la distancia puede ser muy costoso. Por lo tanto, un objetivo de todo método de acceso es ahorrar cálculos de distancia y en su gran mayoría para lograrlo aprovechan que d satisface la desigualdad triangular.

Si se considera que la base de datos $\mathcal{S}$ posee n objetos, cualquier consulta puede responderse de manera trivial realizando n evaluaciones de distancia. Sin embargo, en la mayoría de las aplicaciones sobre grandes volúmenes de datos y siendo las distancias costosas de computar (por ej.: comparación de huellas digitales), no es factible aplicar la solución trivial. Por lo tanto, si el objetivo es responder consultas con la menor cantidad de cálculos de distancia posibles, se debe construir un índice a través del preprocesamiento de la base de datos.

En ciertos casos particulares, es probable que la base de datos, el índice, o ambos, no puedan almacenarse en memoria principal y deban hacer uso de niveles más bajos de la jerarquía de memorias, como la memoria secundaria. Pero ello acarrea altos costos en las operaciones de E/S. Por lo tanto, para lograr eficiencia, se debe minimizar el número de operaciones de E/S, considerar el nivel de la jerarquía de memorias sobre la que se trabaja y en algunos casos admitir respuestas no exactas; utilizando, cuando sea posible, técnicas de computación paralela.

En este contexto, se considera como objetivo principal obtener herramientas de recuperación de información para procesar conjuntos masivos de datos, desarrollando nuevas técnicas y aplicaciones que soporten la interacción con el usuario, diseñando índices que permitan la manipulación eficiente de grandes volúmenes de datos no estructurados y faciliten la realización de diferentes tipos de consultas. De esta manera, se espera contribuir al desarrollo de aplicaciones reales para problemas de big data.

3. Líneas de Investigación

Dado que en esta investigación se pretende contribuir a distintos aspectos de los sistemas de recuperación de información sobre grandes volúmenes de datos no estructurados, se ha considerado el diseño de nuevas medidas de similitud, de nuevos índices y la resolución de distintas consultas sobre estos tipos de bases de datos y cómo lograr eficiencia y escalabilidad para grandes volúmenes de datos.

Nuevas Medidas de Similitud

Existen numerosos algoritmos que permiten resolver eficientemente las búsquedas cuando se consideran espacios de baja dimensión. Sin embargo, su desempeño se degrada a medida que la dimensionalidad intrínseca del espacio crece [5, 13]. Más aún, en bases de datos cuya dimensionalidad intrínseca es alta, el desempeño puede empeorar de tal forma que llega a ser equivalente al de realizar una búsqueda

exhaustiva sobre la base de datos [5]. Por lo tanto, el desafío se encuentra en esa clase de base de datos, donde el histograma de las distancias entre los objetos es muy concentrado; es decir, donde todas las distancias entre pares de objetos de la base de datos son muy similares.

Una aproximación práctica sobre este tipo de base de datos es resignarse a no obtener la respuesta por similitud exacta para las consultas. En su lugar, es posible conformarse con respuestas aproximadas; lo cual significa que se admite que se pierdan algunos objetos relevantes desde el conjunto de objetos de la respuesta o que se reporten en dicho conjunto algunos elementos que no sean relevantes [6]. Así, el objetivo es diseñar métodos eficientes cuya calidad de la respuesta esté dentro de ciertos límites.

En [1], se presenta un nuevo método aproximado para búsquedas por similitud, cuyo desempeño es insuperable en bases de datos de alta dimensión [11]. Sin embargo, se puede mejorar aún más su desempeño si se considera una medida diferente de similitud entre permutaciones [7].

La medida más utilizada de distancia entre permutaciones es *Spearman Footrule*, la cual se define como $S_F(\Pi_u, \Pi_q) = \sum_{1 \leq i \leq m} |\Pi_u^{-1}(i) - \Pi_q^{-1}(i)|$, donde Π_u y Π_q son las permutaciones de u y q.

Las nuevas medidas de similitud entre permutaciones que en particular, como se demuestra en [7], no cumplen con ser una métrica sino una semimétrica, se basan en particionar las permutaciones en trozos y utilizar en los trozos significativos (el inicial y el final) un factor que permita amplificar las grandes diferencias de posiciones de los permutantes y descartar el trozo medio menos significativo.

Índices

Cuando se trabaja con bases de datos que contienen datos no estructurados, los índices métricos resultan apropiados para realizar búsquedas sobre las mismas [5]. Éstos aprovechan una propiedad de la función de distancia, la desigualdad triangular, para ahorrar algunos cálculos de distancia y, de esta manera, ahorrar tiempo. Si se mantienen algunas distancias precalculadas entre los elementos de la base de datos y objetos distinguidos, la desigualdad triangular permitirá estimar la distancia entre cualquier objeto de consulta q y los elementos de la base de datos. Los dos enfoques más comunes se diferencian en si esos objetos distinguidos son *pivotes* o *centros*. Si son pivotes se almacenan las distancias de todos los objetos de la base de datos a ellos y si por el contrario son centros se particiona el espacio en zonas denominadas *particiones compactas*, por cercanía a los centros, almacenando un radio de cobertura que determina la zona de cada centro.

Los aspectos que se consideran de interés al diseñar índices incluyen: dinamismo, en qué nivel de la jerarquía de memorias deben almacenarse, si pueden aplicar técnicas de computación de alto desempeño para mejorar los tiempos de respuesta, si deben proporcionar una respuesta exacta o basta con una respuesta aproximada y la dimensionalidad del espacio métrico considerado.

Como los conjuntos de datos masivos de interés son aquellos que contienen datos no estructurados, los volúmenes de información con los que se debe trabajar (por ejemplo, millones de imágenes en la Web) hacen necesario que los índices sean almacenados en memoria secundaria. En este caso, para lograr eficiencia, no sólo se debe considerar que se realicen el menor número de cálculos de distancia sino también, que efectúen la menor cantidad posible de operaciones sobre el disco (E/S), debido a su costo. Por ello, esta línea se dedica a diseñar índices adaptados para trabajar en memoria secundaria, cuyo desempeño en las búsquedas sea bueno. En particular, se ha diseñado e implementado una versión paralela del *Conjunto Dinámico de Clusters* (DSC) [14]. Este índice, basado en la *Lista de Clusters* (LC) [4], está especialmente diseñado para memoria secundaria y es completamente dinámico, admite inserciones y eliminaciones y tiene un buen desempeño en las búsquedas, principalmente en la cantidad de operaciones de E/S. DSC ha demostrado ser muy competitivo frente a algunas buenas estructuras; por lo tanto, se buscará aplicar y comparar distintas estrategias de paralelización en él.

El *Árbol de Aproximación Espacial Distal* (DiSAT), basado en el *Árbol de Aproxiación Espacial* [12], es un índice estático que no necesita sintonizar ningún parámetro y es muy eficiente gracias a la partición de hiperplanos que define [3]. La raíz elegida para el DiSAT define una partición sobre el espacio, donde las zonas que se obtienen son muy compactas y los hiperplanos que las definen permiten diferenciarlas muy bien. Por ello, se busca aprovechar la información que brindan distintas particiones sobre el espacio para clasificar los elementos de acuerdo a las zonas en las que cada elemento cae en las distintas particiones consideradas. En este caso, a cada elemento se le asigna una secuencia de bits, denominada “sketch’; cada bit indica de qué lado del

hiperplano considerado se encuentra el objeto. Este conjunto de "sketches" constituye el índice en sí mismo. Al momento de una consulta, se calcula el sketch del objeto de consulta q y se lo compara con los sketches de todos los elementos de la base de datos, sin calcular realmente distancias entre objetos sino entre sketches, revisando luego los objetos más prometedores primero. Se espera que un elemento similar a q estará en una partición similar en el espacio, logrando una respuesta aproximada a la consulta por similitud con poco costo.

Una técnica que ha logrado mucho éxito es la de *Algoritmos Basados en Permutaciones* (PBA). Recientemente se desarrolló una forma novedosa de reducir el tamaño del índice sin eliminar ningún permutante, lo que afectaría el desempeño del índice, almacenando en lugar de la permutación de cada elemento, su *firma* con respecto a los pares de permutantes del conjunto [8].

Por otra parte, se está estudiando cómo aprovechar los índices sobre conjuntos masivos de datos, como herramienta de apoyo para solucionar un problema de estacionamiento de vehículos.

DBMS para Bases de Datos Multimedia

A pesar de que las operaciones más comunes sobre bases de datos multimedia son las búsquedas por rango o de k-vecinos más cercanos, la operación de *join* por similitud se considera una de las operaciones que debería brindar típicamente un sistema administrador para bases de datos multimedia [16].

Existen diferentes variantes para el join por similitud, dependiendo del criterio de similitud Φ utilizado, pero tienen en común que se aplican entre dos bases de datos A y B, ambas subconjuntos del mismo universo del espacio métrico $\mathcal{U}$ que modela a la base de datos. El resultado de cualquiera de las variantes de este join por similitud obtendrá el conjunto de pares formados por un objeto de A y otro de B, tales que entre ellos se satisface el criterio de similitud Φ. Las variantes más conocidas son: el join por rango, el join de k-vecinos más cercanos y el join de k pares de vecinos más cercanos; entre otras.

Formalmente, dadas $A, B \subseteq \mathcal{U}$, se define el *join por similitud* entre A y B ($A \underset{\Phi}{\bowtie} B$) como el conjunto de todos los pares (x, y), donde $x \in A$ e $y \in B$; es decir, $(x, y) \in A \times B$, tal que $\Phi(x, y)$ es verdadero (se satisface el criterio de similitud Φ entre x e y). Al resolver el join por similitud es posible que ambas, una o ninguna de la bases de datos posean un índice; o que ambas bases de datos se indexen conjuntamente con un índice diseñado para el join. Calcular cualquiera de las variantes del join por similitud de manera exacta es muy costoso [15], por lo tanto vale la pena analizar posibilidades de obtener más rápidamente una respuesta aproximada al join, buscando siempre buena calidad en la respuesta.

PostgreSQL es el primer sistema de base de datos que permite realizar consultas por similitud sobre algunos atributos, particularmente indexa para búsquedas de k-vecinos más cercanos (índices *KNN-GiST*). Estos índices pueden ser usados sobre texto, comparación de ubicación geoespacial, etc. Sin embargo, los índices *K-NN GiST* proveen plantillas sólo para índices con estructura de *árbol balanceado* (*B-tree*, *R-tree*), pero el "balance" no siempre es bueno para los índices que se utilizan en búsquedas por similitud [2]. Por otro lado, no se dispone de este tipo de consultas para todo tipo de datos métricos. Así, es importante proveer un DBMS para todos los posibles datos métricos y sus operaciones [10].

Más aún, dado que las respuestas a consultas de join suelen ser conjuntos muy grandes de pares de objetos y muchos de esos pares son muy similares entre sí, se planea introducir sobre las operaciones de join la posibilidad de diversificar las respuestas [17]; es decir, un operador de join por similitud que asegure un conjunto más pequeño, más diversificado de respuestas útiles y, de ser posible, más rápido de obtener. Estos desarrollos, entre otros, permitirán tener un DBMS con mayores posibilidades de aplicación en sistemas de información reales.

4. Resultados

Se ha publicado en [7] una familia de medidas de similitud para permutaciones que permiten mejorar el desempeño de los algoritmos basados en permutaciones [1]. Además, se ha publicado en [9, 8] nuevas estructura para búsquedas aproximadas, especialmente diseñadas para trabajar con grandes volúmenes de datos. Por otro lado, se está evaluando experimentalmente la versión paralela del índice *DSC*, que trabaja con grandes volúmenes de datos, diseñada para memoria secundaria, admite inserciones/eliminaciones de objetos y que permitirá responder eficientemente a lotes de consultas por similitud. Además, se encuentra en proceso de evaluación la propuesta de sketches basados en el *DiSAT*. Se continúa trabajando en la extensión de *PostgreSQL* para que brinde facilidades de soporte a más tipos de consultas por similitud, sobre distintos tipos de datos y considere opciones de respuesta aproximada, como

también la posibilidad de diversificación de respuestas para los joins por similitud.

5. Formación de Recursos

En esta línea se están realizando las siguientes tesis de Maestría en Ciencias de la Computación:

1 - "Estructuras Eficientes sobre Datos Masivos para Búsquedas en Espacios Métricos",

2 - "Cómputo Aproximado del Grafo de Todos los k-Vecinos",

3 - "Sistema Administrador para Bases de Datos Métricas".

Además, está en desarrollo un trabajo final de la Ingeniería en Computación.

Referencias

[1] E. Chávez, K. Figueroa, and G. Navarro. Effective proximity retrieval by ordering permutations. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(9):1647–1658, 2009.

[2] E. Chávez, V. Ludueña, and N. Reyes. Revisiting the VP-forest: Unbalance to improve the performance. In *Proc. de las JCC08*, page 26, 2008.

[3] E. Chávez, V. Ludueña, N. Reyes, and P. Roggero. Faster proximity searching with the distal sat. *Information Systems*, 59:15 – 47, 2016.

[4] E. Chávez and G. Navarro. A compact space decomposition for effective metric indexing. *Pattern Recognition Letters*, 26(9):1363–1376, 2005.

[5] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín. Searching in metric spaces. *ACM*, 33(3):273–321, September 2001.

[6] P. Ciaccia and M. Patella. Approximate and probabilistic methods. *SIGSPATIAL Special*, 2(2):16–19, 2010.

[7] K. Figueroa, R. Paredes, and N. Reyes. New permutation dissimilarity measures for proximity searching. *Similarity Search and Applications - 11th International Conference, SISAP, Proceedings*, volume 11223 of *Lecture Notes in Computer Science*, pages 122–133. Springer, 2018.

[8] K. Figueroa and N. Reyes. Permutation's signatures for proximity searching in metric spaces. In *Similarity Search and Applications - 12th International Conference, SISAP*, volume 11807, pages 151–159. Springer, Cham, 2019.

[9] K. Figueroa, N. Reyes, A. Camarena-Ibarrola, and L. Valero-Elizondo. Improving the list of clustered permutation on metric spaces for similarity searching on secondary memory. In *Pattern Recognition*, pages 82–92, Cham, 2018. Springer International Publishing.

[10] F. Kasián and N. Reyes. Búsquedas por similitud en PostgreSQL. In *Actas del XVIII Congreso Argentino de Ciencias de la Computación (CACiC)*, pages 1098–1107, Bahía Blanca, Argentina, 2012. Universidad Nacional del Sur.

[11] B. Naidan, L. Boytsov, and E. Nyberg. Permutation search methods are efficient, yet faster search is possible. *Proc. VLDB Endow.*, 8(12):1618–1629, August 2015.

[12] G. Navarro. Searching in metric spaces by spatial approximation. *VLDBJ*, 11(1):28–46, 2002.

[13] G. Navarro, R. Paredes, N. Reyes, and C. Bustos. An empirical evaluation of intrinsic dimension estimators. *Information Systems*, 64:206 – 218, 2017.

[14] G. Navarro and N. Reyes. New dynamic metric indices for secondary memory. *Information Systems*, 59:48 – 78, 2016.

[15] R. Paredes and N. Reyes. Solving similarity joins and range queries in metric spaces with the list of twin clusters. *JDA*, 7:18–35, March 2009. doi:10.1016/j.jda.2008.09.012.

[16] C. Rong, C. Lin, Y. N. Silva, J. Wang, W. Lu, and X. Du. Fast and scalable distributed set similarity joins for big data analytics. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 1059–1070, April 2017.

[17] L. Santos, L. Carvalho, W. Oliveira, A. Traina, and C. Jr. Traina. Diversity in similarity joins. *Similarity Search and Applications*, volume 9371 of *Lecture Notes in Computer Science*, pages 42–53. Springer International Publishing, 2015.

Herramientas Informáticas para el Estudio de la Biodiversidad utilizando Datos Abiertos Enlazados

Gustavo Samec[1,2,3], María Emilia Diez[1,2,5], Marcos Zárate[1,2,4], Carlos Buckle[1,2], Joaquín Lima[1,2], Rodrigo Jaramillo[1,2,3], Alejandro Sánchez[1,2,6], Renato Mazzanti[1,2,3]

[1] Departamento de Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB), Puerto Madryn.
[2] LINVI, Laboratorio de Investigación en Informática, Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB)
[3] Unidad de Gestión de la Información, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (CCT CONICET-CENPAT)
[4] Centro para el Estudio de Sistemas Marinos, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (CESIMAR) (CCT CONICET-CENPAT)
[5] Laboratorio de Parasitología (LAPA), Instituto de Biología de Organismos Marinos, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (IBIOMAR) (CCT CONICET-CENPAT)
[6] Departamento de Informática, Universidad Nacional de San Luis (UNSL), San Luis.
{gsamec, emiliadiez, zarate, rjaramillo, renato}@cenpat-conicet.gob.ar, cbuckle@unpata.edu.ar

RESUMEN

En la actualidad existen grandes bases de datos globales de biodiversidad con contenidos abiertos a la comunidad científica, la mayoría proveen APIs Web para realizar consultas y recuperar información. A pesar de éstas facilidades frecuentemente no son interoperables, por lo general representan un modelo de datos propietario y carecen de vocabularios de descripciones semánticas formales esenciales para garantizar la integración de los datos. El presente trabajo tiene como objetivo hacer accesible y abiertos los datos, a la comunidad científica, de la base de datos Southwest Atlantic Benthic Invertebrates[1] (SWATL) que registra datos de de invertebrados bentónicos de la región y publicaciones taxonómicas, por medio de Datos Abiertos Enlazados[2] (LOD) y hacerla interoperable con bases de datos de referencia global desarrollando micro-servicios SPARQL[3] como envoltura (wrapper) de las APIs Web que las mismas proveen.

Palabras clave: Web Semántica, Datos Abiertos Enlazados, SPARQL, Biodiversidad.

CONTEXTO

Este trabajo se encuadra dentro del proyecto "Aplicaciones Informáticas para el Estudio de Biodiversidad de Poliquetos Espiónidos en los Golfos Nordpatagónicos", elaborado en el Laboratorio de Investigación en Informática (LINVI) de la Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB) e integrado por docentes investigadores de la Facultad de Ingeniería Sede Puerto Madryn, con participación de estudiantes y graduados de las carreras de dicha Sede. El proyecto fue avalado por el Consejo Directivo de la Facultad de Ingeniería de la UNPSJB y es financiado por la Secretaría de Ciencia y Técnica de la Universidad para llevarlo a cabo durante 2019-2020.

1. INTRODUCCION

La gran diversidad de datos generados por distintas disciplinas (taxonómicos, genéticos, meteorológicos, etc.) y la posibilidad de extraerlos e integrarlos para realizar investigaciones resulta de sumo

[1] http://sistema.cenpat-conicet.gob.ar:8081/rb/
[2] https://www.w3.org/egov/wiki/Linked_Open_Data
[3] https://www.w3.org/TR/rdf-sparql-query/

interés a la comunidad científica [1]. La Web Semántica [2] proporciona soluciones a éstas necesidades a través de Datos Enlazados (Linked Data) [3] donde los objetos de datos se identifican de manera única y las relaciones entre ellos se definen explícitamente.

En nuestro caso, contamos con la base de datos SWATL sobre la cual se requieren desarrollar herramientas informáticas para que sus datos sean accesibles y abiertos a la comunidad científica y por otro lado interoperable con otras bases de datos de referencia global permitiendo hacer consultas y extraer información de ellas.

Si bien es cierto que existen grandes bases de datos con contenidos similares como por ejemplo: Global Biodiversity Information Facility[4] (GBIF), World Register of Marine Species[5] (WoRMS), Species 2000[6] (SP2000), etc., que gestionan listas taxonómicas de especies a partir de la validación de especialistas taxónomos, la importancia de SWATL radica en que es una base de datos regional focalizada a un grupo determinado de especies y publicaciones taxonómicas, alimentada y mantenida por especialistas locales, que en numerosos casos cuenta con más información y está más actualizada que las grandes bases de datos de referencia a nivel global, siendo su contenido de gran interés por quienes desarrollan investigaciones que tienen relación con los datos que almacena la misma.

Para el acceso abierto de los datos de SWATL se propone utilizar LOD, para ello se requiere exportar el contenido de la base de datos relacional a tripletas RDF[7]. La utilización de LOD mejora la capacidad de localización, accesibilidad, interoperabilidad y reutilización de los mismos (principios FAIR - Findable, Accessible, Interoperable, and Reusable) [4], y permite la vinculación con otros conjuntos de datos abiertos enlazados.

Esta vinculación de datos, con el agregado de una mayor integración semántica [5], facilita a los científicos descubrir y utilizar esos datos de manera consistente en sus investigaciones.

A la hora de hacer interoperable SWATL con otras bases de datos encontramos la limitación que, en su gran mayoría, no proveen un endpoint SPARQL. Por lo general los portales de acceso a las bases de datos proveen APIs Web para hacer distintos tipos de consultas y extraer registros de los millones de datos que almacenan. En este sentido realizar consultas interoperables con SWATL demandaría realizar un código *ad-hot* para cada tipo de proveedor.

En este caso la propuesta es desarrollar micro-servicios SPARQL [6,7] como envoltura de las APIs Web que los portales de las bases proveen y traducir su respuesta en tripletas RDF lo cual va a permitir que SWATL interopere con las mismas.

2. LÍNEAS DE INVESTIGACION Y DESARROLLO

Dentro de los trabajos realizados, se rediseño el modelo de la base de datos SWATL, en el mismo se incluyeron datos que facilitan la integración principalmente con WoRMS y bases relacionadas que utilizan la Plataforma Aphia [8] y se reorganizaron sus tablas y relaciones para simplificar las consultas y obtener información con un contenido similar a las ofrecidas por las base de datos de referencia global. Por otro lado se realizó una refactorización de la aplicación que accede a la base de datos para su carga y consulta y los cambios necesarios para adaptarla al nuevo modelo de datos. El desarrollo original corresponde a una aplicación Web desarrollada por capas, todos sus módulos están escritos en Java y utiliza el estándar JavaServer Faces (JSF) a nivel de capa de presentación. Las capas de modelo, acceso a datos y servicios se adaptaron a los cambios requeridos por las nuevas bibliotecas utilizadas además de los cambios correspondientes al nuevo modelo de datos. La capa lógica prácticamente no requirió cambios más allá de la adaptación al nuevo

[4] https://www.gbif.org/
[5] http://www.marinespecies.org/
[6] https://sp2000.org/
[7] https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/

modelo de datos. La capa de presentación demandó un gran esfuerzo, si bien se siguió utilizando el mismo estándar JSF, se cambió la utilización de componentes de ICEFaces[8], por Primefaces[9]. La utilización de componentes más avanzados provisto por Primefaces motivó cambios significativos en esta capa. También se cambió la tecnología utilizada en seguridad, la versión original utilizaba Realm de Tomcat y la nueva versión utiliza SpringSecurity independizándola del servidor de aplicaciones. Las razones que motivaron la refactorización de la aplicación original responde a adaptarlas a las nuevas tecnologías existentes permitiendo que la aplicación actual y los futuros desarrollos sobre la misma capitalicen las ventajas que éstas proveen.

Por otro lado se han realizado experiencias utilizando las APIs que provee Apache Jena[10] y Graphdb[11] para exportar el contenido de la base de datos relacional a LOD y permitir consultarla en un endpoint SPARQL. Las bases de datos taxonómicas y de registros históricos como SWATL tiene una baja tasa de actualización por lo que tener una copia de la base de datos relacional (BDR) en un *triplestore* y actualizarla periódicamente no afectaría a la calidad del servicio, de todas maneras no se descarta evaluar la performance y utilizar servidores que acceden a la BDR y convierten los datos en LOD sin necesidad de almacenarlos *(on-fly)*.

Para interoperar con otras bases de datos (que no posean endpoint SPARQL) la propuesta es desarrollar micro-servicios SPARQL. Para ello se requiere crear un micro-servicio SPARQL por cada API Web que la base provee. Dada una consulta, el cliente SPARQL envía la misma al micro-servicio SPARQL, el cual se encarga de traducirla al formato adecuado y enviar la solicitud a la API Web, cuando la misma responde, el micro-servicio SPARQL se encarga de traducir el resultado en tripletas RDF y enviar las mismas al cliente SPARQL.

El starndard Darwin Core [9] es ampliamente utilizado en bases de datos de biodiversidad (incluido SWATL), el hecho de poseer un vocabulario en común facilita la interoperabilidad.

3. RESULTADOS ESPERADOS

Poner a disposición de la comunidad científica los datos de SWATL y facilitar su interoperabilidad con bases de datos de referencia global.

Aplicar la experiencia adquirida y los productos obtenidos a otros campos donde se requiera interoperabilidad.

Contribuir en la formación de recursos humanos en nuevas tecnologías.

Consolidar un grupo de investigación interdisciplinario en informática para la biodiversidad dentro del grupo LINVI de la UNPSJB.

4. FORMACION DE RECURSOS HUMANOS

En este proyecto participan integrantes de formación docente y académica en las áreas de Ingeniería de Software, Bases de Datos, Inteligencia Artificial y Biología. Cinco de los docentes son del Departamento de Informática y uno del Departamento de Matemáticas de la Facultad de Ingeniería de la UNPSJB Sede Puerto Madryn. La Doctora en biología María Emilia Diez es investigadora del CONICET y es especialista en taxonomía y ecología de poliquetos, uno de los autores están realizando la carrera de doctorado, otro iniciando su post-doctorado con beca del CONICET y otros dos se encuentran desarrollando carreras de especialización y maestrías. También forman parte del grupo de investigación un graduado de la carrera de Licenciatura en Informática y tres alumnos del ciclo superior. Uno de ellos desarrollando su tesina.

[8] http://www.icesoft.org/

[9] https://www.primefaces.org/

[10] https://jena.apache.org/

[11] http://graphdb.ontotext.com/documentation/free/quick-start-guide.html

5. BIBLIOGRAFIA

1. Richard K Lomotey and Ralph Deters. Terms extraction from unstructured data silos. In System of Systems Engineering (SoSE), 2013 8th International Conference on, pages 19-24. IEEE, 2013.

2. Tim Berners-Lee, James Hendler, Ora Lassila, et al. The semantic web. Scientific American, 284(5):28-37, 2001.

3. Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data: The story so far. In Semantic services, interoperability and web applications: emerging concepts, pages 205-227. IGI Global, 2011.

4. Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The fair guiding principles for scientific data management and stewardship. Scientific data, 3, 2016.

5. Paolo Ceravolo, Antonia Azzini, Marco Angelini, Tiziana Catarci, Philippe Cudr_e-Mauroux, Ernesto Damiani, Alexandra Mazak, Maurice Van Keulen, Mustafa Jarrar, Giuseppe Santucci, et al. Big data semantics. Journal on Data Semantics, 7(2):65-85, 2018.

6. Michel F, Faron-Zucker C, Tercerie S, Ettorre A, Olivier G. Assisting Biologists in Editing TaxonomicInformation by Confronting Multiple Data Sources using Linked Data Standards. Biodiversity Information Science and Standards 3, 2019. https://doi.org/10.3897/biss.3.37421

7. Michel F, Faron-Zucker C, Gargominy O, Gandon F. Integration of Web APIs and Linked Data Using SPARQL Micro-Services - Application to Biodiversity Use Cases, 2018. https://doi.org/10.3390/info9120310

8. Nozères, C., Vandepitte, L., Appeltans, W., Kennedy, M. Best Practice Guidelines in the Development and Maintenance of Regional Marine Species Checklists, version 1.0, released on August 2012. Copenhagen: Global Biodiversity Information Facility, 32 pp. 2012 http://www.gbif.org/orc/?doc_id=4712

9. Baskauf S, Wieczorek J, Deck J, Webb C, Morris PJ, Schildhauer M. Darwin Core RDF Guide. Biodiversity Information Standards (TDWG). 2015. http://rs.tdwg.org/dwc/terms/guides/rdf/

Aplicación de Técnicas Descriptivas de Minería de Textos sobre Contenido Digital Realizando Análisis Inteligente

Sánchez Rivero, Víctor David; Farfán, José Humberto; Rodríguez, Mariela Ester; Vargas, Luis Alejandro; Vega, Ariel Alejandro; Garcete, Christian Brian; Llampa, Álvaro Facundo; Ramos, Pablo Nicolás; Contreras, Facundo; Churquina, Cintia Noelia; Águila, Viviana Emilia y Iogna Prat Genzel, Nicolás

Área de Ingeniería Informática - Facultad De Ingeniería – Universidad Nacional De Jujuy

ivansrivero@gmail.com, jhfarfan@hotmail.com, maru972@gmail.com, alevar98@yahoo.com, arielalejandrovega@gmail.com, chrisbriancerrudo@gmail.com, a.facundollampa@gmail.com, pablonicolasr777@gmail.com, facucontreras21@gmail.com, cintia7828@gmail.com, vivi.agui.4@gmail.com, iognapratnicolas@gmail.com

RESUMEN

El presente proyecto pretende implementar técnicas de Minería de Textos o Text Mining en conjunción con técnicas de Minería Web o Web Mining (metodología para la recuperación y extracción de información desde páginas web) para poder realizar un estudio de los Patrones de Escritura empleados para la confección de documentos digitales científicos. Tanto Text Mining como Web Mining se encuadran dentro de las técnicas de Minería de Datos y son técnicas que permiten descubrir patrones usados en grandes volúmenes de texto. El proyecto también incluirá una investigación sobre la aplicación de técnicas o algoritmos orientados al Procesamiento del Lenguaje Natural usados en el análisis de textos o documentos obtenidos de Redes Sociales, por ejemplo, y se persigue, a través de su empleo, la obtención de prototipos de sistema que faciliten el análisis en cuestión.

Los textos o documentos digitales, sobre los cuales se trabajará en este proyecto, se obtendrán principalmente desde la Web, considerando que en la Sociedad del Conocimiento actual, la gestión de la información y conocimiento es un componente estratégico para el análisis inteligente de la información digital, para la clasificación de contenidos y la extracción de conceptos, entre algunos de los principales tópicos que estudia Text Mining.

Palabras clave: Minería de Textos, Minería de Datos, KDD, Análisis Inteligente.

CONTEXTO

El proyecto se encuentra inserto dentro de las siguientes Líneas Prioritarias de Investigación de la Facultad de Ingeniería (LIPIFI) - UNJu:

- Ingeniería del Software.
- Ingeniería de Procesos.

Es un proyecto aprobado de categoría B (Código D/B036) denominado:

"Aplicación de técnicas descriptivas de Minería de textos sobre contenido digital realizando un análisis inteligente".

Financiamiento: Secretaría de Ciencia y Técnica y Estudios Regionales (SeCTER) de la UNJu.

Vigencia del Proyecto: 01/01/2020 al 31/12/2021

1.INTRODUCCION

El actual proyecto pretende implementar técnicas de Minería de Textos o Text Mining (el cual es el proceso de derivar información nueva a partir de textos digitales) en conjunción con técnicas de Minería Web o Web Mining (metodología para la recuperación y extracción de información desde páginas web) para poder realizar un estudio de los Patrones de Escritura empleados para la confección de documentos digitales científicos. Tanto Text Mining como Web Mining se encuadran dentro de las técnicas de Minería de Datos y son técnicas que permiten descubrir patrones usados en grandes volúmenes de texto. El proyecto también incluirá una investigación sobre la aplicación de técnicas o algoritmos orientados al Procesamiento del Lenguaje Natural usados en el análisis de textos o documentos obtenidos de Redes Sociales, por ejemplo y se persigue, a través de su empleo, la obtención de prototipos de sistema que faciliten el análisis en cuestión.

Los textos o documentos digitales, sobre los cuales se trabajará en este proyecto, se obtendrán principalmente desde la Web, considerando que, en la Sociedad del Conocimiento actual, la gestión de la información y conocimiento es un componente estratégico para el análisis inteligente de la información digital, para la clasificación de contenidos y la extracción de conceptos, entre algunos de los principales tópicos que estudia Text Mining.

La Minería de Datos es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de datos [1]. El Grupo de Investigación se inició en el año 2016 con un proyecto de categoría B denominado “Data Mining aplicado a análisis telefónico” (Código SeCTER D/B026). Luego, en 2018, se trabajó en otro proyecto (categoría B Código SeCTER D/B030) denominado “Implementación de técnicas específicas de Minería de datos en aplicaciones web con motores de Base de Datos Relacionales”. Algunos integrantes del Grupo de Investigación también participan de la beca EVC-CIN 2018 como Estímulo a la Vocación Científica, con los trabajos denominados “Detección de Plagio Implícito en Tesis de Grados” (Código 9411) y “Análisis Inteligente de Técnicas de Minería de Texto e Implementación para la Detección de Ciberacoso en Redes Sociales” (Código 12100).

Las temáticas Minería de Datos, Minería de Textos y Minería Web se desarrollan en la materia Aplicaciones de Base de Datos II; asignatura en la que algunos integrantes son docentes de la misma; de tal manera que los resultados de los proyectos de investigación y becas, antes mencionados, se vuelcan en dicha asignatura.

La minería de textos es el proceso de derivar información de alta calidad del texto que podría obtenerse de páginas web, por ejemplo. La información de alta calidad se alcanza, generalmente, a través de la elaboración de patrones y tendencias de medios tales como el

aprendizaje estadístico de patrones. La minería de texto generalmente implica el proceso de estructurar el texto de entrada (típicamente el análisis sintáctico, junto con la adición de algunas características lingüísticas derivadas y la eliminación de otras, y la posterior inserción en una base de datos), derivando patrones dentro de los datos estructurados y, finalmente, la evaluación e interpretación del resultado [2].

Al ser una temática que involucra técnicas de gran auge, desarrollo y relevancia en los últimos años [3], permite, a los alumnos de carreras informáticas de la Facultad de Ingeniería de la UNJu, plantear y desarrollar análisis inteligente sobre textos o documentación obtenida de la web, que se traducirán en propuestas de Proyectos Finales de Carrera.

Se destaca que los seres humanos son hábiles a la hora de producir e interpretar el lenguaje cotidiano porque son capaces de expresar, percibir e interpretar significados complejos en fracción de segundos; sin embargo, no son capaces de describir y comprender eficientemente las reglas que gobiernan el lenguaje natural. Este es el principal motivo por el cual entender y producir el lenguaje por medio de una computadora es un problema difícil de resolver. Este problema, pertenece al campo de estudio de Inteligencia Artificial llamado Procesamiento del Lenguaje Natural [4].

Por las características mencionadas, en el párrafo anterior, resulta razonable experimentar la aplicación del Deep Learning dentro de las posibles técnicas que se desarrollarán en el Text Mining y analizar los resultados obtenidos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación y desarrollo abarcan el estudio de las siguientes temáticas:

- Técnicas descriptivas de Minería de Textos.
- IDE´s o herramientas informáticas de Minería de Datos que empleen las técnicas mencionadas anteriormente.
- Obtención de contenido digital que permitan aplicar dichas técnicas, ya sea con técnicas de Minería Web u otros procesos similares.
- Análisis Inteligente sobre los resultados obtenidos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se espera a través de los resultados alcanzados y los objetivos en curso que los mismos se vuelquen a:

- Tesis de grado de alumnos de las carreras Ingeniería en Informática y Licenciatura en Sistemas de la Facultad de ingeniería de la UNJu, cuyas temáticas se enfoquen a la Minería de Textos, Minería de Datos y Análisis Inteligente.
- Alumnos que desarrollen estas temáticas aplicándolas en becas científicas, como las becas EVC-CIN de Estímulo a la Vocación Científica.
- Favorezcan la aplicación en las prácticas de los alumnos que cursen la cátedra Aplicaciones de Base de Datos II, perteneciente a la carrera Licenciatura en Sistemas.

Los conocimientos adquiridos se pondrán a disposición de cualquier

persona, institución u organización que desee implementar técnicas de herramientas de Minería de Textos para realizar Análisis Inteligente sobre datos existentes o de su propiedad mediante la generación de repositorios de información específicos.

Se destaca la vinculación existente entre la Facultad de Ingeniería y el Gobierno Provincial en relación a proyectos anteriores del Grupo de Investigación, por tal motivo es intención mantener y afianzar esta relación de trabajo e implementar técnicas que faciliten la toma de decisiones, en sus diferentes áreas.

4. FORMACIÓN DE RECURSOS HUMANOS

Cantidad de tesinas de grado en curso en el 2.019: 2 (cuatro integrantes).

Cantidad de tesinas de grado aprobadas en 2.019: 1 (dos integrantes).

El equipo que desarrollará la propuesta posee vasta experiencia en el trabajo docente con cursos de gran diversidad, en el área de tecnologías básicas y avanzadas, con promoción con distintas modalidades. Son sus miembros:

Director del Proyecto: Esp. Ing. Sánchez Rivero, Víctor David.
Investigadores Docentes:

- Esp. Ing. Inf. Farfán, José Humberto: director de beca EVC-CIN 2018 en curso.
- Ing. Rodríguez, Mariela Ester: CoDirectora de beca EVC-CIN 2018 en curso.
- Mg. Ing. Vega, Ariel Alejandro: director de beca EVC-CIN 2018 en curso.
- Ing. Vargas, Luis Alejandro.

Estudiantes Investigadores:

- Ramos, Pablo Nicolás: tesina de grado en curso, beca EVC-CIN 2018 de Estímulo a la Vocación Científica en curso.
- Garcete, Christian Brian.
- Churquina, Cintia Noelia.
- Águila, Viviana Emilia.
- Iogna Prat Genzel, Nicolas.

Egresados

- Llampa, Alvaro Facundo: tesina de grado aprobada, beca EVC-CIN 2018 de Estímulo a la Vocación Científica en curso.
- Contreras, Facundo: tesina de grado aprobada, beca EVC-CIN 2018 de Estímulo a la Vocación Científica en curso.

5. BIBLIOGRAFIA

[1] Maimon, O., & Rokac, L. (2010). ¨Data Mining and Knowledge Discovery Handbook", "O. Maimon, & L. Rokac, Data Mining and Knowledge Discovery Handbook", Nueva York, Springer, 2010, págs. 1-18.

[2] Ortiz A. (2018). "¿Qué es el análisis de texto, extracción de textos o minería de textos?". Obtenido de https://pcweb.info/que-es-analisis-de-texto-extraccion-mineria-de-textos en Junio de 2019.

[3] Ricardo & Barbosa. (2019). "Importancia de la minería de datos en el mundo empresarial actual". Obtenido de https://www.ricardo-barbosa.com/es/importancia-de-datos-mineria-en-hoy-negocios-mundo/ en Junio de 2019.

[4] López Briega R. (2019). IAAR, "Comunidad Argentina de Inteligencia

Artificial". Obtenido en https://iaarbook.github.io/procesamiento-del-lenguaje-natural/ en Junio 2019.

Análisis cienciométrico de la producción en investigación científica y tecnológica

Roberto M. Muñoz, Analía Guzmán, Martín M. Casatti,
Calixto Maldonado, Juan C. Cuevas, Luis E. Damiano

Departamento de Ingeniería en Sistemas de Información
CIDS – Centro de Investigación Desarrollo y Transferencia
de Sistemas de Información
Facultad Regional Córdoba – Universidad Tecnológica Nacional
Maestro Marcelo López esq. Cruz Roja Argentina – Córdoba
0351 – 4686385

rmunoz@frc.utn.edu.ar, aguzman@frc.utn.edu.ar, mcasatti@frc.utn.edu.ar,
cmaldonado@frc.utn.edu.ar, juancarloscue@gmail.com, luis.damiano@gmail.com

Resumen

El proyecto caracterizará la producción científica y tecnológica desarrollada por las universidades de la República Argentina, por medio de la elaboración de una metodología de análisis cienciométrico a partir de la documentación producida por los investigadores, becarios y centros de investigación. Se realizará un análisis de la información contenida en la documentación y se desarrollarán técnicas y herramientas para la extracción de la misma y el almacenamiento en un medio, que posteriormente permita el correcto análisis cienciométrico. Parte de este desarrollo incluye una plataforma web para la gestión administrativa y académica de congresos relacionados a tecnología informática y sistemas de información en una primera etapa, obteniendo así la información de análisis en tiempo real a partir de los artículos enviados por los autores a los congresos.

.

Palabras clave: *cienciometría – redes de colaboración – producción científica – métricas - patrones.*

Contexto

El presente trabajo forma parte del proyecto de investigación y desarrollo que ha sido homologado por la Secretaría de Investigación, Desarrollo y Posgrado de la Universidad Tecnológica Nacional (UTN), reconocido con el código: SIUTNCO0007848. El mismo se desarrolla en el Centro de Investigaciones, Desarrollo y Transferencia de Sistemas de Información – CIDS, de la Facultad Regional Córdoba de UTN.

1. Introducción

La cienciometría (scientometrics) como disciplina se dedica al "estudio de los aspectos cuantitativos de la comunicación científica, las prácticas de investigación y desarrollo, así como las políticas de ciencia y tecnología" [1]. Con la aparición de big data se logra que grandes conjuntos de datos puedan ser tratados y utilizados con fines analíticos y de

predicción, lo que hasta hace poco era imposible de procesar mediante el uso de herramientas tradicionales debido a su tamaño y complejidad. La combinación de "big data", "data mining", bibliometría y redes sociales, ha dado lugar a lo que se ha denominado como science mapping [2], cuyo objetivo es visualizar la estructura y las relaciones de la ciencia en general, y de las áreas científicas, disciplinas, líneas de investigación, autorías, etc., en particular.

Una cita es una referencia a un trabajo, ya sea propio o de otros autores o grupos, que se ha tomado total o parcialmente para construir el propio trabajo de investigación. En cierta medida el tipo y cantidad de citas que un trabajo de investigación registra es un indicador, con mayor o menor grado de fidelidad, del impacto que dicho trabajo produce en la comunidad científica. Un trabajo muy citado es la base para diversos papers científicos y líneas de investigación, mientras que un trabajo raramente citado no produce un efecto importante en la comunidad.

Otro análisis importante que se puede realizar por medio del análisis de citas es el de los patrones de colaboración entre instituciones o entre investiga-dores. Las redes de coautoría son una importante clase de redes sociales. El análisis de estas redes revela las características de las comunidades académicas que ayudan en el entendí-miento de los trabajos científicos colaborativos y en la identificación de los investigadores prominentes [3]. Es importante destacar que el análisis de redes de coautoría puede ayudar a caracterizar una cierta población de investigadores, determinando cuales son las temáticas de interés y la producción total y relativa por área temática. Si se considera que las relaciones entre los investigadores se puede representar como un grafo no dirigido existe una métrica que es de especial importancia a la hora de caracterizar dicho grafo. Dicha métrica es el grado de centralidad. La centralidad puede indicar, en un grafo que represente coautoría de trabajos de investigación, aquellos autores que tienen un grado alto de colaboración con otros autores. Por otra parte, una métrica asociada, la centralidad de intermediación, puede indicar aquellos autores que ofician de nexo entre otros investigadores o grupos de investigación.

Por otra parte, se pueden utilizar las redes de coautoría para evaluar la interrelación entre las instituciones académicas, el sector gubernamental y el sector productivo [7]. De esta manera es posible identificar el impacto generado por esa interrelación y se pueden diseñar políticas tendientes a reforzar esos vínculos para generar resultados beneficiosos a mediano y largo plazo.

Un estudio de extrema importancia, sobre todo en economías en donde la investigación científica y tecnológica no es prioritaria en las políticas de gobierno, es el análisis que establece relaciones entre la producción científica y tecnológica y el ámbito de los investigadores sobre todo desde el punto de vista socioeconómico [1] [4] . Desde este punto de vista también se puede analizar el surgimiento o decadencia de ciertas líneas de investigación.

Todas estas prácticas entran dentro de las incumbencias del denominado mapeo científico o en su nombre original "science mapping" [5].

"El objetivo central de "science mapping" es el de revelar la estructura y dinámica del conocimiento científico …", tal como menciona el trabajo de Chen, Dubin y Schultz [9]. Los análisis cienciométricos sirven como punto de partida para contar con información

cuantitativa, heterogénea y relacionada para poder plantear análisis de science mapping [8].

Hay que mencionar que, a la hora de optimizar el uso de recursos escasos, como pueden ser los lugares físicos para laboratorios, las plazas en las plantas de investigadores permanentes, los recursos económicos, de materiales o de insumos, es fundamental contar con un diagnóstico adecuado de la situación actual del ámbito en donde esos recursos se van a utilizar. En este tema en particular el análisis cienciométrico puede servir para tener una imagen actualizada y confiable de la infra-estructura de I+D+i a los fines de asistir en la toma de decisiones [6].

2. Líneas de Investigación, Desarrollo e Innovación

En el presente trabajo se explorarán las siguientes líneas de investigación, las que pueden ser adaptadas o adecuadas de acuerdo a los hallazgos producto del propio trabajo de investigación:

- Estudio de la representación de información vinculada a la producción científica y redes de colaboración entre investiga-dores y centros de investigación.
- Revisión y/o implementación de mecanismos de extracción de información de mapeo a partir de documentos de difusión de resultados de investigación (papers).
- Estudio de características comunes en diversos tipos de instrumentos y medios de difusión de resultados de investigación y obtención de atributos que definen la temática, metodología y resultados obtenidos en dichos trabajos.
- Caracterización de los participantes de redes de colaboración, ya sea desde el punto de vista personal o institucional y estudio de la representación de dichas características a fines de su consulta.
- Estudio de las bases de datos de grafo como un mecanismo idóneo para la representación de la información necesaria para un "science mapping" exitoso.
- Estudio de los mecanismos de consulta y visualización de la información, teniendo en cuenta la heterogeneidad y el posible gran tamaño de los resultados de las consultas.

3. Resultados esperados

El objetivo general del presente estudio es caracterizar la producción e impacto de las líneas de investigación que llevan adelante las universidades argentinas, por medio de una herramienta desarrollada ad-hoc.
Para ello se han identificado los siguientes objetivos particulares, para obtener los resultados esperados:
1. Modelar la representación de la información asociada a las tareas de I+D, que permita caracterizar todos los trabajos producidos por las universidades argentinas.
2. Estructurar un sistema de almacenamiento cienciométrico que permita la registración y consulta de la información registrada.
3. Procesar los documentos de los congresos, para extraer los atributos para el análisis cienciométrico e inferir lo que no se obtiene directamente de los textos.
4. Diseñar, desarrollar e implementar un sistema de gestión de congresos y conferencias para registrar automáticamente la información en el sistema de almacenamiento, a partir de los papers y artículos presentados, sin necesidad de interacción adicional o carga manual de datos.
5. Desarrollar una herramienta de consulta y análisis que permita obtener métricas asociadas a la producción científica, que posibilite

analizar los diversos patrones que pudieran surgir a partir del análisis de citas, autores y líneas de investigación.

Como se puede apreciar, los objetivos del proyecto se agrupan en dos enfoques muy particulares:

- El análisis y la caracterización de la investigación producida por las universidades argentinas, en base a información preexistente.
- La construcción de una herramienta de software que permita realizar dichos análisis sobre futuros trabajos y líneas de investigación.

4. Formación de Recursos Humanos

El equipo de investigación está formado por: 6 docentes, de diferentes asignaturas de la carrera de Ingeniería en Sistemas de Información, 1 graduado y 3 estudiantes.

Los docentes investigadores explorarán los usos de una metodología, que les permitirá ampliar sus posibilidades, en cuanto a líneas de investigación de interés, tanto local como regional o internacional, y accederán a nombres y trabajos de posibles colaboradores en sus respectivos campos de estudio. De los docentes integrantes del equipo, dos de ellos realizarán su tesis en el marco del presente proyecto, la cual les permitirá finalizar los estudios de posgrado en la Maestría en Ingeniería en Sistemas de Información.

Dos estudiantes de grado, que aún cursan asignaturas y no poseen experiencias en investigación, podrán realizar sus primeras experiencias en este proyecto, y uno, que ya posee experiencia en proyectos, podrá ampliar los conocimientos adquiridos durante el cursado, en el desarrollo de las herramientas informáticas que se describen en los objetivos. Los 3 estudiantes podrán proponer la temática en sus Prácticas Supervisadas, contando con el asesoramiento y seguimiento de los docentes integrantes del PID como Docentes Tutores.

La inclusión del reciente graduado se produce por su experiencia en anteriores Proyectos de Investigación académico y en el Laboratorio de la Carrera, como también por su rendimiento académico.

5. Referencias

[1] RENDÓN, ESMERALDA CERVANTES AND VICTORIANO GARZA-ALMANZA (2015). "La cienciometría como herramienta para analizar el impacto de la investigación científica en una región." Cultura Científica y Tecnológica 48.

[2] M.J. COBO, A.G. LÓPEZ-HERRERA, E. HERRERA-VIEDMA AND F. HERRERA (2011). "Science mapping software tools: Review, analysis, and cooperative study among tools". Journal of the American Society for Information Science and Technology, vol. 62, nº 7, pp. 1382-1402.

[3] UMADEVI, V. (2013). "Case study–centrality measure analysis on co-authorship network" Journal of Global Research in Computer Science, 4(1), 67-70.

[4] LEYDESDORFF, LOET, AND ISMAEL RAFOLS (2011). "Local emergence and global diffusion of research technologies: An exploration of patterns of network formation." Journal of the American Society for Information Science and Technology 62.5: 846-860.

[5] EBENER, STEEVE, et al. (2006). "Knowledge mapping as a technique to support knowledge translation." Bulletin of the World Health Organization 84: 636-642.

[6] MCGRATH, W. (1989). "What bibliometricians, scientometricians and informetricians study; a typology for definition and classification; topics for discussion" International Conference on Bibliometrics,

Scientometrics and Informetrics (Vol. 2). Ontario: The University of Western Ontario.
[7] SÁNCHEZ, M. A., SCHMIDT, M. A., & OBIOL, L. C. (2018). "Redes de coautorías y patrones de colaboración institucional: análisis de un caso en el campo de la ingeniería".
[8] ROBLES, E., & VINCK, D. (2010). "Mapeo de la evolución de la producción de ciencias y tecnologías emergentes: comparación de estrategias lexicales sobre las nanociencias y nanotecnologías". In ESOCITE 2010.
[9] CHEN, C., DUBIN, R., & SCHULTZ, T. (2015). "Science mapping". In Encyclopedia of Information Science and Technology, Third Edition (pp. 4171-4184). IGI Global.

WEB-GIS ORIENTADO AL MONITOREO Y EVALUAR EL CONTROL DE LA POBLACIÓN DE PERROS CALLEJEROS EN PATAGONIA

Diego E. Procopio(1,3), Maria Eugenia de San Pedro(1,2), Marta Lasso(1,2)

Unidad Académica Caleta Olivia - Universidad Nacional de la Patagonia Austral

(1)Universidad Nacional de la Patagonia Austral-Santa Cruz, Argentina (UNPA)

(2)Instituto de Tecnologías Aplicadas, Unidad Académica Caleta Olivia (ITA)

(3) Centro de Investigaciones Puerto Deseado, Instituto de Ciencias del Ambiente, Sustentabilidad y Recursos Naturales (ICASUR)

dproco@hotmail.com , {edesanpedro,mlasso}@uaco.unpa.edu.ar

RESUMEN

La necesidad de brindar accesibilidad a la información estadística espacial provenientes de los Sistemas de Información Geográfica a (GIS) a través de la web, pondera la gestión de algunos campos de la ciencia, siendo relevante y muy utilizadas en la resolución de los problemas ambientales, acentuando en los últimos tiempos su uso mediante la implementación de Webmapping.

Esto ha llevado a reconstruir todo el proceso de toma de decisiones, a diferenciar a un GIS como sistemas de soporte de decisión espacial y de planificación integrado, contando con una plataforma Web para la carga, visualización, manejo y análisis de datos temporales y geográficos.

CONTEXTO

El grupo de investigación que trabaja en este proyecto, es interdisciplinario y está constituido por docentes-investigadores de dos Institutos, el primero desarrolla sus actividades orientadas a la Investigación, Desarrollo, Implementación y Mantenimiento de Sistemas y Tecnologías, el segundo desarrollan sus actividades orientadas al estudio de la biología, ecología y monitoreo de las poblaciones de la fauna silvestre, comenzando a desarrollar estudios de ecología urbana, puntualmente en el monitoreo de perros callejeros.

Este grupo de investigación desde hace unos años está poniendo a prueba la tecnología Web-GIS y se encuentra trabajando en identificar, diversas necesidades funcionales para profundizar en el marco tecnológico; y poder brindar soluciones con una problemática creciente y común a gran parte de los municipios de todo el país, relacionada al crecimiento poblacional de perros callejeros, dando apoyo en particular al municipio de Puerto Deseado (Santa Cruz), utilizándolo como caso de estudio, pero en general buscando estandarizar, coordinar con otros municipios, de manera de obtener diagnósticos verdaderos y eficientes de la situación, para poder estimar algunos parámetros de la población de perros callejeros, evaluar la necesidad de un control poblacional, planificar la intervención, evaluar dicha intervención a realizar y finalmente dejar disponible el acceso a la web-gis a los entes encargados de toma de decisiones para una mejor gestión.

1. INTRODUCCIÓN

La problemática ambiental que estamos abarcando en la zona patagónica está relacionada con el crecimiento poblacional de perros que deambulan libremente por la vía pública (Serpell 1995), siendo este un tema común en todas las ciudades de países en desarrollo y subdesarrollados (Majumber et al. 2014), no escapando esta a los pueblos de la Patagonia,

cuyas consecuencias afectan a la salud pública, bienestar animal y a la salud ambiental.

Por otro lado, la falta de recursos o de interés de los organismos encargados de financiar o apoyar este tipo de iniciativas, ha sido también otro factor a considerar a la hora de elegir la problemática ambiental, al igual que el costo político que tiene trabajar con poblaciones de perros sin dueño, en donde algunas ONGs proteccionistas tratan de cogobernar contraponiéndose a las medidas de manejo, además de minimizar la persistencia de problemas sanitarios como son la transmisión de enfermedades infecciosas y parasitarias, focos insalubres, contaminación del ambiente por deposiciones de los perros (Acha y Szyfres, 1986 , Cardinal et al. 2006, Rubel et al.2003, 2005, Dopchiz et al. 2013), otros como la latencia de la rabia por la existencia de un ciclo silvestre de carácter endémico en murciélagos de las ciudadades (Angeleri 2018), lesiones físicas, accidentes viales, disputas entre vecinos y daños en propiedades (Ashby, 1996) y ataques e interferencias en la ganadería y fauna silvestre, entre otros (Home et al. 2017, ,Vanak AT y Gompper ME 2009,2010,)

El costo-beneficio de los programas de control poblacional, sólo se puede estimar si se conoce el tamaño de la población y algunas características de la misma (Ibarra et al., 1991; Escárate y Briones, 2003) y se cuenta con precisiones sobre los problemas generados.

La falta de información en Puerto Deseado sobre la población de perros callejeros, motivaron que parte de los objetivos de este trabajo consistieran en estimar la demografía de perros callejeros en la ciudad (cantidad y densidad), conocer la distribución de los recursos disponibles para los perros callejeros (alimento, agua y refugio), evaluar y registrar en un futuro los diferentes efectos que puedan tener sobra las densidades de perros callejeros: las campañas de castraciones, registro, adopción, las acciones relacionadas a limitar la presencia de agua, alimento y refugio en la vía pública, las denuncias de las incidencias de los perros callejeros en la vía pública, así como también en la reserva natural y los campos ganaderos, y las acciones realizadas por los inspectores de control canino a través de las actas de infracción.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

La metodología se basa en establecer una combinación de diversos componentes, tales como de GISWeb, Estadística Espacial, Información Georeferenciada y Bases de Datos. Las herramientas seleccionadas se encuentran bajo la licencia de Software Libre. Las mismas se encuadran en proyectos sustentables y en constante evolución e interacción entre ellas, lo que facilita la integración de una forma clara. La Arquitectura de Sistema se orienta a Estructura de Cliente/Servidor. Los resultados obtenidos permiten ser divulgados a través de la Web y se integran en un modelo, que permite graficar y simular el comportamiento del fenómeno de la proliferación de perros callejeros y evaluación del control. La construcción de la Web-Gis se realiza a través de la Suite R (R Development Core Team 2008) permitiendo modelar el fenómeno a estudiar. El rango de aplicaciones es de gran amplitud, desde la importación de datos hasta técnicas avanzadas como el krigeado. Permitiéndonos desarrollar un entorno Web a través del paquete Shiny que permite crear aplicaciones web interactivas, directamente desde Rstudio. Y tiene como finalidad la interacción gráfica con los datos permitiendo a los usuarios interactuar con sus datos sin tener que manipular el código (reactividad). En este entorno Web hemos utilizando algunas librerías de R tales como leaflet, sp, ggplot, raster, rgdal, y maptools entre otros (Figura 1).

En relación al manejo de problemática ambiental, la metodología utilizada para los relevamientos en el terreno, fue a través de planillas electrónicas

con el software libre Cybertracker (CyberTracker Sofware (Pty) Ltd Reg no 97/01908/07, http://www.cybertracker.co.za) para ser utilizadas en teléfonos inteligentes (Smatphones) aprovechando la tecnología GPS para obtener datos georeferenciados, mejorando la recolección de datos en el terreno, haciéndola menos costosa y más simple y facilitando la transferencia de los datos recogidos a bases de datos y a múltiples formatos de exportación para su posterior análisis. Esta planilla electrónica permite la colección de gran cantidad de datos, de forma rápida y sin errores, que pueden posteriormente volcarse a un formato digital a través de la sincronización del dispositivo Android Smartphone con la PC. Por lo que se realizó una planilla electrónica para los censos de perros callejeros anuales y otras dos para los relevamientos diarios realizados por los inspectores, de la direcciones de Control Canino y de Ambiente y Desarrollo Sustentable Municipal.

Para la recolección de datos, la ciudad de Puerto Deseado fue dividida en 12 zonas, teniendo en cuenta las arterias principales para tener una buena organización de las calles durante los recorridos. El esfuerzo en promedio de cada censo de perros callejeros y el registro de los recursos tróficos y de refugio, llevó en promedio 6 días en el mes de abril, se realizaron en vehículo a una velocidad de 5 km/h y recorriendo un promedio de 110 km en cada censo.

En los recorridos cuando un perro callejero era avistado, se registraron: el agrupamiento (solitario o en jauría), raza, edad, sexo, tamaño, si poseía propietario (con collar y sin collar), salud, comportamiento y actividad. En relación a los recursos, se registraron recursos tróficos disponibles (basura, comederos, contenedores, etc.), agua (charcos, grifos abiertos, pérdidas de agua, etc.) y refugio (baldíos, obras abandonadas, terreno con escombros, etc.).

Con los contactos de los perros callejeros georreferenciados y utilizando las librerías de R, la primera intención es determinar los descriptores más básicos de la distribución espacial, como son las estadísticas centro gráficas, sí los signos observados exhiben algún tipo de patrón sistemático, en oposición a una distribución aleatoria por medio índice del vecino más cercano, y a que distancias ocurre esas distribuciones utilizando la estadística K de Ripley. Luego de ésta primera etapa de determinar, se exploró como la población de perros callejeros utilizan el área de estudio y se resolvió estimando la intensidad de uso mediante una interpolación utilizando la función Kernel.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados obtenidos permiten ser divulgados a través del servidor de la Universidad Nacional de la Patagonia Austral-Unidad Académica Caleta Olivia, a través de la dirección www.uaco.unpa.edu.ar/prueba , y se integran en un modelo, que permite graficar y simular el comportamiento del fenómeno de la proliferación de perros callejeros del 2016 al 2019. Hasta el momento se visualizan las capas de las localizaciones de perros callejeros, las capas espaciales estadísticas, capas provistas por Google Map, Google Satelite y capa de Google Map Street. También se avanzó en la interfaz gráfica para la selección de capas, funcionalidad de zoom y desplazamiento, reglas, etiquetas y leyendas, así como también en la dinámica de algunos gráficos referidos a la problemática. Se espera en una segunda etapa incorporar reactividad en la estructura del Web-Gis que permitirá a los usuarios interactuar con los datos sin tener que manipular el código.

Durante los censos de perros callejeros se almacenaron 9400 registros georreferenciados en 4.89 km2, de los cuales 1.058 de los registros, fueron perros callejeros en el 2016, 1372 en el 2017, 1600 en el 2018 y 1451 en el 2019. Los perros callejeros fueron representados en un 68.4 % por perros mestizos en el 2016 disminuyendo al 52% en el 2019, siendo el resto de los perros callejeros de diferentes razas, en orden importancia, compuestos por: Caniche, Pastor

alemán, Border collie, Labrador retriever, Golden retriever y Pequines.

La distribución de los perros callejeros para los diferentes años siempre tendieron a la agregación. Las clases de densidades relativas de perros callejeros, fueron estimadas por la función de densidad Kernel (función cuartica) y con un radio de búsqueda fijo de 42 metros (considerado por la K de Ripley). Se representaron 5 clases de densidades relativas para cada censo que van por ejemplo para el año 2016 de 1 a 95 perros/ Km^2 en las zonas menos densas, a zonas más densas de 864 a 1.793 perros/km^2 o para el año 2019 densidades de 1 a 90 perros/ Km^2 en las zonas menos densas, a zonas más densas de 746 a 1456 perros/km^2 (visitar la página web). Si bien no existe grandes diferencias se pudo observar una tendencia al aumento anual entre el 2016 y 2017 del 29.6%, del 16.6% entre el 2017 y 2018 y una leve disminución del 9.3% entre el 2018 y 2019, indicando que el manejo de la población en el último año comienza a hacer efecto. Esta disminución en la población se debe a una disminución del 15.4% de presencia en las calles de los perros de clase de edad menor a un año y del 12. 1% de los perros adultos, así como también puede estar afectada por la disminución de jaurías de 2 a 5 perros callejeros en un 18.4%, en. En una segunda etapa se espera avanzar con los factores que favorece a la presencia de los perros en las calles y a la medición de las distintas medias de manejo que comenzaron a desarrollarse en la ciudad de puerto Deseado.

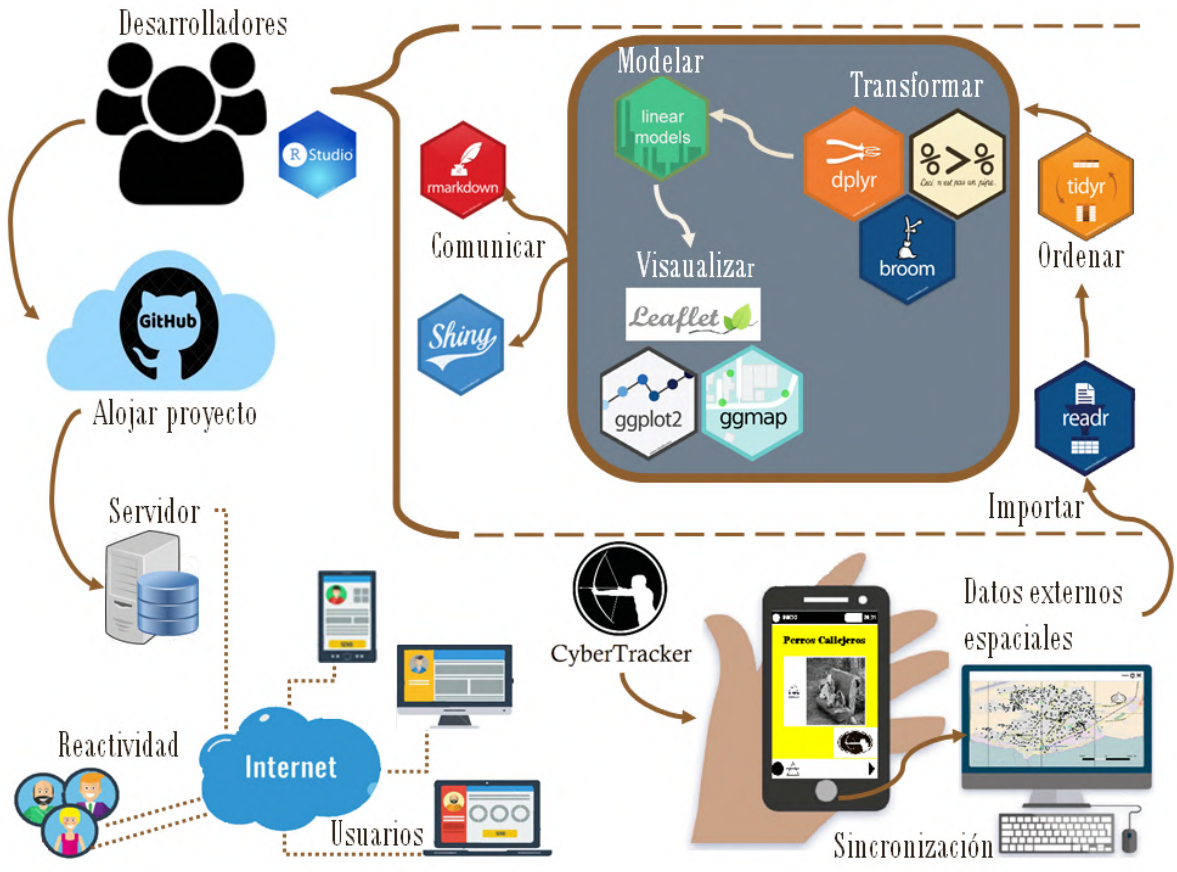


Figura 1 Esquema de Integración GISWeb : mostrando la arquitectura orientada al servicio, señalando algunos de los paquetes de R utilizados y la obtención de datos espaciales externos.

4. FORMACIÓN DE RECURSOS HUMANOS

Un integrante se encuentra en la etapa de finalización de su tesis de maestría.

Anualmente se incorporan al proyecto de investigación, alumnos becarios a través de las becas de iniciación a la investigación que convoca la UNPA.

5. BIBLIOGRAFÍA

Acha, P., Szyfres, B., 1986. Zoonosis y enfermedades transmisibles comunes al hombre y a los animales. Publicación Científica N° 503. Segunda edición. Washington D.C. USA. OPS/OMS. 708 páginas.

Angeleri P.2018. Guia para la prevención, vigilancia y control de la rabia en Argentina. Dirección Nacional de Epidemiologia y Análisis de la Situación de Salud.

Ashby, K., 1996. Dog bites. Hazard, 30: 7-13.

Cardinal , M.V. , Castañera , M.B. , Lauricella , M.A., Cecere , M.C. , Ceballos , L.A. , Vazquez–Prokopec, G.M. , Kitron , U. , and Gürtler , R.E. (2006). A prospective study of the effects of sustained vector surveillance on Trypanosoma cruzi infection of dogs and cats in rural northwestern Argentina. American Journal of Tropical Medicine and Hygiene, 75, 753– 61.

Dopchiz, M., Lavallén, C., Bongiovanni, R., Gonzalez, P., Elissondo, C., Yannarella, F., & Denegri, G. (2013). Endoparasitic infections in dogs from rural aereas in th Lobos District, Buenos Aires province, Argentina. Revista Brasileira de Parasitología Veterinaria, 22:92-97.

Escárate, P., Briones, F., 2003. Académicos de la Unicit presentaron estudio sobre la población de animales domésticos en Santiago. http:www.universia.cl/portada/actualidad/noticia_actualidad.

Ibarra, L., Espínola, F., & Echeverría L., M. (2006). Una prospección a la población de perros existente en las calles de la ciudad de Santiago, Chile. Avances en Ciencias Veterinarias, 21(1-2). doi:10.5354/0719-5273.2010.3953

Ibarra, M., Núnez, F., Cisternas, L., Méndez, M., 1991. Demografía canina y felina en la Comuna de La Granja, Santiago, Chile. Avances en Medicina Veterinaria, 6: 141-145.

Ibarra, L., Morales, M., & Acuña, P. (2003). Aspectos demográficos de la población de perros y gatos en la ciudad de Santiago, Chile. Avances en Ciencias Veterinarias, 18(1-2). doi:10.5354/0719-5273.2010.9163

Majumber S., Chatterjee A., Bhadra A. A Dog's Day with Humans—Time Activity Budget of Free-Ranging Dogs in India. Curr. Sci. 2014;106:874–878.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rubel, D., Wisnivesky, C., 2005. Magnitude and distribution of canine fecal contamination and helminth eggs in two areas of different urban structure, Greater Buenos Aires, Argentina. Veterinary Parasitology, 133: 339-347.

Rubel , D. , Zunino , G. , Santillán , G., and Wisnivesky , C. (2003). Epidemiology of Toxocara canis in the dog population from two areas of different socioeconomic status, Greater Buenos Aires, Argentina. Veterinary Parasitology , 115 , 275 – 86

Vanak AT y Gompper ME (2009). Dietary niche separation between sympatric free-ranging domestic dogs and Indian foxes in central India. J Mamm 90:1058–1065.

Vanak, A. T. y Gompper, M. E. (2010), Interference competition at the landscape level: the effect of free-ranging dogs on a native mesocarnivore. Journal of Applied Ecology, 47: 1225-1232. doi:10.1111/j.1365-2664.2010.01870.x

Computación Gráfica, Imágenes y Visualización

VERIFICACIÓN Y VALIDACIÓN DE REPRESENTACIONES VISUALES Y SUS INTERACCIONES

Martín Larrea, Matías Selzer, Dana K. Urribarri,
M. Luján Ganuza, Clara Casalini, Silvia M. Castro
Laboratorio de I+D en Visualización y Computación Gráfica
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (DCIC-UNS)
Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)
Laboratorio de I+D en Visualización y Computación Gráfica,
(UNS-CIC Prov. de Buenos Aires)
{mll, matias.selzer, dku, mlg, mcca, smc}@cs.uns.edu.ar

RESUMEN

Cada día más decisiones son tomadas en base al análisis visual mediante el uso de representaciones visuales de grandes conjuntos de datos. No sólo hay un aumento cuantitativo sino también cualitativo, las decisiones son más críticas y con mayor impacto en la sociedad, el medio ambiente y los individuos. En este contexto resulta fundamental desarrollar nuevas y mejores metodologías y herramientas que permitan asegurarnos el correcto funcionamiento de las representaciones visuales y sus interacciones.

Palabras Clave: Verificación y Validación, Visualización de Datos, Análisis Visual de Datos

CONTEXTO

Este trabajo se lleva a cabo en el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab, UNS-CIC Prov. de Buenos Aires) del Departamento de Ciencias e Ingeniería de la Computación, de la Universidad Nacional del Sur. Los trabajos realizados bajo esta línea involucran a docentes investigadores, becarios doctorales y alumnos de grado.

1. INTRODUCCIÓN

La toma de decisiones se ha transformado en una herramienta clave en toda organización, pasando de ser un proceso basado en la experiencia y la intuición a uno cada vez más establecido en el análisis de datos. En este contexto, en el último tiempo se ha comenzado a hablar de Big Data; es el concepto que hace referencia a conjuntos de datos tan grandes y complejos donde los sistemas informáticos tradicionales de procesamiento de datos no pueden tratarlos adecuadamente. Por ende, los procedimientos usados para tomar decisiones en base a estos datos son más sofisticados y requieren software especializado. Un tipo de software especializado para Big Data, es el de representaciones visuales para grandes conjuntos de datos. Las representaciones visuales son de suma importancia en la actualidad ya que permiten la exploración efectiva de un conjunto de datos y facilitar la tarea de identificar patrones y extraer conclusiones. Los sistemas de representaciones visuales no sólo ofrecen una vista de los datos, sino también proveen interacciones mediante las cuales el usuario puede manifestar sus necesidades y a partir de las cuales obtener la perspectiva deseada del conjunto de datos. Cada día más decisiones son tomadas en base al análisis visual mediante el uso de representaciones visuales de grandes conjuntos de datos. No sólo hay un aumento cuantitativo sino también cualitativo, las decisiones son más críticas y con mayor impacto en la sociedad, el medio ambiente y los individuos. En este contexto resulta

fundamental desarrollar nuevas y mejores metodologías y herramientas que permitan asegurarnos el correcto funcionamiento de las representaciones visuales y sus interacciones ([1]).

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Basándonos en el estado del arte de las representaciones visuales de grandes conjuntos de datos ([2]), y teniendo en cuenta sus aplicaciones ([3]), los estados intermedios de los datos ([4]) y las posibles interacciones ([5]), es posible definir e implementar metodologías y herramientas que permitan detectar errores y/o problemas en los softwares de representaciones visuales de grandes conjuntos de datos. La principal línea de investigación y desarrollo consiste en el estudio y desarrollo de metodologías para la validación y verificación de representaciones visuales de grandes conjuntos de datos. De esta línea principal se desprenden sublíneas relativas al diseño y desarrollo de herramientas que asistan en el desarrollo de representaciones visuales de grandes conjuntos de datos para asegurar la calidad del producto final. Estas herramientas buscarán eliminar o reducir la incidencia de errores en los sistemas de representación visual. Esta línea de investigación expande nuestro trabajo sobre el área de Visualización ([6, 7, 8, 9, 10]).

3. RESULTADOS OBTENIDOS Y ESPERADOS

Sobre los ejes presentados se han obtenido resultados parciales. Desde el punto de vista de la teoría de base se ha comenzado a realizar un relevamiento de las diferentes metodologías disponibles en el ámbito de la Verificación y Validación del Software que sean aplicables a un dominio tan particular como el de la Visualización de Información. En lo referido al desarrollo de nuevas metodologías y herramientas se ha comenzado a publicar los resultados obtenidos hasta el momento. Tal es el caso de "Black-Box Testing Technique for Information Visualization. Sequencing Constraints with Low-Level Interactions" ([11]) y "White-Box Testing Framework for Object-Oriented Programming based on Message Sequence Specification" ([12]), este último seleccionado como mejor trabajo para la "Communications in Computer and Information Science" de Springer. En el corriente año se pretende avanzar sobre una herramienta específica para el testeo de visualizaciones y la publicación de nuevas metodologías.

4. FORMACIÓN DE RECURSOS HUMANOS

En lo concerniente a la formación de recursos humanos se detallan las tesis en desarrollo y concluidas relaciones con las líneas de investigación presentadas, así también como un proyecto en evaluación

Tesis en Desarrollo: "GENERACIÓN DE DATOS DE PRUEBA PARA VERIFICACIÓN Y VALIDACIÓN DE SOFTWARE EN LA INDUSTRIA MEDIANTE ALGORITMOS GENÉTICOS", tesis de grado para la Ingeniería en Sistemas de Información. Alumna: Constanza Giorgetti. Director: Martín Larrea, Colaborador: Matías Selzer. Para esta alumna también se solicitó una Beca de Inicio en la Investigación de la Universidad Nacional del Sur. La misma se encuentra en evaluación.

Tesis en Desarrollo: "VERIFICACIÓN Y VALIDACIÓN DE SOFTWARE EN ENTORNO SYMFONY", proyecto final de carrera para la Ingeniería en Computación. Alumno: Nehuen Antiman. Director: Martín Larrea.

Proyecto: PGI 2019 “VERIFICACIÓN Y VALIDACIÓN DE REPRESENTACIONES VISUALES Y SUS INTERACCIONES”. Director: Martín Larrea, Miembros: Matías Selzer, Dana K. Urribarri, M. Clara Casalini. Proyecto presentado para acreditación de incentivos, en evaluación.

5. BIBLIOGRAFÍA

[1] Kirby, Robert M., and Cláudio T. Silva. “The need for verifiable visualization”. Computer Graphics and Applications, IEEE 28.5, 78-83, 2008.

[2] Rees, D., and R. S. Laramee. "A Survey of Information Visualization Books." Computer Graphics Forum. 2019.

[3] Börner, Katy, Andreas Bueckle, and Michael Ginda. "Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments." Proceedings of the National Academy of Sciences 116.6 (2019): 1857-1864.

[4] Hagen, Hans. "DFU, Volume 2, Scientific Visualization: Interactions, Features, Metaphors, Complete Volume." Dagstuhl Follow-Ups. Vol. 2. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2012.

[5] Ganuza, M. L. "Thesis Overview: Interactions in Visualization." Journal of Computer Science & Technology 18.2 (2018): 178-179.

[6] Castro, Silvia Mabel, et al. "Métricas, técnicas y semántica para la visualización de datos." XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste).. 2018.

[7] Urribarri, Dana K., and Silvia M. Castro. "Prediction of data visibility in two-dimensional scatterplots." Information Visualization 16.2: 113-125. 2017

[8] Ganuza, María Luján, et al. "Visualización de Datos aplicada a resolver problemas en Ciencias Geológicas." XVI Workshop de Investigadores en Ciencias de la Computación. 2014.

[9] Larrea, Martín Leonardo, et al. "Ontologías y Semántica en el Proceso de Visualización." XVI Workshop de Investigadores en Ciencias de la Computación. 2014.

[10] Selzer, Matías Nicolás, et al. "Modelos de interacción y aplicaciones en realidad virtual mediante dispositivos móviles." XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires). 2017.

[11] Larrea, Martín Leonardo. "Black-Box Testing Technique for Information Visualization. Sequencing Constraints with Low-Level Interactions." Journal of Computer Science & Technology 17. 2017.

[12] Silva, J.I.R., Larrea, M.: White-box testing framework for object-oriented programming based on message sequence specification. In: XXIV Congreso Argentino de Ciencias de la Computación (Tandil, 2018). 2018.

Avances en Línea de Investigación Doctoral: Integración Escalable de Realidad Aumentada Basada en Images y Rostros

Nahuel Mangiarua[1[0000−0003−2674−7324]], Jorge Ierache[1[0000-0002-1772-9186]], Martin Becerra[1], Maria Jose Abasolo[2[0000-0003-4441-3264]]

[1] Universidad Nacional de La Matanza, La Matanza, Buenos Aires, Argentina
nmangiarua@unlam.edu.ar

[2] Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina

Resumen

El presente trabajo resume una de las líneas de investigación doctoral en curso radicada en el Grupo de Realidad Aumentada Aplicada la cual se orienta al desarrollo de una arquitectura y prototipo demostrador que integre de forma escalable la realidad aumentada basada en imágenes y la aumentación de rostros humanos.

Palabras clave: Realidad Aumentada, Reconocimiento de Rostros

Contexto

La investigación es desarrollada en el grupo de investigación de Realidad Aumentada Aplicada del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza, en el marco de un proyecto de tesis doctoral.

Introducción

La realidad aumentada mediante teléfonos móviles es una de las subáreas de investigación de realidad aumentada que más está creciendo en la actualidad. Se define a un sistema de Realidad Aumentada (RA) como un sistema que permite a usuarios interactuar en tiempo real con contenidos virtuales asociados a objetos del entorno real [1],[2]. Podemos ampliar esta definición como un sistema que aumenta los sentidos de la persona (Visión, audio y tacto) para permitir la interacción con contenidos virtuales del mundo digital que son invisibles en el entorno real [3]. El grupo de realidad aumentada aplicada realizó diversos trabajos en el campo de realidad aumentada que permite a usuarios sin mayores conocimientos técnicos de RA poder crear catálogos virtuales aumentados agrupando contenidos virtuales asociados a marcadores de RA

como imágenes impresas[4],[5]. En el área de educación el grupo de investigación ha realizado diversas aplicaciones experimentales [6],[7] orientadas a juegos, como así también aplicaciones experimentales destinadas a la explotación de materiales didácticos para el área educativa [8],[9].

Se presenta en las secciones siguientes la línea de investigación de tesis doctoral, Integración Escalable de Realidad Aumentada Basada en Imágenes y Rostros.

Líneas de Investigación, Desarrollo e Innovación

Integración de RA escalable

La línea de investigación doctoral se orienta al desarrollo de una arquitectura y prototipo demostrador que integre de forma escalable la realidad aumentada basada en imágenes y la aumentación de rostros humanos.

En cuanto a la arquitectura, se define un pipeline basal para la aumentación de imágenes y otro para la aumentación de rostros humanos como se muestra en las figuras 1 y 2.



Figura 1: Pipeline para AR basada en imágenes



Figura 2: Pipeline para reconocimiento y aumentación de rostros

Para la aumentación de imágenes, primero se debe, en forma offline, buscar puntos de interés (POI) y computar descriptores robustos sobre los mismos para cada imagen que se quiera aumentar con contenido virtual. Luego, durante la explotación, se buscan los mismos POI sobre las imágenes que van siendo capturadas desde un flujo de video y se computan los mismos descriptores robustos para los cuales se busca correspondencias contra los cálculos previamente offline para las imágenes a aumentar. Si tenemos coincidencias suficientes, podemos efectivamente detectar la presencia en el video de una de las imágenes a aumentar y luego calcular su posición y rotación relativas a la camara.

En el caso de la aumentación de rostros, análogamente se computan offline descriptores robustos para cada persona cuyo rostro quiera ser reconocido mediante el uso de una red neuronal convolucional (CNN) [10],[11],[12]. Luego se utiliza una otra CNN entrenada para detectar la posición de rostros humanos sobre imágenes que vengan de un flujo de video y para cada detección, se computa un descriptor con la primera CNN. Los descriptores encontrados en las

imágenes provenientes del flujo de video deben entonces ser buscados por coincidencias entre los calculados previamente de forma offline y así identificar al individuo.

Se observa que ambos pipelines comparten un paso en común, la búsqueda de correspondencias entre descriptores robustos. A su vez, si analizamos la complejidad computacional de ambos pipelines, encontramos que mientras la mayoría de los pasos dependen del tamaño de las imágenes a procesar el cual puede ser fijado a un nivel adecuado en cada implementación concreta. Sin embargo, el paso de búsqueda de coincidencias depende del número de imágenes o rostros a aumentar, volviéndose el cuello de botella al momento de escalar el sistema.

Resultados y Objetivos

Bajo estas consideraciones se propone entonces la utilización de algoritmos de búsqueda de coincidencias aproximadas (approximate nearest neighbour, ANN) que ofrezcan una complejidad computacional sub lineal y estabilidad ante el incremento en el volumen de datos a procesar. Existen en la literatura infinidad de algoritmos de ANN y es actualmente un campo de investigación activo.

Se toma como objetivo analizar una muestra significativa de dichos algoritmos para elegir el idóneo para esta tarea particular en el contexto de la RA. A su vez resulta necesario definir y adaptar conjuntos de datos de prueba representativos del contexto de la RA, seleccionando o creando las métricas específicas necesarias.

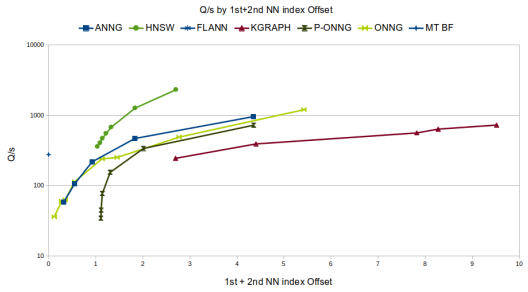


Figura 3: Análisis de la curva precisión sobre velocidad utilizando métricas y juegos de datos propuestos.

Como resultado preliminar observamos en la figura 3 el trade off entre la precisión y la velocidad de distintos algoritmos de ANN. Las mediciones son generadas utilizando una adaptación del set de datos SIFT1M[13] para adecuarlo a las necesidades de la RA mientras que la precisión se mide como el error promedio que comete cada algoritmo al encontrar la primera y la segunda mejor coincidencias para cada elemento.

Referencias

1. Papagiannakis G., Gurminder S. y Nadia M. T., “A survey of mobile and wireless technologies for augmented reality systems.” Comput. Animat. Virtual Worlds, vol. 19, no. 1, pp. 3-22, 2008.

2. Azuma R. T., The Most Important Challenge Facing Augmented Reality, Presence Teleoperators Virtual Environ., vol. 25, n.o 3, pp. 234-238, dic. 2016.

3. Specht M., Ternier S. y Greller W., “Dimensions of mobile augmented reality for learning: a first inventory,”Journal of the Research for Educational Technology (RCET), vol. 7, no. 1, pp. 117-127, 2011

4. Ierache J., Mangiarua N., Verdicchio N., Becerra M., Duarte N., Igarza S. “Sistema de Catálogo para la Asistencia a la Creación, Publicación, Gestión y Explotación de Contenidos Multimedia y Aplicaciones de Realidad Aumentada”. CACIC 2014 Red UNCI ISBN 978-987-3806-05. 2014

5. Ierache J., Mangiarua N., Bevacqua S., Verdicchio N., Becerra M., Sanz D., Sena M., Ortiz F., Duarte N., Igarza S. “Development of a Catalogs System for Augmented Reality Applications”. World Academy of Science, Engineering and Technology, International Science Index 97, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 9(1), 1 - 7. ISSN 1307:6892. 2015

6. Becerra M., Sanz D., Igarza I, Mangiarua N., Ierache J. “Sistema de Catálogo Virtual Aumentado. Integración de Framework Especializado orientado a Material Didáctico” TEYET, pp 350-356, ISBN 978-950-656-154-3. 2015

7. Mangiarua N., Ierache J., Bevacqua S., Becerra M., Verdicchio N., Duarte N., Sanz D., Igarza S. “Herramienta de Realidad Aumentada para la explotación de material didáctico tradicional”. TE&ET. E-Book. ISBN 978-987-24611-1. 2014

8. Verdicchio N., Sanz D., Igarza S., Mangiarua N., Montalvo C., Ierache J. “Sistema de Catalogo Virtual Aumentado Integración de Framework Especializado Orientado a Juegos Didácticos”. TE&ET, pp 597-604,ISBN 978-987-3977-30-5. 2016

9. Ierache J., Mangiarua N., Bevacqua S., Becerra M., Verdicchio N., Duarte N., Sanz D., Igarza S. “Herramienta de Realidad Aumentada para facilitar la enseñanza en contextos educativos mediante el uso de las TICs”.Revista Latinoamericana de Ing de Software,1(1): -3, ISSN 2314-2642. 2014

10. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: BMVC (2015)

11. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embed-ding for face recognition and clustering. CoRRabs/1503.03832 (2015), http://arxiv.org/abs/1503.03832

12. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014. pp. 1701–1708. IEEE Computer Society (2014). https://doi.org/10.1109/CVPR.2014.220, https://doi.org/10.1109/CVPR.2014.220

13. Jégou, H., Douze, M., Schmid, C.: Product quantization for nearest neighbor search (2011)

EVALUACIÓN DE VISUALIZACIONES EFICIENTES EN CIENCIA DE DATOS

Mag. Raúl Oscar Klenzi, Mag. María Alejandra Malberti, Mag. Graciela Elida Beguerí

Instituto de Informática / Departamento de Informática /Facultad de Ciencias Exactas Físicas y Naturales / Universidad Nacional de San Juan

Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", Rivadavia, San Juan, Teléfonos: 4260353, 4260355 Fax 0264-4234980, Sitio Web: http://www.exactas.unsj.edu.ar
e-mail: {rauloscarklenzi,amalberti,grabeda}@gmail.com

RESUMEN

El presente proyecto plantea como objetivo proponer criterios para la evaluación de visualizaciones eficientes en Ciencia de Datos.
A tal fin se tiene pensado investigar sobre distintos aspectos que atañen a una visualización tales como escala, longitud, área, color, entre otros. También examinar herramientas libres con capacidades para visualización de datos e información y lenguajes de propósito general como Python y JavaScript.
De este modo se espera analizar la aptitud de diversas visualizaciones y sugerir métricas para su evaluación.

Palabras clave: Visualización, Ciencia de Datos, Lenguajes de código abierto, Software libre.

CONTEXTO

Los avances tecnológicos, tanto en potencialidades de cómputo, capacidades de almacenamiento, lenguajes de programación, herramientas de software, entre otros, permiten que más y más personas y empresas generen y guarden datos. Esta gran cantidad de datos almacenados requiere no solo de herramientas para procesarlos, sino también de "formas" de entenderlos y percibir rápidamente la información que contienen. Si bien la representación de la información tiene la capacidad de sintetizar y condensar datos, y constituye un enfoque eficiente para el análisis; no resuelve todos los problemas. (Erraissi y otros, 2019)

Si bien la visualización de la información no resuelve todos los problemas, es una parte importante en varias de las etapas de la Ciencia de datos.

La propuesta cuenta con antecedentes logrados en el tema conforme a los sucesivos proyectos aprobados y subsidiados por CICITCA - UNSJ en los que el grupo ha trabajado, siendo estos:

- "Búsqueda inteligente de información no relacionada en grandes bases de datos" 21/E508, período 2003-2005
- "Descubrimiento de conocimiento a través de Data Warehousing y Data Mining, en los datos de la Biblioteca de la Facultad de Ciencias Exactas, Físicas y Naturales" 21/E639, período 2005-2007
- "Búsqueda estratégica de conocimiento en los datos de biblioteca y de alumnos de la FCEFN" 21/E824, período 2008-2010
- "Minería de datos en la determinación de patrones de uso y perfiles de usuarios" 21/E889, período 2011-2013
- "Extracción de Conocimiento en Datos Masivos" 21/E-951, período 2014-2015
- "La Ciencia de Datos en grandes colecciones de datos" 21/E1014, período 2016-2017

- "Visualización y Deep Learning en Ciencia de Datos" 21/E1071, período 2018-2019

1. INTRODUCCIÓN

La Ciencia de Datos es una rama del paradigma de big data, un enfoque en el que se capturan, procesan a velocidades enormes, variedad y volúmenes de datos estructurados, no estructurados y semiestructurados, utilizando un conjunto de técnicas y tecnologías completamente novedosas en comparación con las que se utilizaron en décadas anteriores. Es útil para derivar conocimiento a partir de datos en bruto, esencial para cualquier sistema integral de apoyo a la toma de decisiones y extremadamente importante a la hora de formular estrategias sólidas para la gestión empresarial.
A la Ciencia de Datos se la puede pensar como el dominio científico dedicado al descubrimiento de conocimiento mediante análisis de datos. Es decir, se refiere al sector de la industria o dominio temático que los métodos de Ciencia de Datos utilizan para explorar. Los científicos de datos aplican técnicas matemáticas y enfoques algorítmicos para derivar soluciones a problemas complejos empresariales y científicos. Tanto en los negocios como en la ciencia, estos métodos pueden proporcionar capacidades de toma de decisiones más robustas.

"Visualización es aquella tecnología plural (esto es, disciplina) que consiste en transformar datos en información semántica — o en crear las herramientas para que cualquier persona complete por sí sola dicho proceso — por medio de una sintaxis de fronteras imprecisas y en constante evolución basada en la conjunción de signos de naturaleza icónica (figurativos) con otros de naturaleza arbitraria y abstracta (no figurativos: textos, estadísticas, etc.)". (Cairo, 2011)
En Barcellos y otros (2017) se argumenta que el uso de algunos tipos específicos de visualizaciones, asociados con el análisis de datos y las técnicas de aprendizaje automático, pueden ayudar en la interpretación de aspectos univariados, bivariados y multivariados de los datos.
Al diseñar visualizaciones cuanto más se conozca a los destinatarios, mejor posicionado se estará para comprender cómo resonar en ellos y así formar una comunicación que satisfaga sus necesidades. Comprender el contexto, elegir una pantalla visual adecuada, eliminar el desorden, decidir dónde dirigir la atención, deben ser tenidos en cuenta al diseñar y crear visualizaciones para impartir información. (Knaflic, 2015)

Desde hace varios años KNIME Analytics (herramienta de software libre) permite incorporar, en forma modular, todos los pasos involucrados en el proceso de Ciencia de Datos. Esta herramienta proporciona un repositorio de módulos fáciles de usar. Adicionalmente a las técnicas estándares de Minería de Datos, añade los algoritmos más actuales de análisis tales como los de deep learning. Posee integraciones con Python, JavaScript, R y con otros grandes proyectos de código abierto, dando libertad de mezclar y combinar las herramientas que se deseen, dentro de un entorno uniforme. (Klenzi, R y otros 2019)

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Entre las líneas más importantes de investigación y desarrollo se mencionan:

- Ciencia de Datos, principalmente Visualización de Información
- Deep Learning
- Herramientas de software libre para arquitecturas secuenciales, paralelas y distribuidas
- Lenguajes de programación de código abierto tales como Python y JavaScript.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se tiene como objetivo proponer criterios para la evaluación de visualizaciones eficientes en Ciencia de Datos.

También se pretende:

- Analizar distintos aspectos que atañen a una representación visual
- Realizar búsqueda de datos abiertos y recopilar otras fuentes de datos provenientes de actividades de cooperación
- Evaluar herramientas de software libre en cuanto a sus potencialidades de visualización y eficiencia computacional
- Realizar análisis comparativo de los lenguajes Python y JavaScript como soporte de visualizaciones
- Seleccionar visualizaciones y clasificarlas de acuerdo a tipos de datos que admiten, usuarios a las que estén dirigidas, así como lenguajes y herramientas que las soportan.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de docentes investigadores del proyecto trabaja desde hace varios años en diversas áreas relacionadas con la Ciencia de Datos.
Es de destacar también, que la propuesta incorpora ayudantes y alumnos que cursan diferentes carreras de FCEFN-UNSJ. Es convencimiento de los integrantes del proyecto que la formación de estos jóvenes en el área de la investigación y áreas temáticas inherentes a la Ciencia de Datos enriquecerá las diferentes actividades y son de suma importancia para su futuro profesional y laboral.
Actualmente forman parte del equipo los coordinadores del “Laboratorio de Sistemas Inteligentes para la Búsqueda de Datos” del Instituto de Informática-FCEFN y los docentes titulares de las asignaturas Inteligencia Artificial, Probabilidad y Estadística, Sistemas de Datos, Estructuras de Datos y Algoritmos, Programación Procedural y Programación Orientada a Objetos, de las carreras Licenciatura en Sistemas de Información y Licenciatura en Ciencias de la Computación pertenecientes al Departamento de Informática de la FCEFN-UNSJ.
Se habrán de proponer en el bienio de duración del proyecto, trabajos finales de grado, posgrado y becas de investigación.
En particular, con docentes investigadores del INAUT-FI y personal del INTA-San Juan, se llevará adelante la tesis de maestría de un integrante del equipo de trabajo "Análisis de Fenómenos en Estaciones Agrometeorológicas Mediante Ciencia de Datos".

Equipo de investigación:

Director: Mag. Raúl Oscar Klenzi
Co-Director: Mag. María Alejandra Malberti Riveros

Integrantes docentes-investigadores:

- Mag. Graciela Elida Beguerí
- Lic. Laura Gutiérrez
- Lic. María Isabel Masanet Yáñez
- Lic. Manuel Ortega
- Prog. Luis Olguín
- Lic. Fabrizio Amaya

Integrantes alumnos:

- Joaquín Cortez
- Carolina Olivares
- Malena Páez
- Noelia Pérez

5. BIBLIOGRAFÍA

- Barcellos, R., Viterbo, J., Miranda, L., Bernardini, F., Maciel, C., & Trevisan, D. (2017, June). Transparency in practice: using visualization to

enhance the interpretability of open data. In Proceedings of the 18th Annual International Conference on Digital Government Research (pp. 139-148). ACM.

- Barcellos, R., Viterbo, J., Bernardini, F., & Trevisan, D. (2018, July). An Instrument for Evaluating the Quality of Data Visualizations. In 2018 22nd International Conference Information Visualisation (IV) (pp. 169-174). IEEE.
- Benoit, G. (2019). Introduction to Information Visualization: Transforming Data Into Meaningful Information. Rowman & Littlefield.
- Cady, F. (2017). The data science handbook. John Wiley & Sons.
- Cairo, A. (2011). El arte funcional: infografía y visualización de información. Alamut.
- Castro Almudena M. e Úcar Iñaki (2019). Gráficas para la ciencia y ciencia para las gráficas. Cuaderno de Cultura Científica. https://culturacientifica.com/2019/04/25/graficas-para-la-ciencia-y-ciencia-para-las-graficas/
- Chen, M., Feixas, M., Viola, I., Bardera, A., Shen, H. W., & Sbert, M. (2017). Information theory tools for visualization. AK Peters/CRC Press
- Cielen, D., Meysman, A., & Ali, M. (2016). Introducing data science: big data, machine learning, and more, using Python tools. Manning Publications Co..
- Erraissi, A., Mouad, B., & Belangour, A. (2019, April). A Big Data visualization layer meta-model proposition. In 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO) (pp. 1-5). IEEE.
- Grus, J. (2019). Data science from scratch: first principles with python. O'Reilly Media.
- How to choose a visualization (2019). https://www.kdnuggets.com/2019/06/how-choose-visualization.html
- Irisarri Patricio (2017, Junio). Breve historia de la visualización de datos. http://periodismodigitaliset18.blogspot.com/2017/06/breve-historia-de-la-visualizacion-de.html
- Klenzi, R. O., Malberti, A., & Beguerí, G. (2019, June). Propuesta didáctica inherente al área de ciencia de datos. In *XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de San Juan).*
- Knaflic, C. N. (2015). Storytelling with data: A data visualization guide for business professionals. John Wiley & Sons.
- Laura Igual Muñoz, & Santi Seguí Mesquida. (2017). Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications. Springer.
- Lo, L. Y. H., Ming, Y., & Qu, H. (2019). Learning Vis Tools: Teaching Data Visualization Tutorials. arXiv preprint arXiv:1907.08796.
- Murray, S. (2017). Interactive data visualization for the web: an introduction to designing with D3" O'Reilly Media, Inc.".
- Schutt, R., & O'Neil, C. (2013). Doing data science: Straight talk from the frontline. O'Reilly Media, Inc.
- Schwabish Jonathan A. (2014, February). A Visualization Mapping: Form and Function

https://policyviz.com/2014/02/05/a-visualization-mapping-form-and-function/

- Sosulski, K. (2018). Data Visualization Made Simple: Insights Into Becoming Visual. Routledge.
- Tufte, E. R. (2001). The visual display of quantitative information (Vol. 2). Cheshire, CT: Graphics press.
- VanderPlas, J. (2016). Python data science handbook: essential tools for working with data. " O'Reilly Media, Inc.".
- Wang, J., Hazarika, S., Li, C., & Shen, H. W. (2018). Visualization and visual analysis of ensemble data: A survey. IEEE transactions on visualization and computer graphics.
- Wang, C., & Shen, H. (2011). Information Theory in Scientific Visualization. Entropy, 13, 254-273.
- Ware, C. (2012). Information visualization: perception for design. Elsevier
- Wenqiang Cui (2019) Visual Analytics: A Comprehensive Overview. EEE Xplore Digital Library. Volume 7: 2019 https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8740868
- Why Data Visualization Is The Most Important Skill in a Data Analyst Arsenal (2019). https://www.kdnuggets.com/2019/08/simpliv-data-visualization-data-analyst.html
- Wilke, C. O. (2019). Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. O'Reilly Media.
- Zuk, T., Schlesier, L., Neumann, P., Hancock, M. S., & Carpendale, S. (2006, May). Heuristics for information visualization evaluation. In Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization (pp. 1-6). ACM.

GENERACION GRAFICA DE PARTICULAS ARTIFICIALES DE ARENA PARA SU UTILIZACIÓN EN REDES NEURONAELS

Mg. Ing. Carlos Gustavo Rodriguez Medina [1], Dr. Ing. Oscar Daniel Chuk [2],
Ing. Regina Bertero, Lic. Adriana Luna, Ing. Enrique Núñez. Sr. Darío Quinteros

Instituto de Investigaciones Mineras / Facultad de Ingeniería /
Universidad Nacional de San Juan

Av. Libertador Gral. San Martin 1109 oeste. San Juan
0264-4211700 (int. 285 [1], int. 389 [1,2])

grodriguez@unsj.edu.ar [1], dchuk@unsj.edu.ar [2]

RESUMEN

El presente trabajo expone una alternativa para generar un conjunto de datos, correspondiente a imágenes de partículas de arenas, con el propósito de ser utilizadas en el entrenamiento de una Red Neuronal Convolucional para la clasificación teniendo en cuenta sus características de redondez y esfericidad. La necesidad de medir tales características en estas arenas, se debe a que permite estimar la calidad para su utilización en procesos de fracturación hidráulica (*fracking*) para la extracción de petróleo de manera no convencional.

La utilización de Redes Neuronales posibilita identificar y clasificar elementos de interés, en base a un entrenamiento previo, para ello se debe contar con un conjunto de datos de entrenamiento con un elevado número de elementos (entre 1.000 y 10.000 muestras aproximadamente), para obtener resultados aceptables. Por lo cual esto requiere contar con un conjunto de imágenes de arenas, que sean representativas de todo el espectro posible a identificar, y rotuladas con sus correspondientes valores de redondez y esfericidad.

Se propone una alternativa para el armado de tal conjunto, la que consiste en graficar partículas de arenas, mediante métodos geométricos y matemáticos, por medio de un software matemático especializado, calculando simultáneamente los valores de redondez y esfericidad, lográndose así el armando del conjunto de datos requerido compuesto de imágenes artificiales de partículas de arena, etiquetadas con las características antes mencionadas. Esto se obtiene mediante un algoritmo computacional desarrollado específicamente y que permite el graficado y medición de manera automatizada, prefijando ciertos parámetros.

Palabras clave: Partículas artificiales de arena, Procesamiento de imágenes, Redes Neuronales, Datos de entrenamiento, Fracturación hidráulica.

CONTEXTO

El tema del trabajo propuesto surge como una alternativa de solución a una problemática surgida durante la ejecución del proyecto de Investigación y Creación (PIC – CICITCA) titulado "*Visión artificial aplicada a la determinación de propiedades geométricas de arenas de fracturación para el análisis de calidad*", desarrollado durante 2018 y 2019.

Dicho trabajo de elaborar gráficamente una partícula artificial de arena para poder conformar un conjunto de datos para el uso en Redes Neuronales se continua en la propuesta de Proyecto de Investigación y Creación presentada para el periodo 2020 y 2021 titulado *"Clasificación de partículas de arena*

de fracturación mediante redes neuronales convolucionales".

Los mencionados proyectos se inscriben dentro de la línea de investigación que lleva adelante el grupo de trabajo en la temática de Visión Artificial e Inteligencia Artificial orientado al control automático de minerales, que se viene desarrollando hace varios años en el Instituto de Investigaciones Mineras de la Facultad de Ingeniería de la UNSJ.

1. INTRODUCCIÓN

Con la finalidad de garantizar la capacidad de las arenas con destino a los procesos de *Fracking* para actuar como apuntalantes o propantes, se reconocen como base las normas de referencia del *American Petroleum Institute* (API) [1]. Esta norma recomienda una serie de ensayos a realizar sobre las arenas con el fin de conocer la calidad del material a utilizarse en la fracturación hidráulica. Entre los ensayos a realizarse se encuentran los de redondez y esfericidad de las partículas de arena.

Tales ensayos padecen de una fuerte dependencia del criterio del observador, dado que el procedimiento se realiza mediante la utilización de un microscopio y las mediciones de esfericidad y redondez se realizan por comparación respecto a tablas que contienen una serie de formas provistas por la norma (cartillas de *Krumbein y Sloss* – figura 1). Esto deriva en resultados sustancialmente disímiles que invalidan el procedimiento, aunque es el que se sigue usando por norma.

El uso de técnicas de Visión Artificial se presenta como la alternativa tecnológica que puede permitir obtener medidas objetivas no dependientes del criterio de un operador. [1]

Cada partícula de arena puede presentar diversas variaciones unas a otras, pudiendo ser algunas más redondeadas, o con más puntas, más o menos achatadas, más o menos alargadas, de diferentes tamaños, etc.

Dado que esta situación que se presenta, es un problema relacionado con la forma de las partículas, se sabe que existen herramientas de la inteligencia artificial orientadas a resolver problemas de reconocimiento de formas, patrones y su clasificación; éstas son las conocidas como Redes Neuronales Artificiales [2].

Las Redes Neuronales Artificiales (RNA) son modelos de procesamiento de información, inspirados por la forma en que el cerebro procesa información. Aunque son una simplificación de las redes neuronales biológicas, las RNA son aptas para resolver problemas que la gente puede resolver, pero las computadoras no pueden, o los algoritmos resultan muy complicados. [3]

Las neuronas biológicas son componentes relativamente simples del cerebro, pero cuando millares de ellas se conectan en forma conjunta se hacen muy poderosas (conforman una Red), esto justifica la alta capacidad de cómputo del cerebro. Una red neuronal artificial está constituida por muchos elementos simples de procesamiento, organizados en niveles y muy interconectados. [4]

Aprendizaje de la red neuronal artificial:

El conocimiento es adquirido por la red a través de un proceso que se denomina aprendizaje; dicho conocimiento se almacena mediante la modificación de la fuerza o peso sináptico de las distintas uniones entre elementos de procesamiento. [5].

Todo conocimiento de una red neuronal se encuentra distribuido en los pesos sinápticos de las neuronas. Una red neuronal artificial aprende a través de un proceso de ajuste de sus pesos sinápticos. El aprendizaje es el proceso por el cual una red neuronal modifica sus pesos sinápticos en respuesta a una entrada, para proporcionar la salida adecuada. [3]

Los cambios que se producen durante el proceso de aprendizaje implican la destrucción, modificación y creación de conexiones entre las neuronas. La creación de una nueva conexión consiste en que el peso de la misma pasa a tener un valor distinto de cero, una conexión se destruye cuando su peso pasa a ser cero.

Datos de entrenamiento y problemática presentada:

Dado que para trabajar con una red neuronal se debe contar con un conjunto o *set* de elementos (específicamente imágenes de partículas de arena en este contexto), estos conjuntos deben contener un número elevado de elementos o imágenes, debido a que dicho conjunto se debe dividir en dos subconjuntos, uno de entrenamiento (entre un 60% y un 70% del total de las imágenes), y otro subconjunto para el testeo o evaluación del aprendizaje alcanzado por la red neuronal.

Las redes neuronales aumentan su capacidad de generalizar, a medida que el conjunto de entrenamiento es mayor. Si bien, se podría entrenar con un conjunto reducido (en términos relativos) de imágenes, el desempeño de la red aumenta cuanto mayor sea la cantidad de imágenes de entrenamiento, y con mayor razón cuando el tipo de red neuronal a utilizar es del tipo convolucional (tal como se pretende desarrollar en el proyecto de investigación 2020 – 2021).

Una alternativa (ideal) sería el disponer de un banco de imágenes de partículas de arenas ya clasificadas, etiquetadas o rotuladas con los correspondientes valores de esfericidad y redondez. Esto opción es difícil, a pesar de que en el Instituto de Investigaciones Mineras se realizan este tipo de servicios, pero debido a que dicho trabajo se lleva a cabo de manera manual mediante microscopio e inspección visual humana, donde el laboratorista clasifica los granos de arena por comparación mediante las cartillas de *Krumbein y Sloss* (figura 1), lamentablemente no se dispone de estos resultados digitalizados de manera sistematizada. Ni de las partículas de arena y tampoco de la clasificación realizada.

Tal como se dijo en párrafos anteriores, la norma que se aplica para la determinación de los parámetros de calidad de las partículas de arena para *Fracking*, utiliza las cartillas de *Krumbein y Sloss,* para clasificarlas en cuanto a redondez y esfericidad, tal como se expone en la figura 1. Dicha tabla posee solamente 20 elementos gráficos de acuerdo a distintos valores de redondez y esfericidad, los que resultan extremadamente insuficientes si se quisieran utilizar estos elementos imágenes de partículas de arena para llevar a cabo el entrenamiento de la red neuronal.

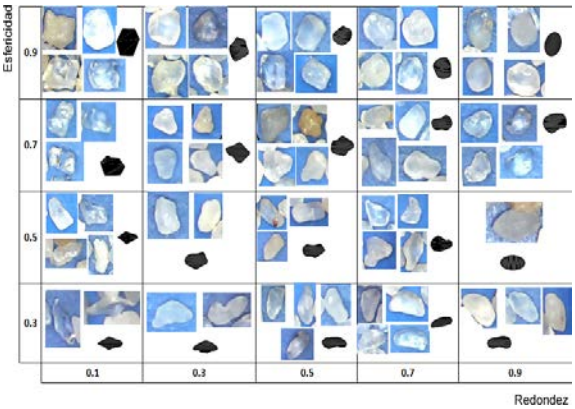


Figura 1. Cartilla de esfericidad y redondez de *Krumbein y Sloss*

Entonces, si no se puede utilizar la tabla de la figura 1 para el proceso de entrenar la red neuronal, y tampoco se cuenta con trabajos de clasificación anteriores, se debe buscar una alternativa para generar un conjunto de datos de entrenamiento y de testeo con una cantidad de elementos los suficientemente grande para lograr que la red pueda clasificar las partículas de arena adecuadamente

Propuesta:

Con el objeto de poder desarrollar un conjunto de datos lo suficientemente grande, de acuerdo a las necesidades de entrenamiento de la red neuronal (para la clasificación de partículas de arena), se propone generar de manera gráfica, mediante el software matemático Matlab [6], partículas de arena que no son imágenes tomadas mediante un microscopio a granos de arena reales, sino que las mismas son elaboradas mediante funciones matemáticas (elementos geométricos básicos tales como elipses y círculos), pudiéndose así hacer referencia como un conjunto de elementos de partículas "*artificiales*" de arena.

Tales imágenes generadas, tienen la particularidad que como están compuestas por formas básicas de elipses y círculos, entonces se puede aplicar con cierta facilidad la ecuación desarrollada por *Wadell* (1932) para

el cálculo de la redondez, tal como se expone en la Ecuación 1. A la vez resulta de importancia aclarar que la cartilla de *Krumbein y Sloss (1963)* se basa en la aplicación de la ecuación de *Wadell*.

$$R = \frac{\frac{\sum_{i=1}^{N} r_i}{N}}{r_{max\text{-}in}} \qquad \textit{(Ecuación 1.)}$$

Donde:

N: Cantidad de círculos inscriptos.
ri: Radios de los círculos.
r max-in: Radio del máximo circulo inscripto

Por lo tanto, a cada partícula generada gráficamente como una partícula artificial se le calcula la redondez mediante la ecuación 1, utilizando *Wadell*, y la correspondiente para el cálculo de esfericidad, por lo que en definitiva se estaría aplicando la referida norma API para la determinación de estas características, utilizando técnicas de inteligencia artificial (procesamiento de imágenes, etc.)

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

En el Laboratorio de Control Automático del Instituto de Investigaciones Mineras (Facultad de Ingeniería de la Universidad Nacional de San Juan) se realiza investigación y desarrollo respecto a la automatización de procesos de molienda de minerales hace casi dos décadas. Hace algunos años se viene desarrollando el Área de Procesamiento de Imágenes.

Actualmente se está trabajando en la caracterización de arenas especiales (calidad, forma, color, tamaños, clasificación del tipo de material, etc.) mediante el procesamiento de imágenes para ser utilizadas en la extracción de petróleo mediante el método de *Fraking*.

De manera complementaria, en búsqueda de dar solución al tema de clasificación de partículas, se ha incursionado en la temática de redes neuronales.

Otra de las líneas de investigación y desarrollo del grupo de trabajo, está orientado al de optimización. Concretamente, se está trabajando en la optimización multiobjetivo, del tipo predictivo, para la explotación de minas.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Se ha elaborado un algoritmo que posibilita la realización grafica de la partícula artificial de arena. Dicho algoritmo se presenta en la figura 2.

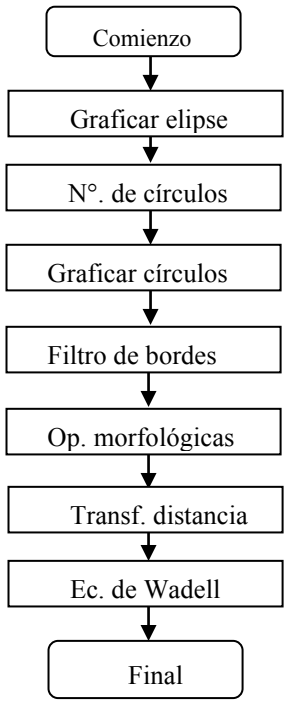
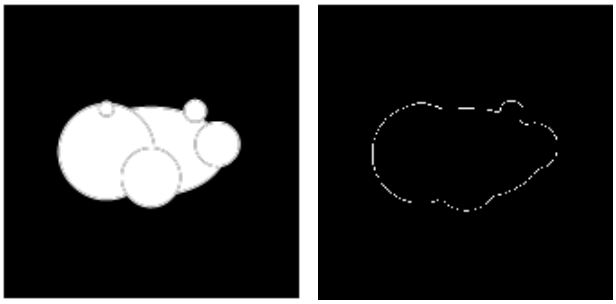


Figura 2. Algoritmo para graficar partículas artificiales de arena.

Sobre una imagen con fondo de color negro se comienza graficando una elipse, determinando las coordenadas y los ejes de la misma. Luego se determina la cantidad de círculos a emplear, relacionado a la cantidad de picos o esquinas de la partícula a generar. En la medida que más círculos o picos se le agreguen a la elipse graficada previamente, menos redondeada será la partícula artificial resultante. Esto se puede apreciar en la figura 3-A. El paso siguiente es aplicar un filtro de bordes, de *Canny* específicamente [7], para poder trabajar solamente con el contorno resultante, esto se observa en la figura 3-B.



A- Partícula artificial de elipses y círculos.

B- Filtro de bordes aplicado.

C- Operación morfológica de dilatación.

D- Operación morfológica de erosión.

E- Transformada de distancia.

F- Radios de círculos para calcular *Wadell*.

Figura 3. Resultados de aplicar el algoritmo para graficar partículas artificiales de arena.

En el diagrama de bloques del algoritmo de la figura 2, se indica que el paso que sigue es el de utilizar operaciones morfológicas, esto se hace con el objeto de lograr un suavizado de la partícula obtenida, ya que hasta esa instancia es muy notoria la presencia de los círculos sobre la elipse. Obsérvese el resultado de aplicar la operación de dilatación y luego la de erosión en la figura 3-C y 3-D, respectivamente.

Como paso siguiente del algoritmo, se realiza el cálculo de la transformada de distancia de la partícula obtenía en la figura 3-D. Esto genera el resultado de la figura 3-E.

Finalmente se calcula la redondez mediante la Ecuación 1 de *Wadell*, utilizando los radios de los círculos inscriptos en las esquinas de la partícula artificial, y tomando el valor de la transformada de distancia como el valor del radio del máximo circulo inscripto en la partícula. Se calcula también la esfericidad.

Como resultado preliminar, los cálculos obtenidos mediante la aplicación de la formula de *Wadell*, utilizando el algoritmo elaborado (figura 2), es coincidente con el que se obtiene de aplicar las cartillas de *Krumbein y Sloss (figura 1)* para la obtención de los valores de redondez y esfericidad (de acuerdo a las normas API, universalmente utilizada y explicado en el apartado 1).

Tal como antes se mencionó, estos representan resultados preliminares pero prometedores, dado que se logra coincidencia en los resultados obtenidos de clasificar una partícula mediante la Norma API (tabla 1) comparados con la generación grafica y cálculo realizado mediante la ecuación 1.

Durante el periodo de investigación 2020 - 2021 se trabajará en la optimización del algoritmo y automatización sobre ciertos aspectos.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de investigación está conformado por Ingenieros Electrónicos, Electromecánicos, en Minas, Metalurgistas y Geólogos, con formación diversa en posgrado, tal como Doctorado en sistemas de control, Maestría en informática, Especialización en gestión y vinculación tecnológica, etc.

Actualmente, una de las integrantes del equipo de trabajo, se encuentra desarrollando su Tesis de Doctorado en Ingeniería en procesamiento de minerales, siendo su director de Tesis también otro de los integrantes del grupo de investigación.

Además, como inicio en la carrera de investigador, otra de las actuales integrantes, comenzó hace cuatro años participando de un proyecto PROJOVI (Proyecto de Jóvenes Investigadores), cuya finalidad es la de

permitir que docentes, becarios, y alumnos de la Universidad, menores de 40 años, con o sin experiencia en investigación lleven a cabo dicha actividad. También participó de un proyecto PIC (Proyecto de Investigación y Creación) en el periodo 2018 – 2019.
Además, un alumno avanzado de la carrera de grado de Ingeniería en Electrónica, comienza a participar de un nuevo proyecto PIC a ejecutarse en el periodo 2020 – 2021, llevando a cabo su tesina de fin de carrera.

5. BIBLIOGRAFIA

[1] Rodriguez Medina, C. G.; Chuk, O. D.; Bertero R.; Luna, A.; Núñez, E.; Trigo, P. 2019. Visión artificial aplicada a la determinación de propiedades geométricas de arenas de fracturación para el análisis de calidad. Universidad Nacional de San Juan. Informe de Final de Proyecto PIC.

[2] Jain, A. K. 2000. Statistical Pattern Recognition: a review. IEEE Transactions on Pattern Análisis and Machine Inteligence. Vol 22, Nro. 1.

[3] Castro G., J. F. 2006. Fundamentos para la implementacion de red neuronal perceptron multicapa mediante software. Tesis de grado. Facultad de Ingenieria, Universidad de San Carlos de Guatemala.

[4] Haykin S. 1994. Neural Networks - A comprehensive foundation, IEEE Press - Macmillan College Publishing Company, Inc.

[5] Long, L. 1990. Introducción a las computadoras y al procesamiento de información. 2ª. Edición; México: Editorial Prentice All Hispanoamericana.

[6] Mathworks. (1994-2019). Image Processing Toolbox.
https://es.mathworks.com/products/
image.html ?s_tid=srchtitle.

[7] Rodriguez Medina, C. G. y Navas, G. S. 2014. Aplicación del filtro de Canny en la esteganografía digital. XVI Workshop de Investigadores en Ciencias de la Computación - WICC 2014 Ushuaia. RedUNCI. p 806 - 811. ISBN 978-950-34-1084-4.

REALIDADES ALTERNATIVAS COMO SOPORTE PARA EL DESARROLLO SOSTENIBLE

G. Rodríguez, N. Jofré, Y. Alvarado, J. Fernandez, R. Guerrero
Laboratorio de Computación Gráfica / Dpto. Informática / FCFMyN
Universidad Nacional de San Luis
Ejército de los Andes 950, Tel: 0266 4420823, San Luis, Argentina
{gbrodriguez, npasinetti, ymalvarado, jmfer, rag}@unsl.edu.ar

RESUMEN

En los últimos tiempos el término Desarrollo Sostenible ha cobrado fuerza. Su objetivo hace referencia a responder a las necesidades del presente de forma igualitaria, sin comprometer las posibilidades de sobrevivencia y prosperidad de las generaciones futuras. Esto lleva a la sociedad actual a enfrentar sus problemas operando de un modo ecológicamente sustentable e intentando mejorar la calidad vida de la mayor cantidad de personas.

Por otro lado, la tecnología ha tenido un crecimiento exponencial que ha impactado en todos los ámbitos de nuestra vida cotidiana. En él se encuentran las tecnologías de Realidad Virtual y Realidad Aumentada, que gracias a su versatilidad y amplitud de posibilidad de aplicación, han logrado captar el interés de diversas áreas del conocimiento. Su transversalidad las convierte en posibles generadoras de iniciativas sostenibles que impactan en la educación, la concientización, el entrenamiento en cualquier ámbito, y muy específicamente en la salud, donde tienen un uso primordial.

Esta propuesta de trabajo expone sobre el potencial de la Realidad Virtual y Realidad Aumentada en el alcance de los Objetivos de Desarrollo Sostenible propuestos por Naciones Unidas, con el fin de satisfacer las necesidades actuales sin comprometer los recursos y posibilidades de las futuras generaciones.

Palabras clave: Desarrollo Sostenible, Tecnología Sostenible, Realidad Virtual, Realidad Aumentada, Objetivos de Desarrollo Sostenible.

CONTEXTO

La propuesta de trabajo se lleva a cabo dentro del proyecto "Realidades Alternativas como lenguaje generativo aplicado a la solución de problemas reales". Este proyecto es desarrollado en el ámbito del Laboratorio de Computación Gráfica de la Universidad Nacional de San Luis.

1. INTRODUCCIÓN

Actualmente, la civilización transita nuevos escenarios cotidianos en los que se pueden identificar cuatro grandes aspectos: la globalización, el permanente cambio de contexto, la valoración del conocimiento y la relevancia del cuidado del medio ambiente.

Debido al vínculo existente entre el desarrollo económico y social, y sus efectos más o menos inmediatos sobre el medio natural, es que se inició hace algunos años un proceso de reflexión, sin precedentes a nivel mundial, impulsado por la Organización de las Naciones Unidas (ONU) y seguido por la Unión Europea (UE), para abordar el problema de la sostenibilidad y el desarrollo sostenible del planeta como uno de los grandes retos a los que se enfrentarían las sociedades del siglo XXI [1, 2, 3].

En los tiempos actuales, el desarrollo sostenible ha llegado a erigirse como la nueva filosofía que podría orientar hacia modelos productivos más racionales con el entorno y

equitativos socialmente. El objetivo básico sería hacer viable y perdurable el bienestar humano, mediante procesos de integración socioambiental.

En Septiembre de 2015, la Asamblea General de la ONU aprobó la *Agenda 2030* para el Desarrollo Sostenible, la cual establece una visión transformadora hacia la sostenibilidad económica, social y ambiental de los 193 Estados Miembros que la suscribieron, siendo guía de referencia para el trabajo de la institución en pos de esta visión durante los próximos 15 años. Para América Latina y el Caribe presentó y aún presenta una oportunidad histórica, dado que incluye temas altamente prioritarios para la región, como la erradicación de la pobreza extrema, la reducción de la desigualdad en todas sus dimensiones, un crecimiento económico inclusivo con trabajo decente para todos, ciudades sostenibles y cambio climático, entre otros [4, 5, 6].

Por otro lado, las nuevas tecnologías han cobrado fuerza en nuestra sociedad actual y pueden ser utilizarlas para mejorar nuestra vida en casi cualquier aspecto cotidiano. Más aún, la confluencia de una serie de tecnologías está posibilitando un cambio profundo en el desarrollo de la forma de vida de las personas y de los modelos de negocio de las empresas que adoptan estas tecnologías. Resulta entonces importante lograr que ese desarrollo sea sostenible.

El desarrollo sostenible implica satisfacer las necesidades actuales sin comprometer los recursos y posibilidades de las futuras generaciones. Su importancia radica en el hecho de evitar circunstancias de zozobra. En consecuencia, las tecnologías sostenibles son aquellas que emplean menos energía para realizar los procesos, emplean una cantidad menor de recursos limitados y, en definitiva, no agotan los recursos naturales tanto en su creación, puesta en marcha o utilización. Entre ellas se pueden destacar: el software para *nuevas maneras de medición*, el *almacenamiento energético*, las *tecnologías exponenciales*, la transparencia en la *cadena de suministros*, entre otros [7].

En particular, las Naciones Unidas señalan que "las **tecnologías de la información** son fundamentales para responder a los desafíos del desarrollo", gracias al "gran potencial" de la conectividad global para estimular el desarrollo humano.

Del conjunto de tecnologías mencionado, las tecnologías exponenciales (tecnologías que evolucionan rápidamente gracias al desarrollo exponencial de la computación) consideradas claves para el desarrollo son: el *Móvil Conectado*, la *Inteligencia Artificial*, la *Realidad Virtual*, la *Realidad Aumentada*, el *Internet de las Cosas*, la *Nube* (cloud), la *Robótica Avanzada*, los *Vehículos Autónomos*, la siguiente generación de la *Genómica*, la *Impresión 3D*, la nueva *Tipología de Materiales*, las *Energías Renovables*, el *Blockchain*, e incluso la *Bio* y *Nanotecnología*.

En los últimos años, la Realidad Virtual (RV) y la Realidad Aumentada (RA) han despertado el interés de diversas áreas del conocimiento, mostrando su versatilidad y posibilidades como tecnologías innovadoras. Sus capacidades de mostrar escenarios virtuales ficticios o insertar objetos virtuales en el espacio real posibilitan la realización de experiencias en primera persona, la que genera una mentalidad más empática en las personas que viven la experiencia, logrando una mayor tendencia a involucrarse en la vida real con lo experimentado. Estas características las han convertido en herramientas muy útiles para presentar determinados contenidos bajo diferentes premisas en diversas áreas [8].

Ambas tecnologías tienen usos que pueden ser muy útiles para el cumplimiento de los Objetivos de Desarrollo Sostenible (ODS), dado que se trata de tecnologías transversales que se pueden aplicar a la Educación, la Sensibilización y el Entrenamiento en cualquier ámbito, y muy específicamente para la Salud, donde tiene un uso primordial.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las realidades alternativas provistas por la Realidad Virtual y la Realidad Aumentada permiten múltiples usos para alcanzar varios de los ODS con un coste marginal mínimo, por lo que se convierten en herramientas de carácter exponencial. Los ODS beneficiarios de la RV y RA son muchos y se basan en los siguientes conceptos: Educación Universal con contenidos atractivos, Tratamientos médicos innovadores, Concientización en todos los ODS y Nuevos modelos más sostenibles que eliminan la pobreza.

El uso de estas tecnologías se extiende a muchas actividades económicas pero también constituye una fuente de riqueza en el propio desarrollo de herramientas y de los contenidos digitales necesarios para su puesta en valor.

Los mayores logros obtenidos a través de las contribuciones de estas tecnologías se refieren a los siguientes ODS:

- **Salud y Bienestar (ODS3): Asegurar vidas saludables y promover el bienestar de todos en todas las edades**.
  Las aplicaciones de RV en salud son muchas, y si además, tienen un gran impacto en la disminución de los costes, pueden ser de aplicación en países en vías de desarrollo o en comunidades sin recursos. Sin embargo, es en la RA donde los médicos encuentran una ayuda inestimable pues les permite insertar información relevante en el momento de la intervención, mediante el guiado o la asistencia adecuada. Algunos ejemplos son: Cirugías [9], Anatomía [10], Asistencia sanitaria [11], Enfermería [12], Educación sanitaria [13].

- **Educación de Calidad (ODS4): Garantizar una educación de calidad inclusiva y equitativa y promover las oportunidades de aprendizaje a lo largo de toda la vida para todos.**
  En el contexto educativo, las aplicaciones de RV y RA existentes actualmente son diversas y se aplican en múltiples áreas. La RV es utilizada mayormente en las disciplinas sociales y exactas para los diferentes niveles educativos. Por su parte, la RA es más comúnmente utilizada como refuerzo del proceso cognitivo de aprendizaje para niveles educativos inferiores. Ejemplos Inglés [14], Geometría [15], Dibujo Técnico [16], Historia [17], Física [18].

- **Trabajo Decente y Crecimiento económico (ODS8): Promover un crecimiento económico sostenido, inclusivo y sostenible, el empleo pleno y productivo y el trabajo decente para todos.**
  La Industria 4.0 tiene en las realidades alternativas un sinfín de aplicaciones que pueden ser trasladadas desde las fábricas más avanzadas a lugares de extrema pobreza. La RV puede ser utilizada para el entrenamiento de profesionales en el diseño, fabricación, uso y reparación de maquinaria avanzada, muchas de esas veces para la simulación de situaciones que pueden ser riesgosas en la vida real. En muchos casos dichos entrenamientos se complementan con el uso de RA de forma que sea posible realizar entrenamientos con el simple uso de un dispositivo móvil. Ejemplos Automotriz [19], Construcción [20], Operaciones militares [21], Robots Industriales[22].

3. RESULTADOS OBTENIDOS/ESPERADOS

Las actividades realizadas hasta el momento por este grupo se han enmarcado dentro de 1 proyecto de Investigación de la UNSL, 1 proyecto de la Comunidad Europea, 4 proyectos de Desarrollo Tecnológico de la

Secretaría de Políticas Universitarias y 3 proyectos de Extensión de Interés Social.

Como consecuencia del trabajo elaborado, se ha logrado desarrollar varios sistemas relacionados a RV con interacción multimodal involucrando aspectos verbales y gestuales, así como también se han elaborado aplicaciones de innovación y desarrollo asociadas a la RV y RA.

Actualmente las acciones se encuentran focalizadas en la incorporación de nuevas estrategias que permitan alcanzar una mejor percepción e interacción al mismo tiempo que se logra incentivar al usuario tanto cognitiva como físicamente.

Como resultante, se analizará el impacto logrado por el uso de sistemas de RV y RA en la mejora y adquisición de habilidades funcionales en procesos físicos y cognitivos asociados a las áreas de la salud y la educación, entre otras. De esta manera, se pretende detectar y evaluar la evidencia científica resultante para determinar la envergadura de dichas intervenciones.

4. FORMACIÓN DE RECURSOS HUMANOS

Los trabajos realizados han permitido la definición de trabajos finales de carrera de la Licenciatura en Cs. de la Computación (4 finalizados), tesis de Especialización en Educación Superior (1 en proceso de evaluación), tesis de Maestría en Cs. de la Computación (2 en ejecución y 1 finalizada) y tesis de Doctorado en Ciencias de la Computación (1 en ejecución).

Adicionalmente se ha obtenido una beca de iniciación a la investigación y una beca de perfeccionamiento en investigaciones otorgadas por la Secretaría de Ciencia y Técnica de la UNSL; y una beca doctoral de CONICET.

5. BIBLIOGRAFÍA

1. Saura Calixto, P. y Hernández Prados, M. Á. (2008). La evolución del concepto de sostenibilidad y su incidencia en la educación ambiental. Teoría de la Educación. Revista Interuniversitaria, 20, 179-204.
2. Sachs, J. (2014). La era del desarrollo sostenible: Nuestro futuro está en juego. Incorporemos el desarrollo sostenible a la agenda política mundial. Comité Editorial/Editorial Board, 101.
3. Menoyo, M. Á. M. (2013). Desarrollo sostenible: problemáticas, agentes y estrategias. Universidad Nacional de Educación a Distancia.
4. Desa, U. N. (2016). Transforming our world: The 2030 agenda for sustainable development.
5. https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/
6. Colglazier, W. (2015). Sustainable development agenda: 2030. Science, 349(6252), 1048-1050.
7. https://www.un.org/development/desa/dpad/publication/estudio-economico-y-social-mundial-2018-tecnologias-de-vanguardia-en-favor-del-desarrollo-sostenible/
8. Moreau, Guillaume & Arnaldi, Bruno & Guitton, Pascal. (2018). Virtual Reality, Augmented Reality: myths and realities.
9. Gomez, Pedro & Willis, Ross & Sickle, Kent. (2014). Development of a virtual reality robotic surgical curriculum using the da Vinci Si surgical system. Surgical endoscopy.
10. Moro, C., Štromberga, Z., Raikos, A., & Stirling, A. (2017). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anatomical sciences education, 10(6), 549-559.

11. Carlson, K. J., & Gagnon, D. J. (2016). Augmented reality integrated simulation education in health care. Clinical simulation in nursing, 12(4), 123-127.
12. Ferguson, C., Davidson, P. M., Scott, P. J., Jackson, D., & Hickman, L. D. (2015). Augmented reality, virtual reality and gaming: an integral part of nursing.
13. Zhu, E., Lilienthal, A., Shluzas, L. A., Masiello, I., & Zary, N. (2015). Design of mobile augmented reality in health care education: a theory-driven framework. JMIR medical education, 1(2), e10.
14. Hsu, T. C. (2017). Learning English with augmented reality: Do learning styles matter?. Computers & Education, 106, 137-149.
15. Lin, H. C. K., Chen, M. C., & Chang, C. K. (2015). Assessing the effectiveness of learning solid geometry by using an augmented reality-assisted learning system. Interactive Learning Environments, 23(6), 799-810.
16. Alvarado, Y., Jofré, N., Rosas, M., & Guerrero, R. (2019). Aplicaciones de Realidad Virtual y Realidad Aumentada como soporte a la enseñanza del Dibujo Técnico. Enseñanza y Aprendizaje de Ingeniería de Computadores. (9). 65-74
17. Efstathiou, I., Kyza, E. A., & Georgiou, Y. (2018). An inquiry-based augmented reality mobile learning approach to fostering primary school students' historical reasoning in non-formal settings. Interactive Learning Environments, 26(1), 22-41.
18. Kaufmann, H., & Meyer, B. (2008). Simulating educational physical experiments in augmented reality. In ACM SIGGRAPH Asia 2008 Educators Programme (pp. 1-8).
19. Lawson, Glyn & Salanitri, Davide & Waterfield, Brian. (2015). Future directions for the development of Virtual Reality within an automotive manufacturer. Applied ergonomics. 10.1016/j.apergo.2015.06.024.
20. Goulding, Jack & Nadim, Wafaa & Petridis, Panagiotis & Alshawi, Mustafa. (2012). Construction Industry Offsite Production: A Virtual Reality Interactive Training Environment Prototype. Advanced Engineering Informatics. 26. 103-116. 10.1016/j.aei.2011.09.004.
21. Livingston, M. A., Rosenblum, L. J., Julier, S. J., Brown, D., Baillot, Y., Swan, I. I., ... & Hix, D. (2002). An augmented reality system for military operations in urban terrain. Naval research lab washington dc advances information technology branch.
22. Pettersen, T., Pretlove, J., Skourup, C., Engedal, T., & Lokstad, T. (2003, October). Augmented reality for programming industrial robots. In The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings. (pp. 319-320). IEEE.

Realidad Virtual: Maximizando Presencia, Inmersión y Usabilidad

Matías N. Selzer, Martín L. Larrea y Silvia M. Castro

Departamento de Ciencias e Ingeniería de la Computación,
Universidad Nacional del Sur (DCIC-UNS)
Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)
Laboratorio de I+D en Visualización y Computación Gráfica, (UNS-CIC Prov. de Buenos Aires)
{matias.selzer, mll, smc }@cs.uns.edu.ar

Resumen

La Realidad Virtual es una tecnología que intenta sumergir a los usuarios en un mundo virtual generado por computadora. Aunque suele pensarse que el realismo gráfico es lo más importante, existen además otros factores que influyen en las sensaciones de presencia e inmersión percibidas, como el tipo de sonido, en campo de visión, la resolución del display, etc. La definición e influencia de dichos factores ha sido muy discutida en la literatura. Esta línea de investigación tiene como objetivo relevar dichos factores y analizar su impacto y relación con distintas técnicas computacionales y estadísticas. Se busca desarrollar una métrica que prediga el nivel de presencia e inmersión percibida por cualquier sistema de Realidad Virtual dado. Esta métrica y la importancia de cada factor brindará una guía para poder seleccionar el sistema y técnicas de Realidad Virtual que resulten más eficientes para la aplicación requerida.

Palabras clave: *Realidad Virtual, Interacción Humano Computadora en Realidad Virtual, Presencia e Inmersión*

Contexto

Este trabajo se lleva a cabo en el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) del Departamento de Ciencias e Ingeniería de la Computación, de la Universidad Nacional del Sur. Los trabajos realizados bajo esta línea involucran a docentes investigadores, becarios doctorales y alumnos de grado.

1. Introducción

La Realidad Virtual (RV) es una rama de las ciencias de la computación que promete sumergir al usuario en un mundo generado por computadora, al mismo tiempo que intenta desvanecer la línea entre lo real y lo ficticio. El avance tecnológico de los últimos años ha facilitado la posibilidad de generar y experimentar entornos virtuales y simulaciones cada vez más realistas. El mercado se encuentra actualmente inundado con juegos y simulaciones para entretener, motivar y divertir a los usuarios. Por el otro lado, el ámbito científico y comercial se enfoca principalmente en la creación de aplicaciones y sistemas muy específicos en áreas como la medicina ([1, 2, 3]), la arqueología ([4, 5]), el entrenamiento militar ([6]), o diferentes tipos de simulaciones ([7]).

En RV, la presencia se define como “la sensación de estar ahí”, es decir, la sensación de que uno de verdad pertenece a el mundo virtual que actualmente está experimentando ([8]). La inmersión, por el otro lado, está más relacionada al hardware utilizado y se define como “qué tanto de

nuestros sentidos está comprendido por el sistema de RV?" ([9]). Por este motivo, la presencia es una medida subjetiva que depende de la percepción de cada usuario; mientras que la inmersión es una medida objetiva que depende del sistema de RV que se esté utilizando. Se suele pensar que en un sistema de RV es fundamental maximizar el nivel de realismo gráfico para obtener la mayor sensación de presencia. Sin embargo, está demostrado que esto no es necesario ([10]).

Desde los inicios de la RV, muchos investigadores han tratado de definir qué aspectos o factores de un sistema de RV verdaderamente influyen en el nivel de presencia percibida. Algunos de los más relevantes incluyen la resolución de la pantalla, la calidad gráfica, el nivel de sonido, entre otras ([11, 12, 13]). Sin embargo, existe mucha discusión entre la verdadera influencia de dichos factores e incluso la relación entre ellos. Lo que es más, la mayoría de dichos trabajos no son actuales, por lo que sus experimentos no aprovechan las posibilidades de las tecnologías modernas.

Por este motivo, es de vital importancia realizar un estudio exhaustivo de todos aquellos factores que pueden influir en la sensación de presencia, y por lo tanto también en la sensación de inmersión producida por cualquier sistema de RV. Dicho análisis revelará los factores más relevantes que deben tenerse en cuenta a la hora de crear un nuevo sistema o a la hora de seleccionar uno de los sistemas existentes. Como la presencia y la inmersión son medidas generales de la percepción de los usuarios, este estudio beneficia cualquier área de aplicación que desee incorporar o complementarse con la tecnología de RV.

2. Líneas de Investigación, Desarrollo e Innovación

Actualmente esta línea de investigación se enfoca en el análisis y relevamiento de todos aquellos factores que influyen en la sensación de presencia e inmersión en un sistema de RV. La cantidad de combinaciones de todos los componentes de dichos factores es muy extensa y no existe una forma directa de calcular el nivel de presencia o inmersión que produce cada una de ellas. Por este motivo, la participación de un usuario es pertinente para poder calificar cada una de dichas combinaciones. Aunque no será posible calificar todas las combinaciones, con un gran número basta para realizar un análisis. A partir de esto, nuestra línea de investigación trata de hallar una métrica que relacione aquellos factores que influyen mayormente en la determinación de estos niveles de presencia e inmersión. Con esta métrica se puede entonces predecir el nivel de presencia o inmersión que generaría cualquier sistema de RV dado.

A partir de las combinaciones de factores generadas con su correspondiente calificación del usuario, se forma un dataset que puede utilizarse para realizar diversos análisis.

- **Predictor de presencia e inmersión:** dado un determinado sistema de RV, resulta importante predecir el grado máximo de presencia e inmersión que éste produce. Esta línea trabaja en

conjunto con técnicas de Machine Learning para la generación de dicho predictor a partir del dataset ya construido.

- **Importancia y optimización de los factores:** como la cantidad de factores que influyen en el nivel de presencia e inmersión de los sistemas de RV es muy grande (al menos 30), resulta difícil optimizar manualmente cada uno de ellos cuando se desea desarrollar un nuevo sistema. Por este motivo, es importante identificar aquellos factores que más influyen en el nivel de presencia e inmersión final. Al reducir considerablemente la cantidad de factores, el diseñador de un nuevo sistema puede enfocarse en la optimización de los mismos, mejorando así el grado de presencia e inmersión del sistema.
- **Análisis estadísticos:** a partir del dataset construido se pueden comparar los factores entre sí para determinar si existe alguna correlación entre los mismos. De esta forma se puede calcular la interacción entre ellos y la influencia que tienen en el valor final de presencia e inmersión.

3. Resultados y Objetivos

Sobre los ejes presentados se han obtenido resultados parciales. Actualmente se está desarrollando una aplicación de RV que genere aleatoriamente un escenario distinto en función de cada uno de los factores identificados. Por ejemplo, un escenario tendrá determinada resolución de pantalla, nivel gráfico, tipo de audio, tipo de tracking, etc.; mientras que otro escenario generado tendrá factores completamente distintos. La aplicación permitirá a los usuarios experimentar un escenario generado durante un tiempo determinado para que finalmente dichos usuarios califiquen el escenario según el nivel de presencia experimentado. Los usuarios podrán realizar esto todas las veces que deseen y cada prueba constituirá una nueva muestra del dataset. Esperamos que en sistema esté online todo el tiempo posible para que el dataset siga creciendo, lo que ayudará a obtener mejores predictores y análisis estadísticos entre las variables.

4. Formación de Recursos Humanos

En lo concerniente a la formación de recursos humanos se detallan las tesis en desarrollo y concluidas relaciones con las líneas de investigación presentadas:

Tesis Concluidas

"Interacción Humano Computadora en Ambientes Virtuales", tesis de posgrado para el Magíster en Ciencias de la Computación. Alumno: Matías Selzer. Director: Martín Larrea.

"La Realidad Virtual como Herramienta para el Desarrollo Arquitectónico", tesis de grado de Ingeniería en Sistemas de Información. Alumnos: Facundo Reissing, Sebastián Vicente. Director: Martín L. Larrea. Colaborador: Matías N. Selzer.

Tesis en Desarrollo

"Métricas de Inmersión para sistemas de Realidad Virtual", tesis de posgrado para Doctor en Ciencias de la Computación. Alumno: Matías Selzer. Director: Silvia M. Castro. Co-Director: Martín L. Larrea.

5. Bibliografía

1. Freedman, S. A., Dayan, E., Kimelman, Y. B., Weissman, H., & Eitan, R. (2015). Early intervention for preventing posttraumatic stress disorder: an Internet-based virtual reality treatment. *European journal of psychotraumatology*, *6*.
2. Rothbaum, B. O., Price, M., Jovanovic, T., Norrholm, S. D., Gerardi, M., Dunlop, B. & Ressler, K. J. (2014). A randomized, double-blind evaluation of D-cycloserine or alprazolam combined with virtual reality exposure therapy for posttraumatic stress disorder in Iraq and Afghanistan War veterans.*American Journal of Psychiatry*.
3. Gorini, A., & Riva, G. (2014). Virtual reality in anxiety disorders: the past and the future. *Expert Review of Neurotherapeutics*.
4. Lynch, J., & Corrado, G. (2014). Arqueología virtual aplicada al sitio Villavil, Catamarca, Argentina Virtual Archaeology applied to the site Villavil, Catamarca, Argentina.
5. Gaugne, R., Gouranton, V., Dumont, G., Chauffaut, A., & Arnaldi, B. (2014). Immersia, an open immersive infrastructure: doing archaeology in virtual reality. *Archeologia e Calcolatori, supplemento 5*, 1-10.
6. Carroll, J. M. (Ed.). (2003). *HCI models, theories, and frameworks: Toward a multidisciplinary science*. Morgan Kaufmann.
7. Schreuder, H. W., Persson, J. E., Wolswijk, R. G., Ihse, I., Schijven, M. P., & Verheijen, R. H. (2014). Validation of a novel virtual reality simulator for robotic surgery. *The Scientific World Journal*, *2014*.
8. Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire.Presence, 7(3):225–240, 1998.
9. Mel Slater. Measuring presence: A response to the witmer and singer presence questionnaire.Presence, 8(5):560–565, 1999.
10. Rheingold, H. (1991). Virtual reality: exploring the brave new technologies. Simon & Schuster Adult Publishing Group.
11. Khanna, P., Yu, I., Mortensen, J., and Slater, M. (2006). Presence in response to dynamic visual realism: A preliminary report of an experiment study. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST '06, pages 364–367,New York, NY, USA. ACM.
12. Jsselsteijn, W., Ridder, H. d., Freeman, J., Avons, S. E., and Bouwhuis,D. (2001). Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence.Presence: Teleoperators and Virtual Environments, 10(3):298–311.
13. Skarbez, R., Brooks, Jr., F. P., and Whitton, M. C. (2017a). A survey of presence and related concepts.ACM Comput. Surv., 50(6):96:1–96:39.

Aplicaciones de la Televisión Interactiva y aplicaciones móviles para el mejoramiento de los pueblos latinoamericanos

María José Abásolo[12], Armando De Giusti[1], Marcelo Naiouf[1], Patricia Pesado[1], Magdalena Rosado[3], Telmo Silva[6], Joaquín Pina[4], Raoni Kulesza[5]

[1] Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática, Universidad Nacional de La Plata (UNLP), Argentina
{mjabasolo, degiusti, mnaiouf, ppesado}@lidi.info.unlp.edu.ar
[2] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA)
[3] Universidad Católica de Santiago de Guayaquil, Ecuador
[4] Universidad Tecnológica de La Habana José Antonio Echeverría (CUJAE), Cuba
jpina@ceis.cujae.edu.cu
[5] Universidade Federal de Paraíba (UFPB), João Pessoa, Brasil
raoni@lavid.ufpb.br
[6] Universidad de Aveiro (UA), Aveiro, Portugal
tsilva@ua.pt

Resumen

Desde el año 2012 diversas universidades iberoamericanas conforman la Red de Aplicaciones y Usabilidad de la Televisión Digital Interactiva (RedAUTI) coordinada por la UNLP. La línea de investigación y desarrollo presentada consiste en estudiar, desarrollar y evaluar aplicaciones de Televisión Digital Interactiva (TVDi) y tecnologías complementarias para el mejoramiento de los pueblos latinoamericanos. Uno de los principales objetivos es la formación de recursos humanos y fortalecimiento de la investigación mediante el trabajo intergrupal entre diferentes instituciones nacionales y extranjeras.

Palabras Clave: Televisión digital, TVDi, aplicaciones móviles, Interactividad

Contexto

La RedAUTI[1] cuenta con la participación de 29 universidades de España, Portugal y 11 países latinoamericanos (Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, Guatemala, Perú, Uruguay, Venezuela). La RedAUTI fue creada y financiada por el Proyecto "512RT0461 Red de Aplicaciones y Usabilidad de la Televisión Digital Interactiva (RedAUTI)" del Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED) durante el período 2012-2015. Actualmente se lleva a cabo el proyecto internacional: "PGTF INT/19/K08: TDTi*ABC La Televisión Digital Terrestre aplicada al mejoramiento de los pueblos de los países en vías de desarrollo" financiado por el Fondo Fiduciario Pérez Guerrero de la Oficina de Naciones Unidas para la Cooperación Sur-Sur. Dicho proyecto es coordinado por la UNLP y se lleva a cabo con la CUJAE (Cuba) y la UFPB (Brasil).

Introducción

Los televisores son parte de la vida cotidiana para la mayoría de las personas. Por lo tanto, este dispositivo tiene un gran potencial en varias áreas. La televisión digital interactiva (TVDi) se enfoca a que las personas tengan acceso a aplicaciones, como es el caso de tele-salud, tele-gobierno, tele-educación, entre otras, permitiendo así explorar todas las posibilidades y las alternativas de

[1] http://redauti.net

interactividad, donde los usuarios de acuerdo a sus requerimientos, utilicen la TVDi, -desde la comodidad de su hogar- induciéndolos a una mejor proactividad en el desarrollo de su vida cotidiana.
En particular el grupo de investigación III-LIDI ha presentado en años anteriores las líneas de investigación relacionadas con la TVDi [1-4].

RedAUTI

Desde el año 2012 diversas universidades iberoamericanas conforman la Red de Aplicaciones y Usabilidad de la Televisión Digital Interactiva (RedAUTI) coordinada por la UNLP.
El objetivo general es apoyar la formación de los recursos humanos, estimular la investigación de forma transdisciplinaria y fortalecer los grupos de investigación de las universidades iberoamericanas en temas relacionados con TVDI, avanzando en el diseño, implementación y despliegue de aplicaciones, servicios y producción de contenidos para TVDI, en sus múltiples plataformas, de código abierto, de interés colectivo, para dar solución a problemas del contexto iberoamericano.
En particular, se está llevando a cabo un proyecto que tiene como objetivo demostrar las potencialidades de la TDT en los países Argentina, Brasil y Cuba mediante contenido multimedia y softwares adaptados a los escenarios priorizados para el desarrollo en las esferas: Educación, Sanidad, Alimentación y Patrimonio. El proyecto es un marco propicio para aplicar otras tecnologías emergentes como Realidad Virtual (RV) /Realidad Aumentada (RA) y dispositivos móviles para enriquecer las funcionalidades y contenidos presentados al televidente. También se tienen en cuenta modernas técnicas de minería de datos y su integración con redes sociales y portales corporativos para llegar al ciudadano con contenido útil, dinámico y personalizado.

Adultos mayores y TVDi

La TVDi supone una alternativa al uso de la computadora e internet, lo cual puede facilitar el acceso a la sociedad de la información en sectores que aún no disponen o no están acostumbrados a la tecnología. Con este propósito se trata que los usuarios más vulnerables, como es el caso del adulto mayor, tengan una atención integral a través de la tecnología para que su envejecimiento sea activo e independiente. En [5] se presenta una revisión de experiencias realizadas con personas de 65 años o más, empleando las TIC con el objetivo de mejorar su auto-asistencia y empoderamiento para el cuidado de su salud. En particular se está trabajando en utilizar la TVDi para ayudar a la realización de actividades físicas que ayuden al adulto mayor a mejorar su marcha.

Dispositivos móviles y TVDi

Por su parte, los dispositivos móviles, por su naturaleza, permiten a los usuarios captar la atención de los usuarios de manera muy eficiente. Sin embargo, aunque estas potencialidades se están explorando ampliamente en diferentes contextos, aún no se están explotando de manera eficiente en el ecosistema de la televisión. En este contexto y cuando se utilizan como complemento del contenido que se muestra en la pantalla grande, pueden aprovechar la diversidad de servicios. Se pretende, en función de este contexto, notificar al usuario en el televisor que, en relación con el contenido que está viendo, puede disfrutar de promociones o participar, por ejemplo, en juegos relacionados con ese contenido, a través de acciones simples que tendrá que realizar en su dispositivo móvil [6].

Líneas de investigación y desarrollo

Las líneas de investigación y desarrollo principales son:

- Contenidos y Aplicaciones Interactivas para TVDi, con especial interés en asistencia a adultos Mayores
- TVDI y las aplicaciones móviles y otras tecnologías como RA/RV
- Herramientas informáticas para el desarrollo de aplicaciones para TVDi
- Experiencias de usuario con la TVDi: usabilidad y accesibilidad

Resultados

Entre los principales resultados de la RedAUTI se resaltan:

- Difusión mediante el portal web y redes sociales con participación de los miembros enviando información para difusión como publicaciones, proyectos, noticias, eventos, etc.

- Organización de un evento científico anual denominado Jornadas de difusión y capacitación de Aplicaciones y Usabilidad de la TVDi desde el año 2012 [7-19]. Durante 2019 se realizó la 8va Conferencia Iberoamericana de Aplicaciones y Usabilidad de la TVDi AUTI 2019, Rio de Janeiro, Brasil, 30 oct-1 nov 2019. Actualmente se está organizando el próximo evento jAUTI 2020 a realizarse en CUAE (Cuba) en el mes de noviembre de 2020.

- Se están desarrollando contenidos y aplicaciones de TVDi para asistir el mejoramiento de las alteraciones de la marcha en personas mayores, con el objetivo de probar los desarrollos realizados en Ecuador

- Se planifica la realización de cursos de doctorado conjunto con profesores de la UNLP y las instituciones extranjeras colaboradoras

- Se planifica una estancia de investigación de dos profesores extranjeros en la UNLP

Formación de recursos humanos

En la actualidad se está desarrollando una tesis de doctorado en el marco de esta línea de investigación dirigida conjuntamente entre docente argentino y docente extranjero:

- Magdalena Rosado “Televisión Digital Interactiva (TVDi) para reducir las alteraciones de la marcha en adultos mayores”. Directores: M.J.Abásolo y T. Silva (tesis de Doctorado en Ciencias UNLP en curso)

Referencias

[1] Abásolo M.J., De Giusti A., Naiouf M., Pesado P., Rosado Alvarez M., Pina J., Kulesza R., Silva T. (2019) *Aplicaciones de la Televisión Interactiva y tecnologías afines para el mejoramiento de los pueblos latinoamericanos.* XXI Workshop de Investigadores en Ciencias de la Computación WICC 2019, Universidad Nacional de San Juan, RedUNCI, ISBN 978-987-3619-27-4

[2] Abásolo M.J., De Giusti A., Sanz C., Pesado P., Martorelli S., Artola V., Naiouf M., Zangara A., Santos G., Casas S. (2016) *Red AUTI Red de Aplicaciones y Usabilidad de la Televisión Digital Interactiva (TVDi)* XVIII Workshop de Investigadores en Ciencias de la Computación WICC 2016, Entre Ríos, Argentina

[3] Abásolo M.J., De Giusti A., Naiouf M., Sanz C., Zangara A.,Perales Lopez F., Santos G., Casas S., González N., Paz M. (2015) *Red AUTI Red de Aplicaciones y Usabilidad de la Televisión digital Interactiva.* XVII Workshop de Investigadores en Ciencias de la Computación WICC 2015, Salta,

Argentina

[4] Abásolo, M.; Mitaritonna, A.; Giacomantone, J.; De Giusti, A.; Naiouf, M.; Perales, F.;Manresa, C.; Vénere, M.; García Bauza,C. *Realidad Virtual, Realidad Aumentada y TVDI.* WICC 2014 XVI Workshop De Investigadores en Ciencias de la Computación, Ushuaia, Argentina

[5] Rosado M., Abásolo M.J., Silva T. (2020) *Revisión de TIC Orientadas al Adulto Mayor y su Envejecimiento Activo.* En Abásolo M.J., Kulesza R. y Pina-Amargós J. (eds) Libro de aplicaciones y usabilidad de la televisión digital interactiva: AUTI 2019. Universidad Nacional de La Plata.

[6] Silva T., Almeida P., Cardoso B., Oliveira R., Cunha A. and Ribeiro C. (2019) *Exploring balance between TV and Smartphone to distribute efficiently discount coupons*. En Abásolo M.J., Kulesza R. y Pina-Amargós J. (eds) Libro de aplicaciones y usabilidad de la televisión digital interactiva: AUTI 2019, Rio de Janeiro, Brasil, 30 oct-1 nov 2019. Universidad Nacional de La Plata

[7] Abásolo M.J., Kulesza R. y Pina-Amargós J. (eds) *Libro de aplicaciones y usabilidad de la televisión digital interactiva : AUTI 2019 8va Conferencia de Aplicaciones y Usabilidad de la TVDi, Rio de Janeiro, Brasil, 30 oct-1 nov 2019.* Universidad Nacional de La Plata

[8] Abásolo M.J., Silva T. and González D. (eds.) (2019) *Applications and Usability of Interactive Television 7th Iberoamerican Conference, jAUTI 2018, Bernal, Argentina, October 16–18, 2018, Revised Selected Papers*. Communications in Computer and Information Science 1004, Springer ISBN 978-3-030-23862-9

[9] Abásolo M.J., González D. (eds.) (2019) *Anales 7ma Conferencia Iberoamericana JAUTI 2018: Televisión Digital, estudios del audiovisual y nuevas plataformas y VIII Jornadas Transversales de TV Digital.* Universidad Nacional de Quilmes

[10] Abásolo, María José; Ferraz de Abreu, Jorge; Almeida, Pedro; Silva, Telmo (eds.) (2018) *Applications and Usability of Interactive Television 6th Iberoamerican Conference, jAUTI 2017, Aveiro, Portugal, October 12-13, 2017, Revised Selected Papers.* Communications in Computer and Information Science 813, Springer ISBN 978-3-319-90170-1 https://www.springer.com/us/book/9783319901695

[11] Ferraz de Abreu, Jorge; Abásolo Guerrero, María José; Almeida, Pedro; Silva, Telmo (Eds.) (2017) *Anales jAUTI'17. 6ª Jornadas de Aplicaciones y Usabilidade de la Televisión Digital Interativa.* UA Editora, ISBN 978-972-789-521-2

[12] Abásolo, María José; Pedro Almeida; Joaquín Pina Amargós (eds.) (2017) *Applications and Usability of Interactive TV 5th Iberoamerican Conference, jAUTI 2016 La Habana, Cuba November 21-15, 2016 Revised Selected Papers.* Communications in Computer and Information Science 605, Springer ISBN 978-3-319-63321-3, julio 2017 http://www.springer.com/in/book/9783319633206

[13] Abásolo Guerrero, María José; Pina Amargós, Joaquín (2017) *Libro de aplicaciones y usabilidad de la televisión digital interactiva. V Jornadas Iberoamericanas sobre Aplicaciones y Usabilidad de la Televisión Digital Interactiva, jAUTI2016. Artículos*

seleccionados. UNLP, ISBN: 978-950-34-1514-6

[14] Abásolo, María José; Perales López, Francisco; Bibiloni Coll, Antoni (eds.) (2016) *Applications and Usability of Interactive TV 4th Iberoamerican Conference, jAUTI 2015 and 6th Congress on Interactive Digital TV, CTVDI 2015 Palma de Mallorca, Spain, October 15–16, 2015 Revised Selected Papers.* Communications in Computer and Information Science 605, Springer ISBN 978-3-319-38906-6, mayo 2016 http://www.springer.com/us/book/9783319389066

[15] A.Bibiloni, F. Perales, M.J.Abásolo (Eds.) (2016) *"Proceedings del VI Congreso de Televisión Digital Interactiva - jAUTI 2015 IV Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva"* celebrado en Palma (España) del 14 al 16 de octubre 2015. Universitat de les Illes Balears (España). ISBN: 978-84-608-5567-5

[16] Abásolo, M.J., Kulesza, R. (Eds.) (2015) *Applications and Usability of Interactive TV Third Iberoamerican Conference, jAUTI 2014, and Third Workshop on Interactive Digital TV, Held as Part of Webmedia 2014, João Pessoa, PB, Brazil, November 18-21, 2014. Revised Selected Papers* ISBN: 978-3-319-22656-9, Communications in Computer and Information Science Springer-Verlag,

[17] Abásolo, M.J., Kulesza, R. (Eds.) (2015) *"Proceedings de jAUTI 2014 III Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva | Workshop WTVI Webmedia 2014"* celebrado en João Pessoa, Paraíba (Brasil), Noviembre 2014, ISBN: 978-950-34-1188-9, Universidad Nacional de La Plata (Argentina), 2015

[18] Abásolo, M.J., Castro, C. de (Eds.) (2014) *"Anales jAUTI 2013 II Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva"* celebrado en Córdoba (España) en septiembre 2013, ISBN: 978-84-697-0302-1. Universidad de Córdoba (España), 2014.

[19] Abásolo, M.J., De Giusti A. (Eds.) (2013) "Anales jAUTI 2012 I Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva" celebrado en La Plata (Argentina) en octubre 2013. ISBN: 978-950-34-0945-9. Universidad Nacional de La Plata (Argentina), 2013.

Aplicaciones de Realidad Extendida y Aplicaciones Móviles

María José Abásolo[1 2] , Armando De Giusti[1], Marcelo Naiouf[1], Patricia Pesado[1], Sebastián Barbieri[3], Wilma Gavilanes[4], Alejandro Mitaritonna[5], Mario Alberto Vincenzi, Bria O.[1], Ronchetti F.[1], Francisco Montero[7], Francisco Perales López[6], Valeria Springer[8]

[1] **III-LIDI Instituto de Investigación en Informática,**
Facultad de Informática – Universidad Nacional de La Plata (UNLP), Argentina
{mjabasolo, degiusti, mnaiouf, ppesado}@lidi.info.unlp.edu.ar
[2] **Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA), Argentina**
[3] **Facultad de Cs.Exactas.Universidad Nacional del Centrode la Pcia.de Bs.As.(UNICEN), Argentina**
[4] **Universidad Técnica de Ambato, Ecuador**
[5] **Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF), Argentina**
[6] **Universidad de las Islas Baleares (UIB), España**
[7] **Universidad de Castilla-La Mancha (UCLM), España**
[8] **Universidad Nacional del Sur (UNS), Argentina**

Resumen

La línea de investigación y desarrollo presentada en este artículo realiza actividades de análisis, desarrollo y evaluación de aplicaciones de Realidad Extendida, siendo este un término para englobar en este contexto aplicaciones de Realidad Virtual, Realidad Aumentada, Interfaces Multimodales, Visión por computador. Uno de los principales objetivos es la formación de recursos humanos y el fortalecimiento de la investigación mediante el trabajo intergrupal entre diferentes instituciones nacionales y extranjeras.

Palabras Clave: Realidad Extendida, Realidad Aumentada, Realidad Virtual, Interfaces multimodales, Visión por Computador

Contexto

Las diferentes líneas de investigación están enmarcadas en diversos proyectos:

- Proyecto IDEAS-CICPBA “AMAPAS Aplicaciones Móviles para la Medición de Agentes Peligrosos para el Ambiente y la Salud” (2019-2020) dirigido por M.J.Abásolo.
- Proyecto acreditado UNLP Programa de Incentivos "Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real" (2018-2021) dirigido por M. Naiouf
- Proyecto acreditado UNLP Programa de Incentivos "Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso" (2018-2021) dirigido por P. Pesado

Introducción

Desde hace varios años el grupo de investigación III-LIDI se dedica al desarrollo de aplicaciones en las áreas de Visión por Computador, Realidad Aumentada y Realidad Virtual [1-7]. En este artículo se presentan los principales avances alcanzados en esas diferentes líneas de investigación fruto de un trabajo conjunto con investigadores de diferentes instituciones del país y España.

El término Realidad Extendida

Realidad Extendida (RA) es un término abarcativo que se refiere a las tecnologías que fusionan el mundo real con mundos virtuales. Engloba Realidad Aumentada (RA)- mundo real visto a través de un dispositivo con información añadida-, Realidad Virtual (RV) - el usuario está inmerso en un mundo virtual-, Realidad Mixta (RM) - una combinación de RA y RV- así como las áreas

relacionadas como Interfaces Multimodales. Chuah [8] señala que pese a la perspectiva prometedora, esta tecnología naciente ha sido envuelta por posibilidades inciertas, haciendo que la adopción sea mucho más lenta de lo esperado. Realiza una revisión actual identificando los principales antecedentes y factores importantes para la adopción de RE, clasificando y mapeando en los modelos de referencia, revelando las áreas en las que falta investigación.

Realidad Aumentada en el Ámbito Militar

En esta línea de investigación se han analizado diferentes proyectos de RA en el ámbito militar, y se ha trabajado sobre la implementación del framework RAIOM (Realidad Aumentada para la Identificación de Objetivos Militares) desarrollado en el CITEDEF (Instituto de Investigaciones Científicas y Técnicas para la Defensa) [9-12]. Se utilizan dispositivos móviles, visión por computador y sensores externos para el reconocimiento, detección, ubicación, identificación y suministro de información contextual mediante gafas de realidad aumentada.

Se diseñó una arquitectura cliente-servidor, en la cual el servidor se implementa utilizando un procesador mini board ODROID-XU3 y el cliente utilizando gafas Epson Moverio. En el cliente se ejecutan los procesos menos intensivos tales como captura de vídeo, sensores de ubicación y tracking, comunicación, mapeo y renderizado de imágenes. Por otra parte, el servidor implementa procesos que soportan cargas intensivas de cómputo.

Las siguientes características funcionales son incluidas en RAIOM:

- Radar de 360 grados donde se visualiza la posición del usuario y los objetivos tácticos y de amenaza mediante el uso de símbolos militares.
- Reconocimiento de edificios y otros tipos de infraestructura de referencia
- Mapas interactivos que se visualizan superpuestos sobre la mano del operador
- Manipulación de menús gráficos mediante el reconocimiento de gestos y voz.

Se elaboró el modelo 3D-SAM [11] para la identificación, selección y clasificación de los requisitos de la conciencia situacional (CS) junto con el diseño de soluciones basadas en RA. Con el modelo 3D-SA se identifican, documentan y clasifican los requisitos de la CS y se elaboran prototipos relacionados con las historias de usuarios de personal militar.

El despliegue se evaluó con personal militar mediante la realización de tareas utilizando el sistema, evaluando mediante entrevistas el nivel de CS y la experiencia del usuario.

Realidad Aumentada para brindar Información al Ciudadano

Las aplicaciones basadas en lectura del código de barras y códigos QR como aplicaciones de Realidad Aumentada, en las cuales los códigos son hiperenlaces a contenidos, sin existir registro de modelos 3D en 3D ni seguimiento de los marcadores. Dado el uso masivo de *smartphones* este tipo de aplicaciones resulta muy útil a la hora de brindar información contextual al ciudadano. Se desarrolló la aplicación móvil ConZoom que, a partir de la captura con el teléfono celular del código de barras presente en el producto, brinda información sobre el producto. En particular se informa la información presente en las etiquetas nutricionales así como el reciclaje post consumo [13].

La aplicación ConZoom se nutre de una base de datos que, hasta el momento, debe ser manualmente cargada. Se pretende utilizar herramientas OCR que permitan semi automatizar la la carga de forma que se reduzca el tiempo de carga y la tasa de errores. En relación a esto Reibring, J. [14] propone un algoritmo que combina *Machine Learning* con procesamiento de imágenes para la detección del texto presente en las etiquetas nutricionales de EEUU.

Entre las implementaciones de OCR disponibles más reconocidas se encuentran

Tesseract OCR, GOCR, Cuneiform, OCRFeeder, Ocrad. Todas las aplicaciones mencionadas son de código abierto y distribución gratuita, y son el punto de partida de nuestro trabajo en el cual se está probando cómo se comportan con diferentes tipos de etiquetas nutricionales de productos disponibles en el mercado. En estas pruebas resulta imprescindible el pre-procesamiento, el cual consiste en una serie de operaciones necesarias para conseguir caracteres fácilmente discriminables que permitan alcanzar una mayor tasa de acierto. Entre las operaciones necesarias se incluye la eliminación de ruido, binarización para discernir carácter de fondo, normalización de textos, rotación, escalado, operaciones morfológicas.

Realidad Aumentada en Educación

La RA es una tecnología innovadora y con importantes oportunidades en los procesos de enseñanza-aprendizaje. En [15] se presenta un resumen de revisiones de investigaciones recientes sobre las posibilidades educativas de la RA.

La incorporación de la RA en los procesos educativos en diferentes disciplinas, incluyendo medicina, arquitectura, urbanismo, matemáticas y geometría, idiomas, arte, historia, ciencias naturales, química, física y geografía. En [16] se presenta una recopilación actual de aplicaciones de RA educativas en diferentes áreas y países. A su vez, se realizó la catalogación de aplicaciones de RA disponibles en el mercado con el objetivo de que sean de utilidad a docentes de los diferentes niveles educativos [17].

En [18] se presenta una experiencia universitaria de innovación de estudiantes universitarios que diseñan objetos de aprendizaje enriquecidos con RA. Se mide el nivel de aceptación tecnológica (TAM), el impacto en el rendimiento académico, la prueba del proceso de capacitación y las herramientas de producción utilizadas. Los resultados mostraron que el uso de AR en la enseñanza universitaria ha suscitado un alto grado de aceptación y motivación, se determinó una mejora en el rendimiento académico y las ventajas y desventajas de las herramientas utilizadas para la producción.

Aplicaciones móviles para información de agentes peligrosos para la salud y el ambiente

Como consecuencia del crecimiento de la población mundial, la producción agrícola-ganadera se ha intensificado notablemente en los últimos años para suplir la progresiva demanda de alimentos por parte de la sociedad. El uso masivo de plaguicidas en prácticas agrícolas y en el ambiente doméstico ha generado en los últimos años una creciente preocupación de carácter socio-económico puesto que sus residuos representan un peligro potencial para la salud pública. Tal es el caso de los plaguicidas organofosforados, ampliamente utilizados para el control de malezas e insectos, cuyos residuos han sido detectados en alimentos y aguas destinados al consumo humano. Por otro lado, debido a la intensificación de las prácticas agropecuarias, se requiere especial atención en la presencia de nitratos en alimentos y aguas naturales. Estos pueden generar nitritos por acción bacteriana con el consecuente efecto nocivo en la salud ya que su efecto es la generación de metahemoglobina que provoca cianosis, y nitrosaminas, las cuales pueden comportarse como agentes cancerígenos.

El desarrollo de métodos de análisis rápidos, sencillos, económicos y portables para la determinación de agroquímicos y nitratos en productos agrícolas y aguas tiene gran importancia e implicancia socio-económica no sólo en nuestro país sino también en el mundo, para el consumo seguro y sustentable. Los dispositivos móviles se presentan como un instrumento analítico al alcance de todos para la detección colorimétrica. López-Ruiz et al [19] presentan una aplicación Android que captura una imagen de un papel que contiene diferentes áreas a sensar, y realiza la

detección de las mismas analizando los cambios de color al colocarse una solución reactiva. Actualmente el grupo de investigación está trabajando actualmente junto a una investigadora del Departamento de Química de la UNS, en el desarrollo de una aplicación móvil que sirva como herramienta de análisis para el control fitosanitario de productos agrícolas y aguas. En particular, se están realizando pruebas de detección de nitratos y nitritos, y plaguicidas plaguicida organofosforado como clorpirifos-metil.

Además, dado que los dispositivos móviles con GPS pueden identificar la ubicación de la prueba y transmitir los resultados a una base de datos geográfica, Sicard et al [20] presentan un sistema que combina dispositivos analíticos basados en papel microfluídico con lectura mediante dispositivos móviles y un sitio web para el monitoreo ambiental de la calidad del agua. En este sentido, se está desarrollando una aplicación móvil que pueda servir como herramienta de gobiernos y ciudadanos al permitir volcar información georeferenciada de análisis realizados (por ej. en pozos de agua de red o de agua de red en los hogares), de forma tal que pueda consultarse el mapa de lecturas.

Sistemas Multimodales

Existen un conjunto de paradigmas de interacción persona-ordenador, tales como interfaces multimodales, afectivas, ubicuas, tangibles, realidad aumentada y mixta, basadas en agentes conversacionales, etc. que pretenden una interacción natural y multimodal embebida en el propio entorno físico del usuario. En el marco de estas tecnologías, los Agentes Sociales son capaces de interaccionar de forma natural con el usuario, a través de un interfaz de humano virtual, o de un dispositivo robot con capacidades sociales y afectivas. Perales et al. [21] propone la implementación de un sistema interactivo dirigido al tratamiento de personas con autismo basado en un *serious games* para la captura de las expresiones en tiempo real, durante la ejecución del juego y mediante el uso de *deep learning* plantea obtener una clasificación de dichas emociones.

Según Pennisi et al. [22] los robots socialmente interactivos se utilizan para comunicar, expresar y percibir emociones, mantener relaciones sociales, interpretar claves naturales y desarrollar competencias sociales. Se utilizan como herramientas para enseñar habilidades a los niños con autismo, jugar con ellos y provocar ciertos comportamientos deseados de ellos. Crean interesantes, atractivos y significativas situaciones de interacción que motivan a los niños a interactuar con ellos.

En [23] se presenta un sistema de interacción multimodal avanzado basado en un robot social que permite a los usuarios replicar y aprender de forma lúdica las expresiones faciales básicas. Este paradigma de interacción pretende fomentar la atención y la motivación de los usuarios, especialmente de las personas con necesidades especiales. El sistema está capacitado para evaluar la expresión realizada por el usuario delante del robot. Se ha utilizado una implementación de red neuronal propia que está entrenada para el reconocimiento de emociones faciales. En base a esta red CNN (Convolutional Neural Network) el sistema hace de entrenador personal de emociones a los usuarios. Finalmente, se puede usar como un sistema de adquisición de expresiones faciales para generar una nueva base de datos de acuerdo a las características de los usuarios. En el diseño del sistema propuesto tenemos dos aproximaciones: una primera basada en una red neuronal convolucional (CNN) previamente entrenada para clasificar expresiones faciales, y una segunda donde esa valoración es realizada por los usuarios. En base a lo anterior se plantean los siguientes objetivos: a) Evaluación de la CNN en entornos reales con usuarios no expertos (idea principal de este experimento). b) Herramienta de valoración de la experiencia de usuario (UX) y que el robot pueda adaptarse según las emociones del usuario

(retroalimentación positiva). c) Herramienta como entrenador de emociones, siendo el robot social un supervisor del nivel de aciertos del usuario respecto a la emoción realizada.
Por otra parte, en Argentina se están llevando a cabo pruebas de entrenamiento en niños con TEA con un robot de bajo costo. Algunos de los objetivos consisten en evaluar la eficacia del tratamiento y comparar sus resultados con los obtenidos por la Universidad de las Islas Baleares con equipamiento de costes muy superiores.

Líneas de investigación y desarrollo

- Aplicaciones de Visión por computador y Realidad Aumentada para brindar información y servicios al ciudadano
- Realidad Virtual y Aumentada aplicada a la educación
- TVDI y aplicaciones móviles

Resultados y Objetivos

Con el objetivo de formar recursos humanos desde el año 2012 se dicta la carrera de postgrado "Especialización en Computación Gráfica, Imágenes y Visión por Computadora", Facultad de Informática de la UNLP. (Nº 11.162/12).

- Se ha realizado el dictado de cursos de doctorado relacionados con la temática, como por ejemplo: "Realidad Aumentada", "Interfaces Avanzadas" y "Tópicos de Procesamiento de Imágenes" (junto a profesores de la Universidad de las Islas Baleares), "Usabilidad" (junto a profesor de la Universidad de Castilla La Mancha).
- Se finalizó el desarrollo y pruebas del framework de Realidad Aumentada en el ámbito militar RAIOM
- En el ámbito de la asistencia al ciudadano se realizaron pruebas del prototipo de aplicación móvil ConZoom basada en la captura de códigos de barra de productos de supermercado, con el objetivo de brindar información
- Se realizó el Repositorio de Aplicaciones Educativas RA [17]

Formación de recursos humanos

Las siguientes diferentes tesis de grado y postgrado se desarrollan en el marco de esta línea de investigación:

- Alejandro Mitaritonna. "Realidad Aumentada para la Identificación de Objetivos Militares" . Directores: Abásolo M.J., Montero F. (Tesis de Doctorado en Ciencias aprobada diciembre de 2019) [9]
- Mario Vincenzi. "La Realidad Aumentada en la educación. Vigencia, proyecciones y límites" Director: Abásolo M.J. (Tesis de especialización aprobada en marzo de 2020) [16]
- Wilma Gavilanes "Metodología para la evaluación del impacto de experiencias con Realidad aumentada en educación superior" Director: Abásolo M.J. (tesis de Doctorado en Ciencias en curso)
- Florencia Puppo "Reconocimiento óptico de caracteres de la etiqueta nutricional de productos alimenticios". Directores: Abásolo M.J., Ronchetti F. (tesina de grado de Licenciatura en Sistemas en curso)
- Luis Franceschi "Aplicación móvil para detección de agroquímicos a partir de colorimetría". Directores: Abásolo M.J., Bria O. (tesina de grado de Licenciatura en Sistemas en curso)
- Agustín De Luca "Sistema de censo y visualización de datos ambientales georeferenciados" Director: Abásolo M.J. (tesina de grado de Licenciatura en Sistemas en curso)

Además, se colabora en la formación de recursos humanos de otras universidades, entre los cuales se enumeran:

- Lucas Benjamin Cicerchia "Detección de enfermedades y falta de nutrientes en cultivos utilizando algoritmos de Active Learning aplicados al sensado remoto" Directores: Claudia Russo (UNNOBA), María José Abásolo (tesis de Doctorado en Ciencias UNLP en curso)

- Sebastián Barbieri "Interfaces multimodales aplicadas a la valoración de niños con TEA y THDA" Directores: Perales F., Abásolo M.J. (tesis doctoral de la Universidad de las Islas Baleares en curso)

Referencias

[1] Abásolo M.J., De Giusti A., Naiouf M., Pesado P., Sanz C., Barbieri S., Boza R., Gavilanes W., Mitaritonna A., Prinsich N., Vincenzi, M., Montero F., Perales López F. *Aplicaciones de realidad virtual, realidad aumentada e interfaces multimodales.* XXI Workshop de Investigadores en Ciencias de la Computación WICC 2019, Universidad Nacional de San Juan, RedUNCI, ISBN 978-987-3619-27-4

[2] Abásolo M.J., Mitaritonna A., Castañeda, S., Sanz C., Boza R., Prinsich N., Silva, T., Rosado, M.; Naiouf, M.; Pesado, P.; De Giusti, A. (2018) *Aplicaciones de visión por computador, realidad aumentada y TVDi.* XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste). p. 389-395, RedUNCI, ISBN: 978-987-3619-27-4

[3] Abásolo, M.J.; Sanz, C.; Naiouf, M.; De Giusti, A.; Santos, G.; Castro, M.;Bouciguez, M.J. G. (2017) *Realidad Aumentada, Realidad Virtual e Interacción Tangible para la Educación.* XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017), pp. 1312-1316, RedUNCI, ISBN 978-987-42-5143-5

[4] Abásolo, M.; Mitaritonna, A.; Bouciguez, M.; Encina, N.; Vincenzi, M.; De Giusti, A.; Naiouf, M.; Giacomantone, J.; Manresa Yee, C. *Realidad Aumentada, Realidad Virtual, Interfaces Avanzadas y Juegos Educativos.* XVIII Workshop de Investigadores en Ciencia de la Computación (WICC 2016); Concordia, Entre Ríos, RedUNCI, ISBN 978-950-698-377-2

[5] Abásolo, M; Mitaritonna A.; Encina N.; Vincenzi M.; Borelli L.; De Giusti A.; Naiouf M.; Giacomantone J. *Realidad Aumentada y Realidad Virtual* XVII Workshop de Investigadores en Ciencia de la Computación (WICC 2015), RedUNCI, ISBN 978-987-633-134-0

[6] Abásolo, M.; Mitaritonna, A.; Giacomantone, J.; De Giusti, A.; Naiouf, M.; Perales, F.;Manresa, C.; Vénere, M.; García Bauza,C. *Realidad Virtual, Realidad Aumentada y TVDI.* WICC 2014 XVI Workshop De Investigadores en Ciencias de la Computación, RedUNCI, ISBN 978-950-34-1084-4

[7] Manresa-Yee, C.; Abásolo, M.J.; Mas Sansó, R.; Vénere, M.. (2011) *Realidad Virtual, Realidad Aumentada e Interfaces Basadas en Visión.* XV Escuela Internacional de Informática, XVII Congreso Argentino de Ciencia de la Computación CACIC 2011. Editorial EDULP, ISBN 978-950-34-0765-3

[8] Chuah, S. (2018). Why and Who Will Adopt Extended Reality Technology. *Literature Review, Synthesis, and Future Research Agenda*.

[9] Mitaritonna, A. (2019) *Empoderamiento de la Conciencia Situacional en Operaciones Militares utilizando Realidad Aumentada.* Tesis doctoral, Facultad de Informática, Universidad Nacional de La Plata

[10] Mitaritonna A., Abásolo M.J. y Montero F. (2019) *Situational Awareness through Augmented Reality: 3D-SA Model to relate Requirements, Design and Evaluation.* Proceedings of the 9th International Conference on Virtual Reality and Visualization, Shenzhen University, China

[11] Mitaritonna, A.; Abásolo, M. J. (2015) *Improving Situational Awareness in Military Operations using Augmented Reality*. Proceedings of WSCG 2015. ISBN Nº:978-80-86943-72-5

[12] Mitaritonna, A.; Abásolo, M. J.(2013) *Mejorando la conciencia situacional en operaciones militares utilizando la realidad aumentada* (2013) Proceedings of XVIII Congreso Argentino de Ciencias de la Computación. ISBN Nº: 978-987-23963-1-2 pp. 356-365.

[13] Boza, R.; Prinsich, N., Abásolo, M.J.; Sanz, C. (2019) *Aplicación móvil para el consumo consciente y responsable.* Directores. Tesis de grado de la carrera Licenciatura en Informática, Facultad de Informática, Universidad Nacional de La Plata, febrero 2019.

[14] Reibring, J.. (2017). *Photo OCR for Nutrition Labels. Combining Machine Learning and General Image Processing for Text Detection of American Nutrition Labels.* Gothenburg, Sweden: Master's Thesis Department of Signals and Systems Chalmers University of Technology.

[15] Gavilanes, W., Abásolo, M. J., Cuji, B. (2018) *"Realidad Aumentada en la Educación: una Revisión desde la Perspectiva Pedagógica",* Revista ESPACIOS. ISSN 0798 1015

[16] Vincenzi M. (2019) *La realidad aumentada en la educación: catalogación de aplicaciones educativas.* Especialista en Tecnología Informática Aplicada en Educación, Facultad de Informática de la Universidad Nacional de La Plata

[17] Vincenzi M. (2019) *Repositorio de aplicaciones educativas RA* https://sites.google.com/view/repositorio-educativo-ra

[18] Gavilanes López W.L., Cuji B.R., Salazar Mera J.V., Abásolo M.J. (2020) *Methodology for the Production of Learning Objects Enriched with Augmented Reality by University Students.* In: Auer M., Hortsch H., Sethakul P. (eds) The Impact of the 4th Industrial Revolution on Engineering Education. ICL 2019. Advances in Intelligent Systems and Computing, vol 1134. Springer, Cham.

[19] Lopez-Ruiz, N., Curto, V. F., Erenas, M. M., Benito-Lopez, F., Diamond, D., Palma, A. J. & Capitan-Vallvey, L. F. (2014). *Smartphone-based simultaneous pH and nitrite colorimetric determination for paper microfluidic devices.* Analytical chemistry, 86(19), 9554-9562.

[20] Sicard, C., Glen, C., Aubie, B., Wallace, D., Jahanshahi-Anbuhi, S., Pennings, K. & Filipe, C. D. (2015). *Tools for water quality monitoring and mapping using paper-based sensors and cell phones.* Water research, 70, 360-369.

[21] Perales, Francisco & Ramis, Silvia & Riquelme, Inmaculada. (2017). *Un videojuego serio para el estudio de expresiones faciales en personas con Autismo.* Cognitive Area Networks. 4. 85.

[22] Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., & Pioggia, G. (2015). *Autism and Social Robotics: A Systematic Review*. Autism Research, 9, 165-183.

[23] Ramis, Silvia & Perales, Francisco & Guerrero, Alejandro & Buades Rubio, Jose Maria. (2019). *Interacción basada en robots sociales para la evaluación de expresiones faciales.* Interacción 2019, June 25–28, 2019, Donostia, Gipuzkoa, Spain

Ingeniería de Software

Desarrollo de requisitos en "Aspect-Oriented Process for a Smooth Transition"

Fernando Pinciroli y Marcelo Palma

Instituto de Investigaciones
Facultad de Informática y Diseño
Universidad Champagnat
Mendoza, Argentina, +54 (261) 424-8443
pincirolifernando@uch.edu.ar; mpalma@civitas.com.ar

Resumen

En este proyecto se está trabajando en el establecimiento de los lineamientos para la fase de requisitos de usuario, del ciclo de vida del desarrollo de software, para el paradigma de orientación a aspectos.

El acrónimo AOP4ST corresponde Aspect-Oriented Process for a Smooth Transition, que es un proceso marco para el desarrollo de software que pretende:

a) Permitir, ante todo, alcanzar los objetivos específicos de cada una de sus fases.

b) Mantener un proceso homogéneo a lo largo de todas las fases del ciclo de vida del desarrollo de software en cuanto a herramientas, estándares, notación, técnicas, etc.

c) Colaborar con la detección de incumbencias en forma progresiva y manteniendo su separación a lo largo del ciclo de vida.

d) Asegurar la separación de incumbencias, su encapsulamiento y composición, como así también la resolución de los conflictos que pudieran ocurrir.

e) Aportar a la definición de las mejores prácticas para la ingeniería de requisitos orientada a aspectos.

Este proyecto es el tercero de la misma línea de investigación de desarrollo de software orientado a aspectos (AOSD) que se vienen desarrollando en forma consecutiva desde el año 2013 en el Instituto de Investigaciones de la Facultad de Informática y Diseño de la Universidad Champagnat, en la provincia de Mendoza, y que continúan del otro llevado a cabo previamente en la Universidad Tecnológica Nacional, Facultad Regional Mendoza.

Palabras clave: orientación a aspectos, early aspects, modelado de requisitos de usuario, requisitos funcionales, requisitos no funcionales, atributos de calidad, reglas de negocio, separación de incumbencias, encapsulamiento de incumbencias, composición de incumbencias, resolución de conflictos, AOP4ST.

Contexto

Dentro de las líneas de investigación del Instituto de Investigaciones de la

Facultad de Informática y Diseño "Prof. Carlos Olivera", de la Universidad Champagnat, se encuentra la línea de "Ingeniería de Software". En ella se han llevado a cabo una serie de proyectos a lo largo de más de seis años, sobre desarrollo de software orientado a aspectos (AOSD), en particular, los siguientes tres proyectos consecutivos:

- *"Procesos de desarrollo de software de calidad basados en aspectos"*, realizado en conjunto con la UTN Facultad Regional Mendoza [1] durante 2011 y 2012.

- *"Definición de criterios para la detección temprana de aspectos en el modelado de negocios y el desarrollo de los requisitos"* llevado a cabo por completo en la Universidad Champagnat, desde 2013 a 2016 y presentado en WICC 2015 [2].

- *"Modelado de procesos de negocio orientados a aspectos con BPMN",* de la Universidad Champagnat y realizado en los años 2016 a 2018, y que fue presentado en WICC 2016 [3] y 2017 [4].

- *"Ingeniería de requisitos orientada a aspectos en AOP4ST"*: este nuevo proyecto aborda la fase de requisitos de usuario y se inició formalmente en junio de 2018.

Durante el año y medio de desarrollo de este proyecto ya se realizaron seis publicaciones: [5] [6] [7] [8] [9] [10].

Este proyecto cuenta con el financiamiento de la Universidad Champagnat y de la empresa Aconcagua Software Factory S.A., una de las principales fábricas de software de la provincia de Mendoza, con oficinas en esa provincia y en Buenos Aires, Santiago de Chile, Madrid, Barcelona y Miami.

Introducción

La definición de un proceso de desarrollo exige abarcar numerosos frentes. Uno de ellos corresponde a la definición específica de cada una de sus fases del ciclo de vida del desarrollo de software. En el presente proyecto se procura establecer los lineamientos de la fase de requisitos de usuario, que en la industria se denomina "gestión de la demanda" [11]. Al mismo tiempo, también se procura que esta fase continúe en forma homogénea y natural de la fase anterior, correspondiente al modelo de procesos de negocio.

En cuanto a las exigencias del paradigma en particular que se está empleando, se espera que en esta definición se tengan en cuenta las actividades propias del enfoque orientado a aspectos, que son la detección de las incumbencias, su separación, su encapsulamiento, la posterior composición y la resolución de los posibles conflictos. Además, como esta fase del ciclo de vida es continuación de una anterior, es imprescindible el aseguramiento de la trazabilidad con las incumbencias detectadas en el modelo anterior.

Finalmente, también se espera establecer buenas prácticas de manera de asegurar los objetivos de la fase del ciclo de vida en cuestión, la calidad del producto software en desarrollo y la obtención de los beneficios que son la razón del empleo del paradigma de orientación a aspectos. Estos beneficios apuntan al desarrollo de un producto de software final, como así también de los productos intermedios que permiten producirlo, más modular, mantenible, reusable, extensible, comprensible, etc.

[12], al administrar en forma separada las incumbencias que están *desparramadas y enredadas* en cada uno de los niveles de abstracción a lo largo del ciclo de vida completo.

Estas incumbencias pueden corresponder a cuestiones que son específicas del dominio del problema y también cuestiones que no lo son, lo que corresponde a los enfoques simétrico y asimétrico de la orientación a aspectos [13] [14].

Desde ya que los principales aportes a los requisitos tempranos (early aspects) son fuentes obligadas de referencia y de las que nos nutrimos en este proyecto [15] [16] [17] [18]. También, los estudios de mapeo sistemático de Parreira [19] y de Narender [20] nos orientan en la organización y selección de las principales fuentes.

Líneas de Investigación, Desarrollo e Innovación

Los proyectos de investigación que se vienen desarrollando en nuestra Facultad presentan cuatro ejes de investigación, donde los tres primeros se ven integrados con el cuarto:

1. *Modelado de procesos de negocio orientados a aspectos.*
2. *Separación y composición de incumbencias con resolución de conflictos.*
3. *Especificación de requisitos y gestión de incumbencias con casos de uso e historias de usuario.*
4. *Procesos de desarrollo de software orientados a aspectos.*

Resultados y Objetivos

Existen resultados alcanzados por los proyectos anteriores que posibilitan los resultados de este proyecto.

- Los resultados provenientes de los proyectos anteriores son la definición del marco del proceso AOP4ST [1] [2] y la descripción detallada de ese proceso para su primer modelo, el modelo de procesos de negocio [3] [4] [21] [22] [23].

A partir de ellos se elaboraron las definiciones para el segundo modelo, el modelo de requisitos de usuario, que constituye el foco del presente proyecto.

Siguiendo las actividades que correspondientes a los modelos orientados a aspectos mencionados en el punto anterior, se avanzó en los siguientes puntos:

Detección y separación de incumbencias: en el modelo de negocio se detectaron incumbencias, que deben trasladarse al modelo de requisitos de usuario y constituyen las primeras incumbencias de este modelo. Al seguir los procesos de negocio del modelo inicial, se deben obtener los requisitos de usuario que pueden ser de tipo funcional, no funcional (atributos de calidad) y reglas de negocio. Los requisitos funcionales son más sencillos de obtener y su gestión cae dentro del enfoque simétrico de AOSD. Ellos se van ubicando dentro de las incumbencias heredadas del primer modelo, pero pueden tener trazabilidad "muchos a muchos" ellas, lo que podría exigir la definición de una nueva incumbencia transversal en este modelo. Algunas ideas al respecto fueron presentadas en [24].

Los atributos de calidad, o requisitos no funcionales, siguen el mismo criterio empleado para los requisitos funcionales, pero por su propia naturaleza tienen

relaciones de contribución positivas y negativas que obligan a rever las incumbencias a las que son asignados. Las consecuencias de estas relaciones de contribuciones fueron presentadas en [25] y [26].

Con respecto a las reglas de negocio, se siguen también los lineamientos descriptos y que, para la detección de incumbencias originadas por los tres tipos de requisitos, se publicaron los avances en [6].

Composición de incumbencias: no hemos avanzado por el momento en este punto, más que nada porque entendemos que con la definición de la trazabilidad se estaría cubriendo, pero entendemos que esto requiere un estudio más profundo y que está pendiente en nuestra agenda de trabajo.

Resolución de conflictos: también está pendiente de avance, aunque en esta fase del ciclo de vida, los conflictos entre requisitos han sido largamente estudiados y, además, el estudio entre las relaciones de contribución positivas y negativas mencionadas entre atributos de calidad, en el que hemos hecho avances importantes, también aportan a los objetivos de esta actividad.

Por último, está pendiente de publicación un estudio de mapeo sistemático sobre herramientas, técnica, notaciones, etc. orientadas a aspectos existentes a lo largo de la vida del paradigma, cuyo protocolo está disponible en ArXiv [9].

Los criterios adoptados para esta fase del ciclo de vida están siendo aplicados en diversos proyectos -y para ser honestos, con diferente suerte, ya que dependemos de los casos que se presentan y no de los que desearíamos tener- de Aconcagua Software Factory S.A., una de las principales fábricas de software de la provincia de Mendoza.

Formación de Recursos Humanos

El proyecto promueve la formación del equipo de profesores, alumnos y egresados de la Universidad Champagnat.

Se están elaborando diferentes tesis sobre el cuerpo de esta investigación: la tesis de Gustavo Albino, de la Maestría en Ingeniería de Software de la Universidad Nacional de San Luis, la tesis doctoral de Fernando Pinciroli, en el Doctorado en Ciencias Informáticas de la Universidad Nacional de San Juan y dos tesinas de grado de la carrera de Licenciatura en Sistemas de Información de nuestra Universidad, de los alumnos Elías Motta y Javier Amutio.

Referencias

[1] F. Pinciroli, “Consideraciones para un proceso de desarrollo de software de calidad orientado a aspectos,” in *Sexto Encuentro de Investigadores y Docentes de Ingeniería EnIDI 2011*, 2011.

[2] F. Pinciroli, “AOP4ST – Aspect-Oriented Process for a Smooth Transition,” in *WICC 2015 - XVII Workshop de Investigadores en Ciencias de la Computación*, 2015.

[3] F. Pinciroli and L. Zeligueta, “El modelo de negocio en AOP4ST,” in *WICC 2016 - XVIII Workshop de Investigadores en Ciencias de la Computación*, 2016.

[4] F. Pinciroli and L. Zeligueta, “Modelado de negocios orientado a aspectos con AOP4ST,” in *WICC 2017 - XIX Workshop de Investigadores en Ciencias de la Computación*, 2017.

[5] F. Pinciroli, “Requisitos de usuario y gestión de la demanda en AOP4ST,” in *XX Workshop de Investigadores en Ciencias de la Computación - WICC 2018*, 2018.

[6] F. Pinciroli, “Concern detection along the requirement development,” in *Congreso Internacional de Ciencias de la Computación y Sistemas de Información - CICCSI 2017*, 2017.

[7] J. L. Barros-Justo, F. Pinciroli, S. Matalonga, and N. Martínez-Araujo, "What software reuse benefits have been transferred to the industry? A systematic mapping study," *Inf. Softw. Technol.*, Jun. 2018.
[8] J. L. Barros-Justo, D. N. Olivieri, and F. Pinciroli, "An exploratory study of the standard reuse practice in a medium sized software development firm," *Comput. Stand. Interfaces*, 2018.
[9] F. Pinciroli, L. Zeligueta, M. I. Lund, and J. L. Barros Justo, "Systematic Mapping Protocol - Coverage of Aspect-oriented Methodologies for the Early Phases of the Software Development Life Cycle," 2017.
[10] F. Pinciroli, "Modeling the Static View in Aspect-Oriented Software Development," no. section 6.
[11] Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, and M. Mendonca Teixeira, *Analysis in Intelligent Systems and Technologies*, vol. 444. Springer, 2016.
[12] A. Mendhekar, G. Kiczales, and J. Lamping, "RG: A Case-Study for Aspect-Oriented Programming," *Tech. Rep. SPL97-009P9710044, Xerox PARC, Feb*, no. February, 1997.
[13] J. Bálik and V. Vranić, "Symmetric aspect-orientation: some practical consequences," *Proc. 2012 Work. Next ...*, pp. 7–11, 2012.
[14] W. Harrison, H. Ossher, and P. Tarr, "Asymmetrically vs. Symmetrically Organized Paradigms for Software Composition," vol. 22685, 2002.
[15] N. Weston, R. Chitchyan, and A. Rashid, "Formal semantic conflict detection in aspect-oriented requirements," *Requir. Eng.*, vol. 14, no. 4, pp. 247–268, 2009.
[16] A. Y. Mohamed, A. E. F. Hegazy, and A. Dawood, "Aspect Oriented Requirements Engineering," *Comput. Inf. Sci.*, vol. 3, no. 4, pp. 135–154, 2010.
[17] A. Rashid, P. Sawyer, A. Moreira, and J. Araújo, "Early aspects: A model for aspect-oriented requirements engineering," *Proc. IEEE Int. Conf. Requir. Eng.*, vol. 2002-Janua, pp. 199–202, 2002.
[18] S. Clarke and E. Baniassad, *Aspect-oriented analysis and design. The Theme approach.* Boston: Addison-Wesley, 2005.
[19] P. A. Parreira Júnior and R. A. Delloso Penteado, "Aspect-Oriented Requirements Engineering. A Systematic Mapping," in *XVI International Conference on Enterprise Information Systems*, 2014, vol. 1, pp. 83–95.
[20] N. Singh and N. Singh Gill, "Aspect-Oriented Requirements Engineering for Advanced Separation of Concerns : A Review," *Inernational J. Comput. Sci. Issues*, vol. 8, no. 5, pp. 288–298, 2011.
[21] F. Pinciroli, J. L. Barros Justo, and R. Forradellas, "Aspect-Oriented Business Process Modeling Approaches: An assessment of AOP4ST," in *46 JAIIO, Jornadas Argentinas de Informática*, 2017.
[22] F. Pinciroli, "Aspect-oriented business process composition rules in AOP4ST," *2016 35th Int. Conf. Chil. Comput. Sci. Soc.*, pp. 1–6, Oct. 2016.
[23] F. Pinciroli, "Considerações acerca da mineração de aspectos," *Perspect. em Ciências Tecnológicas*, vol. 5, no. 5, pp. 83–101, 2016.
[24] F. Pinciroli and J. L. Barros Justo, "Early aspects in 'Aspect-Oriented Process for a Smooth Transition,'" in *XXIII Congreso Argentino de Ciencias de la Computación*, 2017.
[25] F. Pinciroli, "Improving software applications quality by considering the contribution relationship among quality attributes," *Procedia Comput. Sci. 3rd Int. Work. Comput. Antifragility Antifragile Eng. (ANTIFRAGILE 2016)*, vol. 83, pp. 970–975, 2016.
[26] F. Pinciroli, "An HCI quality attributes taxonomy for an impact analysis to interactive systems design and improvement," in *XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016)*, 2016.

Implantación de Sistemas de Software: Nuevas Tendencias

Marisa Panizzi [1,2], Mauricio Davila [1], Agustin Hodes [1], Pablo Vázquez [1], Felipe Ortiz [1], Federico Arana [1], Rodolfo Bertone [2]

[1] Departamento de Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Medrano 951 (C1179AAQ), C.A.B.A, Argentina.
2 Instituto de Investigación en Informática (III-LIDI). Facultad de Informática.
Calles 50 y 120 - La Plata - Bs. As. - Argentina
3 Instituto de Tecnología e Ingeniería. Universidad Nacional de Hurlingham.
Av. Vergara 2222 (B1688GEZ)- Villa Tesei - Bs. As. - Argentina

marisapanizzi@outlook.com, davilamr.80@gmail.com, hodesagustin@gmail.com, vazpablo@gmail.com, ortizfd@gmail.com, federico.arana@ieee.org, pbertone@lidi.info.unlp.edu.ar

Resumen

La implantación de sistemas de software consiste en la entrega final del producto software al cliente. En esta línea de investigación, la cual se encuentra en un constante crecimiento, se están diseñando una serie de aportes. Estos intentan dar respuesta a los inconvenientes que actualmente se desprenden de una implantación desempeñada de manera inadecuada o errónea. Dentro de los objetivos propuestos para esta línea de investigación, se pueden enumerar: a) la definición de un modelo de proceso de implantación de sistemas de software, b) la construcción de métricas para este proceso, c) el diseño de un conjunto de riesgos, así como también los procedimientos para su mitigación, d) una técnica para el estudio de factibilidad del proceso y por último, e) un relevamiento en la industria del software de Argentina respecto al proceso de implantación de sistemas de software. En este artículo se presenta un estado de los avances realizados en esta línea de investigación hasta la actualidad.

Palabras clave: procesos de software, implantación de sistemas de software, desafíos.

Contexto

La investigación que se reporta en este artículo es financiada parcialmente por el Proyecto UTN 6576 de la Secretaria de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional.

Esta línea de investigación cuenta con el asesoramiento científico del grupo de investigación de Ingeniería de Software del Instituto de Investigación en Informática (III-LIDI) de la UNLP y UCLM

Introducción

Como punto de partida de esta línea de investigación a largo plazo, surgió la necesidad de proponer una definición conceptual del proceso de implantación de sistemas de software. Esta necesidad surge porque la literatura existente trata al proceso con terminología y alcance diferente. A continuación, se presenta la Tabla 1, en la cual se enuncian los términos y las metodologías y/o estándares de los que provienen. En esta línea de investigación se considera *¨proceso de implantación de sistemas de software¨* como:

a) conjunto de actividades y tareas necesarias para la transferencia de un sistema de software a su ambiente de utilización,

b) realizadas por recursos humanos,

c) que construyen artefactos de salida y que requieren artefactos de entrada,

d) se considera la aplicación de prácticas y/o técnicas y

e) se hace uso de herramientas.

Tabla 1. Términos equivalentes al proceso de implantación y fuentes de obtención.

Término	Fuente
Flujo de despliegue	Proceso Unificado Rational (Jacobson I. et al., 1999). Proceso Unificado Ágil (AUP)(Ambler S., 2016).
Proceso Transición	Estándar ISO/IEC/IEEE 12207 (ISO/IEEE/IEEE 12207, 2017).
Fase de despliegue	DSDM (Dynamic Systems Development Method) (Agile Business Consortium, 2014).
Fase de Implantación y Aceptación del usuario	Metodología Métrica v3 (Portal de Administración Electrónica - Gobierno de España, 2001).

Con el propósito de identificar, evaluar e interpretar toda la investigación disponible relevante respecto al proceso de implantación de sistemas de software, se ha desarrollado un estudio de mapeo sistemático de la literatura (en Inglés, Systematic Mapping Studies o SMS). Este estudio permitió sintetizar la información existente sobre los modelos, metodologías o métodos que sirven de guía a las empresas productoras de software, para ejecutar la implantación de sistemas de software, como así también identificar la existencia de herramientas, artefactos, técnicas y/o prácticas para el dicho proceso (Marisa Panizzi et al, 2020). Para determinar las metodologías, métodos y estándares que abordan la gestión de riesgos en proyectos de desarrollo de software y evidenciar si existen propuestas de gestión de riesgos para la implantación de sistemas de software, se desarrolló un SMS (Felipe Ortiz et al, 2019). Por último, se realizó un SMS, para identificar la existencia de métricas para el proceso de implantación de sistemas de software.

La definición del modelo de proceso de implantación que se desprende de esta línea de investigación, la cual es parte de la investigación de la tesis doctoral de la primera autora, ha sido presentada en el Simposio Doctoral del Congreso Iberoamericano de Ingeniería de Software, CIbSE 2019 (Marisa Daniela Panizzi, 2019). Una primera definición del modelo de implantación de sistemas de software denominado MoProIMP, el cual tiene por objetivo proporcionar una guía para que las empresas productoras de software logren sistematizar las actividades que componen la implantación. En su definición preliminar cuenta con fases, actividades y tareas (Panizzi M., et al, 2019). Actualmente, MoProIMP se está refinando mediante el diseño de sus artefactos de entrada y salida para cada una de las tareas que lo componen.

De la línea de investigación se desprenden dos sub-líneas, una orientada a la definición de métricas para el proceso de implantación de sistemas de software y la otra orientada a proporcionar un conjunto de riesgos, así como también las acciones para su mitigación. Para el estudio de estas dos propuestas, se ha considerado un análisis del proceso con un enfoque tridimensional. Las dimensiones consideradas son: ¨*Proceso*¨, ¨*Producto*¨ y ¨*Persona*¨. La primera dimensión, se considera dado el interés de las fases o etapas, actividades y tareas que componen al proceso. La segunda dimensión, contempla las características como la complejidad del producto a instalar, los requisitos de instalación para el producto software, la integración con la infraestructura del cliente, entre otras. La última dimensión, se considera por la existencia del peopleware y su impacto en este proceso de software en estudio (Panizzi M. et al., 2016).

Para atender la dimensión ¨*Proceso*¨, se han diseñado propuestas para los riesgos, así como también para las métricas.

Para la definición de riesgos para el proceso de implantación de sistemas de software, la se consideraron las actividades y tareas del proceso de transición del estándar ISO/IEC/IEEE 12207((ISO/IEEE/IEEE 12207, 2017) por ser un estándar internacionalmente reconocido. Además, se utilizó la clasificación de riesgos propuesta en (Jones C., 1994) con algunas adecuaciones. Esta propuesta ha sido validada en el contexto real mediante el método de investigación, estudio de caso (Ortiz F. et al., 2019).

Los avances realizados respecto a la definición métricas consisten en una métrica de estimación de tiempo para el proceso de implantación de sistemas de software (Pablo

Vazquez et al., 2019) y actualmente se está avanzando en una métrica de productividad. En el caso de la métrica de estimación de tiempo, se ha logrado la validación de la misma en la industria de software nacional, mediante el método de investigación de estudio de caso.

Los autores adhieren a que, en los proyectos de software, generalmente se le resta importancia al proceso de implantación o entrega del sistema informático por tratarse de uno de los eslabones finales de la cadena productiva del producto software (Panizzi M. et al., 2016) (Panizzi M. et al., 2017).

Otros autores plantean que en el proceso de implantación de sistemas de software se presentan errores o prácticas inadecuadas. El proceso de implantación comprende prácticas que permiten que el sistema de software se instale, se desinstale o bien se actualice en la empresa-cliente. Estas prácticas vinculadas a la gestión de la configuración del cliente, en algunas ocasiones son afectadas por algunos inconvenientes, como por ejemplo la falta de componentes, instalaciones incompletas o errores de instalación (Jansen et al, 2006). Forbes et al. (Forbes et al., 2003) plantean que el resultado de prácticas de despliegue no estandarizadas e inadecuadas se refleja en los sistemas de información, los cuales son difíciles de mantener y operar. En el estudio de benchmarking realizado por Toufaili et al. se plantean dos puntos significativos, el primero se refiere al hecho de que el proceso de implantación de sistemas de software es un proceso del ciclo de vida de desarrollo de software que no se estudia con tanta frecuencia como los demás procesos, mientras que el segundo se refiere a la necesidad de crear un modelo para la ejecución del proceso, de modo que se pueda sistematizar el proceso como los roles específicos, dada la diversidad de roles que participan en los procesos de las empresas estudiadas (Toufaili et al., 2010).

El grupo de investigación, se plantea las siguientes preguntas de investigación (PI):

PI1: ¿Se puede lograr cubrir la vacancia de un Modelo de Proceso para la implantación de un sistema de software, que integre las actividades y sus tareas, las técnicas, las herramientas, los artefactos, los roles y sus competencias?

PI2: ¿Se puede mejorar la calidad del proceso de implantación mediante la determinación de métricas específicas para el mismo?

PI3: ¿Se puede fortalecer el proceso de implantación mediante la identificación de un conjunto de riesgos, así como los procedimientos para su mitigación?

El desafío de esta línea de investigación consiste en lograr las respuestas a las preguntas presentadas.

Líneas de Investigación, Desarrollo e Innovación

Esta línea de investigación contempla una serie de aportaciones al proceso de implantación de sistemas informáticos en el contexto de la industria nacional. Dentro de estos aportes, se puede mencionar la evolución del **Mo**delo de **Pro**ceso de **IMP**lantación de sistemas informáticos, denominado MoProIMP; su primera aproximación ha sido presentada en el Congreso Argentino de Ciencias de la Computación (CACIC 2018) (Panizzi M. et al., 2018).

En este contexto, se realizará una propuesta de métricas para el proceso en estudio, estas contemplarán una visión tridimensional anteriormente presentada. En concordancia a la visión tridimensional del proceso en estudio, se diseñarán un conjunto de recomendaciones riesgos como así también los procedimientos para su mitigación.

Resultados y Objetivos

En esta línea de investigación en progreso se han logrado una serie de ***resultados*** que se detallan a continuación:

a) Académicos, se han logrados dos trabajos de especialidad en Ingeniería en Sistemas de Información.
b) Producción Científica: se ha presentado la línea de investigación en eventos

científicos de alcance nacional (WICC[1] 2017, WICC 2019, CACIC[2] 2016, CACIC 2018, CACIC 2019) y en el ámbito internacional, CIACA[3] 2017, SEKE[4] 2017, CIbSE[5] 2019, CIbSE 2020 y ICAETT[6] 2019. Además, se lograron dos publicaciones en Springer y en el Brazilian Journal of Development.

c) Proyectos de Investigación: ha finalizado el PID UTNBA4347 titulado: ¨Impacto del factor peopleware en el proceso de implantación de sistemas informáticos¨ (período 2017-2019). Se inicia el PID UTNBA6576 titulado: ¨Estudio del proceso de implantación de sistemas informáticos en el contexto industrial de la República Argentina¨ (período 2020-2021).
d) Formación en investigación: el grupo de investigación se encuentra en un proceso de aprendizaje constante de métodos de investigación de ingeniería de software experimental, revisiones sistemáticas, estudios de casos y encuestas (Genero B. Marcela et al, 2014).

El grupo de investigación se plantea como ***próximos objetivos***:

- Continuar con el diseño y refinamiento de MoProIMP y su validación en el contexto real,
- Extender la definición de riesgos para las dimensiones *¨Producto¨* y *¨Persona¨,*
- Desarrollar una serie de métricas para la dimensión ¨Proceso¨,
- Comenzar con el diseño de una encuesta que permita relevar la situación del proceso de implantación de sistemas de software en la industria nacional.

[1] Workshop de Investigadores en Ciencias de la Computación
[2] Congreso Argentino de Ciencias de la Computación.
[3] Conferencia Iberoamericana de Computación Aplicada.
[4] International Conference on Software Engineering and Knowledge Engineering.
[5] Congreso Iberoamericano en Ingeniería de Software.
[6] Congreso Internacional sobre Avances en Nuevas Tendencias y Tecnologías.

Formación de Recursos Humanos

El equipo de investigación se encuentra conformado por un Director, un investigador formado, dos becarios de grado, dos tesistas de posgrado y un asesor científico-tecnológico

Se estima la formación de dos Magister en Ingeniería en Sistemas de Información y una Doctora en Ciencias Informáticas.

Referencias

Jacobson I, Booch G., Rumbaugh J. The Unified Software Development Process.1 st Edition. Addison Wesley, USA (1999).

Ambler S. 2016. The Agile Unified Process (AUP).http://www.ambysoft.com/unifiedprocess/agileUP.html, last accessed 2019/12/12.

IEEE ISO/IEC/IEEE 12207:2017(E). Systems and software engineering - Software life cycle processes (2017).

Agile Business Consortium. The DSDM Agile Project Framework (2014 Onwards). https://www.agilebusiness.org/resources/dsdm-handbooks, last accessed 2019/12/29.

Portal de Administración Electrónica - Gobierno de España. Metodología MÉTRICA V3. .http://administracionelectronica.gob.es. Último acceso 12 de diciembre 2019 (2001).

Marisa Panizzi, Marcela Genero, Rodolfo Bertone. Software system deployment process: A systematic mapping study. Congreso Iberoamericano en Ingeniería de Software (CIbSE 2020). Track de Ingeniería de Software Experimental (ESELAW). Brasil, Curitiba, 4 al 8 de mayo (2020).

Felipe Ortiz, Mauricio Dávila, Marisa Panizzi, and Rodolfo Bertone. State of the Art Determination of Risk Management in the Implantation Process of Computing Systems. Advances in Emerging Trends and Technologies. Vol 1066, Springer, pp. 23-32. Springer Nature Switzerland

AG (2019).

Marisa Daniela Panizzi. Propuesta de un modelo de implantación de sistemas informáticos. XXII Congreso Iberoamericano en Ingeniería de Software. Simposio Doctoral. Cuba, La Habana, 22 al 26 de abril (2019).

Panizzi, M., Bertone, R., Hossian, A. Proposal for a Model of a Computer Systems Implantation Process (MoProIMP). Communications in Computer and Information Science – CACIC 2018 24th Argentine Congress Tandil, Argentina, October 8–12, 2018. Revised Selected Papers. Springer Nature Switzerland AG, pp. 157-170 (2019).

Panizzi, M., Hossian, A., García-Martínez, R. Implantación de Sistemas: Estudio Comparativo e Identificación de Vacancias en Metodologías Usuales. Libro de Actas del XXII Congreso Argentino de Ciencias de la Computación. Pág. 546-555. ISBN 978-987-733-072-4. Universidad Nacional de San Luis (2016).

Jones C., Assessment and control of software risk. Yourdon Press (1994).

Ortiz, F., Panizzi, M., Bertone, R. Risk determination for the implantation process of software systems. Libro de Actas del XXV Congreso Argentino de Ciencias de la Computación. ISBN en trámite. Universidad Nacional de Río Cuarto, Río Cuarto del 12 al 16 de octubre (2019).

Pablo Vazquez; Marisa Panizzi, Rodolfo Bertone. Estimación del esfuerzo del proceso de implantación de software basada en el método de Puntos de Caso de Uso / Estimating the effort of the software implantation process based on the Use Case Points method. Brazilian Journal of Development, Curitiba, v. 5, n. 2, feb. 2019. http://www.brjd.com.br/index.php/BRJD, pp. 1809-1822. ISSN: 2525-8761 (2019).

Panizzi, M., Bertone R., Hossian A. (2017). Proceso de Implantación de Sistemas Informáticos – Identificación de vacancias en Metodologías Usuales. Libro de Actas de la V Conferencia Iberoamericana de Computación Aplicada CIACA 2017. Pag.207 -215. ISBN 978-989-8533-70-8. Vilamoura, Algarve, Portugal.

Panizzi, M., Hossian, A., García-Martínez, R. (2016). Implantación de Sistemas: Estudio Comparativo e Identificación de Vacancias en Metodologías Usuales. Libro de Actas del XXII Congreso Argentino de Ciencias de la Computación. Pág. 546-555. ISBN 978-987-733-072-4. Universidad Nacional de San Luis.

Jansen, S., Brinkkemper, S. Definition and validation of the key process of release, delivery and deployment for product software vendors: Turning the ugly duckling into a swan. In: Proceedings of the 22nd IEEE International Conference on Software, pp. 166-175. IEEE, Philadelphia (2006).

Forbes, J.A., Baker, E.R. Improving Hardware, Software, and Training Deployment Processes. In: Proceedings of 19th International Conference on Software Maintenance, pp. 377-380. IEEE, The Netherlands (2003).

Toufaili, H., Kajko-Mattsson, M. Benchmarking a Software Transition Process. In: Proceedings of the 11th International Conference on Product Focused Software Development and Process Improvement (PROFES '10), pp. 140-151. (2010).

Panizzi, M., Bertone, R. Hossian, A. Propuesta de un Modelo de Proceso de Implantación de Sistemas Informáticos (MoProIMP). Libro de Actas del XXIV Congreso Argentino de Ciencias de la Computación. ISBN 978-950-658-472-6. Universidad Nacional del Centro de la Provincia de Buenos Aires (2018).

VALIDACIÓN DE LA TÉCNICA DE ANÁLISIS DE DOMINIO (JODA) AL PROCESO UML COMPONENTS (DSBC)

Moyano Ezequiel UNTDF, Urciuolo Adriana UNTDF, Gel Matías UNTDF, Iturraspe Rodolfo UNTDF, Villarreal Martín UNTDF
Instituto de Desarrollo Económico e Innovación, UNTDF

Dir.: Fuegia Basket 251, (9410) Ushuaia. Tierra del Fuego. Tel: ++54-2901-432403
emoyano@untdf.edu.ar,aurciuolo@untdf.edu.ar, mgel@untdf.edu.ar, riturraspe@untdf.edu.ar, mvillarreal@untdf.edu.ar

Resumen

El desarrollo de software basado en componentes se ha convertido en uno de los mecanismos más efectivos para la construcción de grandes sistemas, sobre todo en aquellos de gran complejidad.

Construir una aplicación se convierte en la búsqueda y ensamblaje de componentes. UML Components presenta un gran potencial para construir sistemas basados en componentes en dominios complejos, y permite definir una manera de modelar.

El propósito del Análisis de Dominio (AD) es el de proporcionar la reutilización de la especificación de un dominio específico para aplicaciones similares. En trabajos previos se especificó que una de las técnicas de análisis de dominio que mejor se ajusta a sistemas complejos es JODA (Object-Oriented Domain Analysis Method), ya que trabaja con análisis orientado a objetos, utilizando notación UML.

A los fines de proveer una mejor definición del dominio, se incorpora la técnica JODA (Análisis de Dominio) como primer etapa al proceso de desarrollo de software basado en componentes UML Components, con el objetivo de permitir el modelado de los requerimientos propuesto en el proceso de desarrollo UML Components.

El presente trabajo tiene como objetivo demostrar y validar que los componentes obtenidos (UML Components), a través del análisis previo, son verdaderamente reutilizables.

Palabras clave: Desarrollo de software basado en componentes, Análisis de dominio, Sistemas complejos, Reuso.

Contexto

La línea de investigación del presente proyecto forma parte del "Área de Investigación 8 – Desarrollo Informático", del Instituto de Desarrollo Económico e Innovación de la Universidad Nacional de Tierra del Fuego (UNTDF).

El proyecto se presentó a la convocatoria realizada por la UNTDF en septiembre de 2016 y resultó aprobado luego de haber sido sometido a evaluación externa. (Resolución Rectoral 060/2017 del 10/04/2017). – en Sistemas de Información Ambiental.

Se lleva adelante con un enfoque multidisciplinario por parte de organismos de investigación y gestión en la provincia de Tierra del Fuego y consta de diversos componentes.

1 Introducción

Muchos dominios, como los sistemas de información ambiental, se caracterizan por su complejidad [01], con una gran cantidad de subdominios heterogéneos[02].

Los trabajos de investigación actuales en esta problemática, no se han detenido en el estudio de técnicas de análisis de dominio apropiadas para comprender la complejidad

y sus efectos, por cual no se cuenta con modelos del dominio que puedan reutilizarse para distintos sistemas.

La ingeniería de dominio tiene como objetivo lograr la reutilización del conocimiento y de las especificaciones de un dominio específico[03]. Se constituye en el punto de partida para obtener software reusable en sistemas de gran complejidad.

Las técnicas de Análisis de Dominio (AD) están asociadas a la reutilización, su principal característica es capturar información relacionada con el dominio, y determinar qué y cómo este conocimiento se reutilizará. El AD produce un modelo del dominio[04], enfocado en la reutilización de componentes de software dentro de un conjunto de sistemas que pertenecen a un dominio[05].

La necesidad de contar con sistemas que no demanden grandes periodos de tiempo, dio origen a la idea de contar con componentes que permitieran su reutilización. Esto favoreció el avance del Desarrollo de Software Basado en Componentes (CSBD)[06].

El DSBC se ha convertido actualmente en uno de los mecanismos más efectivos para la construcción de grandes sistemas y aplicaciones de software[07].

En trabajos anteriores (WICC 2019 – San Juan, CICCSI 2019) se demostró los beneficios de aplicar análisis de dominio como etapa previa al proceso de DSBC para obtener componentes reusables y se utilizo la técnica de análisis de dominio JODA como primer etapa, incorporada a UML Components como proceso de DSBC.

El propósito de este trabajo es validar la consistencia y el beneficio de aplicar AD (JODA) al proceso de DSBC UML Components. A tal efecto se toma como punto de partida los resultados obtenidos durante la investigación en curso[08], en el cual se redefinió y completó la etapa de modelado de requerimientos propuesta por UML Components, incorporando técnicas de análisis de dominio JODA en forma previa.

Esto permitió una mejor definición del dominio y lograr comprender sus aspectos comunes y variabilidades, con el propósito de representar el conocimiento y permitir su reutilización.

Con el objetivo de validar el proceso desarrollado y obtener componentes verdaderamente reusables se define como dominio específico de estudio (subdominio de los sistemas de información ambiental) los sistemas de información meteorológica (SIM), aplicar el proceso descripto para encontrar y definir componentes reusables y posteriormente demostrar su reuso a través de dos sistemas representativos (de los SIM), analizando y satisfaciendo los requerimientos de cada uno.

2 Líneas de Investigación y Desarrollo

La línea de investigación general se enfoca en el estudio y desarrollo para incorporar técnicas de análisis de dominio al proceso de desarrollo basado en componentes para dominios complejos. Para tal fin se utilizará la técnica de Análisis de dominio JODA al Proceso UML Components en su primer estapa.

Uno de los principales problemas observados en los procesos de DSBC, es que las técnicas de especificación y modelado de requerimientos no contemplan métodos de análisis del dominio. Se ve la necesidad de completar la etapa de modelado de requerimientos de UML Components, incorporando técnicas de análisis de dominio en forma previa.

Esto permitó una mejor definición del dominio y lograr comprender sus aspectos

comunes, para representar el conocimiento y permitir su reutilización.

Pariendo de los resultados de trabajos anteriores, donde se implementó la técnica de análisis de dominio JODA en el proceso UML Components, se procedió a redefinir la fase de requerimientos de la siguiente manera[08]:

- Redefinir el proceso de Requerimientos de UML Components agregando la construcción de un modelo de dominio como entrada[09].

- Construir el modelo de dominio considerando las etapas de JODA para el modelado del dominio[09].

- Obtener un modelo basado en lenguaje UML para la etapa de modelado del dominio que fue incluida al Proceso de definición de Requerimientos[08].

El siguiente paso consistió en definir un dominio específico que contenga las características de un dominio complejo, se definió como dominio específico los Sistemas de Información Meteorológicos (encargados del estudio de la atmósfera, sus propiedades y de los fenómenos que en ella tienen lugar). Los fenómenos físicos en la atmósfera ocurren en todas las escalas espaciales y temporales y sus impactos son relevantes para muchas actividades[10].

Uno de los principales temas que manejan los sistemas de información meteorológica es obtener, registrar y se procesar los distintos datos que se necesitan para el estudio de estos fenómenos.

2.1 Análisis del Dominio:

JODA divide el proceso del análisis del dominio en tres fases:

Fase de Preparación del dominio: Se trabajó con expertos del dominio para obtener el conocimiento necesario de las aplicaciones, se utilizó información obtenida sobre el funcionamiento de diversos organismos que trabajan en este dominio: El Servicio Meteorológico Nacional, el dpto. de Cs. de la Atmósfera de la UBA, Organización Meteorológica Mundial, CIMA- Centro de Investigación del Mar y la Atmósfera dependiente del CONICET y otras.[11]

Fase de Definición del dominio: Las actividades realizadas en la fase permitieron obtener componentes conceptuales del dominio como: Diagrama de Contexto en JODA, Diagrama de Conjunto-Parte y Parte, Dependencias y Servicios del Dominio, Escenarios principales en el Modelado del Dominio en JODA, Diagrama de Clases /Estructura JODA y Diag. de Actividad – Diag. de Works Units. – Diag. de Packages.

Fase de Modelo del dominio: En esta fase se redefinió el diagrama de clases presentado por JODA por un diagrama de clases en Notación UML.

Siguiendo el proceso de UML Components, se especificó el modelo de casos de uso de un sistema genérico del dominio y su respectiva descripción[11] para la obtención de los componentes.

2.2 Especificación de Componentes:

Del análisis previo se identificaron los siguientes componentes junto a sus responsabilidades e interfaces:

1- Captura datos del ambiente: información relativa a los objetos ambientales medidos, las estaciones e instrumental utilizado.

2- Inventario Meteorológico: tratamiento primario, depuración y almacenamiento de las variables meteorológicas y sus mediciones.

3- Tratamiento Secundario de los Datos: Procesamiento de los datos, obtener

predicciones, estadísticas, alertas, mapas meteorológicos, etc.

4- Modelar Simulaciones Meteorológicas y Climatológicas: clases necesarias para realizar simulaciones: Modelo, Elementos, Método, etc.)

5- Fenómenos Meteorológicos: registro de fenómenos (fecha, hora, tipo fenómeno) asociados a las observaciones que permiten analizarlo o predecirlo.

2.3 Validación

Por último se debe validar que los componentes obtenidos, y sus interfaces, sean verdaderamente reutilizables a través de al menos dos aplicaciones diferentes del dominio específico. Para lo cual se utilizaron dos.

*Sistema de Alertas***:** Su finalidad es advertir fenómenos meteorológicos (inundaciones, temporales, sequías, granizo, etc) que se producen bajo ciertas circunstancias.

Los sistemas de alerta utilizan el monitoreo de las últimas 72 hs para predecir alguna alerta o pronóstico. Los datos son observados y registrados en Tiempo Real, generalmente utilizando Estaciones Automáticas e instrumental de medición, registradas por un observador, sondeos o a través de imágenes satelitales. Los datos son almacenados y verificados comparándolos con medidas "lógicas" a fin de controlar la veracidad de los mismos.

Por último se utilizan modelos simples para el cálculo de índices y se modelan procesos (simulaciones de baja complejidad), con estos índices y simulaciones se puede estimar la intensidad o la ocurrencia de un alerta meteorológico.

Los requerimientos del sistema se pudieron satisfacer, validados por los expertos del domino, utilizando los componentes especificados con las interfaces adecuadas: *Captura datos del ambiente, Inventario Meteorológico, Tratamiento Secundario de los Datos y Modelar Simulaciones Meteorológicas y Climatológicas.*

Sistema de Modelización de Caudales: Su finalidad es conocer los caudales de un río para el usufructo del agua (agua potable, turismo, acuicultura, riego, etc.)

En estos sistemas los datos se necesitan a través de series consistentes de varios años de historia y durante todos los meses, en varias secciones del mismo.

Las precipitaciones, deshielo y altura del río son los datos más importantes. La medición del caudal se realiza a través de un cálculo en función de la velocidad del viento y de la superficie de la sección observada. La información, a diferencia del sistema anteriormente, se obtiene en tiempos diferidos.

Para satisfacer los requerimientos de estos sistemas se utilizaron los siguientes componentes: *Captura datos del ambiente, Inventario Meteorológico, Tratamiento Secundario de los Datos.*

3 Resultados

Como conclusión a partir de aplicar Análisis de Dominio a UML Components permitió obtener un proceso de desarrollo (sobre todo en la etapa de especificación de requerimientos) más detallado, preciso y formalizado, extendiendo UML Components a sistemas de mayor complejidad[11].

Utilizar la notación UML en la técnica JODA, permitió definir un número de modelos y diagramas para la especificación del software con varios grados de abstracción, con lo cual se obtuvo un modelado del dominio más semejante a la realidad. El presente trabajo permite

concluir que los componentes que se obtienen del proceso son verdaderamente reutilizables, aplicando debidamente sus interfaces.

Los resultados mencionados constituyen aportes al conocimiento del dominio, sus características y avances en el desarrollo de una arquitectura referencial.

Como futuros trabajo en el marco de este estudio se espera poder validar otras aplicaciones dentro del dominio, y ampliar el mismo a otros subdominios dentro de las aplicaciones ambientales, a los efectos de encontrar más entidades y procesos comunes modelados en el dominio.

4 Formación de Recursos Humanos

El Equipo de Trabajo está conformado por docentes investigadores de la UNTDF y expertos del dominio. En ésta línea de investigación se encuentran en desarrollo dos tesis de posgrado (Magister en Ing. de Software, UNLP, Ezequiel Moyano y Martín Villarreal).

Colaborará y prestará apoyo un alumno en formación de grado, con el objetivo formarlo y que le sirva para el desarrollo de su futura tesis de grado.

5 Referencias

[01] Günther, O. Springer-Verlag, Environmental Information Systems, Berlín Heidelberg, Germany, 1998.

[02] Arne Koschel, Ralf Kramer, Ralf Nikolai, A Federation Architecture For An Environmental Information System Incorporating GIS, Karlsruhe, Germany, 2000.

[03] Hamar, V. Aspectos Metodológicos Desarrollo y Reutilización de Componentes de Software, Mérida – 2003.

[04] Maarit Harsu. A Survey On Domain Engineering, Institute Of Software SystemsTampere University Of Technology.

[05] Holibaugh R.: Working Group (JIAWG) Object-Oriented Domain Analysis Method (JODA)

[06] Manuel F. Bertoa, José M. Troya Y Antonio Vallecillo Aspectos De Calidad En El Desarrollo De Software Basado En Componentes, 2002.

[07] Cheesman, John; Daniels John, UML Components, Octubre 2000.

[08] Moyano E.Urciuolo A., Iturraspe R., , Incorporar Análisis de Dominio (Joda) en Uml Components (DSBC) WIIC 2019 – San Juan, 2019.

[09] Moyano E. Tesis de Grado: Técnicas De Análisis de Dominio para el Desarrollo de Componentes dn Sistemas Complejos, Caso De Estudio: Sistemas De Información Ambiental, dirigida por Urciuolo A. 2007.

[10] Servicio Meteorológico Nacional https://www.smn.gob.ar

[11] Moyano E.Urciuolo A., Iturraspe R. , Incorporar Análisis de Dominio (Joda) en Uml Components (DSBC) CICCSI 2019 – Mendoza, 2019..

APLICACIÓN DE GESTIÓN DE CONOCIMIENTO A LA FASE DE PRUEBAS DE LA INGENIERÍA DE SOFTWARE

Rozo, Mauricio[1,2]; Casanovas, Inés[1,2]

[1]Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS)
Universidad Tecnológica Nacional. Facultad Regional Buenos Aires.
Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires

[2]Programa de Maestría en Ingeniería de Sistemas de Información
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires.
Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires

{mauricio.rozo.rodriguez, inescasanovas}@gmail.com

RESUMEN

La fase de pruebas en la ingeniería de software es un proceso que genera un gran volumen de conocimiento, considerándose como un factor crítico para la calidad de producto, por lo tanto, exige una creciente demanda sobre cómo mejorar la efectividad en el cumplimiento de esta tarea, es ahí donde el uso de métodos y principios de Gestión de Conocimiento se convierte en la base para gestionarla. En este contexto, el presente trabajo, indaga los desafíos de los modelos de gestión de conocimiento existentes en el domino de las pruebas de software para identificar las falencias de los modelos existentes y a partir de estas proponer una solución para la ingeniería de software en el ámbito de la gestión de conocimiento basado en la utilización de metodologías ontológicas en el dominio de la fase de pruebas.

***Palabras clave*:** Gestión del Conocimiento, Fase de Pruebas, Ontologías, Modelos Ontológicos, Ingeniería de Software.

CONTEXTO

En el marco de las actividades del Grupo de Estudio de Metodologías para Ingeniería en Software y Sistemas de Información (GEMIS), conformado por un equipo de docentes y alumnos dentro del ámbito de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional (UTN-FRBA) con dependencia del Departamento de Ingeniería en Sistemas de Información, se comienza una nueva línea de trabajo en el campo de aplicación de la Gestión del Conocimiento (GC) a la fase de pruebas en la ingeniería de software mediante la definición e implementación de un modelo de GC. De esta manera, se articula dentro de los objetivos de la Universidad Tecnológica Nacional (UTN) que ha definido diversos Programas de Investigación + Desarrollo + Innovación (I+D+i), entre los cuales se encuentra el Programa Tecnología de las Organizaciones, que tiene por finalidad "articular las distintas temáticas relacionadas con la GC, la innovación y los sistemas de gestión de la calidad e integrados aplicables a las organizaciones", incluyendo la GC en las organizaciones como área prioritaria [1].

1. INTRODUCCIÓN

La evolución constante de la sociedad ha permitido nuevos retos que demandan una profunda preparación en herramientas tecnológicas, acelerando los cambios de las tecnologías de la información y comunicación, dando pie al fortalecimiento de la sociedad del conocimiento, en ese sentido, en la actualidad, la GC es sin duda uno de los activos más valiosos de las organizaciones, debido que provee de manera estructurada y sistemática las herramientas para facilitar que el conocimiento generado ayude alcanzar los objetivos estratégicos de las organizaciones y optimizar las decisiones que se tomen, para mejorar los procesos [2].

En [3] la GC es entendida como la creación, descubrimiento y recolección interna de conocimiento y de las mejores prácticas, para compartir y entender las experiencias que la organización puede usar, de esta manera ajustarlas y aplicarlas a nuevas situaciones, en la búsqueda de la mejora en el desempeño organizacional.

Entre tanto, para [4] la GC es el conjunto de principios, métodos, técnicas, herramientas, métricas y tecnologías que permiten obtener los conocimientos precisos de un modo adecuado, en el tiempo oportuno de forma eficiente y sencilla, con el fin de conseguir una actuación lo más inteligente posible. Ahora bien, con la adopción de prácticas de GC en la ingeniería de software, se mejoraría la gestión tanto para la construcción del software como su mantenimiento [5].

Para Morales y Gutiérrez [6], en la ingeniería de software, la GC está desarrollándose en la medida de las necesidades que expresan las organizaciones de este rubro y resulta necesario expandir el uso de la GC dentro de las diferentes fases de la ingeniería de software, debido a la especificidad que exige la evolución de cada uno de los procesos.

Por su parte, Blanquicett, Bonfante y Acosta [7], consideran qué a través, de la historia de la ingeniería del software y la industria, las pruebas de software (PS) se han convertido en una importante herramienta para el aseguramiento de la calidad del producto final, lo cual permite identificar si satisface los requisitos iniciales del cliente.

La PS en la ingeniería de software, es un proceso que genera un gran volumen de conocimiento, considerándose como un factor crítico para la calidad de producto, por lo tanto, exige una creciente demanda sobre cómo mejorar la efectividad en el cumplimiento de esta tarea, es ahí donde el uso de métodos y principios de GC se convierte en la base para gestionarla.

Por otro lado, Souza et al. [8], entienden que las pruebas de software son un proceso intensivo en conocimiento, y se hace necesario proporcionar soporte computarizado para las tareas de adquisición, procesamiento, análisis y diseminación de conocimiento para su reutilización.

Wnuk y Garrepalli [9] identificaron varios desafíos para las PS entre los que destacan, selección y aplicación de técnicas inadecuadas, baja tasa de reutilización del conocimiento de PS, barreras en la transferencia de conocimientos de PS, baja posibilidad de lograr rápidamente la distribución más eficiente de recursos humanos durante las PS. Además, enuncian que las debilidades en las PS se deben a que existe una pérdida significativa de capital intelectual debido a la rotación del personal y su conocimiento limitado. Al mismo tiempo, la gestión de los recursos de prueba o el conocimiento sobre los casos de prueba o el código de prueba ha sido muy poco representada. La obtención, difusión, adquisición, evolución y empaquetamiento del conocimiento reciben poca atención debido a que el conocimiento se maneja principalmente durante las PS dentro de un proyecto u organización y se dedica menos atención al intercambio de conocimientos.

Al respecto, Durán et al. [10], listan algunos desafíos de la PS:

- Baja tasa de reutilización del conocimiento de PS.
- Barreras de transferencia de conocimiento de PS.
- Falta de un entorno adecuado para el intercambio de conocimiento del proceso.
- Alto nivel de rotación de personal.

En el ámbito de la aplicación de la GC a la ingeniería de software, específicamente en la fase de PS, Souza et al [11] a través de un estudio de mapeo, propone identificar la distribución de los estudios seleccionados a lo largo de los años, con foco de investigación desde la perspectiva de la PS y foco desde la perspectiva de la GC además del tipo de investigación, el cual permitió identificar que la utilización de la GC en PS es un tema de investigación reciente, donde el principal problema en las organizaciones es la baja tasa de reutilización del conocimiento y las barreras en su transferencia; además, la reutilización del conocimiento de las PS es el objetivo principal de la aplicación de GC en las PS.

De acuerdo con lo anterior, existe gran preocupación con el conocimiento explícito, en particular, artefactos de casos de prueba, aunque el conocimiento tácito también se ha reconocido como un elemento de conocimiento muy útil y las tecnologías avanzadas utilizadas para proporcionar sistemas de GC en PS incluyen sistemas de recomendación y mapas de conocimiento.

Para [4] el propósito de la GC es aprovechar el conocimiento para generar valor a nivel organizacional, basado en los procesos, herramientas y actividades; desde la ingeniería de software, las PS, garantizan la calidad del producto. En esta perspectiva, en la articulación de la GC a las PS identifica como problemáticas, la falta de reutilización del conocimiento, la alta rotación de personal, modelos con enfoque corporativo y compleja implementación. Para mitigar lo anteriormente descripto, proponen un modelo de implementación para gestionar el conocimiento en las PS, permitiendo obtener una mejora en el desempeño del equipo de PS.

Souza et al. [8], en su investigación, se proponen identificar los problemas relacionados con el conocimiento en las PS, los propósitos de las organizaciones de aplicar GC en PS, los tipos de elementos de conocimiento típicamente manejados en el contexto de PS, los beneficios y problemas informados sobre la implementación de iniciativas de GC en PS, y los mecanismos o tecnologías utilizados para proporcionar GC en PS.

Las contribuciones luego de aplicar el mapeo sistemático realizado por [8] identifican que los principales problemas en las organizaciones son: las barreras en la transferencia de conocimiento con mayor representatividad, la reutilización del conocimiento de las pruebas como objetivo principal de la aplicación de la GC en las PS, y el conocimiento tácito.

Sin embargo, Souza et al. [8] indican que, aunque reconocida como un instrumento importante por la comunidad GC, las ontologías no se utilizan ampliamente en las iniciativas de GC en PS. Para Flores y Hadfeg [12], como González, Sánchez y Montejano [13] una ontología es una especificación explícita de una conceptualización, en esta se modela el vocabulario del dominio, básicamente usando los conceptos, características y sus relaciones.

A la vista de lo anteriormente descripto en [14] propone estudiar cómo las ontologías se pueden utilizar para gestionar el conocimiento en el dominio de PS en la ingeniería de software mediante un modelo de aplicación de GC.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En el ámbito de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional (UTN.BA) se ha conformado en el año 2009 el Grupo de Estudio de Metodologías para Ingeniería en Software y Sistemas de Información (GEMIS) integrado por un equipo de docentes, alumnos y graduados con interés en la sistematización de conocimientos y su promoción sobre el campo de la Ingeniería en Sistemas de Información y la Ingeniería de Software incluyendo sus aplicaciones y abordajes metodológicos en todo tipo de escenarios.

En el marco de lo anteriormente descripto y con base en las problemáticas presentes en la GC en la industria del software, se ha

considerado necesario proponer una nueva línea de trabajo que tenga como objetivo aplicar la Gestión de Conocimiento a la fase de pruebas de la Ingeniería de Software que indaga los desafíos de los modelos de GC existentes en el domino de las pruebas de software para identificar las falencias de los modelos existentes proponiendo una solución para la ingeniería de software en el ámbito de la GC en el dominio de la fase de pruebas, de esta manera, se articula dentro de los objetivos de GEMIS en el campo de la Informática, la generación de nuevos conocimientos en el área de la Ingeniería de Software.

3. RESULTADOS ESPERADOS

La finalidad de este trabajo es iniciar una línea de investigación dándole continuidad a los objetivos de GEMIS mediante la definición, proposición e implementación de un modelo de aplicación de Gestión de Conocimiento a la fase de pruebas de la ingeniería de software.

De acuerdo con lo anterior, se espera: a) identificar qué modelos de GC son aplicables en la ingeniería de software en el dominio de la fase de pruebas; b) determinar qué modelos Ontológicos son aplicables en la ingeniería de software en el dominio de la fase de pruebas; c) adaptar o desarrollar un modelo de GC, incorporando un modelo Ontológico aplicable en la ingeniería de software en el dominio de la fase de pruebas; d) validar el modelo de GC propuesto; e) exponer las conclusiones sobre el tema y posibles aportes futuros.

La consecución de los objetivos definidos en el proyecto permitirá la definición y proposición de un modelo de aplicación de GC que incorpore una ontología en el dominio de la fase de pruebas de la ingeniería de software que asista a los equipos en la identificación de los escenarios y/o casos de prueba existentes, además de facilitar su documentación durante el desarrollo de las actividades necesarias para la correcta ejecución de esta fase en la ingeniería de software. Este modelo será validado mediante el uso de la técnica de juicio de expertos. Esta técnica en [15][16][17] se define como una opinión informada de personas con trayectoria en el tema, que son reconocidas por otros como expertos cualificados en éste, y que pueden dar información, evidencia, juicios y valoraciones [18].

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo se encuentra conformado por investigadores formados, tesista de maestría, graduados de grado y alumnos de la carrera de Ingeniería en Sistemas de Información. Con esta nueva línea de trabajo se busca tanto la obtención de nuevos conocimientos en el ámbito de la ingeniería de software como la motivación de los implicados para que se desarrollen dentro de la carrera de investigadores, además de fomentar la aplicación de las temáticas en su propia actividad profesional; y plantea la integración de alumnos avanzados de grado y posgrado con posibilidades de articular Proyectos Finales, Trabajos Finales Integrador de Especialidad y Tesis de Maestría de las carreras del área.

5. BIBLIOGRAFÍA

[1] SeCTyP (2017). Programas I+D+I. UTN. Secretaría de Ciencia, Tecnología y Posgrado.

[2] Avendaño Perez, V. y Flórez Urbáez, M. (2016). Modelos teóricos de gestión del conocimiento: descriptores, conceptualizaciones y enfoques. Entreciencias: diálogos en la Sociedad del Conocimiento, 4(10), 201-227.

[3] Marulanda Echeverry, C.E.; López Trujillo, M.; Castellanos Galeano, J.F. (2016). La cultura organizacional y su influencia en las buenas prácticas para la gestión del conocimiento en las Pymes de Colombia. AD-minister, (29), 163-176.

[4] Durán, D. E. S.; Gamboa, A. X. R.; Builes, J. J. (2017). Aplicación de la Gestión de Conocimiento al proceso de pruebas de software. Ingenierías Revistas USBMed, 8(2), 6-13.

[5] De Vasconcelos, J. B.; Kimble, C.; Carreteiro, P.; Rocha, Á. (2017). The application of knowledge management to software evolution. International Journal of Information Management, 37(1), 1499-1506.

[6] Morales, F. J. L.; Gutiérrez, H. A. (2015). La gestión del conocimiento: Modelos de comprensión y definiciones. Revista de investigación en ciencias estratégicas, 2(2), 84-111.

[7] Blanquicett, Luis A., Bonfante, María C., & Acosta-Solano, Jairo. (2018). Prácticas de Pruebas desde la Industria de Software. La Plataforma ASISTO como Caso de Estudio. Información tecnológica, 29(1), 11-18.

[8] Souza, E. F. ; Falbo, R. A. ; Vijaykumar, N. L. (2013). Knowledge management applied to software testing: A systematic mapping. (pp. 562-567).

[9] Wnuk, K., &Garrepalli, T. (2018). Knowledge Management in Software Testing: A Systematic Snowball Literature Review. e-Informatica Software Engineering Journal, 12(1), 51-78.

[10] Durán, D. E. S.; Gamboa, A. X. R.; Builes, J. J. (2017). Aplicación de la Gestión de Conocimiento al proceso de pruebas de software. Ingenierías Revistas USBMed, 8(2), 6-13.

[11] Souza, É. F.; Falbo, R. A.; Vijaykumar, N. L. (2015). Using the findings of a mapping study to conduct a research project: a case in knowledge management in software testing. In Software Engineering and Advanced Applications (SEAA), 2015 41st Euromicro Conference on (pp. 208-215).

[12] Flores, V., &Hadfeg, Y. (2017). Un método para generar explicaciones de resultados de un Sistema Experto, usando Patrones de discurso y Ontología. RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação, (21), 99-114.

[13] González Gola, F., Sánchez, A., & Montejano, G. A. (2017). Asistencia dirigida por ontologías al diseño arquitectónico de videojuegos. In XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires).

[14] Rozo Rodríguez M., & Casanovas I. (2018). La Gestión de Conocimiento Aplicado a la Fase de Pruebas de la Ingeniería de Software - Revisión Sistemática. 6to Congreso Nacional de Ingeniería Informática / Sistemas de Información. CoNaIISI..

[15] Cruz Ramirez, M. et al. (2012) Perfeccionamiento de un instrumento para la selección de expertos en las investigaciones educativas. REDIE (Revista Electrónica de Investigación Educativa), 14(2).

[16] Escobar-Pérez, J. & Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. Avances en Medición, 6, 27-36.

[17] Crespo, T. (2007). Respuestas a 16 preguntas sobre el empleo de expertos en la investigación pedagógica. Lima, Perú: San Marcos.

[18] Scapolo, F. & Miles, I. (2006). Eliciting experts' knowledge: A comparison of two methods. Technological Forecasting & Social Change, 73, 679–704.

REVISIÓN SISTEMÁTICA PARA EL DISEÑO DE UNA METODOLOGÍA DE REQUERIMIENTOS REFERENCIALES EN DOMINIOS COMPLEJOS.

Moyano Ezequiel UNTDF, Urciuolo Adriana UNTDF, Czelada Alejandro UNTDF, Izarra Emilio UNTDF, Rigoni Brian UNTDF
Instituto de Desarrollo Económico e Innovación, UNTDF

Dir.: Fuegia Basket 251, (9410) Ushuaia. Tierra del Fuego. Tel: ++54-2901-432403
emoyano@untdf.edu.ar, aurciuolo@untdf.edu.ar, eaczelada@untdf.edu.ar, eizarra@untdf.edu.ar, brian.rigoni1@gmail.com

Resumen

Los sistemas que modelan el comportamiento de la naturaleza se caracterizan por su complejidad, la mayoría desarrollados por expertos del dominio, sin considerar el uso de métodos y herramientas de Ingeniería de software; así el conocimiento adquirido queda solo en ese ámbito. Para el desarrollo de nuevos sistemas el conocimiento debe obtenerse nuevamente. El reuso del conocimiento del dominio resulta esencial.

Para lograr el reuso del conocimiento es necesario contar con métodos de especificación de requerimientos a nivel de dominio (Requerimientos Referenciales), que puedan ser utilizados en futuros sistemas del dominio, reduciendo tiempos y esfuerzos.

A pesar de la importancia de la etapa de requerimientos en el desarrollo de software, existen pocos antecedentes sobre métodos para la formulación de Requerimientos Referenciales como paso previo al Análisis de Dominio.

El presente trabajo presenta una revisión sistemática, en base al protocolo recomendado por Kitchenham, *respecto al estado del arte en el uso de una metodología para la especificación de Requerimientos Referenciales en dominios complejos.*

Palabras clave: Revisión Sistemática, Requerimientos Referenciales, Ingeniería de Requerimientos, Análisis de Dominio, Reuso,

Contexto

La línea de investigación se desarrolla en el Instituto de Desarrollo Económico e Innovación (IDEI) de la Univdersidad Nacional de Tierra del Fuego, por parte de un Grupo de docentes-investigadores que lleva adelante proyectos en la temática de Sistemas de Información Ambiental – Hidroinformática desde hace más de 10 años.

El proyecto se presentó a la convocatoria realizada por la UNTDF en septiembre de 2018 y resultó aprobado luego de haber sido sometido a evaluación externa. (Resolución Rectoral 514/2019), para el período de. (Abr/2019-Mar/2021).

1. Introducción

El estudio de los sistemas que modelan la naturaleza adquiere cada vez mayor importancia, dada la función preponderante que los mismos cumplen en la toma de decisiones para el manejo sustentable de los recursos naturales y el medio ambiente. Los modelos ambientales son abstracciones de la naturaleza y de su comportamiento, que permiten la toma de decisión a partir de predicciones[1].

La complejidad de estos sistemas está dada por la necesidad de manejar grandes volúmenes de datos, objetos espacio / temporales, factores ambientales, políticas de gestión, normativas legales, etc[2].

El reuso de conocimiento del dominio en tales situaciones resulta esencial.

Estos sistemas en gran parte fueron desarrollados por expertos del dominio en forma aislada, sin considerar el uso de métodos y herramientas de Ingeniería de software; con lo cual el conocimiento adquirido queda solo en el ámbito en el cual se obtuvo (organismo o institución).

Para el desarrollo de nuevos sistemas (en el mismo dominio), el conocimiento debe obtenerse nuevamente, lo cual involucra un gran esfuerzo. De allí la importancia del reuso de conocimiento en dominios complejos.

Los sistemas de información ambiental, como los de modelado de cuencas, se caracterizan por su complejidad. Rozan una gran cantidad de subdominios diversos de conocimiento, generalmente heterogéneos[3], tanto técnica como semánticamente; se relacionan con el manejo de los datos de distintos componentes interactuantes del ambiente: el suelo, el agua, el aire y las especies existentes.

La mayoría de los fenómenos ambientales se debe modelar considerando cinco dimensiones: su localización en el espacio (latitud, longitud y altitud), el tiempo en que se obtiene y los fenómenos particulares que son analizados.

La ingeniería de dominio tiene como objetivo lograr la reutilización del conocimiento y de las especificaciones de un dominio específico, de manera de poder ser utilizadas en diversas aplicaciones del mismo ámbito.[4]

Las técnicas de análisis de dominio están asociadas a la reutilización, su principal característica es capturar información relacionada con el dominio, y determinar qué y cómo este conocimiento se reutilizará en el desarrollo de futuras aplicaciones[5].

No obstante existen escasas técnicas de especificación de requerimientos a nivel de dominio (Requerimientos Referenciales) que permitan, para sistemas de tal complejidad, obtener un modelo de dominio completo y consistente.

Así como en la ingeniería de software tradicional la primer etapa es la ingeniería de requerimientos (donde existen un gran desarrollo)[6], en la ingeniería de domino, como paso previo al análisis de dominio, se deben obtener requerimientos referenciales[7].

Para lograr el reuso del conocimiento es necesario contar con métodos de especificación de requerimientos a nivel de dominio (*requerimientos referenciales*), con el propósito de que la información (identificada, capturada y organizada) sea reutilizable en la creación de nuevas aplicaciones de software en ese dominio, en otros subdominios similares, reduciendo tiempos y esfuerzos de considerable magnitud.

A pesar de la importancia de la etapa de requerimientos en el desarrollo de software[8], existen pocos antecedentes sobre métodos para la formulación de requerimientos referenciales como punto de partida para el Análisis de Dominio, especialmente en dominio complejos donde adquiere más trascendencia para el reuso.

El presente trabajo tiene como objetivo fundamental realizar una revisión sistemática [9] respecto al desarrollo de metodologías de especificación de requerimientos inferenciales en dominios complejos, y sobre todo en dominios ambientales.

Los resultados obtenidos se enmarcan dentro de las actividades preliminares relacionadas con el proyecto de investigación “Diseño de una Metodología para Requerimientos Referenciales en Dominios Complejos. Caso De Estudio: Aplicaciones De Modelado De Cuencas” (PIDUNTDF) que se encuentra actualmente en desarrollo y será parte del marco teórico y del estado del arte de la temática a desarrollar.

2. Línea de Investigación

Como parte de la línea de investigación que el equipo de investigadores lleva adelante en temáticas vinculadas a los Sistemas de Información Ambiental – Hidroinformática, se

plantean los siguientes objetivos a alcanzar en el presente proyecto:

Promover el reuso de conocimiento en dominios complejos a través del desarrollo de una metodología de análisis y especificación de Requerimientos Referenciales, que permita reducir tiempo y esfuerzo en la construcción de aplicaciones de un dominio de características complejas.

3. Revisión Sistemática.

Para realizar la revisión sistemática respecto al diseño de metodologías para el uso de requerimientos referenciales en dominios complejos se utilizo como base el protocolo recomendado por Kitchenham[9], definido como una guía para conducir el proceso de revisión sistemática de la literatura en ingeniería de sotware.

PASO 1: El primer paso de protocolo es establecer la formulación de las preguntas sobre investigación, el objetivo que se persigue es poder determinar de la literatura que se obtenga, la existencia de estudios en el campo o temática a fin del proyecto. No solo aspectos teóricos sino también estudios que definan estrategias sobre casos reales.

Se definen una pregunta central y subpreguntas para profundizar el tema de investigación. Estas preguntas nos ayudaran posteriormente (sobre la muestra final de artículos) determinar el grado de interés respecto al proyecto.

Preg. 1: ¿Existen estudios o evidencias relacionadas a la obtención y/o especificación de requerimientos referenciales en sistemas complejos?.

SubPreg. 1.1: "¿Cuáles son los principales aportes de los trabajos relacionados al objetivo del proyecto?"

SubPreg. 1.2: "¿Que característica demuestra el tipo de estudio de la literatura?"

PASO 2: A los efectos de realizar la búsqueda de literatura se deben determinar dos aspectos importantes:

El Conjunto de librerías digitales como fuentes de información y Cadenas de búsqueda.

Para el primer caso se utilizaron[10]:

1. *Google Scholar.*
2. *ACM Digital Library.*
3. *IEEEXplore.*
4. *Springer Link.*
5. *Science Direct.*
6. *SEDICI.*

Para el el segundo caso se definieron 4 criterios de búsqueda (concatenados según el criterio de cada librería)

A. Requerimientos + referenciales + sistemas + complejos

B. Ingenieria + Requerimientos + análisis o ingenieria + dominios

C. requirements + referential + systems + complex.

D. engineering + requirements + analysis or engineering + domain.

Los resultados obtenidos fueron ordenados por relevancia, según el criterio de cada librería digital.

PASO 3: Es importante considerar, más allá de los criterios de búsqueda utilizados, que varios de los estudios seleccionados no aporten evidencia relacionada a las preguntas de investigación propuestas. Para obtener un subconjunto más efectivo se definieron un conjunto de criterios[11][12], de inclusión (CI) y exclusión (CE).

Es importante aclarar los articulos son incluidos si cumple al menos uno de los CI y serán excluidos los que verifiquen al menos de uno de los CE[10].

CI1: Trabajos actuales, comprendidos entre 2005 y 2020.

CI2: Contengan conceptos relevantes del proyecto en su título, abstract y/o palabras claves.

CI3: Demuestren ser de interés y/o aporte del uso de requerimientos referenciales en sistemas complejos.

CE1: Sean revisiones sistemáticas que no generan contribuciones de relevancia.

CE2: Trabajos duplicados o similares en toda su concepción.

CE3: Trabajos que si bien puedan ser sobre requerimientos no están aplicados a sistemas complejos.

CE4: Trabajos no sean artículos científicos publicados.

PASO 4: Por último se debe determinar la calidad de los artículos incluidos en la revisión sistemática, para limitar posibles sesgos y guiar la interpretación de los resultados [13].

A efectos de obtener una valoración de calidad de los artículos seleccionados se utilizó una evaluación cuantitativa basada en cuatro preguntas recomendadas en Kitchenham y Charters[14].

Las preguntas de verificación definidas en la presente revisión sistemática son:

PC1: ¿El artículo contempla adecuadamente los objetivos planteados en el proyecto?

PC2: ¿El artículo presenta una clara relación entre la ingeniería de requerimientos aplicada a dominios complejos o su uso como etapa en el análisis de dominio?

PC3: ¿Se describieron adecuadamente los datos / estudios básicos?

PC4: ¿Los autores evaluaron la validez de los resultados obtenidos y/o el beneficio de especificar los requerimientos del dominio?

4. Resultados.

A continuación se presentan los resultados obtenidos obtenidos luego de aplicar el protocolo de Kitchenham.

En la tabla 1 se muestran los resultados del procedimiento, la segunda columna indica la cantidad total de artículos encontrados, la tercera los artículos seleccionados para su análisis (las búsquedas se ordenaron por relevancia y se seleccionaron, para cada una de las cuatro, 15 en español y 25 en ingles)

La última columna indica la cantidad de artículos que nos quedan luego de aplicar los criterios de inclusión y exclusión (ver paso 3).

Librería	Artículos Totales	1er Filtro	2do Filtro
IEEE	3.753	75	**37**
Google Schooll	502.500	80	**24**
ACM	1.060.300	80	**10**
Springer	28.969	80	**24**
Science Direct	210.850	80	**12**
Sedici	3.056	80	**27**
Total	**1.809.428**	**475**	**134**

Tabla 1. búsqueda y selección de artículos.

Para obtener los artículos de real calidad para el estudio del estado del arte del proyecto, se aplica una evaluación cuantitativa (ver paso 4 de la revisión sistemática).

Para la valoración se utilizó el criterio del Centro para Revisiones y Difusión (CDR) Base de datos de resúmenes de revisiones de efectos (DARE)[15] de la Universidad de York. donde a cada pregunta se asigna un valor: Y (total) 1, P (parcial) 0,5 y N (Nulo) 0.

Y se define un umbral del 60 % o más, es decir aquellos que sumen una calificación de 2.5 o más. La tabla 2 refleja la cantidad de artículos que fueron considerados.

Valor alcanzado	Cantidad de Artículos
4 Pts.	**7**
3.5 Pts.	**11**
3 Pts.	**11**
2.5 Pts.	**10**
2 Pts.	54
1.5 Pts.	15
1 Pts.	11
0.5 Pts.	5
0 Pts.	10
Total Considerados	**39**

Tabla 2. Validación de calidad.

5. Conclusión

El protocolo utilizado para la revisión sistemática de la bibliografía permitió obtener un conjunto concreto de publicaciones a fines a la linea de investigación del proyecto, e identificar las actuales iniciativas que se encuentran en la literatura disponible.

Se puede determinar que no existe una evidencia empírica concreta sobre una metodológica de especificación de requerimientos referenciales. Existen pocos antecedentes (a nivel dominio) como paso previo al Análisis de Dominio, sobre todo en sistemas ambientales. Si bien en la última década se ha mostrado una preocupación más acabada sobre esta problemática.

Las actuales investigaciones dirigen su preocupación en la definición de y lenguajes específicos de dominio, elevando el nivel de abstracción mediante la especificación de los conceptos del dominio, sin considerar la necesidad de definir requerimientos referenciales que permitan el reuso del conocimiento.

6. Formación de Recursos Humanos

El Equipo de Trabajo está conformado por docentes investigadores de la UNTDF, licenciados en informática y alumnos de la carrera de Sistemas.

El Proyecto incluye docentes en etapa de formación de postgrado, el Director del proyecto está realizando su tesis de magister en Ingeniería de Software en la UNLP, denominada: "Reuso de Requerimientos Referenciales en Dominios Complejos. Caso de Estudio: Aplicaciones de Modelado de Cuencas." Dirigida por la Mgs. Adriana Urciuolo y co-dirigida por el Dr. Leandro Antonelli de la UNLP.

Participaran Asistentes que realizarán sus primeras en actividades de investigación para desarrollar sus estudios de postgrado en temáticas afines (A. Czelada y E. Izarra). Colaborará el alumno Brian Rigoni, con el objetivo formarlo en los aspecto generales de la investigación y que le sirva para el desarrollo de su futura tesis de grado.

A través del proyecto se espera consolidar un equipo de trabajo de la UNTDF en la temática.

7. Referencias

1. Moyano E, Urciolo A, Iturraspe R, Una Metodología para Requerimientos Referenciales en dominios complejos. WICC 2019, San Juan.
2. Urciuolo A. Sistemas de Hidroinformática Oo + Soluciones Evolutivas 2005.
3. Rol Of Information Technology In Environmental Information System, 2003
4. JanisOsis, Erika Asnina, Model-Driven Domain Analysis and Software Development Riga Technical University, Latvia. 2010.
5. Emeline, Regis; Gustavo, Tondello; Análise de Domínio,Universidad Santa Catarina, 2000
6. Olga De Troyer and Erik Janssens, A Feature Modeling Approach for Domain - Specific Requirement Elicitation. Universiteit Brussel.
7. Torres Pérez L., Delgado Dapena M., Entorno de ingeniería de requisitos aplicado para producir software en una universidad, 2014.
8. Rossi G, Antonelli L, Sampaio JC, Oliveros A, Buenas prácticas en la especificación del dominio de una aplicación, WER 2013.
9. Kitchenham, B., Procedures for performing systematic reviews. UK, Keele University, 2004.
10. Pinto N., Acuña C., Tomaselli G., Impacto de las emociones del usuario en la percepción de la calidad del software: Una revisión sistemática, UTN - Facultad Regional Resistencia, 2019.
11. Noruzi, A. The New Generation of Citation Indexes. http://dx.doi.org/10.1515/LIBR.2005.170 , 2005.
12. Petticrew M, Helen R. Systematic reviews in the social sciences: A practical guide. Blackwell Pub; 2006.
13. Cascaes da Silva, F., Valdivia A., Gutierres F., Escalas y listas de evaluación de la calidad de estudios científicos. Revista Cubana de Información en Ciencias de la Salud. 2013.
14. Kitchenham B., Charters S., Guidelines for

performing systematic literature reviews in software engineering, Keely University.

15. Centre for Reviews and Dissemination, What are the criteria for the inclusion of reviews? http://www.york.ac.uk/inst/crd/faq4.htm, 2007

Desagregación de la información en los escenarios futuros

Gladys Kaplan[1], David Tua[2], Gabriel Blanco[1]

[1]Departamento de Ingeniería e Investigaciones Tecnológicas. Universidad Nacional de La Matanza. Argentina
[2]Programa de Maestría en Ingeniería en Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina
gkaplan@unlam.edu.ar, dtua@frba.utn.edu.ar

RESUMEN

Los escenarios, como la mayoría de los modelos de la Ingeniería de Requisitos, utilizan el lenguaje natural para asegurar una buena comunicación con los clientes-usuarios. El proceso de construcción de escenarios aprovecha las ventajas del lenguaje natural y genera mecanismos para controlar algunas de sus debilidades. Estas últimas son mitigadas con el uso de un glosario, la sintaxis, entre otras. Una vez completos los escenarios, se vuelven a controlar en la Actividad Verificar. Si bien la verificación es exhaustiva, omite analizar la riqueza semántica de las expresiones, las que pueden esconder conocimiento relevante del contexto. En el presente proyecto de investigación se analiza la granularidad de las expresiones utilizadas fundamentalmente en la descripción de los episodios, las restricciones y las excepciones, donde puede existir soterramiento de información que debe ser desagregada. Se ha comprobado empíricamente que los diferentes niveles de detalle utilizados en los escenarios perjudican la comprensión de la propuesta de solución y es una de las causas de omisiones o malos entendidos en los requisitos del software.

Palabras clave: Ingeniería de Requisitos, Escenarios, Lenguaje Natural, Completitud.

CONTEXTO

La línea de investigación que se presenta se encuadra dentro del proyecto de investigación "Aspectos no funcionales en los procesos de requisitos" de la Universidad Nacional de La Matanza (UNLaM), del cual se desprende una tesis de Maestría de la Escuela de Posgrado de la Universidad Tecnológica Nacional Facultad Regional Buenos Aires (UTN-FRBA).

1. INTRODUCCIÓN

La especificación de requisitos de software (ERS) debe ser, entre otros atributos, lo más completa posible, tanto en términos de negociación contractual como de documentación técnica. Alcanzar este objetivo es necesario para asegurar la construcción de un sistema de software que satisfaga las necesidades del cliente. Como es conocido, el costo de corrección de errores y omisiones varía de acuerdo al momento en el cual se lo detecta [1] [2], por lo tanto, es de suma importancia asegurar la calidad de los requisitos del software lo más temprano posible.

El proceso de requisitos basado en escenarios [3], en el que se basa la presente investigación, ha avanzado de diferentes maneras en algunas problemáticas asociadas al uso del lenguaje natural (LN) [4]. En el proceso de requisitos se utilizan básicamente dos modelos, el Léxico Extendido del Lenguaje (LEL) [5] [6] y los escenarios [7]. La construcción de un glosario como primera actividad en el proceso es fundamental para asegurar la comunicación. El LEL contiene el vocabulario propio del dominio de la aplicación y es utilizado en todas las descripciones para mejorar la precisión y disminuir la ambigüedad [8].

Los escenarios son narrativas en LN con un mínimo de formalismo, representan las situaciones del contexto. Como ya se mencionó, tanto la sintaxis como la utilización de un glosario mejoran la calidad de estas descripciones.

La diferencia sustancial entre los escenarios actuales y los futuros es su punto de vista [9]. La semejanza o disparidad entre un conjunto de escenarios y el otro depende de la cantidad de cambios que se espera en el contexto futuro. A menor cantidad de cambios, por ejemplo una informatización, existe una gran similitud entre ambos conjuntos de escenarios. A medida que se incorporan cambios en los procesos del negocio, por ejemplo, en una reingeniería de procesos del negocio, las diferencias son cada vez mayores.

Título : ASIGNAR FECHA DE ANÁLISIS
Objetivo: definir el momento de analizar
Contexto:
Ubicación Geográfica: Oficina técnica
Ubicación Temporal:
Precondición: debe haber un insumo con etiqueta amarilla
Recursos: cronograma de trabajo_1, insumo, fecha de análisis
Actores: jefe de control de calidad, **Sistema de Software**
Episodios:
1. el jefe de control de calidad ingresa fecha probable de análisis
2. el **Sistema de Software** busca los insumos sin fecha de análisis
3. #el **Sistema de Software** muestra el cronograma de trabajo_1 de la fecha ingresada y los insumos ordenados por prioridad de análisis
4. el **Sistema de Software** muestra la disponibilidad de cada técnico #
5. **Si** no hay disponibilidad de técnicos en esa fecha determinada **entonces** REASIGNAR FECHA DE ANALISIS
6. el jefe de control de calidad determina la fecha de análisis
7. ingresa en el cronograma de trabajo_1 el insumo y el técnico seleccionado

Excepciones:
Cuando existen conflictos por igual prioridad de análisis el jefe de control de calidad determina el orden según su criterio.

Fig.1 – Ejemplo de un escenario futuro

Como puede observarse en la Fig.1, los escenarios son representaciones con una gran riqueza en conocimiento. Esto se logra en parte, concediendo la mayor libertad posible en la expresividad de las descripciones, pero deben existir algunas restricciones para asegurar la consistencia de la información. De esta manera, el modelo de escenarios presenta una sintaxis simple, pensada con dos objetivos fundamentales:

1) No entorpecer la construcción de cada escenario.
2) Ser transparente para el lector.

En la Fig.2 se puede observar una parte del modelo de escenarios que define la sintaxis utilizada en los episodios, restricciones y excepciones, foco del presente proyecto.

Sintaxis (usando BNF parcial):
<episodios> ::= <serie grupo> | <serie episodios>
<serie grupo> ::= <grupo> <grupo> | < grupo no secuencial> | <serie grupo> <grupo>
<grupo> ::= <grupo secuencial> | < grupo no secuencial>
<grupo secuencial> ::= <sentencia básica> | <grupo secuencial > <sentencia básica>
<grupo no secuencial> ::= # <serie episodios> #
<serie episodios> ::= <sentencia básica> <sentencia básica> | <serie episodios> <sentencia básica>
<sentencia básica> ::= <sentencia simple> | <sentencia condicional> | <sentencia optativa>
<sentencia simple> ::= <sentencia episodio> CR
<sentencia condicional> ::= Si <condición entonces <sentencia episodio> CR
<sentencia optativa>::= [<sentencia episodio>] CR
donde <sentencia episodio> se describe:
(([Actor | Recurso] + Verbo + Predicado) | ([Actor | Recurso] + [Verbo] +Título)) + {Restricción}
Excepciones: generalmente reflejan la falta o mal funcionamiento de un recurso. Una excepción impide el cumplimiento del objetivo del Escenario. Eventualmente se puede indicar el tratamiento de la excepción a través de un Escenario.
Sintaxis: Causa [(Solución)]
donde Causa es:
Frase | ([Sujeto | Actor | Recurso] + Verbo + Predicado)
donde Solución es:
Título

Fig.2 – Sintaxis de los episodios

Para asegurar la calidad de los escenarios, no ha sido suficiente la definición de una sintaxis particular, la utilización del LEL, los patrones de descripción [10] y la heurística de construcción, también se debió incorporar un proceso de verificación que garantice la consistencia de toda la información representada.

En la Tabla 1 se puede observar en un estudio comparativo, que existen diferentes técnicas de verificación dentro de la IR que aplican en diferente momento del proceso.

Documentos	Técnicas de Verificación	Aplican a LN
ERS	45%	56%
Previos a la ERS	55%	**33%**

Tabla 1 – Técnicas de verificación en la IR

El proceso de requisitos en el cual se basa el presente proyecto, se encuentra en el 55% de las técnicas para documentos previos a la ERS y dentro del 33% que aplican al LN. Entre las técnicas más utilizadas para la verificación se encuentran las Inspecciones [11] [12], las cuales son utilizadas en este proceso para detectar inconsistencias, tanto sintácticas como semánticas. A continuación, se describe el proceso de verificación existente, el cual se ejecuta con formularios de inspección [13].

Pese a la gran flexibilidad con que se pueden escribir los escenarios, éstos son enormemente ricos en cuanto a su posibilidad de contrastar unos contra otros y contra el LEL, es así que se pueden definir dos tipos de consistencias posibles aplicables a un conjunto de escenarios:

I. Consistencia interna de escenarios

I.1 Verificación sintáctica

I.2 Interrelación con el LEL

I.3 Interrelación entre componentes

II. Consistencia entre escenarios

II.1 Relación entre escenarios

II.2 Superposición entre los escenarios

II.3 Cubrimiento del LEL

La Consistencia interna de escenarios se ocupa de contrastar las distintas partes constitutivas de los mismos. La Consistencia entre escenarios se ocupa de analizar la integridad del conjunto de escenarios. En la Fig. 3, se esquematizan ambos procesos.

Consistencia interna de escenarios: se ocupa de analizar cada escenario en forma individual. El proceso de consistencia interna de escenarios utiliza para su desarrollo los siguientes elementos adicionales: el modelo de escenarios y el LEL. Al finalizar el proceso, se obtiene un conjunto de escenarios mejorados y un grupo de dudas que se deben resolver en el UdeD, a partir de las cuales se genera una nueva versión de escenarios y eventualmente un nuevo proceso de consistencia interna de escenarios.

Como se muestra en la Fig.3, el análisis de cada escenario comprende la verificación sintáctica, la interrelación con el LEL y la interrelación de los componentes. La verificación sintáctica comprueba que cada componente haya sido correctamente escrito.

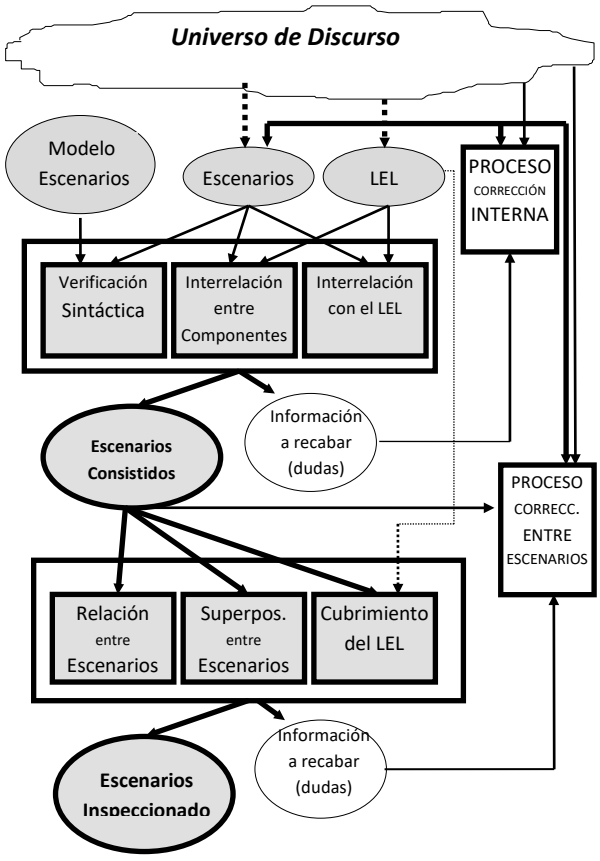


Fig.3 – Inspección de escenarios

La interrelación con el LEL verifica que los símbolos del LEL estén correctamente utilizados y que toda frase destacada como símbolo del LEL efectivamente sea parte del mismo. La interrelación de componentes verifica entre otras cosas, que todos los actores de la lista de actores cumplan un rol en algún episodio y que todo sujeto de los episodios esté en la lista de actores.

Consistencia entre escenarios: este proceso debe realizarse luego de finalizado el proceso de consistencia interna de escenarios. Al finalizar el proceso, se obtiene un conjunto de escenarios mejorados y un grupo de dudas que se deben resolver en el macro sistema, a partir de las cuales se genera una nueva versión de escenarios y eventualmente un nuevo proceso de consistencia interna de escenarios o de consistencia entre escenarios.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La verificación existente de escenarios, cubre algunas debilidades del propio proceso de construcción y otras relacionadas con el uso del lenguaje natural. Ninguna de estas inspecciones controla la granularidad de la información presente en cada componente del escenario. La presente línea de investigación se encuadra dentro de la consistencia interna de escenarios (ver Fig.3) pero en este caso, desde una perspectiva cualitativa. El objetivo es obtener un conjunto de escenarios más homogéneos y completos. Para lograrlo es necesario desagregar algunas expresiones para identificar nuevo conocimiento del contexto y detectar omisiones y faltantes de información que no fueron identificadas por las inspecciones existentes.

Se espera clasificar los tipos de soterramiento de conocimiento, como ser la hiperonimia, la polisemia, etc. que aparecen en general en los verbos que no están respaldados por el LEL y generar una heurística que detecte las expresiones que requieran ser desagregadas.

Entre los temas a analizar son:

1) Diferencia de granularidad en Episodios, Restricciones y Excepciones.
2) Diferencia de granularidad en los componentes Actores y Recursos.
3) Inconsistencia entre un episodio y el episodio anterior y posterior.
4) Inconsistencia entre un sub escenario con el episodio anterior y el siguiente.
5) Inconsistencia entre un episodio y su restricción.
6) Inconsistencia entre la excepción y el objetivo del escenario.
7) Sobre o sub cubrimiento del objetivo del escenario en el conjunto de episodios de un escenario.

Desagregar la información de los escenarios permite incrementar la comprensión de cada situación del contexto, detectando en algunos casos, otras situaciones involucradas que no habían sido visibilizadas. En el caso de los escenarios actuales, bajar la granularidad de la información, mejora la visión del contexto en estudio, mientras que para los escenarios futuros, asegura la validación de la propuesta para el nuevo sistema de software. La desagregación de conocimiento en los episodios mejora cuantitativa y cualitativamente la obtención de los requisitos funcionales, ya que estos se encuentran empotrados en los episodios. La desagregación en las restricciones y las excepciones puede aumentar la identificación de los requisitos no funcionales.

Al desagregar se puede identificar información faltante como sobrante. Incluir información faltante, incrementa la completitud de los escenarios. Eliminar información sobrante, mejora la legibilidad. En este último caso, puede existir una retroalimentar al LEL, ya que cierta información que aparece en los escenarios corresponde al glosario. De esta manera, la mejora en la granularidad de los escenarios puede repercutir en la completitud del LEL.

3. RESULTADOS OBTENIDOS/ESPERADOS

En un estudio empírico de aproximadamente 20 casos de estudio, se pudo observar que la presencia de diferentes niveles de detalle depende en gran medida del conocimiento previo que tiene el ingeniero de requisitos acerca del contexto en estudio, observando que cuanto mayor es su conocimiento del dominio, mayor es la tendencia a obviar detalles. También se detectó una relación con las fuentes de información utilizadas, ya que durante una entrevista con personal jerárquico la tendencia fue abstraer conocimiento, utilizando expresiones genéricas propias de su rol en la organización. Mientras que con el personal

operativo, en la mayoría de los casos, apareció un mayor nivel de detalle. No sucedió lo mismo con la literatura del dominio, las cuales mantienen una cierta regularidad. También es importante resaltar la importancia de la disponibilidad del usuario para disipar todas las dudas que la literatura del dominio no pueda aclarar. Finalmente, la cultura del ingeniero de requisitos se ve reflejada en el estilo de redacción de los escenarios, esto sucede por la ausencia de un vocabulario mínimo que regule todo aquello que no está definido en el LEL. Se ha comprobado que algunas expresiones parecen encerrar otros conceptos, pero al analizarlas con mayor profundidad resulta en un problema de redacción. En algunos casos, es detectado por la verificación existente, donde se analiza la estructura sintáctica de cada componente. En otros casos, el texto es confuso por una mala interpretación del contexto o por falta de información. Los problemas de redacción, a pesar de existir una sintaxis para cada caso, son muy comunes en descripciones en lenguaje natural y dificultan seriamente la consistencia de los escenarios y la comprensión del problema. Muchos de estos casos se resuelven simplemente con reemplazar parte del texto.

También se detectó que un mejor nivel de detalle genera la aparición de estados y jerarquías del contexto en los escenarios, que suelen quedar ocultos en las descripciones de los episodios, principalmente cuando la granularidad es alta. Esto mejora significativamente la precisión de la descripción de cada situación.

4. FORMACIÓN DE RECURSOS HUMANOS

Esta línea de investigación es parte directa de la tesis de Maestría del Ing. David Tua, en la Escuela de Posgrado de la UTN FRBA.

5. BIBLIOGRAFÍA

[1] Boehm, B. W., & Papaccio, P. N. (1988), "Understanding and controlling software costs", IEEE transactions on software engineering, 14(10), 1462-1477.

[2] Sharif, Bushra & Shoab, A & Khan, & Bhatti, Muhammad. (2012), "Measuring the Impact of Changing Requirements on software Project Cost: An Empirical Investigation", International Journal, Computer Science. Capítulo 9.

[3] Leite, J., C., S., P., Doorn, J. H., Kaplan, G., Hadad, G., D., S., Ridao, M., N., (2004) "Defining System Context using Scenarios" in Perspectives on Software Requirements, Kluwer Academic Publishers, cap.8, pp. 169-199.

[4] Vallez M. y Pedraza-Jimenez R. (2007), "El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines". Hipertext.net, núm. 5. https://www.upf.edu/hipertextnet/numero-5/pln.html.

[5] Leite J.C.S.P., Franco, A.P.M., (1990), "O Uso de Hipertexto na Elicitaçao de Linguagens da Aplicaçao", Anais de IV Simpósio Brasilero de Engenharia de Software, SBC, pp. 134-149.

[6] Hadad, G.D.S., Doorn J.H., Kaplan G.N. (2008), "Creating Software System Context Glossaries", in Encyclopedia of Information Science and Technology, Second edition, Mehdi Khosrow-Pour (ed.), Idea Publishing.

[7] Leite, J. C. S., Hadad, G. D. S., Doorn, J. H., & Kaplan, G. N. (2000), "A scenario construction process", Requirements Engineering, 5(1), 38-61.

[8] Berry, D. and Erik Kamsties (2004), "Ambiguity in Requirements Specification" in Perspectives on Software Requirements, Kluwer Academic Publishers, cap.2, pp. 7-39.

[9] Doorn J.H., Hadad G.D.S., Kaplan G.N. (2002), "Comprendiendo el Universo de Discurso Futuro", Workshop en Ingeniería de Requisitos, Valencia-España.

[10] Ridao, M., Doorn, J., Leite, J.C.S.P., "Incorporación de patrones al proceso de construcción de Escenarios" (2001), Workshop on Requirements Engineering (WER), Buenos Aires, Argentina.

[11] Fanmuy, G., Fraga, A., & Llorens, J. (2012), "Requirements verification in the industry" in Complex systems design & management (pp. 145-160). Springer, Berlin, Heidelberg.

[12] Leite, J.C.S.P., Doorn, J.H, Hadad, G.D.S., Kaplan, G.N. (2003), "Using Scenario Inspections on Different Scenarios Representations", Monografias em Ciência da Computação, Departamento de Informática, PUC-Rio, N.33/03.

[13] Leite, J.C.S.P., Doorn, J.H., Hadad, G.D.S., Kaplan, G.N. (2005), "Scenario Inspections", Requirements Engineering Journal, Vol.10, Nº 1, SpringerVerlag London., Gran Bretaña, pp.1-21.

Reducción de la Subjetividad en los procesos de requisitos

Gladys Noemí Kaplan[1], Jorge Horacio Doorn[2,3]

[1]Departamento de Ingeniería e Investigaciones Tecnológicas. Universidad Nacional de la Matanza. Argentina
[2]Escuela de Informática, Universidad Nacional del Oeste
[3]Departamento de Ingeniería, Universidad Nacional de Tres de Febrero
gkaplan@unlam.edu.ar, jdoorn@exa.unicen.edu.ar

RESUMEN

Los procesos de requisitos son esencialmente cognitivos, ya que consisten en generar un nuevo sistema de software a partir de la información que se adquiere, del conocimiento existente y de la valoración y elaboración de estos elementos. A medida que se avanza en el proceso, los modelos ya construidos pasan a ser una fuente de información del próximo modelo. De esta manera se avanza más rápidamente y se ayuda al ingeniero de requisitos a evitar el conocido "bloqueo del escritor" que es un tipo de parálisis cognitiva que se genera al comenzar un nuevo modelo. El proceso de requisitos utilizado en el presente artículo, construye como primera actividad, un glosario denominado LEL, desde el cual se genera una primera versión de escenarios actuales. Esta derivación, que toma de manera textual la información del glosario, no ha valorado correctamente que el LEL es esencialmente declarativo y que los escenarios son esencialmente procedurales. El presente proyecto propone generar un nuevo mecanismo de derivación que aproveche la información disponible en el LEL ampliando el conocimiento con toda la documentación disponible, revisando la misma desde una mirada de procesos. De esta manera se espera obtener escenarios mucho más completos y reales.

Palabras clave: Proceso de Requisitos, Lenguaje Natural, Escenarios, Derivación.

CONTEXTO

La línea de investigación que se presenta es parte de los proyectos de investigación "Aspectos no funcionales en los procesos de requisitos" de la Universidad Nacional de La Matanza (UNLaM).

1. INTRODUCCIÓN

Este proyecto está enmarcado en una estrategia de Ingeniería de Requisitos [1] [2] [3] orientada al cliente, ya que se basa en la construcción de modelos en lenguaje natural (LN) [4]. Su objetivo es alcanzar una profunda comprensión del dominio antes de definir el sistema de software. A tal efecto, requiere de dos grandes etapas bien distinguibles: una de aprendizaje y la otra de definición. De existir conocimiento previo, la primera etapa, se convierte en una actividad confirmatoria. Los modelos que se utilizan en esta estrategia son:

1) un modelo léxico, LEL (Léxico Extendido del Lenguaje) [5] [6], el cual describe el vocabulario utilizado en el dominio de la aplicación.

2) un modelo organizacional que describe los procesos actuales del negocio, denominado Escenarios Actuales [7].

3) un modelo organizacional que describe los procesos del negocio proyectados incluyendo el sistema de software a desarrollar [8] [9], denominado Escenarios Futuros.

Los escenarios futuros tienen empotrados los requisitos del software. El documento de especificación de requisitos de software (ERS), se obtiene extrayendo los requisitos del software desde estos escenarios [10].

Debido a la importancia que tiene la comprensión del contexto en el presente proceso de requisitos, es que para definir una propuesta adecuada a las necesidades de cada

cliente, es necesario comprender lo mejor posible el contexto y sus particularidades. La heurística de derivación [11], no contribuye a mejorar esa comprensión ya que toma de manera casi directa la información desde el LEL, con un mínimo de procesamiento. Luego, con este conjunto de escenarios candidatos, se retorna al UdeD para completarlos, rectificando o ratificando la información del LEL.

La actividad Derivar se realiza teniendo en cuenta el conocimiento del LEL, el modelo de escenarios y la heurística de derivación. Consta de tres sub actividades: Identificar los actores del UdeD, Identificar los Escenarios y por último Crearlos usando el LEL. De esta manera se crea una lista inicial de escenarios candidatos con los que se comienza a trabajar. Los pasos son:

a) **IDENTIFICAR ACTORES**

Los símbolos que representan actores en el UdeD son identificados en el LEL. Ellos corresponden a los símbolos clasificados con el tipo Sujeto.

b) **IDENTIFICAR ESCENARIOS**

Cada impacto de los símbolos Sujeto representan un posible Escenario candidato y son incorporados a la lista de Escenarios Candidatos. El título del escenario está compuesto por la acción (verbo) incluido en el impacto pero presentado en infinitivo.

Cuando diferentes actores ejecutan la misma acción, puede producir que dos o más escenarios de la lista puedan compartir el título; en este caso es recomendable no eliminar ninguno hasta asegurarse que realmente son iguales.

c) **CREAR**

Aquí se intenta construir los Escenarios con la mayor cantidad posible de información extraída del LEL, aplicando la heurística de creación de Escenarios. El producto son los Escenarios candidatos derivados.

El contenido de cada impacto es analizado para encontrar en ellos, símbolos del LEL del tipo Verbo.

Si existe un símbolo Verbo en el impacto:

El objetivo es definido en base al título del Escenario y la noción del símbolo Verbo. Los actores y recursos son identificados en la información del símbolo Verbo, mirando los símbolos Sujeto y Objeto existentes en él. Los episodios son derivados de cada impacto del símbolo Verbo. La precondición del Contexto es derivada de las situaciones que impiden que la acción se realice del símbolo Verbo.

Si no existe un símbolo Verbo en el impacto: Se verifican los símbolos del impacto en cuestión y se toman como posible fuente de información. El objetivo se define en base al título del Escenario. Posibles actores y recursos son seleccionados de los símbolos en cuestión leyendo toda su descripción. Los actores son derivados de los símbolos Sujetos que le dieron origen y los recursos de los símbolos Objetos. No se derivan episodios desde el LEL, ellos serán detectados en los pasos siguientes.

En ambos casos, la ubicación geográfica y la ubicación temporal del contexto puede ser extraída de los impactos del símbolo que originó el Escenario (símbolo Sujeto). También se puede obtener información relevante de los símbolos relacionados con el símbolo Sujeto.

Al completar estos escenarios a partir de la nueva información obtenida desde el UdeD, algunos escenarios se transforman en episodios, otros se dividen en varios escenarios o se unifican en uno solo. Cabe destacar que estos escenarios candidatos actúan como un corsé en la búsqueda de nuevo conocimiento, porque el ingeniero de requisitos ya tiene identificadas las situaciones que debe analizar (lista de escenarios candidatos). El paso siguiente en el proceso, describir los escenarios, se reduce prácticamente a elicitar conocimiento para completar estos escenarios y eliminar la etiqueta de candidatos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La derivación existente tiene varias debilidades que paradójicamente, en vez de ayudar al ingeniero de requisitos a avanzar eficientemente, genera un conjunto de escenarios con problemas estructurales y potencialmente incompletos. Este último punto es particularmente relevante porque para construir el LEL se cuenta con mucho conocimiento del dominio, que no fue utilizado en la construcción del glosario. Este conocimiento existente del dominio se encuentra en documentación del contexto (manuales, protocolos, etc.) y principalmente en las transcripciones de las entrevistas realizadas para elicitar el léxico y luego, para validar. La derivación desde el LEL, solo toma el conocimiento existente en el mismo glosario, ignorando el resto del conocimiento con el que ya se cuenta. Por lo tanto, el retorno al UdeD es, en muchas ocasiones, elicitar conocimiento que ya se posee.

Puntualmente, las debilidades de la derivación existente son:

1) La generación de una lista inicial de escenarios candaditos

2) El pasaje de la información casi directa desde el LEL, sin una revisión de la misma con una visión procedural.

3) No se utiliza todo el conocimiento existente en el LEL para construir los escenarios, es decir utiliza solo algunos símbolos.

4) La heurística existente ignora los símbolos de tipo Estado, dejando un aspecto relevante del dominio sin analizar.

5) La heurística existente no contempla las jerarquías conceptuales [12] y los puntos de vista del contexto, los cuales contienen aspectos específicos del dominio.

Estas situaciones obtenidas desde el LEL son efectivamente reales del contexto, pero al ser analizadas desde una mirada procedural, pueden tener una granularidad diferente a la existente en el dominio. Esto se debe a que hay un salto conceptual entre la perspectiva propuesta en el LEL y la de los escenarios. Una de ellas observa el UdeD buscando léxico y la otra se ocupa de comprender los procesos del negocio. En los escenarios, el LEL es utilizado para reducir la ambigüedad y mejorar la precisión de las descripciones. Puede observarse que el objetivo del LEL es transversal a los escenarios, siendo difícil que se pueda obtener una mirada de los procesos con solo trasladar conocimiento de un modelo a otro, como propone la derivación actual. Debe observarse que las acciones relevantes del dominio no suscitan un tratamiento profundo. Al crear los escenarios, estas situaciones tienen la misma granularidad del LEL, heredando la dispersión de conocimiento adecuado para describir los términos de un glosario, pero no cuando se desea visualizar los procesos. En este contexto, la aparición de nuevas situaciones queda casi a merced de que el cliente-usuario las mencione o de la habilidad del ingeniero de requisitos en reconocerlas. Si bien este corsé es difícil de aflojar, no es imposible. Se puede analizar la literatura del dominio para reestructurar el orden propuesto por el LEL y comprender los procesos del negocio.

También hay que tener en cuenta que la lista de escenarios candidata es dependiente de la calidad del LEL [13] y esto puede perjudicar o beneficiar el procedimiento de derivación. Un LEL más completo permitirá identificar más situaciones del dominio, aunque no necesariamente la estructura correcta.

La presente línea de investigación propone generar un nuevo mecanismo de derivación que contemple las características propias de los modelos intervinientes, utilizando todo el conocimiento ya existente del dominio, especialmente las transcripciones de las entrevistas, eliminando toda la subjetividad que tiene la derivación actual. Este nuevo mecanismo se concentrará en incluir a todos los símbolos del LEL, sin excepciones, con un adecuado tratamiento de los estados, las jerarquías conceptuales y los puntos de vista del dominio. De esta manera el conjunto de escenarios derivados serán esencialmente más

representativos de la realidad y su completitud aumentará significativamente.

3. RESULTADOS OBTENIDOS/ESPERADOS

Con los trabajos preliminares que sirven de antecedentes al presente proyecto se han podido detectar algunas regularidades que se han transformado en hipótesis que se esperan corroborar.

La presencia de un problema en la derivación se constató empíricamente al revisar los escenarios obtenidos en varios casos reales. Esta comparación consistió en controlar el pasaje de los escenarios candidatos a los definitivos. Durante muchos años se consideró exitosa una derivación cuando existía un alto porcentaje de coincidencia. En la Tabla 1 se puede observar que los casos donde la lista final es menor a la de los candidatos, fila 1, 3 y 5, corresponden a LELs donde los impactos de los Sujetos representan acciones simples en vez de actividades. En todos los casos, sea que los escenarios se separen o unan, la coincidencia semántica entre ambas listas es muy alta. Es de destacar que los casos presentados en la Tabla 1 fueron seleccionados al azar de una lista mucho más numerosa.

Casos de Estudio	Cant. Esc. Derivados	Cant.Esc. Finales
1	16	10
2	23	25
3	21	11
4	17	18
5	28	15

Tabla 1- Diferencias entre la lista de escenarios candidatos y la lista final

Otro aspecto a observar es que la granularidad del LEL repercute en la granularidad de los escenarios e induce a un retrabajo significativo, siendo necesario juntar o separar los escenarios una vez derivados.

Una vez confirmada esta debilidad inicial de la derivación con respecto a generar una lista inicial, se pasó a analizar la semántica de los escenarios derivados con mayor profundidad, y se observó que solo se incorporaba detalle a las mismas situaciones originales. Por lo tanto, esta similitud resultó no siendo una buena indicación de la calidad del proceso. Esto fue lo que alertó sobre la posibilidad de que el problema se concentrara en la misma actividad de derivación. Se concluyó que el problema comienza con la construcción de una lista inicial de situaciones, ya que esta lista limita la elicitación de conocimiento, debido a que el ingeniero de requisitos va en busca de conocimiento con el objetivo de completar esas situaciones ya determinadas. Luego, retorna al UdeD con escenarios conformados para completarlo, y solo aparecieron nuevas situaciones en algunos casos donde existió un conocimiento previo del dominio o donde hubo un cliente-usuario en alerta.

Para asegurar que esta visión acerca de que la derivación resulta poco apropiada, se analizaron los escenarios derivados pasándolos a diferentes tipos de gráficos. En primer lugar, se realizó una red semántica donde se comprobó que los escenarios se podían mejorar incluyendo información de otros símbolos del LEL. Estos escenarios mejorados eran semánticamente similares a los de la derivación original. Esto se debe a que la red semántica utilizaba todos los símbolos del LEL pero sigue sin incorporar nueva información.

Con el uso de las redes semánticas se identificó que el problema central de los escenarios derivados es que no reflejan procesos, o lo hacen de una manera inadecuada. Esto llevó a estudiar los escenarios con representaciones graficas de procesos, como un diagrama de actividad, BPMN y otros. Se realizó la representación de los escenarios con la derivación original. Luego, a este grafico se le fue incorporando información del contexto, releyendo las transcripciones de las entrevistas y notando

una mejora sustancial en la completitud de los escenarios construidos y en la comprensión del problema. Se pudo observar en la comparación de los escenarios derivados originalmente con los mejorados, que los escenarios derivados tenían algunos problemas de secuencia y muchas omisiones de información. Como ya se mencionó, esta falta de información en el LEL es aceptable para comprender el vocabulario del dominio, pero no lo es para los escenarios.

En esta revisión de la derivación, se pudo detectar algunas omisiones y errores en el LEL. Por lo tanto, aporta una validación diferente al glosario construido, retroalimentando conocimiento.

Estos resultados preliminares han confirmado la necesidad de modificar la heurística de derivación, que tome inicialmente el conocimiento del LEL pero que luego se nutra con toda la información que se posee, principalmente de las transcripciones de las entrevistas, para construir la primera versión de los escenarios. Este mecanismo de derivación debe estar orientado a los procesos y contemplar todas las particularidades del contexto expresadas en el LEL, asegurando un conjunto de escenarios lo más completos posible que reflejen desde el comienzo, los procesos del negocio reales.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación presentada es parte directa de la tesis doctoral de la Mg. Gladys Kaplan, colabora en la tesis de maestría de la Lic. Renata Guatelli y en las tesis doctorales de la Ing. Andrea Vera y de la Lic. María Pepe.

5. BIBLIOGRAFÍA

[1] Davis, A. M (1993), Software Requirements: Objects, Functions and States, Prentice Hall, Englewood Cliffs, NJ.

[2] Elizabeth Hull, Ken Jackson, Jeremy Dick (2010), "Requirements Engineering", Springer, Third Edition, ISBN-10: 1849964041, ISBN-13: 978-1849964043.

[3] Pohl, K. (2010), "Requirements Engineering: fundamentals, principles, and techniques", Springer Publishing Company, Incorporated,.

[4] Leite, J., C., S., P., Doorn, J. H., Kaplan, G., Hadad, G., D., S., Ridao, M., N., (2004) "Defining System Context using Scenarios", in Perspectives on Software Requirements, Kluwer Academic Publishers, cap.8, pp. 169-199.

[5] Leite J.C.S.P., Franco, A.P.M., (1990) "O Uso de Hipertexto na Elicitaçao de Linguagens da Aplicaçao", Anais de IV Simpósio Brasilero de Engenharia de Software, SBC, pp. 134-149.

[6] Hadad, G.D.S., Doorn J.H., Kaplan G.N. (2008), "Creating Software System Context Glossaries", in Encyclopedia of Information Science and Technology, Second edition, Mehdi Khosrow-Pour (ed.), Idea Publishing.

[7] Leite, J. C. S., Hadad, G. D. S., Doorn, J. H., & Kaplan, G. N. (2000), "A scenario construction process", Requirements Engineering Journal, Springer-Verlag London Ltd., Vol.5, Nº1, pp. 38-61.

[8] Doorn J.H., Hadad G.D.S., Kaplan G.N. (2002), "Comprendiendo el Universo de Discurso Futuro", WER'02: Workshop de Engenharia de Requisitos, Valencia-España.

[9] Kaplan, G.N., Doorn, J.H., Gigante, N.C. (2013), "Evolución Semántica de Glosarios en los Procesos de Requisitos", XIX Congreso Argentino de Ciencias de la Computación, CACIC 13, Mar del Plata.

[10] Hadad, G., Doorn J., Kaplan G. (2009) "Explicitar Requisitos de Software usando Escenarios", WER'09: Workshop de Engenharia de Requisitos, Medellín, Colombia.

[11] Souza, R.A.C., Cysneiros Filho, G.A.A., Batista, G.H.C. (2015), "A Heuristic Approach for Supporting Innovation in Requirements Engineering", WER'15: XIX Workshop de Engenharia de Requisitos, Perú.

[12] Kaplan G. y Doorn J. (2017), "Jerarquías Naturales en el Contexto del Proceso de Requisitos", WER'17: IXX Workshop de Engenharia de Requisitos, UCA, Bs As.

[13] Ridao, M., Doorn, J.H. (2006), "Estimación de Completitud en Modelos de Requisitos Basados en Lenguaje Natural", WER'06: IX Workshop on Requirements Engineering, Brasil, pp. 151-158.

Tratamiento de entrevistas en la Ingeniería de Requisitos

María C. Elizalde[1], Jorge H. Doorn[1,2], Graciela D.S. Hadad[1]

[1]Escuela de Informática, Universidad Nacional del Oeste
[2]Universidad Nacional de Tres de Febrero
{melizalde, jdoorn, ghadad}@uno.edu.ar

RESUMEN

La mayoría de los autores de Ingeniería de Software, en general, y de Ingeniería de Requisitos, en particular, señalan la entrevista como la técnica primordial de elicitación de información durante el proceso de requisitos. Estos autores recomiendan cómo preparar la entrevista e incluso cómo llevar adelante la misma, aunque poco indican acerca de cómo se registra la información obtenida, proponiendo en algunos casos la redacción de *minutas* o resúmenes. Por otro lado, en áreas de las Ciencias Sociales las entrevistas son utilizadas como el mecanismo de abordaje por excelencia de sus trabajos de campo. Casi sin excepción, las recomendaciones respecto de la preparación y desarrollo de las entrevistas exigen la grabación o filmación de las mismas y la transcripción a texto de los audios o videos resultantes. También son altamente difundidas las recomendaciones acerca del control de calidad de dichas transcripciones y las quejas acerca del esfuerzo que requieren. La mera comparación de la diferencia entre estos dos abordajes amerita la ponderación de las ventajas e inconvenientes que emergen de incorporar la grabación o filmación de las entrevistas y su posterior des-grabación en el proceso de requisitos. En el presente proyecto se propone evaluar dichas ventajas e inconvenientes.

Palabras Clave: *Ingeniería de Requisitos, Elicitación, Entrevistas, Investigación Cualitativa, Psicología Cognitiva, Reconocimiento de Voz*

CONTEXTO

La propuesta que se presenta es parte del proyecto de investigación "Incorporación Pragmática de Visiones Lingüístico-cognitivas en el Proceso de Requisitos" de la Universidad Nacional del Oeste.

1. INTRODUCCIÓN

Varias disciplinas de las Ciencias Sociales comparten con la Ingeniería de Requisitos la adquisición de conocimiento acerca del objeto de estudio mediante el uso de las entrevistas. Sin embargo, las recomendaciones, protocolos o heurísticas que se utilizan en ambas actividades difieren radicalmente.

Por un lado, en las Ciencias Sociales se establece en forma taxativa [1] que el tratamiento posterior de la entrevista debe ser lo más objetivo posible, recomendándose la grabación o filmación de la misma para ser luego transcripta. Es así que, se suelen incluir pautas acerca de cómo evitar los efectos negativos que esa grabación o filmación pueda tener sobre el o los entrevistados [2, 3].

Por otro lado, en el contexto de la Ingeniería de Requisitos sólo se realizan sugerencias acerca de cómo seleccionar el entrevistado [4, 5] y cómo preparar la entrevista [4, 6], pero no se instala la exigencia de un procesamiento posterior lo más objetivo posible. En algunos casos se sugiere la toma de notas durante la entrevista y/o la realización de resúmenes luego de la misma [4, 6]. Eventualmente, se sugiere el uso de grabador como un complemento a la toma de notas [4]. En [7] se manifiesta la existencia de diferencias entre las notas propias y las hechas por terceros, siendo poco frecuente el uso de anotaciones de terceros. En dicho artículo también se enfatiza la diferencia entre las notas y las minutas, siendo más frecuente el uso de las notas como

fuente secundaria de información. Las conclusiones de este trabajo [7] fueron corroboradas en un trabajo posterior [8], donde se obtuvo que en general el ingeniero utiliza sus propios productos de las entrevistas, ya sea sus notas o sus minutas.

Resulta evidente que, dada la forma de procesar las entrevistas en la Ingeniería de Requisitos, el resultado neto de las mismas es doblemente subjetivo. En primer lugar, porque las notas o los resúmenes estarán influenciados por las experiencias previas, tanto personales como profesionales del entrevistador. En segundo lugar, cuando utilice sus notas o resúmenes, complementará las mismas con sus recuerdos, agregando nuevas subjetividades. Algunos autores recomiendan que un tercero o un escriba tome notas [4]. A menos que este escriba sea un taquígrafo, su introducción puede ser una fuente de nuevas subjetividades ya que la toma de notas siempre involucra juicios de valor acerca de qué registrar y qué omitir.

Por su parte, en las Ciencias Sociales se indica que toda elaboración, conclusión u observación que se realice sobre el objeto de estudio debe estar sustentada específicamente por un componente preciso de la entrevista [9]. Esto es muy similar a lo que se pregona, pero no siempre se lleva a cabo, en la Ingeniería de Software acerca de la rastreabilidad entre los distintos artefactos del proceso [10]. En otras palabras, esto implica que los componentes de los modelos de requisitos deben ser rastreables a las correspondientes partes del texto de la entrevista.

La incorporación de grabaciones o filmaciones en el procesamiento de las entrevistas no es para nada gratuita, ya que la des-grabación es una actividad tediosa, que requiere un gran esfuerzo. En [11] se comprobó que la tarea de transcripción puramente manual requiere entre 6 y 7 horas por cada hora de grabación. Puntualmente, en el análisis de la conversación, los lingüistas dan pautas bien definidas en cuanto a cómo realizar la transcripción manual, considerando no solo lo dicho sino la forma en que se habla, denominados datos contextuales no lingüísticos, tales como gestos, movimientos, cambios en tonos de voz y pausas, entre otros [3]. En el dominio de las Ciencias Sociales, se presupone que la des-grabación la realiza el propio entrevistador [12]. Este es uno de los aspectos centrales en el presente proyecto, ya que se aspira a reducir al mínimo el esfuerzo de des-grabación mediante el uso de técnicas de reconocimiento automática de la voz (speech to text) [13].

En resumen, se apunta a evaluar si las recomendaciones para realizar y procesar entrevistas en los trabajos de campo de las Ciencias Sociales son aplicables en la Ingeniería de Requisitos para adquirir un conocimiento más acabado del objeto de estudio, en comparación con las recomendaciones actuales en la literatura. Es decir, se plantea la posibilidad de lograr una adquisición de conocimiento más eficaz y eficiente en las entrevistas realizadas en un proceso de construcción de requisitos.

2. LINEAS DE INVESTIGACIÓN y DESARROLLO

A partir de la forma en que se realice la entrevista en un proceso de requisitos respecto del soporte elegido para registro de la información obtenida y del tratamiento posterior de la información que se obtenga a partir de la misma se pueden plantear diferentes caminos (ver Fig. 1).

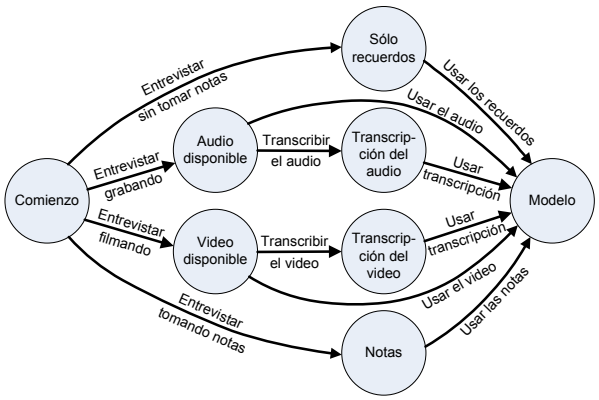


Fig.1. Tratamientos posibles de las entrevistas

Ignorando la combinación de recursos como

podría ser, por ejemplo, grabar y tomar notas o filmar y tomar notas, se pueden distinguir 6 posibles caminos:

1: Entrevistar sin tomar notas y construir el modelo basándose sólo en los recuerdos.

2: Entrevistar grabando y construir el modelo usando el audio.

3: Entrevistar grabando, transcribir la grabación y construir el modelo usando las transcripciones.

4: Entrevistar filmando y construir el modelo usando el video.

5: Entrevistar filmando, transcribir el video y construir el modelo usando las transcripciones.

6: Entrevistar tomando notas y construir el modelo usando las notas y los recuerdos.

Indudablemente el grado de objetividad de los caminos 2, 3, 4 y 5 es superior al de los caminos 1 y 6. Una consecuencia directa de esto es que los caminos 1 y 6 deben ser realizados por la misma persona, mientras que los caminos 2, 3, 4 y 5 admiten que haya una división de tareas. Es así que, en los caminos 2 y 4, podrían participar dos personas y en los caminos 3 y 5 hasta tres personas,

Siguiendo el camino 1, el consejo sería apresurar la construcción del modelo, mientras se tienen aún los recuerdos de lo captado durante la entrevista. En este punto es importante considerar los aspectos cognitivos de los involucrados, ya que es posible que, debido a los propios procesos de percepción, atención y memoria [14], se altere la información realmente recibida en función de los recuerdos que se tienen. Sin embargo, el camino 6 puede adolecer de los mismos inconvenientes, al llevar incluso a editar las notas en función de los recuerdos [4].

En forma preliminar se puede afirmar que existen ventajas e inconvenientes en la posible opción entre la participación o no de más de una persona. Por un lado, es razonable suponer que el grado de aprendizaje que se logra durante la entrevista y eventualmente durante la transcripción incremente la capacidad de percepción del contexto de aplicación y, por tanto, se mejore la elaboración de información. Sin embargo, este aprendizaje e internalización de información puede teñir de subjetividad la tarea final de construcción del modelo. No existe a la vista nada que permita sospechar a priori cuál es el efecto dominante en esta situación.

A partir de estos elementos surge la necesidad de analizar las alternativas posibles para determinar cuál o cuáles son las más apropiadas y si existe alguna interacción significativa con el propósito, la duración y el tipo de entrevista. Con respecto a esto último, la respuesta es inmediatamente afirmativa, ya que en varias instancias del Proceso de Ingeniería de Requisitos resulta necesario realizar consultas de menor cuantía, las que formalmente pueden materializarse mediante entrevistas. Esta sobrecarga de la palabra entrevista hace necesario restringir el uso de la misma, en el contexto de estas consideraciones, a situaciones donde el grado de conocimiento previo del entrevistador responsable es bajo o medianamente bajo y el volumen de información que se espera obtener es significativo.

Dos conclusiones surgen de lo anterior: i) el uso de notas como técnica aislada o el no disponer de ningún registro de la entrevista es indudablemente menos provechoso que las restantes alternativas, y ii) las consideraciones acerca de las grabaciones o las filmaciones son casi coincidentes, con algunas ventajas por parte de estas últimas.

En este punto es necesario abordar la comparación entre el camino 2 y el camino 3 o entre el camino 4 y el camino 5. Aquí es necesario ponderar el impacto de las transcripciones del material de audio o video sobre la calidad de la información obtenida. Nuevamente, existe una corriente de opinión muy fuerte en las Ciencias Sociales [12, 15, 16, 17] enfatizando la notable importancia de realizar efectivamente las transcripciones como mecanismo ineludible para proveer un grado de objetividad alto en los resultados obtenidos. Las pocas voces en disidencia no rechazan frontalmente la realización de las transcripciones, sino que proponen formas de trabajo mixtas que garantizarían una calidad

similar a la transcripción en sí propia. Por ejemplo, en [18] se propone el enriquecimiento de las notas tomadas durante la entrevista con las grabaciones, procurando llevar a cabo la combinación de ambos registros tan pronto como sea posible. Desafortunadamente, no se han reportado estudios que comprueben la eficacia de este enfoque.

Notablemente, la bibliografía de referencia en Ingeniería de Requisitos [4, 6, 7, 8, 19] coincide en que la fuente de información más importante son las personas y que la técnica de elicitación de uso más frecuente es la entrevista, sin hacer consideración alguna acerca de los registros fieles de las mismas y menos aún de la importancia de las transcripciones. Es más, se puede mencionar al menos un autor que recomienda no grabar las entrevistas ya que muestra habilidades deficientes para escuchar [6].

En esta instancia, aun adhiriendo a la estrategia de obtener registros fieles de las entrevistas, surge la cuestión de si el camino 3 o el camino 5 deben ser llevados a cabo por el mismo actor o es posible considerar aceptable o beneficiosa la participación de otras personas, y si además es una alternativa viable el uso de técnicas automáticas de transcripción.

Podría ocurrir que la secuencia en ambos caminos sea ejecutada por tres personas diferentes. En las Ciencias Sociales, siempre la primera actividad y la tercera actividad son realizadas por la misma persona. En lo que se refiere a la transcripción de las entrevistas, no se enfatiza acerca de quien las debe realizar, aunque abundan las menciones acerca que dicha tarea es un mal necesario que retrasa significativamente el proceso de las entrevistas. Esto sugiere que en una gran cantidad de casos las transcripciones son también realizadas por el entrevistador [2]. En el contexto de la Ingeniería de Requisitos el esfuerzo de realizar transcripciones resultaría también un inconveniente significativo.

En las Ciencias Sociales, el proceso de extracción de información está basado en el procesamiento del texto resultante de la transcripción, y toda información que de alguna manera hubiera sido percibida durante la entrevista o la transcripción sólo adquirirá el estatus de información aceptable cuando se la ubique en la transcripción [1]. Como ya se mencionó, esto es coincidente con la construcción de rastros, actividad importante en el proceso de requisitos y en el proceso de desarrollo de software como un todo. En [7] se menciona la dificultad en disponer de una pre-rastreabilidad adecuada de los requisitos debido a la carencia de rastros hacia las fuentes de información, básicamente esto puede solucionarse justamente estableciendo esos rastros hacia las transcripciones de las entrevistas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Todo lo anterior sugiere que la forma de mejorar la elicitación de información a partir de las entrevistas consistiría en aprovechar las ventajas de la transcripción automática. Una mera observación superficial de las consecuencias de su uso permite destacar tres aspectos relevantes: i) por un lado la transcripción automática difícilmente podrá identificar cada hablante a lo largo de la entrevista, ii) aun en el mejor de los casos la transcripción contendrá un número significativo de errores, especialmente cuando ocurran superposiciones en la conversación o interrupciones, y iii) no habrá registro alguno de aspectos cualitativos importantes como afirmaciones enfáticas, frases dubitativas o pausas. En otras palabras, la transcripción automática requiere una intervención humana posterior.

Entonces, la cuestión a resolver es si la revisión de una transcripción automática ofrece ventajas sobre una transcripción puramente manual.

Transcribir manualmente una grabación o una filmación, ya sea usando un grabador convencional o un computador, involucra una cantidad muy grande de paradas y retrocesos que constituyen uno de los aspectos más desagradables de la tarea. En cambio, en una revisión de una transcripción automática de

calidad razonable es de esperar que la cantidad de estas maniobras se reduzca significativamente.

El objetivo del presente proyecto es analizar el impacto de la inclusión de transcripciones automáticas de entrevistas en la obtención de información durante el proceso de elicitación de información, como así también el beneficio o el perjuicio por la intervención de distintas personas en este proceso.

4. FORMACIÓN DE RECURSOS HUMANOS

En el presente proyecto participan tres investigadores, uno de ellos en formación, un becario de grado y tres alumnos colaboradores.

5. BIBLIOGRAFIA

[1] Edwards, R., Holland, J., "What is qualitative interviewing?", Blomomsbury Academic, 2013, Londres, pág. 43 a 52.

[2] Roulston, K., deMarrais, K., Lewis, J.B., "Learning to Interview in the Social Sciences", Qualitative Inquiry, Vol. 9, Nro. 4, 2003, pp. 643-668.

[3] Tusón Valls, A., "El análisis de la conversación: entre la estructura y el sentido", Estudios de Sociolingüística, Vol. 3, Nro. 1, 2002, pp. 133-153.

[4] Wiegers, K., Beatty, J., "Software Requirements", 3º edición, Microsoft Press, Washington, 2013.

[5] Leite, J.C.S.P., Moraes, E.A., Castro, C., "A Strategy for Information Sources Identification", 10th Workshop on Requirements Engineering, Toronto, Canadá, Mayo 2007, pp.25-34.

[6] Whitten, J., Bentley, L., "Systems Analysis and Design Methods", 7º edición, Mc Graw-Hill Education, Nueva York, capítulo 5, 2005.

[7] Antonelli, L., Oliveros, A., "Fuentes Utilizadas por desarrolladores de Software en Argentina para Elicitar Requerimientos", 5th Workshop on Requirements Engineering, Valencia, España, Noviembre 2002, pp. 106-116.

[8] Antonelli, L., Oliveros, A., "Técnicas de elicitación de Requerimientos", XXI Congreso Argentino de Ciencias de la Computación, Buenos Aires, 2015.

[9] Buriro, A.G., Awan, J., Lanjwani, A.R., "Interview: A Research Instrument for Social Science Researchers", International Journal of Social Sciences, Humanities and Education, Vol. 1, 2017, pp. 1 a 14.

[10] Pinheiro, F., "Requirements Traceability". En: Perspectives on Software Requirements, Kluwer Academic Publishers, cap. 5, 2004, pp. 91-113.

[11] Britten, N., "Qualitative research: Qualitative interviews in medical research", British Medical Journal, Vol. 31, 1995, pp. 251-253.

[12] Roulston, K., "Reflective interviewing: A guide to theory and practice", Sage Publications, Londres, 2010.

[13] Rabiner, L.R., Juang, B.H., "Speech Recognition by Machine". En: The Digital Signal Processing Handbook, CRC Press, IEEE Press, cap. 47, 1998.

[14] Sternberg, R.J., Sternberg, K., "Cognitve Psicology", 6° edición, CENGAGE Learning, Boston, 2009, pp. 185-227.

[15] Gillham, B., "Research Interviewing: The range of techniques: A practical guide", Open University Press, New York, 2005.

[16] Dortins, E., "Reflections on phenomenographic process: Interview, transcription and analysis", Higher Education Research and Development Society of Australasia, 2002, pp. 207-213.

[17] Zowghi, D., Coulin, C., "Requirements Elicitation: A Survey of Techniques, Approaches, and Tools". En: Engineering and Managing Software Requirements, Springer, cap. 2, 2005, pp. 19-34.

[18] Halcomb, E.J., Davidson, P.M., "Is Verbatim Transcription of Interview Data Always Necessary?", Applied Nursing Research, Vol. 19, pp. 38-42, 2006.

[19] Carrizo, D., Dieste, O., Juristo, N., "Study of elicitation techniques adequacy", 11th Workshop on Requirements Engineering, Barcelona, España, 2008, pp. 104-114.

Equipos de trabajo 4.0: Nuevas configuraciones

Germán Gaona[2], Paula Lima[2], Verónica A. Bollati[1]
(germanxgaona, limapaulabelen, vbollati)@gmail.com

[1]Facultad Regional Resistencia
Universidad Tecnológica Nacional – CONICET
Resistencia, Chaco - Argentina

[2]Facultad Regional Resistencia
Universidad Tecnológica Nacional
Resistencia, Chaco - Argentina

RESUMEN

Este trabajo presenta las líneas de investigación del proyecto SIUTNRE00065 80: "Equipos de trabajo 4.0: Nuevas configuraciones", acreditado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional (UTN) (Resol. Cons. Dir. 148/2019). Los integrantes pertenecen a la UTN - Facultad Regional Resistencia (UTN-FRRe), al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y a la Universidad Rey Juan Carlos de España.

El objetivo fundamental es caracterizar nuevas configuraciones de trabajo en equipos de desarrollo de software que se ajusten al contexto del mercado de trabajo actual.

Palabras clave: *soft skills, management, gig economy, generaciones, personas, empresas*

CONTEXTO

El proyecto pretende, en primer lugar, determinar los factores que influyen en la manera de trabajar de los equipos de desarrollo de software (SW), específicamente en organizaciones públicas y privadas del Nordeste Argentino (NEA), analizando su impacto en su productividad con el objetivo de proponer nuevas estrategias que permitan maximizar la productividad de los equipos.

Además, continúa la investigación realizada en el Proyecto "Desarrollo de Software Dirigido por Modelos en Entornos Ágiles" por investigadores pertenecientes al Centro de Investigación Aplicada en Tecnologías de la Información y las Comunicaciones (CInApTIC), el cual tenía por objetivo aplicar los principios de la Ingeniería Dirigida por Modelos (*Model Driven Engineering*, MDE) al modelado de procesos propuestos, y las herramientas que los soporten, en diferentes técnicas de desarrollo de software ágiles, en vistas a definir de un *framework* que facilite la adopción de prácticas ágiles por parte de las PyMEs del NEA.

1. INTRODUCCIÓN

En plena revolución digital, el inicio del éxito se origina en el conocimiento, esto generó que, en los últimos años, la economía del conocimiento (EC) adquiera mayor importancia ya que no está sujeta ni a las crisis, ni a ciclos económicos, como tampoco lo está a las minorías económicas [1]. La EC, es el sector en el que se emplea el conocimiento o la información para generar resultados, valores, mejoras y utilidades. Todo ello con el propósito de producir un cambio positivo en la

vida de las personas, creando valor añadido a los productos y servicios en los que participa ya sea para su creación o su transformación, generando un cambio estructural en las actividades económicas, sociales y políticas. Reconfigurando así, las relaciones sociales de producción y distribución en el mundo entero.
En la actualidad, la Gestión del Conocimiento (GC) es primordial para el desarrollo de nuevas tecnologías. Con la masificación de internet, el ritmo de producción de conocimiento es mayor que nunca, por lo que no es fácil asimilar toda la información a la que nos exponemos.
El debate que acompaña a la GC trae asociada, casi por necesidad, la discusión acerca del papel de los Equipos de Trabajo (ET) en organizaciones donde se trabaja en base a la gestión de ese recurso. La formación de ET está directamente relacionada con la misma esencia de la GC pues existe la necesidad de compartir conocimientos y colaborar entre los empleados de una organización para potenciar el resultado final [2], especialmente en desarrollo de SW.
Desde la firma del manifiesto ágil en el 2001 [3], el ámbito del desarrollo del SW ha ido adaptándose a nuevas maneras de trabajar. Uno de los principales cambios fue el organizacional, donde las personas, y el equipo que forman, adquieren mayor protagonismo. La primera vez que se comienza a hablar del tema fue en el año 1987 en el libro de "*Peopleware: Productive projects and teams*" [4], donde se popularizaron estudios que hablaban de la importancia del papel humano en el desarrollo del SW. Se planteaba que una de las claves principales para el éxito o fracaso de un proyecto son las personas (y los equipos que conforman), su motivación, las habilidades personales y las relaciones entre las mismas que impactan en el resultado final [5].
Desde el nacimiento de la agilidad han surgido diferentes prácticas para mejorar la manera de trabajar. De aquellas que se enfocan en las personas, una de las que más adeptos ha logrado es la del Management 3.0 [6], mediante la cual se pasa de considerar a los trabajadores del conocimiento como simples recursos, a considerarlos como la clave del éxito de un proyecto [7] y el componente no lineal de primer orden en el desarrollo del SW [8]. El Management 3.0 se contrapone a la visión clásica de gestionar (Management 1.0) y propone una visión orientada a las personas y equipos en estructuras organizativas pensadas para ello. Aglutina a un conjunto de ideas y "prácticas" para potenciar a las personas y los equipos con el objetivo de incrementar la productividad de estos.
En la última década han surgido diversos factores que influyen en la manera de trabajar y que afectan a la productividad de los equipos de desarrollo. Por un lado, el auge de la economía digital está cuestionando el concepto de empleo, la inteligencia artificial y la robótica no solo redefinirán los tipos de empleo sino la forma de trabajar, con entornos laborales más interactivos, colaborativos y simplificados [9]. Esta automatización está redefiniendo el conjunto de habilidades requeridas, entre las que destacan las *soft skills* (habilidades emocionales o interpersonales) y las competencias transversales.
Existe un proceso paralelo y estrechamente unido a la aplicación de las nuevas tecnologías: la reingeniería empresarial, un cambio en la cultura organizativa para utilizar las nuevas tecnologías mejorando y simplificando la forma en la que los empleados interactúan entre si y con la información. La tecnología no sólo ha cambiado la composición del empleo

por ocupaciones y cualificaciones, también la intermediación laboral y cómo se relacionan los trabajadores y las empresas, con nuevas figuras de empleo como los autónomos dependientes, los *freelancers*, o la "uberización" del empleo. Además, las plataformas digitales están contribuyendo al auge de la *gig economy* (GE) por lo que trabajadores han evolucionado hacia empleos flexibles basados en proyectos y las empresas tienen cada vez más acceso a un mercado mundial.

La *GE* nació en Estados Unidos hace algo más de una década [10]. Surge como una alternativa a la contratación tradicional, donde la flexibilidad y la comunicación *'online'* son los pilares básicos del funcionamiento. La deslocalización, es decir, trabajar para un empleador que se encuentra a kilómetros es otra de las características de la *GE* que es posible gracias a los avances tecnológicos.

En los últimos años, las empresas basadas en la *GE* han ido en aumento. McKinsey [11] revela que entre el 20 y el 30% de la población activa en EE. UU. y Europa participa en la *GE* en diferentes grados.

Este nuevo modelo laboral tiene muchas ventajas para empleadores y para empleados, aunque estos podrían también verse en alguna situación de inestabilidad laboral al depender exclusivamente de proyectos puntuales que encajen con sus habilidades. En épocas pasadas, los trabajadores tenían un ingreso estable con un sueldo a tiempo completo. En la actualidad, este tipo de trabajos son cada vez más escasos en campos como el desarrollo de SW, donde cada vez son más las compañías que ofrecen contratos por proyectos o por un tiempo definido [12].

La *GE* está cobrando mucha importancia en el ámbito del desarrollo del SW. Las empresas están dejando de contratar programadores para puestos permanentes y en su lugar, están comenzando a trabajar con equipos temporales de programadores independientes de cualquier parte del mundo [13], esto trae algunas desventajas obvias, como el hecho de perder la comunicación, por la deslocalización, con el resto del equipo, no lograr la madurez del equipo por el hecho de no trabajar juntos en el tiempo, entre otras cosas.

Estos cambios en el mundo laboral sumado a la prolongación de la expectativa de vida y la gente buscando el disfrute en lo que hace son factores que determinan que haya cada vez más personas iniciando o transitando su proceso de reinvención [14]. La reconversión laboral es un proceso de redefinición de la identidad. En el siglo XX, el trabajo era entendido como sacrificio, en el siglo XXI el trabajo es entendido como disfrute y plenitud, estamos en el siglo del bienestar, de la democratización del bienestar, donde la idea es encontrar el placer en el deber [15]. Una de las razones de esto, es la coexistencia de las diferentes generaciones compartiendo los mismos espacios de trabajo. La generación conocida como *baby boomers* entendía la dedicación al trabajo como el tiempo que estaba en el trabajo, su identidad se construía en ese ámbito, en tanto que, en el modelo actual, impulsado por la generación *millennial*, la dedicación está relacionada con los resultados que se obtienen y la identidad se construye en base a la vida personal, los gustos y el trabajo. Los *millenials* buscan encontrar en el trabajo mayor integración con su proyecto de vida.

2. LÍNEAS DE INVESTIGACIÓN

El objetivo general del Proyecto es "caracterizar nuevas configuraciones de trabajo en equipos de desarrollo de SW que se

ajusten al contexto del mercado de trabajo actual".

En cuanto a las líneas de investigación abordadas en este proyecto se pueden enunciar dos líneas. En primer lugar, desde el punto de vista de la agilidad, se propone el análisis de las diferentes prácticas propuestas por el Management 3.0 con el objetivo de determinar cómo las mismas pueden ser adaptadas a las nuevas maneras de trabajar de los equipos de desarrollo, teniendo en cuenta el contexto en el que se encuentran inmensos.

Por el otro lado, desde el punto de vista de la Economía 4.0, el análisis de los factores sociales, culturales y económicos y cómo impactan en la productividad de los equipos de desarrollo de SW.

3. RESULTADOS ESPERADOS

Este proyecto busca aportar a un área específica de la Ingeniería del Software (IS), como es el desarrollo de SW mediante una caracterización de los ET teniendo en cuenta los factores económicos, sociales y culturales que afectan a la productividad de estos.

En el contexto universitario, la concreción de los objetivos del proyecto permitirá el desarrollo de actividades de investigación en un área considerada de vacancia en la región, y permitirá la inserción y el crecimiento en el camino de la investigación de alumnos y recientes graduados de las carreras de Informática. Contribuyendo a mejorar la formación de los profesionales de Informática dado que la generación de nuevos conocimientos y la actualización resultante de las tareas previstas en el proyecto, se trasladarán a los alumnos a través de las cátedras y capacitaciones que se dictan en las carreras de la disciplina involucradas. Por otro lado, el trabajo con empresas de desarrollo de SW favorecerá la formación de Recursos Humanos insertos en la industria local.

En este sentido, es importante mencionar, que en la región se encuentran consolidados 2 Polos Informáticos: el Polo IT Chaco desde el año 2005 y el Polo IT Corrientes desde el 2007. Ambos nuclean un total de 44 empresas PyMEs de SW. Estas empresas serán las destinatarias de los resultados del proyecto por medio de la adopción de los mismos, permitiendo que aumenten su productividad, lo que impactará en su competitividad y generará valor para la zona mediante la creación de puestos de trabajo y el mejoramiento de las condiciones laborales, lo que a su vez posibilitará la inserción de los profesionales formados en la universidad, evitando el desarraigo de los jóvenes graduados y la migración a los grandes centros urbanos, cambios que se dan por la falta de oportunidades de desarrollo profesional, situación que afecta la calidad de vida de las personas.

Los mecanismos para promover la adopción de los resultados por parte de las empresas que se proponen son: capacitación por medio de workshops y talleres a los directivos y personal de las empresas, demostraciones de las propuestas para que las empresas las conozcan y analicen su adopción, trabajo *in situ* en algunas empresas para colaborar en el proceso de adopción.

4. FORMACIÓN DE RECURSOS HUMANOS

La formación de recursos humanos es un aspecto prioritario del proyecto, por lo que se generarán espacios para promover el desarrollo de Prácticas Supervisadas de la carrera Ingeniería en Sistemas de Información (ISI), proyectos de fin de carrera, Trabajos Finales y Tesis de las carreras de Posgrado que

se dictan en la región, especialmente la Especialización en Ingeniería Gerencial, la Maestría en Administración de Negocios y, a partir de 2020, del Doctorado en Informática recientemente aprobado, a dictarse con las Universidades Nacionales del Nordeste y de Misiones. También Trabajos Finales y Tesis de Posgrado de docentes y graduados de ISI que cursen sus estudios de postgrado en otras Universidades.

El equipo de trabajo del CInApTIC de UTN - FRRe está integrado por la Directora (Doctora, Categorizada como Docente Investigadora de UTN), una Co-Directora (Magister, Categorizada en el Programa de Incentivos y como Docente Investigadora de UTN), dos Docentes Investigadores de Apoyo (Ingenieros, Categorizados como Docentes Investigadoras de UTN) y un Investigador Formado Externo (Doctor, España). Se debe mencionar que, actualmente existen dos Tesis de posgrado radicadas en el CInApTIC.

Además, el equipo no sólo incluye investigadores formados, sino que incorpora a jóvenes investigadores, algunos de los cuales colaboran como docentes en diversas cátedras y otros se encuentran insertos laboralmente en PyMEs de SW del NEA. Se cuenta, así, con un Becario BINID, y dos Becarios alumnos de Rectorado, de los cuales, una realiza su Práctica Supervisada en este proyecto y un Becario alumno de Investigación.

AGRADECIMIENTOS

Este trabajo ha sido financiado en forma conjunta por CONICET y la UTN. Además, ha sido parcialmente financiado por el Ministerio de Economía y Empresa del Gobierno de España bajo el proyecto MADRID (TIN2017-88557-R)

5. BIBLIOGRAFÍA

1.Fernández, H. “La economía del conocimiento y su importancia en la competitividad de un país” Disp. en: https://economiatic.com/economia-del-conocimiento/.

2. García Fernández, F., Cordero Borjas, A. E. (2008). “La gestión del conocimiento y los equipos de trabajo: fundamentos teóricos”. U. Autónoma de Tamaulipas, México. U. de Carabobo, Venezuela.

3. Beck, K. (2001). “Manifesto for Agile Software Development,” The Agile Manifesto. Disp. en: http://agilemanifesto.org/.

4. DeMarco, T., Lister, T. R. (1999). “Peopleware: productive projects and teams”. Dorset House Pub.

5. Garzás Parra, J. (2018). “Peopleware y Equipos Ágiles”. 233 Grados de TI.

6. Appelo, J. (2011) “Management 3.0: leading Agile developers, developing Agile leaders”. Addison-Wesley

7. Davis, M. A. (1995). “201 principles of software development”.McGraw-Hill Inc. U.S

8. Cockburn, A. (2000) “Characterizing people as non-linear, first-order components in software development”. 4th Int. Multi-conference on Systems, Cybernetics and Informatics. Orlando, Florida

9. Melle, M. (2018). “Los retos del empleo en la economía 4.0”. Disp. en: https://cadenaser.com /ser/2018/11/20/economia/1542717537_089946.html.

10. Otero, M. (2018). "Qué es la gig economy?”.Disp. en: https://www.bbva.com/es/que-es-la-gig-economy/.

11. (2016). “Independent Work: Choice, Necessity, and the Gig Economy. Mckinsey Global Institute”. Disp.en:https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Employment%20and%20Growth/Independent%20work%20Choice%20necessity%20and%20the%20gig%20economy/Independent-Work-Choice-necessity-and-the-gig-economy-Full-report.ashx.

12. Sykes, N. (2018). “How Programmers Can Survive the Gig Economy”. Disp. en: https://simpleprogrammer.com/programmers-gig-economy/.

13. Warne. H. (2017). “Software Development and the Gig Economy”. Disp.: https://henrikwarne.com/2017/01/22/software-development-and-the-gig-economy/.

14. Molinari. P. (2017) “Transitando la ReInvención”. Disp. en: http://www.pmolinari.com/2017/06/06/transitando-la-reinvencion/.

15. Molinari, P. “Tu trabajo ideal”. Disp. en: http://www.pmolinari.com/videos/.

EVALUACIÓN DEL IMPACTO DE LAS EMOCIONES EN LA CALIDAD DE SOFTWARE DESDE EL PUNTO DE VISTA DEL USUARIO

Cesar J. Acuña, Noelia Pinto, Gabriela Tomaselli, Nicolás Tortosa
Centro de Investigación Aplicada a las Tecnologías de la Información y Comunicaciones (CInApTIC) Facultad Regional Resistencia – Universidad Tecnológica Nacional
French 414, Resistencia, Chaco
{csr.acn, ns.pinto, gabriela.tomaselli. nicotortosa}@gmail.com

RESUMEN

En este trabajo se describen las líneas de investigación y desarrollo que se realizarán en el marco del proyecto referido en la sección de contexto, pero que, a su vez, forman parte de una línea de investigación más amplia y madura dentro del Centro de Investigación CInApTiC, relacionada a la Ingeniería y Calidad de Software orientado a Pymes de la región NEA.

En proyectos anteriores de este Centro de Investigación se ha abordado el desarrollo de un Framework para la Evaluación de la Calidad de Procesos Agiles que permita a las empresas de software de la región, evaluar a través de un modelo de calidad especifico y una herramienta que lo gestiona, la calidad de sus procesos ágiles. Sin embargo, existen aspectos relacionados con el aseguramiento de la calidad que aún deben ser tratados. Uno de ellos es la relación que existe entre factores como experiencia de usuario y el impacto de las emociones en la calidad percibida por parte del usuario al interactuar con el software. El objetivo principal del proyecto que aquí se describe es diseñar, desarrollar e implementar una propuesta de medición y evaluación del impacto de las emociones en la calidad de software percibida por el usuario.

CONTEXTO

Las líneas de I/D presentadas en este trabajo forman parte de las actividades definidas en el marco del proyecto denominado “Evaluación del Impacto de la Emociones en la Calidad de Software desde el punto de vista del usuario”, proyecto de Investigación homologado y financiado por la Secretaria de Ciencia y Técnica de la Universidad Tecnológica Nacional (UTN), bajo el código SIUTIRE0005517TC (2020-2022), el cual se desarrolla en conjunto con el resto de proyectos de investigación pertenecientes Centro de Investigación Aplicada a las Tecnologías de la Información y Comunicaciones (CInApTIC), que reviste las características de Centro de Investigación UTN y que funciona en dependencias de la Facultad Regional Resistencia. Además, este proyecto, cuenta con el patrocinio de las empresas ESSENTIT Data Engineering y 42Mate pertenecientes al Polo de Empresas Tecnológicas del Chaco.

1. INTRODUCCION

La gestión de la calidad en aquellas organizaciones dedicadas al desarrollo de proyectos de software ofrece una ventaja competitiva puesto que de esta forma aseguran que sus productos y servicios cuenten con estándares mínimos de calidad en el mercado.

Así, en la búsqueda de la mejora continua y enfocados en asegurar la calidad del software, las organizaciones, ante la naturaleza cambiante de los entornos y la necesidad de un rápido retorno de inversión, optan por enfoques ágiles de desarrollo de software.

A pesar de las ventajas que ofrece el desarrollo de software bajo un enfoque ágil, existen aspectos relacionados con el aseguramiento de la calidad que aún deben ser tratados. Uno de ellos es la relación que existe entre factores como experiencia de usuario y el impacto de las emociones en la calidad percibida desde el usuario al interactuar con el software.

La Experiencia del Usuario (UX) es definida por Arhippainen y Täh [1] sencillamente como la experiencia que obtiene el usuario cuando interactúa con un producto en condiciones particulares. Muchas veces esta experiencia se

ve afectada por emociones del usuario, por lo que el análisis de ambos factores combinados es fundamental para analizar el impacto de cierto producto, servicio, etc. en el mercado.
Diversos estudios han demostrado que las emociones se producen por la interacción de una respuesta fisiológica ante un estímulo, que produce cambios en el sistema biológico de un individuo generando un cambio emocional [2]. Así, tal como lo expresaran Paul Ekman y Harrieh Oster en [3], las expresiones faciales pueden facilitar la identificación de emociones humanas, a través no solo de la observación de la expresión sino de la medición de la actividad facial resultante ante diversos estímulos. Además, si consideramos a los avances tecnológicos y la introducción de sistemas interactivos a la cotidianeidad, se puede observar cómo se ha incrementado la demanda de interfaces más "naturalmente amigables" para comunicarse con los dispositivos electrónicos. Se debe considerar que la computación afectiva y la interacción inteligente son una tecnología emergente clave que se enfoca en innumerables aspectos del reconocimiento, la comprensión y la expresión de los estados afectivos y emocionales de los dispositivos electrónicos [4,5].
Además, si tenemos en cuenta que desde hace años, las investigaciones científicas sobre el comportamiento del ser humano se enfocan en estudiar solamente la parte racional del mismo, dejando en segundo plano las emociones [6]. Sin embargo, en la actualidad, y cada vez con más frecuencia, los estudios han comenzado a considerar fuertemente la influencia que las emociones tienen sobre los procesos de pensamiento racional, y cómo éstas afectan en gran medida el proceso de toma de decisiones en los seres humanos. Relacionado a ésto, toma fuerza el término de Computación Afectiva, acuñado por primera vez gracias a la Dra. Rosalind Picard del MIT [7], en su libro "Affective Computing", en el mismo que menciona la necesidad de tomar en cuenta los estados emocionales de los usuarios a la hora de desarrollar software.
En [8], los autores del presente artículo realizan una revisión sistemática sobre el impacto de las emociones del usuario en la percepción de la calidad del software. Los resultados de esta investigación refuerzan y ponen de manifiesto la relevancia del análisis de la interacción emocional de los usuarios en la evaluación de calidad del software, demostrando, a partir del estudio empírico de diversos casos, que ya no basta solo con estudiar características del proceso de desarrollo y del producto final para medir la calidad del software.
Además, si bien este estudio demuestra la evidencia de iniciativas que aborden las emociones de los usuarios y la calidad del software concluye que, el conocimiento existente no aporta propuestas que permitan la definición de modelos o estrategias que evalúen el impacto de las emociones en la calidad de software percibida por el usuario.

Por tanto, el desafío de este proyecto será, entonces, el diseño de proceso de evaluación de calidad del software que contemple las emociones y la interacción del usuario con el producto final. Para ello se trabajará en el desarrollo de un modelo y herramientas que permitan la automatización al evaluar la calidad de software según el impacto de las emociones de usuarios finales.
Finalmente, resulta necesario que la evaluación se realice desde el enfoque de las características del proceso ágil, inclusive, de forma tal de recomendar mejoras tanto a nivel de producto de software como del proceso de desarrollo para su obtención.

2. LINEAS DE INVESTIGACION y DESARROLLO

Las líneas de I+D de este proyecto parten de y, a su vez, complementan esfuerzos y resultados de investigación en proyectos de investigación previos, principalmente del proyecto predecesor denominado "Un Modelo de Evaluación de la Calidad de Procesos Agiles de Desarrollo de Software" (2017-2019). Cuyos principales resultados de investigación componen un marco de trabajo denominado AQF (Agile Quality Framework) [9] que incluye, por un lado, modelo de calidad de software para procesos agiles de desarrollo, denominado QuAM (Quality Agile Model) [J,K] y, por el otro lado, incluye desarrollo de una plataforma tecnológica que, a través de una

herramienta de software denominada QuAGI [10] permite gestionar los elementos del modelo de calidad y evaluar fácilmente la calidad del proceso ágil asociado.

Centrándonos ahora, en el proyecto que aquí se presenta, la principal línea de investigación se enfoca en el análisis de factores emocionales de grupos de personas y en la toma de decisiones por software basada en la misma. En particular se realizarán estudios sobre grupos de usuarios de software en distintos escenarios, a fin de determinar su estado emocional al usar el producto, evaluar el impacto en la calidad percibida luego y ejecutar acciones que permitan mejorar la calidad del proceso de desarrollo y del producto de software.

Para la detección y posterior análisis de las emociones se utilizarán métodos automáticos y semiautomáticos tales como: reconocimiento de expresiones faciales, análisis de audio y variables fono-métricas, etc. Y se complementarán con métodos no automáticos supervisados, tales como observación de protocolos de uso del software y emociones emergentes; y no supervisados, tales como, encuestas de uso. Cabe aclarar que el desarrollo de nuevas técnicas de reconocimiento de emociones queda fuera del alcance de este proyecto pues, como hemos dicho anteriormente, nos centraremos en el impacto de las emociones en la calidad percibida desde el punto de vista del usuario y en cómo esta medición permite tomar acciones que mejoren la calidad del proceso y del producto de software.

Los resultados de investigación de este proyecto, se integrarán y pasarán a formar parte de las prestaciones de AQF, constituyendo a partir de esto, un framework que, además de tener en cuenta la calidad desde el punto de vista del proceso, tendrá en cuenta la calidad desde el punto de vista de las *personas,* y a futuro se expandirá con técnicas, métodos y procesos que incluyan la calidad desde el punto del *producto*.

3. RESULTADOS ESPERADOS

Dentro de las líneas de trabajo mencionadas, se esperan obtener los siguientes resultados:

- Conocimiento actualizado en cuanto a las herramientas, modelos, técnicas, procesos, etc. relacionados con la captura de emociones de los usuarios de aplicaciones software.
- Análisis a través de revisiones sistemáticas de bibliografía que estudie la relación entre Computación Afectiva y la Calidad de Software.
- Análisis del impacto de la perspectiva del usuario (User Experience - UX) en la evaluación de Calidad de Software.
- Diseño de un enfoque integrado de medición, evaluación y mejora de calidad teniendo en cuenta los factores emocionales que inciden en la User Experience (UX), que se integre dentro de AQF.
- Desarrollo de una herramienta que permita automatizar el proceso de medición y evaluación de calidad Validar los resultados obtenidos, que se integre dentro de AQF.
- Vinculaciones con otras instituciones, universitarias, gubernamentales o empresariales (especialmente las PYMES regionales), para la validación de los resultados de investigación.
- Especialización de recursos humanos en los temas propios de la evaluación de calidad de software a través de la elaboración de tesis de carreras de posgrado y trabajos de investigación realizados en el marco de becas de investigación para alumnos y/o graduados.
- Realización de transferencia tecnológica, registros de propiedad intelectual y capacitación en temas relacionados con la temática del proyecto.

A su vez, se pretende mejorar la formación de los profesionales de la Informática dado que la generación y la actualización de los conocimientos, se trasladará a los alumnos a través del proceso de enseñanza y aprendizaje, en cada una de las asignaturas dictadas por los integrantes del proyecto.

4. FORMACION DE RECURSOS HUMANOS

En este sentido, distinguimos dos vertientes:

1) *Formación de alumnos y transferencia de resultados de investigación al aula.* Tiene que ver directamente con la vinculación de las actividades y resultados de investigación y su interacción con los alumnos en el aula contribuyendo a su formación e incentivando las vocaciones científicas. En este sentido las actividades y resultados del proyecto se vinculan directamente con las asignaturas electivas del 4to Nivel de Ingeniería en Sistemas de Información (ISI): (i) Calidad del Producto y Proceso de Software y (ii) Técnicas de Desarrollo de Software Ágiles. Asimismo, se compartirán los resultados obtenidos durante el proyecto con la asignatura Ingeniería de Software, obligatoria del 4to nivel, donde los alumnos podrán aplicar de forma empírica los conocimientos aprendidos durante el curso. Además, el proyecto se vincula con asignaturas pertenecientes al 5to nivel de la carrera. Por un lado, con la asignatura Sistemas de Gestión, pues los estudiantes podrán conocer nuevos enfoques de procesos de negocio que promueven la Open Innovation en las empresas a través de la necesidad de estudiar factores emocionales y su relación con la Calidad de Software. Por último, el proyecto se relaciona con la asignatura Inteligencia Artificial pues pretende a futuro el desarrollo de herramientas tecnológicas que impliquen el conocimiento de Sistemas Inteligentes y su implementación a través de la Computación Afectiva. Cabe aclarar la mayoría de estas asignaturas están a cargo de docentes-investigadores pertenecientes al CInApTIC, al cual pertenece el proyecto aquí descrito. La participación de alumnos de la carrera de ISI posibilitará la formación de Recursos Humanos en esta área de vacancia regional, tanto a nivel científico como profesional.
2) *Formación de recursos humanos en investigación.* El proyecto cuenta actualmente con dos plazas para Becarios de Investigación y Servicios financiadas por la Secretaría de Asuntos Universitarios (UTN). Una plaza para Becario Alumno de Rectorado (BAR) y una plaza de Beca de Iniciación a la Investigación y Desarrollo (BINID) para graduados financiadas por la Secretaria de Ciencia y Tecnología (UTN). Todas ellas actualmente en proceso de selección de los candidatos. Para concluir con la formación de alumnos y egresados en el área de investigación, es importante destacar que, se ha propuesto un becario de investigación a la convocatoria 2019 de Becas Estimulo de las Vocaciones Científicas del Consejo Interuniversitario Nacional (EVC-CIN).

   En cuanto a la formación de Postgrado de los docentes investigadores que participan en el proyecto, se destaca la realización de la tesis doctoral de la Ing. Pinto cuyo tema de tesis, si bien está estrechamente relacionado con un proyecto anterior, ha dado origen a la formulación del proyecto actual. Se espera concluir una tesis de la Maestría en Ingeniería del Software (UNLP), una la tesis de la Maestría en Ingeniería de Calidad (UTN-FRRe) e iniciar una tesis de Maestría en Informática en Salud del Instituto Universitario del Hospital Italiano.

Otro aspecto positivo y destacable en cuanto a la formación de recursos humanos tiene que ver con la reciente acreditación por parte de CONEAU del Doctorado en Informática a dictarse en conjunto entre la Universidad Tecnológica Nacional, Facultad Regional Resistencia; Universidad Nacional de Misiones, Facultad de Ciencias Exactas, Químicas y Naturales; y Universidad Nacional del Nordeste, Facultad de Ciencias Exactas y Naturales y Agrimensura. El objetivo de la carrera será promover la formación científico-tecnológica de más alto nivel para los ingenieros, la investigación y el desarrollo aplicado a las ciencias informáticas, favoreciendo el desarrollo socio-económico de nuestra región. Incrementando sustancialmente las vocaciones científicas dado que se trata del

primer Doctorado en Informática en la Región del Nordeste Argentino.

5. AGRADECIMIENTOS

Este trabajo se enmarca en las actividades financiadas por los proyectos: "Mejora de la Competitividad de Empresas de Software del NEA" (PDTS CIN CONICET IP 253); "Modelo de Evaluación de la Calidad en Procesos Ágiles de Desarrollo de Software" UTN-PID 4445) y "Evaluación del impacto de las emociones en la calidad de software desde el punto de vista del usuario" (UTN-PID 5517)

6. BIBLIOGRAFIA

[1] Arhippainen, Leena; Tähti, Marika. Empirical evaluation of user experience in two adaptive mobile application prototypes. En MUM 2003. Proceedings of the 2nd International Conference on Mobile and Ubiquitous Multimedia. Linköping University Electronic Press, 2003. p. 27-34.

[2] Fernandez, Ana María; Dufey, Michele; Mourgues, Catalina. Expresión y reconocimiento de emociones: un punto de encuentro entre evolución, psicofisiología y neurociencias. Revista Chilena de Neuropsicología, 2007, vol. 2, no 1, p. 8-20 USA

[3] Ekman, Paul; Oster, Harrieh. Expresiones faciales de la emoción.

[4] Picard, R. W., & Klein, J. (2002). Computers that recognise and respond to user emotion: theoretical and practical implications. Interacting with computers, 14(2), 141-169.

[5] Tao, J., & Tan, T. (2005, October). Affective computing: A review. In International Conference on Affective computing and intelligent interaction (pp. 981-995). Springer, Berlin, Heidelberg.

[6] Bosquez, V., Sanz, C., Baldassarri, S., Ribadeneira, E., Valencia, G., Barragan, R., ... & Shauri-Romero, J. (2018). La computación afectiva: emociones, tecnologías y su relación con la educación virtual. Revista de Investigación Talentos, 5(1), 94-103. 7

[7] Picard, R. W. (2013). Affective computing: challenges. I1nternational Journal of Human-Computer Studies, 59(1-2), 55-64.

[8] Noelia Pinto; César J. Acuña; Gabriela Tomaselli; Nicolás Tortosa. Impacto de las emociones del usuario en la percepción de la calidad del software: Una revisión sistemática. 7mo Congreso Nacional de Ingeniería Informática – Sistemas de Información (CONAIISI 2019). 14 y 15 de noviembre de 2019 – Universidad Nacional de La Matanza – San Justo, Buenos Aires, Argentina

[9] Noelia Pinto; Nicolás Tortosa; Blas Cabas Geat; Lucas Ibáñez; Verónica Bollati "Quality evaluation of agile processes: Measurement of requirements management using AQF v2" Proceedings of 11th International Conference on the Quality of Information and Communications Technology (QUATIC), Coimbra, 2018, pp. 15-20.

[10] Pinto, N., Acuña, C., & Cuenca Pletsch, L. R. (2016). Quality Evaluation in Agile Process: A First Approach. In XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016)

[11] Cuenca Pletsch, L., Acuña, C. J., Tomaselli, G., & Pinto, N. (2017). QUAGI: Una propuesta para el seguimiento y evaluación de proyectos de Software Ágiles. V SABTIC, VIII STIN e XVIII Fórum

[12] Pinto, N., Tortosa, N., Cabas Geat, B., Ibáñez, L., & Acuña, C. (2018, November). Validación de la reingeniería aplicada sobre la primera versión de Agile Quality Framework. In XIX Simposio Argentino de Ingeniería de Software (ASSE)-JAIIO 47 (CABA, 2018).

DESARROLLO DE VIDEOJUEGOS INNOVADORES

Ing. Sebastián Blanco, Adriel Arández, Giuliano Taliano, Mauro Wehner, Guido Gallego, Santiago Cuiñas

Centro de Altos Estudios en Tecnología Informática
Facultad de Tecnología Informática
Universidad Abierta Interamericana
Sebastian.Blanco@uai.edu.ar,{Adriel.Arandez, GiulianoDolce.TalianoSantoro, MauroEzequiel.Wehner, GuidoMartin.GallegoPaulicci, Santiago.Cuinas}@alumnos.uai.edu.ar

RESUMEN

El objetivo de este proyecto es sentar las bases dentro del ámbito educativo para realizar videojuegos innovadores en términos de mecánicas, estéticas e interfaces hombre-máquina poco comunes.

Palabras clave: Desarrollo de Videojuegos, Mecánicas, Estéticas, Interfaces hombre-máquina

CONTEXTO

El presente proyecto será radicado en el Centro de Altos Estudios en Tecnología Informática (CAETI), dependiente de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana. El mismo se encuentra inserto en la línea de investigación "Ingeniería de Software". El financiamiento está dado por la misma Universidad Abierta Interamericana.

1. INTRODUCCIÓN

Durante la última década, la industria de los videojuegos atravesó grandes cambios significativos:

- Se popularizaron los smartphones como dispositivos para jugar en forma ubicua.
- Surgieron nuevos dispositivos electrónicos para aumentar la experiencia de juego: cascos de realidad virtual/aumentada/mixta, cámaras infrarrojas, sensores y una miríada de controles alternativos.
- Se desarrollaron entornos multijugador online en donde compiten hasta 150 personas en tiempo real.
- Surgieron nuevas plataformas de distribución de juegos y streaming, así como también canales masivos de contenidos relacionados.
- Se llegó a un nivel de fotorrealismo y complejidad nunca visto, incluso con motores gráficos de uso gratuito.
- Se profesionalizó al jugador de deportes electrónicos y se crearon nuevas competencias a nivel mundial.
- Se profundizó el movimiento de desarrolladores independientes cuyos juegos ofrecen propuestas diferentes a los realizados por los grandes actores de la industria.
- Se avanzó en Machine Learning y Deep Learning para facilitar no sólo los procesos de desarrollo sino también aspectos particulares del contenido de los videojuegos.

En el año 2019 salieron 8290 títulos en Steam, la plataforma digital más grande de distribución de videojuegos para PC. Por otro lado, el catálogo de juegos actualizado a enero de 2019 de las consolas de última generación posee la siguiente cantidad de títulos: 1857 en XBOX One, 1856 en Playstation 4 y 1286 en Nintendo Switch. Es importante destacar que el proceso de aprobación para un juego correspondiente a cualquiera de estas consolas es riguroso, así como también la calidad

esperada. A estos datos se les debe sumar la cantidad de juegos que se lanzan actualmente en plataformas de distribución como Google Play, cuyo catálogo alcanza 315.418 juegos en el tercer trimestre de 2018. Otras plataformas son App Store, Itch.io, Gamejolt, Kongregate y demás portales online.

Los datos mencionados anteriormente implican que la oferta de videojuegos a escala global es alta a nivel cuantitativo y cualitativo. Se lanzan al mercado más juegos de los que uno puede jugar aun dedicando las 24 horas del día. Además, algunos de estos juegos son creados por empresas que tienen personal altamente calificado y presupuestos millonarios tanto para lo técnico como para sus estrategias de marketing. Esto representa no sólo un problema de visibilidad para los títulos de los desarrolladores repercutiendo directamente en su economía sino también un desafío para las instituciones académicas a la hora de formar profesionales aptos para destacarse.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

La solución propuesta en el presente documento es ampliar el mercado a través de la innovación para facilitar la creación de nuevos nichos que otorguen más visibilidad al desarrollador y a las instituciones. Se busca que dicha innovación surja en el ámbito académico ya que convergen estudiantes de varias disciplinas que buscan un complemento formativo a sus carreras mientras se enfocan en expandir el medio lúdico. Es por esta razón que se hará hincapié en la implementación de prototipos cuyas mecánicas giren en torno a la transmisión de ideas originales mediante la interacción con interfaces poco comunes y explorando sentimientos complejos para delimitar la estética de la experiencia. Estas implementaciones permiten, desde la Ingeniería de Software, concebirse como sistemas basados en reglas formales que evocan en el jugador la toma de decisiones significativa en un ambiente mecánica y estéticamente consistente.

Las definiciones tradicionales de metodología en las investigaciones no son aplicables a un proyecto de desarrollo tecnológico. Sin embargo, se exponen a continuación los pasos a seguir por el equipo de investigación y desarrollo:

Se profundizará el estudio del estado del arte para cubrir con mayor amplitud la totalidad de juegos que hay disponibles en la actualidad y sus reglas principales, así como también las estéticas e interfaces de vanguardia que se utilizan para interactuar con sistemas lúdicos y las tendencias futuras de la industria. Un estado del arte realizado adecuadamente otorga un punto de referencia para la creación de mecánicas innovadoras. Se hará hincapié en la observación de la práctica experimental de desarrolladores que sean independientes.

A partir de lo relevado en el análisis previo, se determinarán ideas y estéticas particulares a explorar con sus especificaciones correspondientes. Es importante destacar que el alcance de las ideas estará intrínsecamente relacionado con los tiempos establecidos del proyecto y las capacidades de los integrantes que recibirán formación. Se buscará una comunicación fluida semana a semana presencial para guiarlos y verificar los avances durante las mentorías. También se les enseñará aspectos particulares del desarrollo de videojuegos y patrones como complemento a sus carreras universitarias. Por otro lado, se tratará de que los integrantes del proyecto sean autónomos en la búsqueda del conocimiento y resilientes.

Luego, se procederá a familiarizarse con herramientas requeridas: motores gráficos, software de modelado, herramientas de edición y API's de hardware específico para llevar a cabo el desarrollo. Se utilizarán

principalmente herramientas estándares en la industria como Unity Engine 2020, Unreal Engine 4 y Blender.

Se comenzará la exploración mediante el diseño e implementación iterativo e incremental de prototipos lúdicos de acuerdo con las ideas y estéticas delimitadas anteriormente. Puede haber más de un prototipo en torno a la misma idea. Durante el transcurso del proyecto los prototipos serán evaluados por un grupo heterogéneo de jugadores para mejorar el desarrollo. A modo de complemento, se diseñará y construirá un control alternativo para uno de los prototipos. Un control alternativo es un dispositivo de hardware personalizado para interactuar con un juego.

Además, se documentará formalmente la metodología empleada y sus resultados con el fin de facilitar el desarrollo futuro de prototipos innovadores. Por último, se realizará un catálogo virtual que nucleará a todos los prototipos para que éste pueda ser descargado y presentado en eventos de relevancia para las universidades.

3. RESULTADOS Y OBJETIVOS

El objetivo final de este proyecto es desarrollar un catálogo que nuclee prototipos de videojuegos innovadores en términos de mecánicas, estéticas e interfaces hombre-máquina. El catálogo se irá expandiendo gracias al aporte de cada una de las comisiones que se involucren en el proyecto. Por esta razón se formará en la academia un equipo multidisciplinario que sea consciente del estado del arte actual de los videojuegos, sea capaz de utilizar herramientas estándares en la industria y siga una metodología de diseño basada en conceptos propios de la Ingeniería de Software.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está conformado por un investigador adjunto del Centro de Altos Estudios en Tecnología Informática (CAETI) quien ejerce el rol de director del proyecto y cinco alumnos ayudantes de pregrado correspondientes a las carreras de Ingeniería en Sistemas y Licenciatura en Producción de Simuladores y Videojuegos de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana. Se buscará también la inclusión de estudiantes de otras facultades como Arquitectura y Psicología para el proyecto.

5. BIBLIOGRAFÍA

- Hunicke, R., Leblanc, M., & Zubek, R. (2004). MDA: A Formal Approach to Game Design and Game Research.
- Upton, B. (2015). The Aesthetics of Play. MIT Press.
- Rusch, D. (2016). Making Deep Games: Designing Games with Meaning and Purpose. Focal Press.
- Johnson E., Sullivan A. (2018) Facilitating Undergraduate Experimental Game Design: A Pilot Study with Celestial Harmony. In: Beck D. et al. (eds) Immersive Learning Research Network. iLRN 2018. Communications in Computer and Information Science, vol 840. Springer, Cham.
- Kim, H. (2016). A Study of a Framework for Experimental Games with Sensors in IoT. 2016 IEEE International Conference on Smart Cloud (SmartCloud), 1-4.
- Alves G.F., Souza E.V., Trevisan D.G., Montenegro A.A., de Castro Salgado L.C., Clua E.W.G. (2018) Applying Design Thinking for Prototyping a Game Controller.
- Tan, C.T., Leong, T.W., Shen, S., Dubravs, C. and Si, C. (2015). Exploring Gameplay Experiences on the Oculus Rift. CHI PLAY.
- Smith, J.D., & Graham, T.C. (2006). Use of eye movements for video game control. Advances in Computer Entertainment Technology.
- Marshall, D., Coyle, D., Wilson, S., & Callaghan, M.J. (2013). Games, Gameplay, and BCI: The State of the Art. IEEE Transactions on Computational

Intelligence and AI in Games, 5, 82-99.

- Aaron Oldenburg, A. Sonic Mechanics: Audio as Gameplay. Game Studies, The international journal of computer game research, vol. 17, issue 1.
- Holopainen, J., & Nummenmaa, T. (2010). Modelling Experimental Game Design.
- Annika Waern and Jon Back. 2015. Experimental game design. In Game Research Methods, Petri Lankoski and Staffan Björk (Eds.). ETC Press, Pittsburgh, PA, USA 341-353.
- Fullerton, T. (2006). Play-centric games education. Computer, 39, 36-42.
- Fullerton, T., Chen, J., Santiago, K., Nelson, E., Diamante, V., Meyers, A., Song, G., & DeWeese, J. (2006). That cloud game: dreaming (and doing) innovative game design. Sandbox @ SIGGRAPH.

Internet del futuro: aplicaciones de IoT en la Patagonia Austral

Osiris SOFIA, Karim HALLAR, Esteban GESTO, Daniel LAGUIA, Leonardo GONZALEZ

Unidad Académica Río Gallegos (UARG)
Universidad Nacional de la Patagonia Austral (UNPA)
{osofia, khallar, egesto, dlaguia, lgonzalez}@uarg.unpa.edu.ar

RESUMEN

Internet se ha convertido en la gran revolución tecnológica de principios de Siglo XXI. Los avances y desafíos de esta nueva tecnología son vertiginosos, así como sus aplicaciones e impacto en la sociedad. En los últimos años se ha avanzado con una nueva tecnología disruptiva, que se afianza como integración de tecnologías anteriores con distintos grados de desarrollo, y tiende a convertirse en el nuevo paradigma de interconexión: Internet de las Cosas, o IoT por sus siglas en inglés (Internet of Things).

"Internet de las Cosas" es la columna vertebral que sostiene la interconexión de objetos físicos. En este escenario, las redes de sensores inalámbricos (Wireless Sensor Network, WSN) constituyen actores claves de las mismas, sin embargo el desarrollo tecnológico actual no permite una fácil integración con las redes móviles por la falta de estandarización de protocolos y fundamentalmente por no contar con un modelo de red. El análisis de tráfico en dichas redes, por ejemplo, permitirá proponer un modelo de red apropiado en el marco de IoT.

Son numerosas y variadas las líneas de investigación y desarrollo posibles, ya que se trata de una temática en continua evolución, tanto en el aspecto técnico como de desarrollo de software, con un campo de aplicación vasto y de impacto tecnológico y social aún por determinar.

Palabras clave: IoT, Sistemas, Automatización, Internet, Desarrollo de Software.

CONTEXTO

Este trabajo ha sido realizado en el marco del Proyecto de Investigación PI 29/A425-1 "Internet del futuro: aplicaciones de IoT en la Patagonia Austral" de la Universidad Nacional de la Patagonia Austral (2019-2021).

Si bien el proyecto ha sido acreditado por la UNPA, se está trabajando estrechamente con el INTA en líneas de investigación y desarrollo.

1. INTRODUCCIÓN

Internet de las Cosas (IoT por sus siglas en inglés: "Internet of Things") nace de la fusión de dos dominios completamente separados hasta hace no mucho tiempo: los sistemas embebidos, e Internet. Los sistemas embebidos, en auge actualmente dentro de la comunidad informática, no son nuevos: nacieron con los primeros microprocesadores desarrollados a principios de los años '70. Internet, por otra parte, es un fenómeno un cuarto de siglo más moderno que ya lleva más de dos décadas entre nosotros. La asociación entre estos dos feudos iba a resultar inevitable.

El concepto de computación ubicua (ubiquitous computing) introducido en 1991

por Mark Weiser en "The Computer of the 21st Century"(Weiser 1991) prefiguró la revolución que IoT produciría en años posteriores. En su artículo Weiser bosquejó un futuro en el que multitud de dispositivos interconectados, por medios convencionales o inalámbricos, resultaría habitual y omnipresente en las sociedades avanzadas; para mejorar la vida de las personas no se requeriría así de una revolución en el campo de la inteligencia artificial, sino sólo de una multitud de pequeñas computadoras y sensores incorporados al entorno habitual de los usuarios.

El término The Internet of the Things es posterior y se debe, presumiblemente, a Kevin Ashton quien en 1999 habría defendido la idea de asociar las etiquetas RFID para identificación de productos, con la floreciente tecnología de Internet, en beneficio de las cadenas de suministro de Procter & Gamble (Ashton 2009).

Según la Unión Internacional de Telecomunicaciones (UIT), la Internet de las Cosas (también llamada Internet de los Objetos) es “la infraestructura mundial para la sociedad de la información que propicia la prestación de servicios avanzados mediante la interconexión de objetos (físicos y virtuales) gracias a la interoperatividad de tecnologías de la información y la comunicación presentes y futuras” (ITU Corporation 2015).

El impacto que IoT tendrá sobre la sociedad durante los próximos años es difícil de vaticinar, pero sin duda será enorme. IoT propone una nueva revolución de Internet que trascenderá el aspecto meramente tecnológico al modificar el comportamiento de los usuarios desde la dimensión social, económica y ética, tal y como ya lo hiciera hace más de dos décadas la Red de redes. En efecto, el paradigma IoT está cambiando nuestras vidas porque, por definición, se imbrica en nuestro mundo cotidiano al punto de pasar inadvertida (Weiser 1991). Los objetos IoT recolectan ingentes cantidades de datos que difunden automáticamente hacia otros objetos, aplicaciones, servicios y sistemas para hacer nuestras vidas más llevaderas.

En el contexto de los usuarios particulares el nuevo paradigma tendrá un protagonismo singular en escenarios como la domótica, la e-health, el deporte, la asistencia doméstica, el aprendizaje mejorado, etc. Desde la perspectiva del mundo empresarial IoT tendrá un enorme impacto en campos como la automatización y fabricación, la logística, la gestión de negocios y procesos, el transporte inteligente de personas y mercancías (ITU Corporation 2015), el cuidado del medioambiente, la Industria 4.0, el mantenimiento predictivo, procesos de control mejorados, industria verde (Pacis, Subido Jr., and Bugtai 2017), etc.

Internet de las Cosas es un dominio joven, vasto y transversal; por lo tanto plantea enormes desafíos que requieren de la imaginación y el concurso de especialistas e investigadores de áreas muy diversas, tales como las TICSs, producción, la logística, la gestión, la innovación, salud, higiene, educación, medioambiente, ecología, arquitectura, urbanismo, jurídico y legal, ética, y seguridad, por mencionar sólo algunas.

Se requerirán esfuerzos adicionales en la obtención de modelos arquitecturales de referencia adaptados a aspectos particulares de IoT (Kramp, van Kranenburg, and Lange 2013; Romero 2012), tales como modelos de dominio, de información, funcionales, la respuesta en tiempo real y el modelado

propiamente dicho (lenguajes y extensiones (Eterovic et al. 2015), definiciones, etc.).

En referencia al ciclo de vida del desarrollo de IoT es evidente el interés en la Ingeniería de Sistemas basada en modelos (MBSE) (Mazzini, Favaro, and Baracchi 2015; Papke 2017), el empleo de ciclos de vida iterativos e incrementales (BBC 2018; Mazzini, Favaro, and Baracchi 2015) y la especificación, verificación y análisis mediante métodos formales (Afzaal and Zafar 2017; Houimli, Kahloul, and Benaoun 2018; Latif, Afzaal, and Zafar 2017).

Además de las problemáticas mencionadas, Internet de las cosas presenta otros desafíos a resolver, como por ejemplo: la cuestión de la escalabilidad, que permita la operación eficiente tanto a pequeña como gran escala, los mecanismos de auto organización, que son cruciales para los objetos (a diferencia de las computadoras personales), los servicios de descubrimiento, que permitirán que los aplicaciones y servicios puedan conocer y utilizar los objetos IoT disponibles, la tolerancia a fallos, que dotará a las redes de una capacidad de adaptación mayor en condiciones cambiantes (Kamal et al. 2017), entre otros.

En cuanto a la arquitectura, un modelo IoT genérico se representa con 5 capas:

1) Capa de objetos o percepción: Esta capa recolecta e intercambia información entre los sensores de los objetos físicos. Sensores y actuadores proveen las diferentes funcionalidades. Los datos son transferidos a la capa de red.

2) Capa de red o transmisión: Esta capa transfiere de forma segura la información de los dispositivos sensores al sistema de procesamiento de información (capa de middleware).

3) Capa de middleware: Esta capa es responsable de la gestión del servicio y tiene un enlace a la base de datos. Recibe la información de la capa de red y la deposita en la base de datos. Realiza el procesamiento de información y la computación ubicua, toma decisiones automáticas en función de los resultados.

4) Capa de aplicación: Esta capa proporciona una administración global de la aplicación basada en la información de los objetos procesados en la capa de middleware.

5) Capa de Negocio: Esta capa es responsable de la administración del sistema general de IoT, incluidas las aplicaciones y los servicios. Construye modelos de negocio, gráficos, diagramas de flujo, etc. en función de los datos recibidos de la capa de aplicación. El verdadero éxito de la tecnología IoT también depende de los buenos modelos de negocio. Basado en el análisis de resultados, esta capa ayudará a determinar las acciones futuras y las estrategias comerciales.

En el diseño de un lenguaje de dominio específico para IoT se debe tener en cuenta que los usuarios serán desde personas comunes sin experiencia en ingeniería hasta desarrolladores de software. Trabajos recientes han demostrado que UML es un lenguaje visual de modelado de dominio que puede extenderse para IoT ya que tiene la suficiente potencia para el profesional y simplicidad para ser entendido por un usuario final.

2. LINEAS DE INVESTIGACION

Las investigaciones que se realizan se vinculan al sector productivo y de servicios convenientemente adaptados al ámbito económico y social de nuestra zona. Un

ejemplo de ello es el proyecto de extensión "Automatización de siembra de plantines como herramienta fundamental para la producción frutihortícola de Santa Cruz" (Res Nº 029-17-CS-UNPA) que se llevó a cabo en colaboración con el INTA y está sirviendo como plataforma para la aplicación de conceptos de la esfera IoT a la realidad local. (Gonzalez et al, 2019).

Actualmente la línea de investigación tiende al diseño y desarrollo de un módulo hidropónico cerrado y automatizado para su implementación en climas extremos, con enfoque en la automatización y monitoreo del mismo, así como en el desarrollo de soluciones software para sistemas embebidos y para la toma de decisiones.

3. OBJETIVOS PROPUESTOS

Objetivo Principal

Impulsar estrategias que permitan el desarrollo del Internet de las Cosas y su aplicación para dar solución a problemas reales en el ámbito regional.

Objetivos Específicos

- Realizar Investigación aplicada sobre la Arquitectura de Internet de las cosas
- Diseñar y construir un prototipo de una máquina automática con funcionalidad extra de monitoreo y medición de variables ambientales recolectadas a través de dispositivos con tecnología IoT,
- Investigar técnicas de recolección y análisis de datos provenientes de sensores adecuados para IoT instalados en nuestro prototipo
- Identificar, caracterizar y evaluar los distintos componentes (hardware, software y arquitectura) de IoT adecuados a nuestra zona geográfica.

4. FORMACIÓN DE RECURSOS HUMANOS

Es de destacar que este grupo de investigación, de reciente formación, está conformado por docentes que cubren un campo de conocimientos multidisciplinario con experiencia en requerimientos, análisis, diseño, codificación, verificación y validación, gestión de proyectos, comunicaciones, electrónica y hardware de sistemas embebidos.

El proyecto fue acompañado por la postulación de un becario alumno que fue merituado para la beca. Por último, un alumno avanzado ha presentado recientemente el proyecto de tesina de grado para optar por el título de Licenciado en Sistemas con el proyecto "IoT aplicados a los sistemas de medición de protección catódica".

5. BIBLIOGRAFÍA

- Afzaal, Hamra, and Nazir Ahmad Zafar. 2017. "Modeling of IoT-Based Border Protection System." In 2017 First International Conference on Latest Trends in Electrical Engineering and Computing Technologies (INTELLECT), IEEE, 1–6. http://ieeexplore.ieee.org/document/8277639/ (August 9, 2018).
- Ashton, Kevin. 2009. 22 RFID Journal In the Real World, Things Matter More than Ideas. http://www.rfidjournal.com/articles/pdf?4986 (August 4, 2018).
- BBC. 2018. "Fitness App Strava Lights up Staff at Military Bases." BBC News. http://www.bbc.co.uk/news/technology-42853072 (August 9, 2018).
- Eterovic, Teo et al. 2015. "An Internet of Things Visual Domain Specific Modeling Language Based on UML." In 2015 25th

International Conference on Information, Communication and Automation Technologies, ICAT 2015 - Proceedings, IEEE, 1–5. http://ieeexplore.ieee.org/document/7340537/ (August 8, 2018).

- Houimli, Manel, Laid Kahloul, and Sihem Benaoun. 2018. "Formal Specification, Verification and Evaluation of the MQTT Protocol in the Internet of Things." In Proceedings of the 2017 International Conference on Mathematics and Information Technology, ICMIT 2017, IEEE, 214–21. http://ieeexplore.ieee.org/document/8259720/ (August 9, 2018).
- ITU Corporation. 2015. "Internet of Things Global Standards Initiative." Internet of Things Global Standards Initiative 2060(July 2015): 1. http://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx (August 4, 2018).
- Kramp, Thorsten, Rob van Kranenburg, and Sebastian Lange. 2013. Enabling Things to Talk: Designing IoT Solutions with the IoT Architectural Reference Model Introduction to the Internet of Things.
- Mazzini, Silvia, John Favaro, and Laura Baracchi. 2015. "A Model-Based Approach Across the IoT Lifecycle for Scalable and Distributed Smart Applications." In IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, IEEE, 149–54. http://ieeexplore.ieee.org/document/7313125/ (August 8, 2018).
- Pacis, Danica Mitch M., Edwin DC. Subido Jr., and Nilo T. Bugtai. 2017. "Research on the Application of Internet of Things (IoT) Technology Towards A Green Manufacturing Industry: A Literature Review." In DLSU Research Congress 2017, , 1–11. http://www.dlsu.edu.ph/conferences/dlsu-research-congress-proceedings/2017/SEE/SEE-II-024.pdf (August 11, 2018).
- Papke, Barry L. 2017. "Enabling Design of Agile Security in the IOT with MBSE." In 2017 12th System of Systems Engineering Conference, SoSE 2017, IEEE, 1–6. http://ieeexplore.ieee.org/document/7994938/ (August 9, 2018).
- Romero, Gregorio Martin. 2012. "Internet of Things Architecture IoT-A Project Deliverable D2.2 – Concepts for Modelling IoT-Aware Processes." file:///home/esteban/Descargas/zdoc.site_internet-of-things-architecture-meet-iot.pdf (August 11, 2018).
- Weiser, Mark. 1991. "The Computer for the 21st Century." Scientific American 265(3): 94–104. http://www.nature.com/doifinder/10.1038/scientificamerican0991-94 (August 4, 2018).
- Leonardo Gonzalez, Daniel Laguía, Eder dos Santos , Jorge Birgi, , Karim Hallar, Esteban Gesto, and Osiris Sofía. "LibreSeed: una sembradora de plantines con hardware y software libre", 48 JAIIO - CAI, Congreso Argentino de AgroInformática. ISBN 2525-0949, 2019.

APORTE DE LOS SISTEMAS DE INTELIGENCIA DE NEGOCIOS A LOS SISTEMAS DE INFORMACIÓN ORGANIZACIONALES PARA LA TOMA DE DECISIONES

Romagnano, M.[1,2], Pantano, J. C. [1,2], Ganga, L.[1], Herrera, M.[1,2], Becerra, M.[1], Sarmiento, A.[1], Aballay, A.[1], Gordillo, M. L.[1] , Lépez, H.[1]

[1]Departamento de Informática, FCEFN, Universidad Nacional de San Juan
[2]Instituto de Informática, FCEFN, Universidad Nacional de San Juan

{maritaroma, mherrera}@iinfo.unsj.edu.ar, juancruz871@hotmail.com, {leonelganga, mcbecerra2008, prof.alicia, mgordillo13}@gmail.com,{adriva2005, lepezhr}@yahoo.com.ar

RESUMEN

En la actualidad, en todos los estratos sociales, los datos, la información y el conocimiento se han convertido en uno de los recursos más valiosos para la toma de decisiones. Ante una inquietud o consulta de cualquier tipo, se puede acceder rápidamente a grandes volúmenes de datos. Sin embargo, el almacenamiento, procesamiento y su posterior análisis, representan uno de los problemas más críticos debido al gran volumen de datos.

En las organizaciones, esto representa un desafío, ya que tienen que lidiar diariamente con grandes cantidades de datos que a menudo se generan en las operaciones del día a día. Dichos datos deben ser procesados y convertidos en información, la cual se utilizará para tomar decisiones sobre estrategias a seguir, inversiones a realizar, entre otras acciones. Si no se recolectan los datos adecuados o más relevantes, la información generada no será precisa, los resultados probablemente serán erróneos y, en consecuencia, cualquier decisión tomada no será la mejor ni la más adecuada.

Ante esta problemática planteada, algunas ciencias interdisciplinarias, como Sistemas de Información (IS), Inteligencia de Negocios (BI), Minería de Datos (DM), Big Data (BD), Analítica de Negocios (BA) e Ingeniería del Conocimiento (KE), han fusionado sus saberes y esfuerzos de procesamiento para dar apoyo a la toma de decisiones en las actuales organizaciones; que presentan algunas características tales como: almacenar y gestionar grandes cantidades de datos, adecuarse rápidamente al mercado, tomar decisiones de forma casi inmediata, etc.

Por lo tanto, atendiendo a las necesidades por las cuales transitan actualmente las organizaciones, y observando la debilidad en la actual currícula académica, para apoyar al medio local, es que la presente contribución propone determinar cómo los Sistemas de Inteligencia de Negocios (BIS) aportan a los Sistemas de Información Organizacionales (OIS), para la toma de decisiones.

***Palabras Claves*:** Adquisición del Conocimiento, Analítica de Negocios, Big Data, Inteligencia Artificial, Inteligencia de Negocios, Minería de Datos, Sistemas de Información.

CONTEXTO

El presente trabajo se encuentra enmarcado en el proyecto con igual denominación, presentado en la convocatoria del Consejo de Investigaciones Científicas y Técnicas y de Creación Artística (CICITCA) de la Universidad Nacional de San Juan, para ser desarrollado durante el período comprendido entre 01/01/2020 al 31/12/2021. En este momento se encuentra en proceso de evaluación por parte de evaluadores externos.

Las tareas de investigación se desarrollarán en el ámbito de la Facultad de Ciencias Exactas, Físicas y Naturales, con el apoyo del Departamento e Instituto de Informática.

El grupo de investigación se encuentra conformado por integrantes de distintas disciplinas (Informática, Matemática, Estadística, Administración de Empresas y Abogacía) que, en su mayoría, cuentan con una experiencia de más de 20 años en la disciplina que les compete.

Desde el año 1995 hasta el 2010 la línea de investigación estuvo orientada a la gestión de los IS. Luego, en el año 2011 se puso énfasis en los fundamentos conceptuales de los IS, pudiendo identificar las restricciones que existen en los IS. Se continuó estudiando la crisis por la que transitaba la disciplina al confundir los SI con Tecnologías de Información (IT). Posteriormente se trabajó en la construcción de un modelo disciplinar, basado en constructos. Desde el año 2016 se vienen identificando y caracterizando los modelos conceptuales que aportaran conocimiento a los IS.

Si bien el eje central del equipo de trabajo siguen siendo los IS, actualmente se ha redireccionado la investigación a estudiar el aporte de la Inteligencia Artificial, específicamente la adquisición de conocimiento a través de los Sistemas de Inteligencia de Negocio, a los Sistemas de Información Organizacionales, lo cual permite a la organización tomar decisiones acertadas y oportunas.

1. INTRODUCCIÓN

Desde el año 2005 los Sistemas de Inteligencia de Negocios se han convertido en una apuesta tecnológica para organizaciones y empresas líderes. Así, la Inteligencia de Negocios ha surgido como un concepto integrador para dar respuesta a retos informativos en las organizaciones (Roldan, Galán, Cepeda-Carrión, 2012).

En (Sun, Zuo y Strang, 2015) se define a la Inteligencia de Negocios como un proveedor de tomadores de decisiones con información valiosa y conocimiento, obtenido a través de la nivelación de varias fuentes de datos así como de información, tanto estructurada como no estructurada. Se refiere a una colección de Sistemas de Información y tecnologías que soportan tomadores de decisiones gerenciales de control operacional a través de proveer información en operaciones internas y externas. La Inteligencia de Negocios se presenta como un conjunto de procesos, aplicaciones y tecnologías que facilitan la obtención rápida y sencilla de datos provenientes de distintos Sistemas de Información. Por lo tanto, los datos analizados e interpretados, se transforman en conocimiento apropiado para la toma de decisiones. Las posibilidades de procesamiento que ofrece la Inteligencia de Negocios hacen que las organizaciones puedan desempeñarse en un ambiente global altamente competitivo, el cual exige una constante toma de decisiones estratégicas para poder permanecer de manera exitosa (Ahumada Tello y Perusquia Velasco, 2015).

Por otra parte, en (Russell y Norving, 2003) se menciona: "El campo de la Inteligencia Artificial (AI) intenta comprender entidades inteligentes. Por lo tanto, una de las razones para estudiar es aprender más sobre nosotros mismos. Pero a diferencia de la filosofía y la psicología, que también tienen que ver con la inteligencia, la AI se esfuerza por construir entidades inteligentes, así como a entender..." Además en el libro los autores definen a la AI desde cuatro perspectivas distintas:

- Sistemas que piensan como seres humanos: "La automatización de actividades que asociamos con el pensamiento humano, actividades tales como la toma de decisiones, resolución de problemas, el aprendizaje..." (Bellman, 1978).
- Sistemas que actúan como seres humanos: "El arte de la creación de máquinas que realizan funciones que requieren de inteligencia cuando es realizada por la gente" (Kurzweil, 1990).
- Sistemas que piensan racionalmente: "El estudio de los cálculos que hacen posible percibir, razonar y actuar" (Winston, 199).
- Sistemas que actúan racionalmente: "La rama de la Informática que se refiere a la automatización de comportamiento inteligente" (Luger y Stubblefield, 1993).

Dentro de esta amplia disciplina se han especificado sub disciplinas, algunas de las cuales se detallan a continuación:

Por su parte, la Minería de Datos consiste en el proceso de descubrimiento de patrones o de conocimiento útil desde fuentes de datos tales como bases de datos, textos, imágenes, la Web, etc. Representa un campo multidisciplinario que involucra aprendizaje automático, estadísticas, bases de datos, inteligencia artificial, recuperación de la información y visualización. Entre las tareas más comunes se tiene el aprendizaje supervisado o clasificación, aprendizaje no supervisado o clustering, reglas de asociación, patrones secuenciales (Han y Kamber, 2006).

El Big Data se refiere a una vasta cantidad de datos, que las tecnologías convencionales de Data Warehouse no pueden almacenar, administrar o analizar, pero que es requerida por las organizaciones para proveer mejores percepciones cuando se evalúan nuevas oportunidades de negocios y para mejores tomas de decisiones. Los tres atributos claves del Big Data son volumen, velocidad y variedad. La recopilación de grandes cantidades de datos y la búsqueda de tendencias dentro de los datos permiten que las organizaciones se muevan mucho más rápidamente, sin problemas y de manera eficiente (Heavin, Daly, y Adam, 2014).

La Ingeniería del Conocimiento se ha relacionado con el desarrollo de sistemas de software por computadora en el cual el conocimiento y el razonamiento juegan un papel primordial. Sin embargo, recientemente el alcance de sus técnicas se ha ampliado para ser usadas en la gerencia del conocimiento, la modelación de empresas y los procesos de reingeniería del negocio. A su vez, dentro de la KE se debe mencionar a la Adquisición del Conocimiento (KA), como el proceso de entender cómo una persona lleva a cabo alguna actividad de modo que esa misma actividad pueda ser automatizada (Henao Cálad, 1998).

Además, otro aspecto muy importante a tener en cuenta es la Analítica de Negocios, considerada como la técnica que explora grandes volúmenes de datos para asistir a los ejecutivos en la toma de decisiones (Chahal, H., Jyoti, J., y Wirtz, J., 2019). Se enfoca en recabar y combinar una gran cantidad de datos con el objetivo de derivar en ideas que no llegan a ser percibidas en una escala menor (Orrillo, 2018).

En cuanto a Sistemas de Información, Peña los define como un conjunto de elementos interrelacionados con el propósito de prestar atención a las demandas de información de una organización, para elevar el nivel de conocimientos que permitan un mejor apoyo a la toma de decisiones y desarrollo de acciones (Peña, 2006). Otros autores lo definieron como un conjunto de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa o negocio. En un sentido amplio, un Sistema de Información no necesariamente incluye equipos electrónicos (hardware). Sin embargo, en la práctica se utiliza como analítica de sinónimo de "sistema de información computarizado" (Cohen y Asín Lares, 2005).

Por último, como estos sistemas se encuentran inmersos en la organización, cabe mencionar su definición. Así Chiavenato (2000) la define como la coordinación de diferentes actividades de contribuyentes individuales, con la finalidad de efectuar intercambios planteados con el ambiente.

Por lo tanto, la Inteligencia de Negocios, Business Intelligence o simplemente BI, considerada como una especifidad de Ciencia de los Datos, puntualmente como una especificación de la Inteligencia Artificial para el Negocio, viene a ofrecer esta gestión del conocimiento. Se presenta como un conjunto de procesos, aplicaciones y tecnologías que facilitan la obtención rápida y sencilla de datos provenientes de distintos Sistemas de Información Organizacionales. Los datos analizados e interpretados, se transforman en conocimiento apropiado para la toma de decisiones.

Las posibilidades de procesamiento que ofrece la Inteligencia de Negocios hacen que las organizaciones puedan desempeñarse y responder dinámicamente en un entorno

universal altamente impetuoso, el cual exige una constante toma de decisiones para mantenerse en el medio. Además, uno de los objetivos de los Sistemas de Información es brindar información que provea soporte a la toma de decisiones en la organización. A su vez, uno de los objetivos de Inteligencia de Negocios es la recolección y procesamiento de datos, de tal manera que puedan ser utilizados para generar y mantener estrategias que brinden ventajas competitivas. Se puede decir que los Sistemas de Información utilizan Inteligencia de Negocios para procesar la información y luego recibir una respuesta concreta y acertada del mismo. Al mismo tiempo, los Sistemas de Inteligencia de Negocios utilizan los Sistemas de Información como una herramienta para poder, justamente, procesar estos datos que recolecta (Dedić N., y Stanier C., 2016). Entonces, los Sistemas de Inteligencia de Negocios son, a fin de cuentas, Sistemas de Información con un objetivo principal de apoyar la toma de decisiones.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Actualmente se habla de un ambiente global, turbulento y altamente competitivo en el cual deben trabajar y mantenerse las organizaciones. Hoy en día, es que se les exige una constante toma de decisiones estratégicas para poder permanecer en el mercado, de manera exitosa y rentable. Esta situación las lleva a enfrentar a los mercados con prudencia al optimizar sus recursos. La conversión de datos en información y de información en conocimiento, requiere la sincronía de muchos especialistas y de especialidades dentro de la organización. Este conocimiento logrado, y su posterior gestión para la toma de decisiones puede ser la principal ventaja competitiva con la que cuenta una organización. Heavin, Daly y Adam mencionaban que la habilidad de una organización para administrar el conocimiento es esencial en términos de su desarrollo y como un activo estratégico (Heavin, Daly y Adam, 2014).

Luego, el equipo de investigación, considera la temática propuesta de relevancia social, organizacional y que además posee implicancias prácticas, ya que es altamente propicia debido a que, actualmente, Argentina cuenta con políticas de estado tales como Argentina Innovadora 2020 y 2030, Ley de Promoción de la Economía del Conocimiento, Industria 4.0, Plan de Inteligencia Artificial, entre otras. Precisamente, San Juan posee polos tecnológicos como Casetic San Juan, Servicios Mineros y San Juan TEC, lo cual puede tomarse como puntapié inicial para trabajar conjuntamente con el estado en la sustentabilidad local, aportando conocimiento y casos de estudio desde la academia.

Por lo antes mencionado y debido a que el proyecto propuesto presenta características interdisciplinarias, sus integrantes se encuentran trabajando en las siguientes líneas de investigación y desarrollo:

- Inteligencia y analítica del negocio.
- Análisis y procesamiento estadístico de los datos.
- Preservación y legislación de los datos.
- Sistemas de Información organizacionales.

3. RESULTADOS OBTENIDOS/ ESPERADOS

Se espera poder determinar cuáles son los aportes que los Sistemas de Inteligencia de Negocios hacen a los Sistemas de Información Organizacionales, para una toma de decisiones oportuna y acertada.

Para lograr este propósito se han propuesto:

- Especificar los distintos tipos de conocimientos que aportan los Sistemas de Inteligencia de Negocios a los Sistemas de Información Organizacional, para la toma de decisiones.
- Describir las estrategias que comparten los Sistemas de Inteligencia de Negocios con los Sistemas de Información Organizacional, para la toma de

decisiones.

- Identificar como los conocimientos y las estrategias adquiridas por los Sistemas de Información Organizacional conllevan a oportunas y acertadas decisiones.

Los resultados del proyecto tienen una inmediata transferencia al medio local, principalmente a la alta gerencia de organizaciones locales de cualquier tipo, y a la comunidad científica. Además, se trabajará en experimentación y difusión de resultados con investigadores de la Facultad de Ingeniería de la Universidad de Quindío, Colombia y de la Escuela de Ingeniería en Transporte, Pontificia Universidad Católica de Valparaíso, Chile.

4. FORMACIÓN DE RECURSOS HUMANOS

Dentro del grupo de investigación se identifican integrantes formándose en:

- Licenciatura en Ciencias de la Computación, Facultad de Ciencias Exactas, Físicas y Naturales, UNSJ.
- Maestría en Estadística Aplicada, Escuela de graduados de la Facultad de Ciencias Económicas, UNC.
- Estancia de formación de docencia e investigación a través del I Convocatoria para el Programa de Movilidad Internacional de la UNSJ.
- Doctorado en Ingeniería, Facultad de Ingeniería, UNCuyo.
- Doctorado en Ciencias de la Informática, Facultad de Ciencias Exactas Físicas y Naturales, UNSJ.

5. BIBLIOGRAFÍA

Ahumada Tello, E. y Perusquia Velasco, J. M. (2015). Business intelligence: Strategy for competitiveness development in technology-based firms. Disponible en www.sciencedirect.com, marzo de 2019.

Chahal, H., Jyoti, J., y Wirtz, J. (2019). Business Analytics: Concept and Applications. In Understanding the Role of Business Analytics (pp. 1-8). Springer, Singapore.

Chiavenato I. (2000). Administración de recursos humanos. El capital humano de las organizaciones. Octava edición. ISBN 970-10-6104-7. McGRAW-HILL/INTERAMERICANA EDITORES, S.A. DE C.V.

Cohen, K., y Asín Lares, E. (2005). Sistemas de información para los negocios: un enfoque de toma de decisiones.

Dedić N., y Stanier C. (2016). Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting. Lecture Notes in Business Information Processing. Springer International Publishing. Volume 268, pp. 225-236.

Han, J., y Kamber, M. (2006). Data Mining: Concepts and Techniques. Segunda Edición. Elsevier. ISBN 13: 978-1-55860-901-3 ISBN 10: 1-55860-901-6. USA.

Heavin, C., Daly, M., y Adam, F. (2014). Small data to Big Data: The Information Systems (IS) continuum. KMIS 2014 - Proceedings of the International Conference on Knowledge Management and Information Sharing. 289-297.

Henao Cálad, M. (1998). La Adquisición del Conocimiento dentro de la Ingeniería del Conocimiento. Revista Universidad Eafit. Abril - Mayo - Junio 1998.

Orrillo, J. (2018). Analítica de negocios: Una estrategia basada en la información. Recuperado de: https://www.esan.edu.pe/conexion/actualidad/2018/03/22/analitica-negocios-informacion/, Julio de 2019.

Peña Ayala, A. (2006). Ingeniería de software: Una guía para crear sistemas de información. México DF (México): Instituto Politécnico Nacional.

Roldán Salgueiro, J. L., Cepeda-Carrión, G., y Galán González, J. L. (2012). Los sistemas de inteligencia de negocio como soporte a los procesos de toma de decisiones en las organizaciones - Papeles de Economía Española, Nº 132, 2012. ISSN: 0210-9107.

Russell, S., y Norvig, P. (2003). Artificial Intelligence: A Modern Approach. 2nd Edition. Pearson. Cloth. ISBN-13: 9780137903955. Recuperado de: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.259.8854&rep=rep1&type=pdf, Mayo de 2019.

Sun, Z., Zou, H., y Strang, k. (2015). Big Data Analytics as a Service for Business Intelligence. Recuperado de: https://www.researchgate.net/publication/282867145_Big_Data_Analytics_as_a_Service_for_Business_Intelligence, Junio de 2019.

Modelo Ontológico como garantía de completitud funcional en la elicitación de requerimiento de software

Rebeca Yuan[*], Carlos Salgado[+], Alberto Sánchez[+], Mario Peralta[+]
[*]Ingeniería de Software, Departamento en Ingeniería en Sistemas de Información
Facultad Regional San Francisco – Universidad Tecnológica Nacional
e-mail: ryuan@sanfrancisco.utn.edu.ar
[+]Departamento de Informática Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina
e-mail: {csalgado, alfanego, mperalta }@unsl.edu.ar

Resumen

El protagonismo de los datos, en esta era de la transformación digital, continúa exigiendo a la industria del software la generación de productos de calidad, capaces de favorecer la toma de decisiones y proyectar a las organizaciones hacía nuevos desafíos. Esta exigencia no proclama el uso de nuevas herramientas o cambios de paradigmas para el proceso de desarrollo del software; sino una construcción "consciente" del mismo. Asentar el desarrollo del software en Normas de Calidad, eleva el valor del producto; sumar a esto un modelo ontológico como sistema de representación de un dominio para la elicitación de requerimientos, establece bases sólidas para la continuidad de las distintas etapas del desarrollo del software. El presente trabajo busca validar un modelo ontológico para la educción de requerimientos basado en la Completitud Funcional de la Norma ISO 25010, con el fin de establecer una herramienta que logre identificar datos de gran impacto para las organizaciones.

Palabras Claves: *Calidad de Software, Modelos Ontológicos, Educción de Requerimientos, Adecuación Funcional, Completitud Funcional.*

CONTEXTO

El presente es un trabajo realizado en conjunto entre el Grupo de Investigación sobre Aplicaciones Inteligentes (GISAI), de la Universidad Tecnológica Nacional Facultad Regional San Francisco bajo Resolución de CD N°1126/2015 y el Proyecto de Investigación: Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto PROICO Nº P-031516.

1. INTRODUCCIÓN

Las nuevas generaciones ya no encuentran diferencias entre la vida real y la virtual, vuelcan sin mesura toda su información a la red y ésta sin piedad, consume gran cantidad de datos, para nutrir distintos algoritmos inteligentes que aprenden de sus consumos y predicen preferencias. Esta realidad naturalizada, no solo interpela lo social sino distintos dominios; como industrias, organizaciones, educación, etc. Hoy, bajo el marco de la Cuarta Revolución Industrial, se van evidenciando cambios profundos (como toda revolución), los datos comienzan a jugar un

papel importante y primordial en esta transformación digital [1].

Los sistemas de información no son ajenos a esta revolución, es de ellos mismos, donde surgen los datos para futuras tomas de decisiones. Este avance no proclama nuevos ciclos de vida de desarrollo, ni nuevas técnicas que acompañen al mismo; pero llama a hacer un alto y observar qué funciones del dominio generan información que genere un valor agregado en la organización. Una vez más, el foco está en el análisis de requerimientos del software, etapa inicial en el proceso de desarrollo del mismo.

"Obtener los requisitos del sistema por medio de la observación de los sistemas existentes, discusión con los usuarios potenciales y proveedores, el análisis de tareas, etcétera, puede implicar el desarrollo de uno o más modelos y prototipos del sistema que ayuden al analista a comprender el sistema a especificar" [2]. El análisis de requerimientos basa su desarrollo en la actividad de elicitación de requerimientos, esta última, es la etapa de mayor interacción con el usuario. La extracción de requerimientos requiere el uso de distintas herramientas para construir y perfilar el alcance de un sistema [3]. Sobre esta tarea, se sientan los problemas y objetivos, necesidades y restricciones devenidos a requerimientos, que el analista debe saber interpretar y comunicar [4]. En el trabajo de investigación realizado, el 5% de los proyectos de software de las empresas PyMEs expresan los requerimientos a modo de Historias de Usuario (utilizan metodologías ágiles), con el formato COMO <rol> QUIERO <evento> PARA <funcionalidad>, el resto utiliza texto plano para la especificación.

El tipo de lenguaje utilizado, el uso de homónimos en contextos poco definidos, son algunos de las peligros que se presentan en la etapa de elicitación de requerimientos; una herramienta que proclama y proporciona un vocabulario común son las ontologías [5]. La función de la ontología es facilitar un entendimiento común del conocimiento a los miembros de un equipo de desarrollo de software, "una especificación formal, explícita de una conceptualización compartida" [7,8]. Por las características enunciadas, se propone un modelo ontológico que sirva de base para la elicitación.

Para que el modelo sea capaz de garantizar una descripción completa del dominio que se analiza, se contempló la incorporación, en el desarrollo de la herramienta, a la norma de calidad ISO/IEC 25010. El modelo de calidad definido por la norma se encuentra compuesto por 8 características: Adecuación Funcional, Eficiencia de Desempeño, Compatibilidad, Usabilidad, Fiabilidad, Seguridad, Mantenibilidad y Portabilidad. Una sub característica de la Adecuación Funcional, es la Completitud Funcional, grado en el cual el conjunto de funcionalidades cubre todas las tareas y los objetivos del usuario especificados[9,10]. El modelo bajo estudio, busca garantizar que el modelo ontológico cumpla la Completitud Funcional adaptándose a cualquier dominio que se aplique.

El desarrollo de una ontología se puede realizar a través de distintas metodologías [4,7]. Así, para el presente desarrollo se utilizó el Método 101, propuesto por Natalya F. Noy y Deborah L. McGuinness. [11,12].

El método propone enumerar los términos del dominio con el que vamos a trabajar, en este caso el producto de software. Estos términos y propiedades, conformaran las clases y las jerarquías del modelo.

Se definen las clases: Actor, Función Principal, Función de Apoyo, Interesado,

Organización, Producido, Rol, Rubro, Documentos; sobre este punto surgieran distintos modismos para referenciar el mismo objeto, esto permitirá establecer políticas sobre documentos y lograr un entendimiento por parte de todos los actores que participen del proyecto. Luego se definen las propiedades de las clases, sus relaciones.

Por último, se comienza con la etapa de implementación de la herramienta, que contempla la generación de instancias y etapa de evaluación del modelo. Se busca evaluar el modelo en distintos tipos de proyectos (gestión, control, sitios web).

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Siguiendo la línea de investigación correspondiente a lo expuesto en este trabajo, se llevan a cabo actividades relacionadas con las siguientes áreas temáticas:

- Ingeniería de Software.
- Calidad de Software.
- Ciencia de Datos.

El presente proyecto de investigación cuenta con varias etapas: 1) Estado del arte de la Ontología, 2) Análisis de la norma de calidad ISO 25010 para garantizar la Adecuación Funcional; 4) Desarrollo de un modelo ontológico para la elicitación de requerimientos del software; 5) Validación de la propuesta ontológica.

3. RESULTADOS ESPERADOS

Para el desarrollo del modelo ontológico se utilizó la herramienta Protégé, la misma es de código abierto [12].

Existen distintas metodologías y herramientas para el desarrollo de los modelos ontológicos, sin embargo, el resultado final debe ser consecuente a los requisitos que la propiciaron [13]. La calidad debe medirse evaluando; por este motivo, varios autores establecieron distintos criterios de evaluación, como ser, evaluar: Taxonomía, Lenguaje, Aplicación, Vocabulario, Arquitectura, Requerimientos, Aceptación Social, Razonamiento Automático y Software [14]. Otro método para la evaluación es el propuesto por [13]; quienes establecen el Método de Evaluación Triangular (Figura 1), en el mismo se integran buenas prácticas de otros métodos existentes.

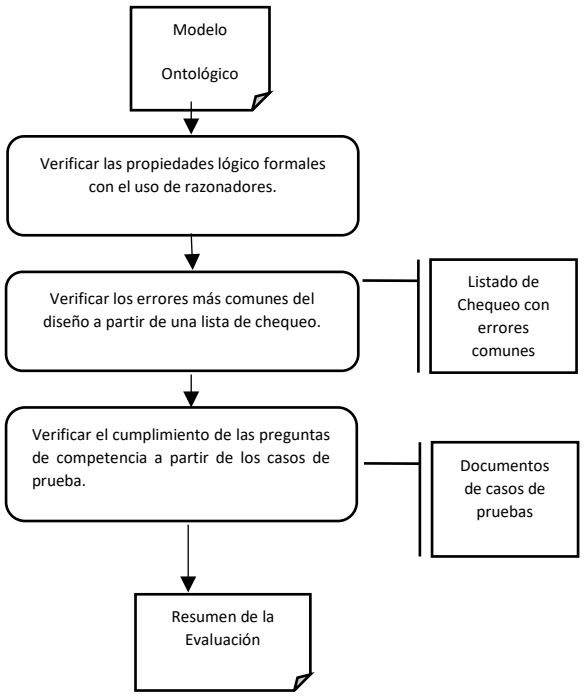


Figura 1: Método para evaluación de Ontologías[13]

Al desarrollar el modelo con la herramienta Protégé, la validación de las propiedades lógico formales, se establecerán con los Razonadores de la herramienta HermiT 1.3.6[14]. Para atacar los errores de diseño, se entregará a las empresas, un listado con los posibles errores a

encontrar, como, por ejemplo: uso excesivo de la relación es-un; existencia de más de un concepto principal; existencia de términos repetidos; falta de estandarización, claridad conceptual, facilidad de entendimiento, facilidad de trazabilidad de requerimientos, donde se dará una valoración a cada uno de estos puntos. Por último, y con el apoyo de las actividades anteriores, se deberá probar en el modelo las siguientes preguntas de competencia: ¿Cuál es el objetivo de la organización?, ¿Qué tipo de productos o servicios comercializa?, ¿Quiénes estarían interesados en relacionarse con la organización?, ¿Que funciones lleva a cabo?, ¿Con qué actores voy a trabajar?, ¿Todos los actores tienen asignada una función?, ¿Cada función tiene un actor asignado? ¿Las funciones de apoyo, son utilizadas por las funciones principales? ¿Existen funciones abiertas (Aquellas que no generan producidos)? ¿Las funciones enunciadas corresponden al objetivo de la organización?

El objetivo de esta validación es garantizar que la herramienta permita la extracción de todas las funciones que debe realizar la organización. Como lo establecieran otros autores, es necesario validar la metodología de construcción y la herramienta/modelo resultante [15].

Las acciones a seguir; además de continuar validando el modelo presentado, es la posibilidad de aunar otra ontología de dominios existentes al modelo para inferir nuevo conocimiento [16]. De esta forma, dar comienzo a herramientas de base inteligente, para obtener información que de soporte a la toma de decisiones del negocio.

4. FORMACIÓN DE RECURSOS HUMANOS

En el marco de esta línea de investigación, se está desarrollando una tesis para la Maestría en Calidad de Software dictada por la Universidad de San Luis bajo Resolución ME N° 1589/13. En el ámbito de dicho proyecto se propone la investigación sobre el impacto del modelo ontológico para la educción de requerimientos del Software basado en la Norma de Calidad 25010, con el objetivo de ofrecer completitud funcional. En el ámbito de la Inteligencia Artificial y Ciencia de Datos, la etapa de análisis de requerimientos es fundamental para el desarrollo del producto de software, tener identificadas todas las funciones y poner énfasis en aquellas que generen “datos”; fuente de información para las distintas herramientas de aprendizaje automático como regresión o CART [17], son el desafío bajo estudio.

En forma paralela, estamos trabajando en el grupo de investigación CIPI 4.0 (Capacitación, Investigación, Paradigma, Industria 4.0) analizando el nivel de digitalización de las empresas de nuestra zona. Valiéndonos de la información adquirida por distintos talleres y autodiagnósticos, logramos analizar en qué nivel de digitalización se encuentran las organizaciones, según el Modelo Acatech [18]; cuál de ellos genera bases para implementar herramientas predictivas.

Como iniciativa de este trabajo, se prevé la capacitación y formación de recursos humanos, por medio de las siguientes actividades:

- Transferencia de conocimiento y resultados a otras áreas de la facultad y a la industria local.
- Incorporar el conocimiento adquirido en las cátedras referentes a la temática planteada.
- Convocar e introducir a los alumnos de las carreras de Ingeniería en Sistemas de Información, a la realización de actividades de investigación y desarrollo.

5. BIBLIOGRAFÍA

[1] K. Schwab, *La Cuarta Revolución Industrial*. *Penguin Radom House.* ISBN: 978-84-9992-699-5 2016.

[2] I. Sommerville, *INGENIERÍA DE SOFTWARE. Novena edición*. Pearson Education. ISBN: 9786073206037. 2011.

[3] M. A. Chaves, "La ingeniería de requerimientos y su importancia en el desarrollo de proyectos de software," *InterSedes Rev. las Sedes Reg.*, vol. VI, no. 10, pp. 1–13, 2005.

[4] M. Griselda Báez and S. I. B. Brunner, "Metodología DoRCU para la Ingeniería de Requerimientos," pp. 210–222.

[5] C. M. Z. Jaramillo, G. L. Giraldo, and G. A. U. Giraldo, "Las ontologías de la ingeníeria de software: un acercamiento de dos grandes áreas del conocimiento," *Rev. Ing. Univ. Medellin*, vol. 9, no. 16, pp. 91–99, 2010, doi: 10.1017/CBO9781107415324.004.

[6] G. N. Aranda and F. Ruiz, "Clasificación y ejemplos del uso de ontologías en Ingeniería del Software," *XI Congr. Argentino Ciencias la Comput.*, 2005.

[7] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications," 1993.

[8] F. Piattini Velthuis, Mario G.; García Rubio, Félix O.; García Rodriguez de Guzmán, Ignacio; Pino, *Calidad de Sistemas de Información*. 2012.

[9] J. A. M. López, "Adaptación del proceso de desarrollo software para cumplimiento de la adecuación funcional según ISO/IEC 25000," 2017.

[10] Natalya F. Noy and Deborah L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," *Sustain.*, vol. 9, no. 12, pp. 1–25, 2017, doi: 10.3390/su9122317.

[11] O. Lassila and D. Mcguinness, "The Role of Frame-Based Representation on the Semantic Web Introduction : Frame-based Representation Systems," 1992.

[12] S. University, "Protégé." [Online]. Available: https://protege.stanford.edu/.

[13] Y. A. López Rodriguez, Y. Hidalgo-delgado, and N. Silega, "Un método práctico para la evaluación de ontologías," no. May, 2018.

[14] R. C. Esmeralda Ramos, Haydemar Núñez, "Esquema para evaluar ontologías únicas para un dominio de conocimiento," *Art War*, vol. 6, no. 3, p. 32 ، 117 ص, 2007, doi: 10.23943/9781400889877.

[15] M. Marciszack, M. Pérez, L. Antonelli, and M. Cardenas, "Construcción de una ontología utilizando Protégé para la elicitación de requerimientos," p. 8, 2009.

[16] N. S. M. Yoan Antonio López Rodríguez1*, Yusniel Hidalgo Delgado1, "Método para la integración de ontologías en un sistema para la evaluación de créditos," *Revista Cubana de Ciencias Informáticas*, vol. 10, no. 4. pp. 97–111, 2016.

[17] C. E. Timón, ""Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso," *26/27*, vol. I, no. Principio activo y prestación ortoprotésica, p. 67, 2017.

[18] M. Jacquez-Hernández and V. López, "Modelos de evaluación de la madurez y preparación hacia la Industria 4.0: una revisión de literatura," *Año*, vol. 11, no. 20, pp. 61–78, 2018.

Construcción de grafos de glosarios guiada por el estilo del discurso

Marcela N. Ridao[1], Jorge H. Doorn[2,3], Gladys N. Kaplan[4]

[1] Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires
[2] Escuela de Informática, Universidad Nacional del Oeste
[3] Departamento de Ingeniería, Universidad Nacional de Tres de Febrero
[4] Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de La Matanza
mridao@exa.unicen.edu.ar, jdoorn@uno.edu.ar, gkaplan@unlam.edu.ar

RESUMEN

La visualización de grafos es un recurso cada vez más utilizado en diferentes dominios de las ciencias de la computación. En la Ingeniería de Requisitos (IR), esta estrategia permite detectar núcleos semánticos específicos en algunos de los modelos utilizados, en forma más eficaz y eficiente que el mero estudio de los documentos. El uso concreto de grafos en glosarios del Universo del Discurso (UdeD)[1] ha permitido detectar una influencia del estilo de la narrativa de la fuente de información, en los grafos resultantes. Aparentemente, si se ignora esa influencia, la misma se transforma en un factor perturbador que deteriora la calidad de los grafos obtenidos. En el presente proyecto se planifica aprovechar el conocimiento del estilo de la narrativa de la fuente de información, para construir grafos que permitan visualizar más eficazmente la información contenida en el glosario.

Palabras clave: Ingeniería de Requisitos. Complejidad estructural, Grafos, Ubicación de nodos, Estilo del discurso

CONTEXTO

La propuesta que se presenta es este trabajo forma parte del proyecto de investigación "Desarrollo y Aplicación de Técnicas de Extracción de Información en Data Science" de la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN), del proyecto de investigación "Incorporación Pragmática de Visiones Lingüístico-cognitivas en el Proceso de Requisitos" de la Universidad Nacional del Oeste (UNO) y del proyecto de investigación "Aspectos no funcionales en los procesos de requisitos" de la Universidad Nacional de La Matanza (UNLaM).

1. INTRODUCCIÓN

Las disciplinas dedicadas al estudio de fenómenos donde el aspecto dominante es la complejidad estructural y no la complejidad esencial de los elementos involucrados, han recurrido crecientemente al uso de grafos para visualizar, y de esa manera estudiar esta complejidad estructural [3] [5]. Existen numerosos ejemplos, en diversas áreas de la ciencia, donde la detección de agrupamientos y la relación entre los mismos, representan una contribución significativa a la mejor comprensión del fenómeno estudiado [19] [21] [22]. En particular, la detección de agrupamientos es utilizada con diversos fines en la Ingeniería de Requisitos. Por ejemplo, en [6] los requisitos se agrupan en clusters, con el objetivo de priorizarlos.

Algunos de los modelos de la Ingeniería de Requisitos pueden ser estudiados desde el punto de vista estructural. Este trabajo se basa en un proceso de IR, que utiliza modelos construidos en Lenguaje Natural. Estos modelos son: el Léxico Extendido del Lenguaje (LEL), los Escenarios Actuales y Futuros (EA y EF), el LEL de Requisitos (LELr), y la Especificación de Requisitos (ERS). Existen

[1] UdeD: todo el contexto en el cual el software se desarrolla e incluye todas las fuentes de información y todas las personas, relacionadas con el software. Es la realidad acotada por el conjunto de objetivos establecidos por quienes demandan una solución de software [14].

vínculos explícitos e implícitos dentro de los modelos y entre ellos [11]. Por esto, cada modelo podría ser visto como un grafo, donde los nodos son los elementos del modelo, y los arcos los vínculos entre ellos. Más aún, el conjunto formado por dos o más modelos podría ser representado mediante un único grafo, considerando las conexiones entre los elementos de los modelos como arcos.

Dentro de los modelos mencionados, el que se ha estudiado desde este punto de vista, hasta el momento, es el LEL [15]. Este modelo registra el vocabulario del UdeD, mediante la descripción de los términos utilizados por el cliente-usuario, postergando la comprensión del problema. En la descripción de los símbolos se debe maximizar el uso de otros símbolos del léxico. De esta manera, el conjunto de símbolos determina una red, que puede ser navegada para conocer todo el vocabulario del dominio.

Si se observa un LEL bajo la óptica estructural se puede construir un grafo donde los símbolos son los nodos y las menciones a otros símbolos, arcos dirigidos. Desde este punto de vista, el LEL puede visualizarse como una suerte de red lingüistica con una estructura compleja. Es así que, además de la información explícita almacenada en cada uno de los nodos, existe una información implícita empotrada en la estructura de las relaciones entre ellos. Epistemológicamente, este enfoque es muy similar al utilizado en minería de datos [2], en el sentido que se hace visible información oculta mediante el uso de una técnica notoriamente diferente a la utilizada comúnmente para la lectura del modelo.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las fuentes de información que se utilizan para construir el LEL tienen un estilo narrativo propio. Ese estilo es parte de la cultura organizacional y del propio proceso del negocio. Es así que el LEL resultante depende en cierta medida de ese estilo narrativo y todo estudio que se realice sobre el modelo debería considerarlo como un factor relevante.

Estilo de la Narrativa

Aun con una observación superficial, es posible percibir que el estilo del discurso de la información adquirida en diferentes casos y de diferentes orígenes, suele tener diferencias relevantes. Estas diferencias dependen de la fuente de información en si misma, pero también de la propia estructura del UdeD.

Sin ser los únicos, los principales estilos son: orientado a los procesos, orientado a las entidades y sus relaciones, orientado a los productos y orientado a los actores.

Obviamente, la importancia de los diferentes símbolos del LEL y de sus relaciones depende del estilo del discurso y por lo tanto el grafo que se construya debe procurar destacar los símbolos de mayor importancia. Por ejemplo, en un sistema de apoyo a la gestión de pacientes agudos, la narrativa estará centrada en los protocolos o procesos, y por lo tanto, la construcción de los grafos deberá centrarse en la visualización lo más clara posible de los verbos.

Trabajos previos [16] [17] [18] han permitido detectar la presencia de agrupamientos en el grafo construido a partir del LEL, mediante técnicas automáticas de trazado de grafos. Sin embargo, los agrupamientos detectados no muestran las relaciones entre actores, procesos u objetos. Se propone, entonces, modificar el algoritmo de detección de agrupamientos, poniendo énfasis en dichas relaciones.

Trazado de Grafos: Métodos dirigidos por fuerzas

La Teoría de Grafos tiene diversidad de aplicaciones, y es usada en la representación de circuitos eléctricos, carreteras, moléculas orgánicas, ecosistemas, relaciones sociológicas, y muchas áreas de las ciencias de la computación [4] [9] [10]. El Trazado de Grafos, como una rama de la Teoría de Grafos, aplica topología y geometría para derivar representaciones de grafos en dos dimensiones.

La generación automática del trazado de grafos tiene importantes aplicaciones en muchas de las ciencias antes mencionadas, y tiene gran impacto en la visualización de información en general (por ejemplo diagramas de flujo, mapas esquemáticos o toda clase de diagramas) [12].

La amplia variedad de familias de grafos ha hecho que los algoritmos de trazado desarrollados varíen según el tipo de grafos que permiten visualizar. Entre ellos, existen algoritmos para trazado de grafos generales, destacándose una familia de métodos conocidos como "dirigidos por fuerzas". Estos métodos dan buenos resultados, son sencillos de implementar, y son muy flexibles, por lo que pueden ser fácilmente adaptados a aplicaciones concretas con requerimientos de visualización específicos [1] [13] [20].

Los métodos de trazado de grafos dirigidos por fuerzas modelan al grafo como un sistema físico, y el trazado es obtenido buscando el equilibrio del mismo. Los modelos físicos más comunes son los que consisten en un sistema de fuerzas que actúan entre los vértices del grafo. El objetivo es encontrar un equilibrio para este sistema de fuerzas, es decir, una posición para cada vértice, de manera que el total de la fuerza ejercida en cada vértice sea cero.

Entre los primeros autores aplicando analogías con sistemas físicos para el trazado de grafos, se destaca el "Spring Embedder" propuesto por Eades [7], que se basa en reemplazar los nodos por anillos de acero y cada arco con un resorte para formar un sistema físico. Los nodos son ubicados en una disposición inicial arbitraria, y se dejan actuar las fuerzas de los resortes hasta lograr un estado de equilibrio.

Fruchterman y Reingold [8] proponen un método derivado principalmente del Spring Embedder, basado en la física de partículas. Los nodos ejercen fuerzas de atracción y de rechazo sobre los demás, que inducen movimiento. El método propone que sólo los nodos conectados se atraigan entre sí, mientras todos los vértices se repelan unos a otros.

Fuerzas dirigidas en la visualización del LEL

Con el fin de detectar agrupamientos de símbolos, se aplicó una modificación del algoritmo propuesto por Fruchterman y Reingold [8] en la visualización de los grafos correspondientes a los Léxicos de diferentes casos de estudio. Para ello, cada símbolo del LEL fue representado mediante un nodo, y las menciones a otros símbolos incluidas en su definición, se representaron como arcos dirigidos a los nodos respectivos.

En el algoritmo, los nodos son ubicados al azar en el marco de trabajo, y posteriormente se va modificando su ubicación en forma iterativa.

- Se calcula el efecto de las fuerzas de atracción sobre cada nodo.
- Se calcula el efecto de las fuerzas de rechazo.
- Se calculan las nuevas posiciones de los nodos.

La nueva posición se calcula desplazando los nodos una distancia proporcional a la fuerza neta resultante sobre cada uno de ellos.

En trabajos anteriores [16] [17] [18], se utilizaron tres conjuntos de fórmulas diferentes para fa (fuerza de atracción) y fr (fuerza de rechazo). En la Tabla 1 se presentan las fórmulas utilizadas en cada caso.

Fruchterman - Reingold [8]	Eades [7]	Ridao - Doorn [18]
$fa(d) = \frac{d^2}{k}$ (1) $fr(d) = \frac{k^2}{d}$ d: distancia entre los vértices k: radio vacío alrededor de un nodo.	$fa(d) = c1 * \log\left(\frac{d}{c2}\right)$ $fr(d) = c3 * \sqrt{d}$ (2) d: distancia entre los vértices c1, c2, c3: ajustadas experimentalmente, de acuerdo a sus efectos sobre la visualización de los agrupamientos.	$fa(d) = c1 * \log\left(\frac{d}{c2}\right)$ $fr(d) = d^{\alpha}$ (3) d: distancia entre los vértices c1, c2: ajustadas experimentalmente α: parámetro que permite modificar la relación entre fr y fa.

Tabla 1. Fórmulas utilizadas en [8], [7] y [18]

La aplicación del primer sistema de fuerzas permitió detectar agrupamientos de símbolos para algunos de los casos analizados. Con las fórmulas propuestas por Eades [7], los clusters que se habían detectado con (1) se visualizaron de una forma mucho más clara, y se comenzaron a detectar grupos para casos de

estudio donde las pruebas con el primer sistema de fuerzas no arrojaban resultados claros.

Con el fin de obtener mejores representaciones, se propusieron variaciones a las fórmulas originales [18]. Para ello, se consideró la relación entre las magnitudes de fr y fa. La idea central fue hacer crecer más lentamente la fuerza de rechazo con la distancia, suponiendo que esto traería como consecuencia una mayor cercanía entre los nodos de un cluster. Con estas fórmulas, y una adecuada relación entre las magnitudes de las fuerzas, la visualización de los agrupamientos se hizo mucho más clara.

En las figuras 1 y 2 se observa la evolución en la visualización de los agrupamientos para dos de los casos de estudio analizados.

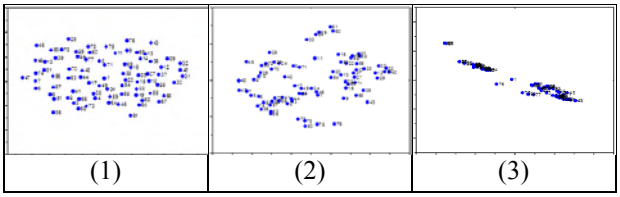


Fig. 1. Distribución de nodos del caso LEL y Escenarios, luego de aplicar los tres conjuntos de fórmulas

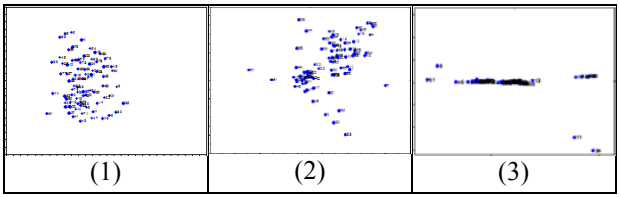


Fig. 2. Distribución de nodos del caso Biblioteca, luego de aplicar los tres conjuntos de fórmulas

En casos donde se observaba la presencia de grupos, pero con límites poco claros, la aplicación de las fuerzas (3) permitió visualizarlos con claridad. En otros, donde no se observaban agrupamientos, se pudo comprobar que estos existían.

La disposición de los grupos en el grafo resultante permite, además, mostrar la relación o falta de ella, entre los agrupamientos. Es así que, para algunos casos de estudio, se observó una alineación entre los clusters, formando una suerte de cadena que denota la ausencia de relación entre los grupos de los extremos, y una relación estrecha entre los grupos vecinos.

3. RESULTADOS ESPERADOS

Se espera mejorar la visualización del grafo asociado al LEL, incorporando la información del estilo de la narrativa como parámetro. Es así que, en forma preliminar se proponen dos tipos de grafos: a) orientados a sujetos u objetos y b) orientados a verbos.

Orientados a sujetos u objetos

A continuación, se presentan sólo las reglas que se proponen para visualización centrada en sujetos. Meramente intercambiando sujetos y objetos, se pueden obtener las reglas para la visualización centrada en objetos.

Regla 1: Ignorar los símbolos verbos.

Regla 2: Si existe una referencia de un sujeto a otro sujeto, partiendo desde la noción, incrementar por un factor M_1 la fuerza de atracción.

Regla 3: En toda otra relación entre sujetos, incrementar la fuerza de rechazo por un factor M_2.

Regla 4: Si existe una referencia de un objeto a otro objeto, partiendo desde la noción, incrementar por un factor M_3 la fuerza de atracción.

Regla 5: Si existe una referencia de un estado a un sujeto, partiendo de la noción, incrementar la fuerza de atracción por un factor M_4.

Regla 6: En toda otra relación entre estados y sujetos, incrementar la fuerza de rechazo por un factor M_5.

Regla 7: Si existe una referencia de un estado a un objeto, partiendo de la noción, incrementar la fuerza de atracción por un factor M_6.

Regla 8: En toda otra relación entre estados y objetos, incrementar la fuerza de rechazo por un factor M_7.

Regla 9: Si existe una referencia transitiva entre sujeto y objeto o viceversa, a través de un verbo considerarla como una referencia directa.

Orientados a verbos

Regla 1: Si existe una referencia de un sujeto a otro sujeto, partiendo desde la noción, incrementar por un factor M1 la fuerza de atracción.

Regla 2: Si existe una referencia de un objeto a otro objeto, partiendo desde la noción, incrementar por un factor M2 la fuerza de atracción.

Regla 3: Si existe una referencia de un verbo a otro verbo, partiendo desde la noción, incrementar por un factor M3 la fuerza de atracción.

Regla 4: En toda otra relación entre verbos, incrementar la fuerza de rechazo en un factor M4.

Se planifica determinar los mejores valores para los factores M_i. Las primeras pruebas se realizarán utilizando el mismo valor para todos los factores indicados.

4. FORMACIÓN DE RECURSOS HUMANOS

El presente proyecto es parte directa de la tesis doctoral de la Mag. Marcela Ridao y contribuye a la tesis doctoral de la Mag. Gladys N. Kaplan.

5. BIBLIOGRAFÍA

[1] Aiello, A., Silveira, A.: Trazado de grafos mediante métodos dirigidos por fuerzas: revisión del estado del arte. Tesis de Licenciatura en Ciencias de la Computación. Departamento de Computación. Facultad de Ciencias Exactas y Naturales. UBA. (2004)

[2] Artz, J.M. "Data Driven vs. Metric Driven Warehouse Design" en John Wang "Encyclopedia of Data Warehousing and Mining, ISBN 1-59140-557-2, pag 223 a 227, tomo I, Idea group reference. (2006)

[3] Barabasi, A.: Linked, The New Science of Network. Perseus publishing. (2002)

[4] Brandes, U., Kenis, P., Wagner, D.: Communicating centrality in policy network drawing. IEEE Trans. on visualization and computer graphics, 9(2), 241-253. (2003)

[5] Dorogovtsev, S., Mendes, J.: Evolution of networks: From biological nets to the Internet and WWW. Oxford University Press, Oxford. (2003)

[6] Duan, C., Laurent, P., Cleland-Huang, J., y Kwiatkowski, C., Towards automated requirements prioritization and triage, Req. Engineering Journal, 14(2), 73-89. (2009)

[7] Eades, P.: A heuristic for graph drawing. Congressus Numerantium 42, 149–160. (1984)

[8] Fruchterman, T., Reingold, E.: Graph Drawing by Force-directed Placement. Software-Practice and Experience 21(11), 1129-1164. (1991)

[9] Gross, J., Yellen, J.: Editors. Handbook of Graph Theory. CRC Press. (2003)

[10] Gross, J., Yellen, J.: Editors. Graph Theory and Its Applications, Second Edition (Discrete Mathematics and Its Applications). Chapman& Hall/CRC. (2006)

[11] Kaplan, G., Doorn, J., Gigante, N. Evolución semántica de glosarios en los procesos de requisitos. XIX Congreso Argentino de Ciencias de la Computación, CACIC. Mar del Plata, Argentina. (2013)

[12] Kaufmann, M., Wagner, D.: (eds.) Drawing graphs: methods and models, LNCS, vol 2025. Springer-Verlag. (2001)

[13] Kobourov, Stephen G.: Spring Embedders and Force-Directed Graph Drawing Algorithms, arXiv:1201.3011. (2012)

[14] Leite, J., Doorn, J., Kaplan, K., Hadad, G., Ridao, M.: Defining System Context Using Scenarios. In: Leite, J., Doorn, J (eds.) Perspectives on Software Requirements. Kluwer Academic Press, pp. 169-199. (2004)

[15] Leite, J., Franco, A.: O Uso de Hipertexto na Elicitaçao de Linguagens da Aplicaçao. IV Simpósio Brasilero de Engenharia de Software, SBC, pp. 134-149. Brazil. (1990)

[16] Ridao, M., Doorn, J.: Semántica Oculta en Modelos de Requisitos. XV Workshop de Investigadores en Ciencias de la Computación, WICC. Paraná, Argentina. (2013)

[17] Ridao, M., Doorn, J.: Agrupamientos en Glosarios del Universo de Discurso. Tecnología y Ciencia - Revista de la Universidad Tecnológica Nacional - Edición Especial: CoNaIISI 2014, 13(27), 5-16. (2015)

[18] Ridao, M., Doorn, J., Visualización de Núcleos Semánticos en Glosarios del Universo de Discurso. IV Congreso Nacional de Ingeniería en Informática/sistemas de Información, CoNaIISI. Salta, Argentina (2016)

[19] Rosy Das, Jugal Kalita, Dhruba K. Bhattacharyya, A pattern matching approach for clustering gene expression data, Int. J. Data Mining, Modelling and Management, 3(2), 130-149 (2011)

[20] Walshaw, C.: A multilevel algorithm for force-directed graph-drawing. Journal of Graph Algorithms and Applications 7(3), 253-285 (2003)

[21] Yijun Mo, Zuo Cao, Bang Wang, Occurrence-Based Fingerprint Clustering for Fast Pattern-Matching Location Determination. Communications Letters, IEEE 16(12), 2012 - 2015 (2012)

[22] Zimmermann, M., Ntoutsi, I., Siddiqui, Z., Spiliopoulou, M., Kriegel, H.P. Discovering Global and Local Bursts in a Stream of News. 27th Annual ACM Symposium on Applied Computing. SAC '12, pp. 807-812. Italy (2012).

Plataforma para la definición y ejecución de actividades orientadas a la recolección y análisis de datos, con intervención humana

Jose Arcidiacono [1], Alejandra B. Lliteras [2,4], Patricia Bazán [3]
[1] UNLP, Facultad de Informática, [2] UNLP, Facultad de Informática, LIFIA [3] UNLP, Facultad de Informática, LINTI, [4] CICPBA
josefine2_94@hotmail.com, alejandra.lliteras@lifia.info.unlp.edu.ar, pbaz@info.unlp.edu.ar

Resumen

En este trabajo se propone una plataforma para que usuarios finales definan y ejecuten actividades orientadas a la recolección y análisis de datos. Dichas actividades serán implementadas mediante el concepto de workflow. La plataforma se basará en una arquitectura distribuida basada en microservicios.

Palabras Claves: Recolección de Datos - Workflow – HCI (Human Computer Interaction) - Arquitectura Distribuida – Microservicios - Análisis de Datos

Contexto

El presente trabajo surge a raíz de una iniciativa del Proyecto de Extensión de la Facultad de Informática, llamado "Recicla tu Compu-Recicla tu Mundo"[1] [1], con el fin de generar herramientas basadas en el uso de la tecnología para provocar en niños, jóvenes y adultos la apropiación de los conocimientos fundamentales que ayuden en los procesos educativos y de concientización en la temática abordada en dicho proyecto.

Para dar respuesta a esta iniciativa se tomaron como base los resultados de los proyectos de investigación presentados en [2] y [3] donde los autores aplicaron conceptos de Ingeniería de Software en conjunto con la Interacción Humano-Computador (entre otras disciplinas), para proponer diferentes estrategias que permitieran llevar adelante actividades (mediadas por tecnología) con intervención humana, en contextos educativos para llevar adelante lo que se conoce como *Aprendizaje Móvil* [4] y, adicionalmente, para ser utilizado como soporte en la recolección y análisis de datos en lo que se conoce como *Ciencia Ciudadana* [5], [6].
Lo propuesto en [2] y [3] generó un enfoque para trabajar con actividades que implican intervención humana para ser llevadas adelante por personas y, en base al mismo, se propuso una herramienta web de autor [7] que permitió realizar la configuración de una aplicación móvil educativa con el objetivo de, en ese caso en particular, concientizar respecto a los Residuos de Aparatos Eléctricos y Electrónicos (RAEE), como experiencia de uso para un

[1] Aprobado y subsidiado en la convocatoria de Proyectos de Extensión Universitaria 2018, de la UNLP. El mismo está destinado a promover la educación ambiental sobre los RAEE en escuelas de la región en el período 2018 al 2019.

Programa de extensión de la UNLP llamado E-Basura.

En este contexto, se observó la necesidad de proponer una nueva herramienta que profundice y mejore lo propuesto en [7], promoviendo mayor robustez para dar soporte a más elementos del enfoque que la misma adopta como base ([2], [3]). Mediante esta propuesta, se espera favorecer la construcción de aplicaciones más generales y aptas para llevar adelante otro tipos de actividades que impliquen la intervención de personas para realizarlas, como por ejemplo la recolección y análisis de datos ya no solo en contextos educativos. Dicha herramienta seguirá los conceptos de reuso esperable para la construcción de este tipo de aplicaciones según lo propuesto en [8].

1. Introducción

Una actividad, de acuerdo a lo propuesto en [2], está formada por un conjunto de tareas (un ejemplo de tarea sería una pregunta de opción múltiple) planificadas de una manera particular (lineal, grafo o conjunto). En esta propuesta de trabajo, la planificación será implementada desde la perspectiva de lo que se conoce como *workflow*. Siguiendo la definición propuesta en [9] donde los autores presentan a un workflow como un conjunto de tareas que deben ser llevadas a cabo para realizar un trabajo, y en donde se plantea que las tareas pueden ser de diferentes tipos e involucrar diferentes formas de interacción por parte de las personas.

Por otro lado, se sabe que un sistema de manejo de workflow permite definir, crear y manejar la ejecución de flujos de trabajo a través del uso de software. Asimismo, la tecnología de workflow posibilita la representación de procesos en los que hay seres humanos activamente involucrados favoreciendo su colaboración.

En el marco de los sistemas de workflow existen ya numerosos estudios acerca de los denominados “patrones de workflow” [10], que fueron definidos para representar un conjunto completo de estructuras de control de flujo posibles.

Por su parte, según [11] se puede definir un subconjunto de patrones más frecuentemente usados. A continuación se los menciona:

- Secuencia
- Bifurcación en Paralelo (*and split*)
- Fusión de Sincronismo (*and join*)
- Selección Exclusiva (*exclusive or-split*)
- Selección Múltiple (*inclusive or split*)
- Fusión Simple (*exclusive or-join*)

En este sentido, la propuesta del presente trabajo incluye por un lado, patrones para Secuencia, Selección Exclusiva y Fusión Simple. y por otro, constructores para iniciar y finalizar el curso del flujo, incluyendo la posibilidad de contar con múltiples inicios para un mismo workflow.

El uso de workflows para actividades con intervención humana ha sido explorado en diversos dominios, por ejemplo en análisis de datos [9] donde los autores estudian aspectos relacionados a la experiencia de usuario a partir de diferentes formas de recorrer las tareas que componen un workflow. Otro uso de workflows se manifiesta en la recolección de datos por parte de personas y asistidos por tecnología [12]

En esta línea de investigación, se propone realizar una plataforma basada en un

sistema distribuido y con una arquitectura de microservicios, tal como se muestra en la Figura 1, donde se presentan las cuatro capas que la componen.
A continuación se detalla brevemente cada una de las capas presentadas en la Figura 1.
La Capa *de Clientes* (*Clients layer*) aloja las aplicaciones que permiten la interacción entre los usuarios y el resto de la plataforma. Existirán diferentes tipos de usuario: autores, usuario web y usuario móvil. Cada uno de ellos con acciones definidas acordes al perfil.

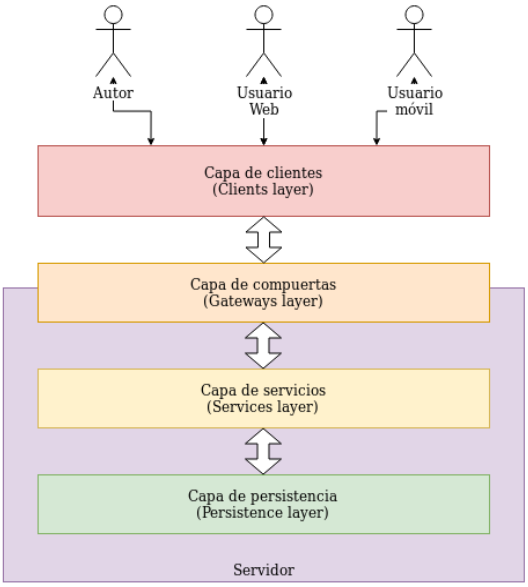


Figura 1 - Arquitectura propuesta.

Luego se presenta una *Capa de Compuertas (Gateways layer)* que incluye a los componentes tecnológicos que re-direccionarán las peticiones entre clientes y los servicios propiamente dichos, contenidos en la *Capa de Servicios (Services Layer)*.
Por último, la *Capa de Persistencia (Persistance Layer)* se ocupa de hacer permanentes los cambios producidos en los datos que gestiona el sistema.

2. Líneas de Investigación

Las líneas de investigación presentes en este trabajo son variadas y confluyen para la producción de la plataforma propuesta.
La plataforma presentada tiene características de una arquitectura distribuida, tanto en su definición, como en las soluciones que genera. La misma se basará en servicios e integrará diversas tecnologías. Durante su implementación, se llevarán adelante "*Test Cases*" tanto para el "backend" como para el "front end".
Adicionalmente se estudiarán aspectos de usabilidad de la plataforma y de la aplicación móvil que se genere a través de la misma. Aspectos de accesibilidad de la aplicación móvil serán considerados, Un ejemplo de esto es la importancia de contar con un lector de pantalla en la misma en caso de disminución visual.

La plataforma podrá ser usada como una herramienta de autor, para generar configuraciones de actividades que podrán ser ejecutadas desde una aplicación móvil o bien desde la misma plataforma web.

3. Resultados Obtenidos/Esperados

Entre los resultados esperados de este trabajo se encuentra el de presentar una plataforma mediante la cual sea posible definir y ejecutar actividades orientadas a la recolección y análisis de datos que requieran intervención humana. El desarrollo de esta plataforma permitirá a usuarios finales, diseñar y crear workflows de manera dinámica que puedan ser ejecutados tanto en tecnología móvil como web, y que contribuyan a actividades de recolección y análisis de datos.

4. Formación de Recursos Humanos

La diversidad de líneas de investigación enunciadas así como la gran variedad de aplicación de los resultados esperados, propone un foco de investigación claro y concreto para la formación de recursos humanos, que encuentran en esta propuesta diversas aristas de trabajo.

La propuesta aborda temas de arquitecturas distribuidas, ingeniería de software, HCI, accesibilidad y usabilidad web, así como aspectos relacionados a herramientas de autor, test cases, workflow con intervención humana, recolección y análisis de datos.

Adicionalmente, la propuesta encuentra su caso de aplicación en el Proyecto de Extensión de la Facultad de Informática "Recicla tu Compu-Recicla tu Mundo" [4].

En este sentido, el presente trabajo describe la línea de investigación que dará lugar a que en el año 2020 se formule una propuesta de tesina de grado para el diseño, construcción y aplicación de una plataforma que permita a usuarios finales obtener aplicaciones para llevar adelante experiencias con usuarios mediante actividades que involucren recolección y/o análisis de datos con intervención humana implementada con workflows.

5. Bibliografía

[1] Proyecto Recicla tu Compu-Recicla tu mundo. Proyecto de extensión de la UNLP, Facultad de Informática, LINTI. http://www.extension.info.unlp.edu.ar/articulo/2018/12/18/la_unlp_acredito_cinco_proyectos_de_extension_presentados_por_el_linti_2019

[2] Lliteras, A. B. (2015). *Un enfoque de modelado de actividades educativas posicionadas que contemplan elementos concretos*. Tesis de maestría. Facultad de Informática, UNLP.

[3] Lliteras, A. B., Grigera, J., dal Bianco, P. A., Corporaal, F. M., & Gordillo, S. E. (2018). Challenges in the design of a customized location-based mobile learning application. In *2018 XIII Latin American Conference on Learning Technologies (LACLO)* (pp. 315-321). IEEE.

[4] Traxler, J. (2009). Learning in a mobile age. International Journal of Mobile and Blended Learning (IJMBL), 1(1), 1-12.

[5] Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience,59(11), 977-984.

[6] Strasser, B. J., Baudry, J., Mahr, D., Sanchez, G., &Tancoigne, E. (2019). "Citizen Science"? Rethinking Science and Public Participation. Science & Technology Studies, 52-76.

[7] Dal Bianco, P. A., Mozzon Corporaal, F., Lliteras, A. B., Grigera, J., & Gordillo, S. E. (2019). MoLE: A web authoring tool for building mobile learning experiences. In *XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019, Universidad Nacional de Río Cuarto).*

[8] Lliteras, A. B., Challiol, C., & Gordillo, S. E. (2017). Location-based mobile learning applications: a conceptual framework for co-design. In *2017 Twelfth*

Latin American Conference on Learning Technologies (LACLO) (pp. 1-8). IEEE.

[9] Sprinks, J., Wardlaw, J., Houghton, R., Bamford, S., & Morley, J. (2017). Task Workflow Design and its impact on performance and volunteers' subjective preference in Virtual Citizen Science. International Journal of Human-Computer Studies, 104, 50-63.

[10] van Der Aalst, Wil MP, et al. "Workflow patterns." Distributed and parallel databases 14.1. Pág 5-51. (2003).

[11] Kunze, Matthias, and Mathias Weske. Behavioural models: From modelling finite automata to analysing business processes. Springer, 2016.

[12] Steinberg, M., Schindler, S., & Klan, F. (2019). Software solutions for form-based, mobile data collection–A comparative evaluation. *BTW 2019–Workshopband*.

Automatización de la Medición de Software para Flotas Dinámicas mediante un Modelo de Calidad Mixto para la Movilidad en Smart Cities

Alejandro Rivoira (+); Giselle Cavallera (+); Carlos Salgado (+); Alberto Fernández Gil (*); Alberto Sánchez (+); Mario Peralta (+)
(+) Departamento de Informática - Facultad de Ciencias Físico-Matemáticas y Naturales - Universidad Nacional de San Luis
Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina
e-mail: {rivoira.ale, giselle.cavallera}@gmail.com, {csalgado, alfanego, mperalta}@unsl.edu.ar

(*) Universidad Rey Juan Carlos
(*) Grupo de Inteligencia Artificial (GIA). Centro para las Tecnologías Inteligentes de la Información y sus Aplicaciones (CETINIA) – Madrid – España
e-mail: alberto.fernandez@urjc.es

RESUMEN

La movilidad se ha convertido en un asunto central del desarrollo urbano. Su relación con los temas de sostenibilidad y su capacidad para generar competitividad y calidad de vida nos sitúa ante la necesidad de replantear su futuro. Los nuevos sistemas de uso compartido de vehículos son parte del desafío que plantean las Smart Cities (Ciudades Inteligentes). Ya que el objetivo de los mismos es la mejora de la movilidad humana y la reducción de costos, su análisis y evaluación es de gran utilidad para lograr su optimización. Para ello, se debe pensar en un planteo más crítico de infraestructura y mejoras en los servicios de transporte público.

Por tanto, se ha definido un modelo de calidad mixto para software responsable de la gestión de la movilidad urbana con el objetivo de producir información cuantitativa sobre ciertas características de calidad para poder tomar decisiones de cambio o mejoras si fuesen necesarias. Como soporte a este modelo se ha completado el desarrollo de una herramienta que posibilita la medición y evaluación de calidad de este tipo de software. En estos momentos nos encontramos ampliando el conjunto de métricas e indicadores para aplicaciones móviles teniendo en mente las características de infraestructura, nuevas tecnologías e impactos en el medioambiente.

Palabras clave: Modelo de Calidad Mixto, Smart Cities, Flotas Dinámicas, Medición y Evaluación de Calidad Software.

CONTEXTO

El presente trabajo se enmarca en el proyecto de investigación: Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto Nº P-031516. Además, se encuentra reconocido por el programa de Incentivos.

La tarea se efectúa en forma colaborativa con el grupo de investigación de Inteligencia Artificial (GIA), de la Universidad Rey Juan Carlos de Madrid- España. Dicho proyecto es la continuación de diversos proyectos de investigación sobre la gestión de flotas dinámicas y calidad aplicada a sistemas software, a través de los cuales se ha logrado un importante vínculo con la mencionada universidad internacional.

1. INTRODUCCIÓN

El concepto ciudad~~es~~ inteligente~~s~~ nace como idea global de la gestión de los recursos de una ciudad dirigidos a mejorar la calidad de vida de los ciudadanos. También se puede describir como aquella ciudad que aplica las tecnologías de la información y de la comunicación (TIC) con el objetivo de proveerla de una infraestructura que garantice:

- Un desarrollo sostenible.

- Un incremento de la calidad de vida de los ciudadanos.
- Una mayor eficacia de los recursos disponibles.
- Una participación ciudadana activa.

Para determinar si una ciudad es o no inteligente, generalmente, se toma como punto de partida un modelo en torno a seis características: *Economía, Gestión de gobierno, Ciudadanía, Entorno, Calidad de vida y Movilidad*. A partir de este modelo, se han desarrollado objetivos agrupados en una serie de servicios para los que se han descrito las principales tecnologías que disponemos en la actualidad. De esta forma, se obtuvo una serie de plataformas de servicios, que se deben integrar y conectar entre sí, facilitando que el ciudadano forme parte activa del proceso de gestión de su ciudad [1]. En otras palabras, se busca modernizar la gestión de las ciudades, fomentando una mayor interacción entre las instituciones, los ciudadanos y los sistemas informáticos [2].

Actualmente vivimos en la convergencia de dos fenómenos importantes en la historia de la humanidad: la aceleración de la urbanización a nivel mundial y la revolución digital. De hecho, todas las aglomeraciones urbanas presentan diversos retos. Uno de sus retos consiste en abordar el proyecto de diseñar una ciudad inteligente manteniendo su ADN. Cada urbe tiene sus propias señas de identidad y sus propios valores que se expresan en las formas arquitectónicas, la configuración de los espacios y, sobre todo, las relaciones entre sus habitantes, y entre los ciudadanos y su gobierno [3]. Por ello, cada vez son más importantes la planificación urbana y el desarrollo de mecanismos de decisión dinámicos que tomen en cuenta el crecimiento y la inclusión de procesos de participación ciudadana.

El empleo masivo de productos software a escala global, prácticamente en todos los ámbitos del desempeño humano, favorecido en el último lustro por la enorme tasa de crecimiento de los servicios relacionados a la red Internet, han convertido a los usuarios en potenciales referentes para evaluar la calidad de productos de software. Además, una extraordinaria demanda de aplicaciones produjo un nuevo cambio en el desarrollo de software. La respuesta a esta demanda, por parte de la industria y la comunidad de desarrolladores fue importante en tecnología y escala de producción. Sin embargo, no es extraño encontrar hoy que, las cuestiones relacionadas a la calidad -especialmente la calidad que perciben los usuarios reales respecto a las aplicaciones software que emplean- han sido postergadas o directamente ignoradas, entre otras razones por las urgencias de entregar los productos al mercado, eventuales ahorros o falta de recursos humanos capacitados [4].

La preocupación por la calidad de los sistemas basados en software ha aumentado a medida que éste se integra en cada aspecto de la vida cotidiana. La dependencia de las organizaciones respecto de este recurso es crítica. La automatización de las actividades y la generación y disponibilidad de información para la toma de decisiones, logradas a través del software, son claves para el logro de los objetivos y supervivencia de las mismas [5]. Los fallos de software afectan a todos los sectores y a todos los países, como puede constatarse en la recopilación de incidentes documentados por Peter Neumann desde 1987 [6]. Lamentablemente, estos problemas pueden ser mucho más graves si ocurren en sistemas críticos, es decir, aquellos cuyo fallo puede provocar terribles pérdidas económicas o problemas ambientales o sociales e, incluso, la pérdida de vidas humanas.

Basados en lo que dice Tom Demarco [7], "No se puede controlar lo que no se puede medir", que después complementan Fenton y Pfleeger [8] "No se puede predecir lo que no se puede medir", es indispensable y altamente necesario para la vida de una ciudad inteligente poder realizar una gestión de calidad integral en los sistemas informáticos que interactúan en ella. Según la norma ISO 9000 [9], la calidad es el "grado en el que un conjunto de características inherentes cumple con los requisitos". Como los requisitos dependerán de las diferentes partes interesadas (stakeholders), la calidad es un

concepto multidimensional, ya que puede ser sinónimo de eficiencia, flexibilidad, corrección, confiabilidad, facilidad de mantenimiento, portabilidad, facilidad de uso, seguridad, integridad, etc. En la evolución experimentada por la calidad de los sistemas informáticos se ha pasado de un tratamiento centrado fundamentalmente en la inspección y detección de errores en los programas, a una aproximación más sistemática, dada la importancia que ha adquirido la calidad en la ingeniería de sistemas y en la ingeniería del software. La demanda de software por parte de las organizaciones y, en general, de la sociedad ha crecido mucho más deprisa que la capacidad de la industria para producir software de calidad, haciendo crónica la denominada "crisis del software" [10].

Una de las razones principales del incremento masivo en el interés por las métricas software ha sido la percepción de que son necesarias para la mejora de la calidad del proceso [11]. Para poder asegurar que un proceso o sus productos resultantes son de calidad, o poder compararlos, es necesario asignar valores, descriptores, indicadores o algún otro mecanismo mediante el cual se pueda llevar a cabo dicha comparación. Para ello, es necesario llevar adelante un proceso de medición del software, que en general, persigue tres objetivos fundamentales: ayuda a entender qué ocurre durante el desarrollo y el mantenimiento, permite controlar qué es lo que ocurre en los proyectos y otorga la posibilidad de mejorar los procesos y productos [8]. En efecto, las métricas son un buen medio para entender, monitorizar, controlar, predecir y probar el desarrollo de software y los proyectos de mantenimiento [12] y pueden ser utilizadas por profesionales e investigadores para tomar mejores decisiones.

El modelo de calidad representa la piedra angular en torno a la cual se establece el sistema para la evaluación de la calidad del producto. Cada modelo cuenta con características, subcaracterísticas y métricas asociadas. Las métricas son una medida cuantitativa del grado en que un sistema, componente o proceso posee un atributo dado [13]. Los indicadores, por su parte, son métricas o combinación de métricas que proporcionan una visión profunda, del proceso, del proyecto o del producto software [14]. Las métricas e indicadores son activos claves de una organización que proveen datos e información útiles para los procesos de análisis [15]. A su vez, también, son un buen medio para entender, monitorizar, controlar, predecir y probar el desarrollo del software y los proyectos de mantenimiento [16].

Para implementar las mediciones sobre el software de movilidad en una ciudad inteligente, fue necesario definir un método de evaluación y medición acorde a las necesidades propias del dominio de trabajo e investigación. El método es una combinación de buenas prácticas, métodos de evaluación y estándares para abordar lo que refiere puntualmente al paradigma de las ciudades inteligentes. Consta de 5 fases, las cuales se describen a continuación.

FASE 1: Seleccionar un modelo de calidad para Software de Movilidad: para poder hacer el análisis, estudio, medición y evaluación del software de movilidad había que adoptar un modelo de calidad, métricas e indicadores que permitiesen tener el marco donde poder hacer el estudio. Este modelo puede perfeccionarse, adaptarse, actualizarse o hasta cambiarse completamente si es necesario.

FASE 2: Aplicación/Definición de Métricas e Indicadores: una vez establecido y definido el Modelo de Calidad para aplicaciones de movilidad, se debe, como segundo paso:

- Calcular las métricas de cada atributo especificado en el Modelo de Calidad.
- Calcular los indicadores elementales.

FASE 3: Generación de la estructura de Agregación utilizando la lógica de operadores de la lógica continua para:

- Permitir detallar el peso y operador para cada atributo.
- Calcular los indicadores derivados.

FASE 4: Interpretar los datos/información y generar reportes.

FASE 5: Generar repositorios de información de cada proyecto de Medición y Evaluación (M&E) para análisis y

comparación posterior. Es decir, ir hacia la mejora continua.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

En base a lo explicado anteriormente, en esta línea de investigación se viene trabajando en la definición de un modelo de calidad mixto, contextualizado y adaptado, para medir y evaluar la calidad de un producto software responsables de la gestión de la movilidad, y por ende del tráfico, en una ciudad inteligente (específicamente Madrid). El modelo de calidad mencionado se denomina mixto porque, si bien tiene como base principal el Modelo ISO/IEC 25010, también se utilizaron indicadores o métricas del Modelo de Pressman, y algunas propias del dominio del problema. Además, se viene trabajando en el desarrollo de una herramienta software que permite la recolección de datos desde una aplicación de bicicletas compartidas llamada Ecobike, desarrollada dentro de un proyecto denominado Ecobike Solutions [17], perteneciente a alumnos en Ingeniería Informática de la Universidad Rey Juan Carlos de Madrid, España.

3. RESULTADOS Y OBJETIVOS

Hoy la tecnología nos brinda muchas herramientas para solucionar las problemáticas de las ciudades. Pero la preocupación está puesta en la calidad de esos sistemas basados en software, que ha aumentado a medida que estos se integran en cada aspecto de nuestras vidas cotidianas. Por ello, para poder gestionar y mejorar las aplicaciones software se deben controlar a través de mediciones y evaluación continua. Es así que se hace indispensable contar con herramientas que den soporte tecnológico a este proceso de medición y evaluación, para alivianar tareas, disminuir errores y obtener resultados adecuados de manera eficiente y eficaz.

Entre los resultados obtenidos que se proponen, se encuentra un modelo de calidad mixto que intenta determinar el impacto de las deficiencias en un agente software dirigido a la gestión de la movilidad en ciudades inteligentes, teniendo como base la ciudad de Madrid. En relación a esto, se propondrán algunos elementos de solución que ayuden a crear un pensamiento consciente en las personas/usuarios.

A su vez se propone la herramienta que mide este tipo de software aplicando dicho modelo.

La aplicación del modelo en el caso de estudió sirvió para validar el modelo en un caso práctico real y para ver una tendencia en el resultado de la evaluación que sirva de análisis global

Se espera cumplir algunos objetivos adicionales en trabajos futuros:

1) Ampliar el conjunto de métricas e indicadores para aplicaciones móviles teniendo en mente las características de infraestructura, nuevas tecnologías e impactos en el medioambiente.
2) Dotar a la herramienta de M&E de un usuario administrador que pueda redefinir y adaptar métricas nuevas provenientes de los nuevos escenarios por venir.
3) Proveer una base de información, a partir de las definiciones, mediciones y evaluaciones logradas, para automatizar la medición en tiempo real a través de un sistema informático.
4) Brindar satisfacción a usuarios actuales, y generar atracción a nuevos usuarios a través de:
    a. Protección y seguridad.
    b. Adaptación y personalización.
    c. Mejoras en usabilidad, rendimiento, eficiencia.
5) Incremento de la productividad y satisfacción al trabajo de los profesionales afines al desarrollo de software.
6) Mejora en la planeación eficaz de los agentes software.
7) El acercamiento a cero defectos.
8) Finalmente, servirá para concientizar sobre la importancia de la calidad en la creación, desarrollo y uso del software.

4. FORMACIÓN DE RECURSOS HUMANOS

Bajo esta línea, en el grupo de investigación, se han desarrollado dos tesis de maestría en Calidad de Software. En una de ella se ha definido el modelo de calidad que cumple con los requerimientos de una aplicación software de movilidad para una Ciudad Inteligente. La otra consiste en la definición e implementación de una herramienta para la medición y evaluación de la calidad de software de gestión de tráfico. Además, se están llevando a cabo trabajos de grado con relación a la temática por alumnos de la carrera de Ingeniería Informática de la Universidad Rey Juan Carlos, Madrid.

Como así también se están llevando a cabo algunas tesinas de grado para la Licenciatura en Computación, y trabajos finales de carrera de la Ingeniería en Informática e Ingeniería en Computación de la Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis.

5. REFERENCIAS

[1]. Sergio Colado, Abelardo Gutiérrez, Carlos J. Vives, Eduardo Valencia. SMART CITY: Hacia la gestión inteligente. Editorial: S.A. MARCOMBO.

[2]. Fundación País Digital. País Digital: Smart Cities. ttp://www.paisdigital.org/PD/smart-cities/.

[3]. Eduard Martin. Economía Digital: http://www.expansion.com/economia-digital/protagonistas/2017/12/15/5a0b3846ca4741123c8b45fe.html. España. 2017.

[4]. R. Baskerville, L. Levine, J. Pries-Heje, S. Slaughter. How Internet Software Companies Negotiate Quality. IEEE Computer, 2001.

[5]. Roger S. Pressman. "Ingeniería del Software: un enfoque práctico. Séptima Edición". Ed: McGraw-Hill, 2006.

[6]. Peter G. Neumann. Illustrative Risks to the Public in the Use of Computer Systems and Related Technology desde 1987 de sitio web http://www.csl.sri.com/neumann/illustrative.html

[7]. Tom Demarco. Software Project Management: A Guide for Service Providers. Editorial: Pearson Education (Us), 1982.

[8]. Fenton y Pfleeger. "Software Metrics: A Rigorous and Practical Approach". Ed: PWS PUBLISHING COMPANY, 1997.

[9]. UNE-EN ISO 9000:2000. Sistemas de gestión de la calidad Fundamentos y vocabulario. España. 2000.

[10]. Piattini, M., García F., Caballero, I. "Calidad de los Sistemas Informáticos". Editorial Alfaomega. España. 2007.

[11]. Fenton, N. "Viewpoint Article: Conducting and Presenting Empirical Software Engineering". 2001.

[12]. Briand, L., Morasca, S. y Basili, V. "Property-Based Software Engineering Measurement". 1996.

[13]. IEEE - Software Engineering Standards, Standard 610.12-1990, 1993.

[14]. Ragland, B. Measure, Metric or Indicator: What is the Difference? Ed: Crosstalk. 1995.

[15]. Olsina L., Pesotskaya E., Covella G., Dieser A. Bridging the Gap between Security/Risk Assessment and Quality Evaluation Methods. Publicada en ASSE Argentine Symposium on Software Engineering. 2012-2013.

[16]. Ian Sommerville. "Ingeniería de Software Séptima edición". Ed: PEARSON Addison Wesley. 2005.

[17]. Alberto Fernandez, Holger Billhardt. Dynamic Co-ordination of Open Fleets in Urban Environments. SURF: Intelligent System for integrated and sustainable management of URban Fleets. Centre for Intelligent Information Technologies (CETINIA). Proyecto: Ecobike Solutions: Gómez Pérez, Manuel y López González, Julio. Gestión de sistemas de alquiler de bicicletas. 2016.

EVALUACIÓN DE SEGURIDAD DE APLICACIONES WEB

Ana Funes, Aristides Dasso

SEG / Departamento de Informática / Facultad de Ciencias Físico-Matemáticas y Naturales / Universidad Nacional de San Luis

Ejército de los Andes 950, D5700HHW San Luis, Argentina

+54 (0) 266 4520300, ext. 2126

{afunes, arisdas}@unsl.edu.ar

RESUMEN

Aquí presentamos los objetivos, lineamientos generales y resultados esperados de una línea de investigación sobre la creación de modelos de evaluación de seguridad informática en organizaciones. Como parte integral del desarrollo de modelos de evaluación de sistemas complejos, y considerando que la evaluación de la estructura y metodología de implementación de un sistema de desarrollo de software, que incluya las tareas necesarias para la implementación de medidas de seguridad en el resultado final del sistema, implica una evaluación de un sistema complejo, es que esta investigación tiene como objetivo la creación, puesta a punto y aplicación de diversos modelos que permitan obtener indicadores del nivel alcanzado en la implementación de medidas de seguridad de aplicaciones web.

La metodología a seguir para el desarrollo de dichos modelos de evaluación está basada en la aplicación del método Logic Score of Preference (LSP) [8]. Asimismo, hemos tomado como referencia para la creación del modelo, estándares reconocidos como la Web Security Testing Guide de la OWASP [11], que sirven de guía a las organizaciones para formular e implementar estrategias para la seguridad del software.

Palabras clave: Redes. Seguridad Web. Métodos de Evaluación. Logic Score of Preference (LSP).

CONTEXTO

El trabajo de investigación aquí presentado se encuentra enmarcado dentro del ámbito del SEG (Software Engineering Group), de la Universidad Nacional de San Luis, ejecutándose dentro de una de las líneas de investigación del Proyecto de Incentivos código 22/F222 “Ingeniería de Software: Conceptos, Prácticas y Herramientas para el Desarrollo de Software de Calidad”, dirigido por el Dr. Daniel Riesco y co-dirigido por el Dr. Roberto Uzal. El mismo se encuentra acreditado con evaluación externa y financiamiento de la Universidad Nacional de San Luis.

INTRODUCCIÓN

En una organización, una vez elegido un proceso de desarrollo de software que implique la aplicación de metodologías de seguridad, resulta imperioso tener un modelo que permita, dentro de dicho proceso de desarrollo de software, la evaluación de las normas y reglas de seguridad a implementar. Este modelo, debe permitir conocer y controlar el grado de la puesta en práctica con la cual la organización lleva adelante sus políticas, actividades, usa sus herramientas y métodos, etc., en pos de su seguridad.

Si bien la construcción de modelos de evaluación de sistemas complejos, entre los que se encuentran los sistemas de seguridad informática, constituye una necesidad importante, no es una tarea sencilla. Múltiples aspectos deben ser considerados en esta tarea, teniendo en cuenta no solo los aspectos físicos, tales como las instalaciones y sus políticas de acceso, sino también medidas de seguridad del software, así como firewalls, permisos, codificación en línea, etc.

Por lo tanto, para una organización preocupada en su seguridad informática, resulta necesario contar con estándares así como con herramientas apropiadas para evaluar el grado de adecuación con dichos estándares.

En este sentido, existen en la literatura y en la web múltiples propuestas. En [4] hay una interesante revisión de la literatura del tema. Así, por ejemplo, empleando un sistema interactivo basado en grafos, desarrollado en el contexto de estándares de codificación segura para el manejo de vulnerabilidades, en [6] los autores esperan superar los posibles errores humanos que podrían aparecer cuando se aplican dichos estándares de codificación segura.

Khairul Anwar Sedek et al. [10] emplean el Open Source Security Testing Methodology (OSSTMM) usando las Top Ten Critical Vulnerabilities definidas por la OWASP, y crean un modelo aditivo para evaluar la performance basada en la propuesta de la Open Source Security Testing Methodology (OSSTMM).

Jun Zhu et al. [9] encaran el problema a través de un análisis estático interactivo, que integra análisis estático dentro de IDEs (Integrated Development Environment) para proveer, in-situ, programación segura de manera de ayudar a los desarrolladores en prevenir las vulnerabilidades en la etapa de construcción del código.

En [13] se presenta una revisión de las distintas vulnerabilidades de seguridad que se usan para asegurar la capa de aplicación web, los abordajes y técnicas empleadas en el proceso, así como las fases del desarrollo de software en las cuales se enfatizan dichas técnicas junto con las herramientas y mecanismos empleados en detectar vulnerabilidades.

Para poder ayudar a reducir el riesgo presentado por el desarrollo de aplicaciones web en un entorno dado, los autores de [12] se basan en un análisis de las políticas de seguridad empleadas en otras instituciones similares, así como en el análisis de las regulaciones y el estado del arte de la seguridad en aplicaciones web.

Como se mencionó más arriba, pueden encontrarse en la web muchas otras propuestas sobre el tema, algunas, por ejemplo en [14], [7], [5].

Por nuestra parte, en este trabajo de investigación, nos proponemos como objetivo construir un modelo que, siguiendo el check list de la Web Security Testing Guide (WSTG) de OWASP [11], permita evaluar la seguridad de una aplicación web comprobando su adhesión a los puntos establecidos en la mencionada lista de control.

LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Este trabajo se enmarca dentro de una línea de investigación dentro del SEG (Sofware Engineering Group) de la Universidad Nacional de San Luis; creemos que el mismo reviste un gran interés y potencial de desarrollo. En este sentido, cabe aclarar que se trata de una extensión de una línea de investigación más amplia y consolidada dentro del grupo, que se ocupa de la aplicación y desarrollo de técnicas de Evaluación de Seguridad en aplicaciones de software (ver p.e. [1], [3], [2]).

En este caso, en particular, nos enfocamos concretamente en producir un modelo que sirva para evaluar el cumplimiento de las normas establecidas en el Web Security Testing Guide (WSTG) de OWASP, siguiendo para ello su Check List [11]. A modo de ilustración, en la Tabla 1 se muestra

el primer ítem y sus correspondientes diez sub ítems. La tabla completa consiste de diez ítems en el primer nivel y un total de más de ciento diez ítems.

Para construir el modelo de evaluación, la tabla completa del Check List será parte integral del mismo, permitiendo, a quien decida adoptar el Web Security Testing Guide como parte de su desarrollo de software web, controlar el nivel de cumplimiento de la norma. Para ello, siguiendo los lineamientos del método LSP, cada sub ítem podrá ser evaluado y a su vez agregado bajo una función de agregación que brinda un indicador parcial de satisfacción; los sub ítems podrán ser agregados nuevamente de forma iterativa hasta obtener un indicador global de satisfacción.

Tabla 1. Primer item del check list (Information Gathering) con 10 sub ítems.

Information Gathering	Test Name	Description
OTG-INFO-001	Conduct Search Engine Discovery and Reconnaissance for Information Leakage	Use a search engine to search for Network diagrams and Configurations, Credentials, Error message content.
OTG-INFO-002	Fingerprint Web Server	Find the version and type of a running web server to determine known vulnerabilities and the appropriate exploits. Using "HTTP header field ordering" and "Malformed requests test".
OTG-INFO-003	Review Webserver Metafiles for Information Leakage	Analyze robots.txt and identify <META> Tags from website.
OTG-INFO-004	Enumerate Applications on Webserver	Find applications hosted in the webserver (Virtual hosts/Subdomain), non-standard ports, DNS zone transfers
OTG-INFO-005	Review Webpage Comments and Metadata for Information Leakage	Find sensitive information from webpage comments and Metadata on source code.
OTG-INFO-006	Identify application entry points	Identify from hidden fields, parameters, methods HTTP header analysis
OTG-INFO-007	Map execution paths through application	Map the target application and understand the principal workflows.
OTG-INFO-008	Fingerprint Web Application Framework	Find the type of web application framework/CMS from HTTP headers, Cookies, Source code, Specific files and folders.
OTG-INFO-009	Fingerprint Web Application	Identify the web application and version to determine known vulnerabilities and the appropriate exploits.
OTG-INFO-010	Map Application Architecture	Identify application architecture including Web language, WAF, Reverse proxy, Application Server, Backend Database

RESULTADOS Y OBJETIVOS

El objetivo principal, en líneas generales, que nos hemos propuesto, es poder construir un modelo que permita, a quien adopte la Web Security Testing Guide [11], conocer el grado de cumplimiento de los lineamientos de dicha guía así como decidir sobre qué aspectos de la misma se pondrá mayor o menor énfasis. Esto es posible debido al método de evaluación adoptado que va más

allá de un simple método de evaluación aditivo. En efecto, el método de evaluación elegido para la construcción del modelo de evaluación, el método Logic Score of Preference (LSP), permite que puedan darse no solo mayor o menor peso a distintos ítems, sino que algunos de estos ítems puedan ser considerados opcionales y otros mandatorios, pudiendo cada uno contar con un grado determinado de esa cualidad, permitiendo de esta manera adaptar el modelo a las necesidades particulares de la organización.

FORMACIÓN DE RECURSOS HUMANOS

Dentro del SEG (Software Engineering Group), en el ámbito de la Universidad Nacional de San Luis, en el que se ejecuta el Proyecto de Incentivos código 22/F222 "Ingeniería de Software: Conceptos, Prácticas y Herramientas para el Desarrollo de Software de Calidad", se vienen llevando a cabo numerosas tesis de grado y de posgrado.

En este sentido, creemos que la línea de investigación aquí descripta, la cual es una extensión de una línea más amplia sobre aplicación y desarrollo de Análisis de Seguridad en aplicaciones de Software, seguirá dando sus frutos, tanto en publicaciones nacionales e internacionales como en la formación de recursos humanos (2 tesis de maestría presentadas más una tesis de maestría en ejecución). Asimismo, de momento, se ha encarado la posibilidad de la ejecución de una nueva tesis de maestría basada en los objetivos que aquí nos hemos propuesto.

REFERENCIAS Y BIBLIOGRAFIA

[1] Ana Funes, Aristides Dasso, Germán Montejano, Daniel Riesco. "A SAMM-based model for assessing Cybersecurity Implementations", actas de CoNaIISI 2018, 29 y 30 de Noviembre de 2018, Mar del Plata, Buenos Aires, Argentina.

[2] Aristides Dasso y Ana Funes, "Threat and Risk Assessment Using Continuous Logic", Encyclopedia of Organizational Knowledge, Administration, and Technologies, 1st. edition. IGI Global. Aceptado para su publicación en 2020.

[3] Aristides Dasso, Ana Funes, Germán Montejano, D. Riesco, R. Uzal, Roberto, N. Debnath; "Model Based Evaluation of Cybersecurity Implementations". ITNG 2016. Las Vegas, Nevada, USA, 11-13 abril 2016. In S. Latifi (ed.), Information Technology New Generations, Advances in Intelligent Systems and Computing 448. DOI: 10.1007/978-3-319-32467-8_28. Springer International Publishing, Switzerland 2016.

[4] Bala Musa Shuaibu, Norita Md Norwawi; Mohd Hasan Selamat; Abdulkareem Al-Alwani. "Systematic review of web application security development model". Artificial Intelligence Review February 2015 https://doi.org/10.1007/s10462-012-9375-6

[5] Cody Arsenault- 11 Web Application Security Best Practices. Updated on March 4, 2019. https://www.keycdn.com/blog/web-application-security-best-practices

[6] Divya Rishi Sahu & Deepak Singh Tomar. "Analysis of Web Application Code Vulnerabilities using Secure Coding Standards". Computer Engineering and Computer Science. Arabian Journal for Science and Engineering volume 42, pages 885–895 (2017)

[7] guru99. Web Application Testing: 8 Step Guide to Website Testing. https://www.guru99.com/web-application-testing.html

[8] Jozo Dujmović. "Soft Computing Evaluation Logic. The LSP Decision Method and Its Applications". Wiley, IEEE Press. Hoboken, NJ : John Wiley & Sons, 2018

[9] Jun Zhu, Jing Xie, Heather Richter Lipford, Bill Chu; “Supporting secure programming in web applications through interactive static analysis”. Cairo University. Journal of Advanced Research. 2013 Production and hosting by Elsevier B.V. on behalf of Cairo University. http://dx.doi.org/10.1016/j.jare.2013.11.006

[10] Khairul Anwar Sedek, Norlis Osman, Mohd Nizam Osman, Hj. Kamaruzaman Jusoff, “Developing a Secure Web Application Using OWASP Guidelines”. Vol. 2, No. 4 (2009), Computer and Information Science.

[11] OWASP Web Security Testing Guide (WSTG). https://owasp.org/www-project-web-security-testing-guide/

[12] S Vargas, M Vera and J Rodriguez. “Security strategy for vulnerabilities prevention in the development of web applications”. Published under licence by IOP Publishing Ltd. Journal of Physics: Conference Series, Volume 1414, V International Conference Days of Applied Mathematics 15–17 May 2019, Barranquilla, Colombia.

[13] Sajjad Rafique, Mamoona Humayun, Zartasha Gul, Ansar Abbas, Hasan Javed. “Systematic Review of Web Application Security Vulnerabilities Detection Methods”. Journal of Computer and Communications, 2015, 3, 28-40. Published Online September 2015 in SciRes. http://www.scirp.org/journal/jcc. http://dx.doi.org/10.4236/jcc.2015.39004.

[14] Top 30+ Web Application Testing Tools In 2020 (Comprehensive List). https://www.softwaretestinghelp.com/most-popular-web-application-testing-tools/ Last Updated: March 19, 2020

HERRAMIENTA DE BÚSQUEDA EN REPOSITORIOS ACADÉMICOS BASADA EN WEB SEMÁNTICA Y SISTEMAS NOSQL

María del Pilar Gálvez, Sergio L. Martínez, Nélida R. Cáceres, Ana C. Tolaba, Laura R. Villarrubia, Felipe F. Mullicundo, José R. Quispe, Marcelo R. Sanguezo Ballon, Iván L. Sandoval, Jairo J. M. Quispe, Daniel A. Lamas

Facultad de Ingeniería - Universidad Nacional de Jujuy
Ítalo Palanca 20 San Salvador de Jujuy – 0388 4221576
mdpgalvezdiaz@fi.unju.edu.ar

RESUMEN

Las instituciones académicas buscan exponer su producción científica/académica a través de repositorios digitales en Acceso Abierto. Sin embargo, presentan limitaciones para lograr los objetivos ya que los datos publicados pueden resultar insuficientes, o bien no se cuenta con términos adecuados que puedan ser relacionados para realizar búsquedas más integrales y eficaces de forma de obtener mejores resultados. Los metadatos empleados para la descripción de los datos publicados, al ser semiestructurados no permiten explotar la información de mejor manera porque hay conocimiento implícito que favorece la descripción de nuevas relaciones entre los datos explicitados que no está siendo usado.

Actualmente el sistema de consultas provisto por SIBUNJU permite ver información de los trabajos finales de grados resumida por: Título, autor, idioma, editorial, ISBN y biblioteca depositaria. Este proyecto plantea la definición de una herramienta de búsqueda en un repositorio creado específicamente para la Facultad de Ingeniería, que facilite el acceso y la explotación de los datos residentes en el mismo a través de la incorporación de tecnologías de web semántica y sistemas NoSQL. Esta herramienta será útil tanto para alumnos, egresados, docentes, investigadores y la sociedad puesto que permite crear y compartir conocimiento, además de facilitar su transferencia al sector productivo.

Palabras clave: *NoSQL, Repositorios Digitales, Semántica, Búsquedas Mejoradas.*

CONTEXTO

La presente investigación corresponde al proyecto "*Desarrollo de Herramienta de Búsqueda utilizando Web Semántica y Sistemas NoSQL*". El mismo fue aprobado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de Jujuy como proyecto categoría A (código D/0168) y se encuentra bajo incentivo.

1. INTRODUCCIÓN

Repositorios Digitales

En la actualidad se observa que la cantidad de repositorios digitales se ha incrementado notablemente, debido principalmente a la tendencia a exponer la producción de las instituciones académicas a través de repositorios digitales de Acceso Abierto [1]. En Argentina la creación de repositorios de Acceso Abierto fue impulsada desde el Ministerio de Ciencia y Tecnología que creó el Sistema Nacional de Repositorios Digitales en CyT (SNRD) y elaboró un proyecto de ley que fue aprobado a fines de 2013. La ley 26899 "Creación de repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos" establece la obligatoriedad del

acceso abierto a la producción financiada con fondos públicos a nivel nacional a través de repositorios digitales que las instituciones deberán crear, mantener e integrar al SNRD [2].

Los repositorios tienen como propósito recopilar, catalogar, gestionar, acceder, difundir y preservar información, permitiendo entre cosas el acceso libre y gratuito a todos los recursos que los conforman[3]. La implementación de un repositorio digital ofrece diferentes beneficios tanto para investigadores, estudiantes, así como al resto de la sociedad ya que permiten crear y compartir conocimiento, y facilitan la transferencia de conocimiento al sector productivo [4].

Metadatos

Los metadatos son datos que describen otros datos y dan información que permite describir, identificar, recuperar o gestionar recursos de información [5]. Desde el punto de vista de las Ciencias de la Documentación y la Información, los metadatos son un conjunto de atributos de catalogación en los documentos que permite su identificación sin tener que realizar la apertura del documento digital para conocer su contenido [6].

Existen diferentes estándares para la implementación de metadatos como Dublin Core Metadata Initiative (DCMI), Learning Objects Metadata (LOM), entre otros. Es necesario que un repositorio cuente con metadatos precisos, completos y con un formato homogéneo, esto le permitirá interoperar con otros repositorios para realizar intercambio de información además de crear servicios de valor añadido [7].

Web Semántica

Berners-Lee [8] define la Web semántica como una extensión de la web actual, en la que la información tiene un significado bien definido posibilitando a los humanos y las computadoras trabajar en cooperación. La web semántica permite el acceso inteligente y preciso a grandes repositorios de datos, esto favorece a la difusión del contenido de los repositorios [9]. Dentro de las tecnologías de la Web semántica se dispone de RDF (Resource Description Framework) que permiten dotar de significado los datos y transacciones de datos en la Web [10]. Una declaración RDF constituye la forma más simple de expresión de un metadato, mediante grafos de tripletas compuestas por un sujeto, un predicado y un objeto. Donde un sujeto es el recurso que se está describiendo. El predicado es la propiedad o relación que se desea establecer acerca del recurso. Por último, el objeto es el valor de la propiedad de lo que se describe sobre el sujeto o el otro recurso con el que se establece la relación [11].

Otra tecnología que la Web Semántica ofrece son las ontologías, estas son estructuras más completas que permiten una representación formal de un concepto, además de la representación semántica y sintáctica del mismo [12].

Búsqueda Semántica

La recuperación de información es obtenida con a través de motores de búsqueda de propósito general. Estos motores de búsqueda se basan en términos de indexación sin tener en cuenta la semántica de los contenidos y el contexto [13].

La búsqueda semántica se refiere a una búsqueda de conceptos no solo basada por la comparación de palabras (búsqueda sintáctica), sino por deducciones lógicas que consideran la intención y el significado contextual de las palabras empleadas en la búsqueda [14].

Bases de Datos NoSQL

Los sistemas NoSQL se emplean cada vez más para el manejo de datos semánticos, es decir, modelos de datos que incluyen información semántica. El objetivo de los modelos de datos semánticos es capturar el significado de los datos mediante la integración de conceptos relacionales con conceptos de abstracción más poderosos.

Las bases de datos NoSQL se caracterizan por: ser no relacionales, distribuidas, de código abierto y escalables horizontalmente [15] [16].

NoSQL corresponde a una estrategia de persistencia que no siguen el modelo de datos relacional, y que no utilizan SQL como

lenguaje de consulta [17] en otras palabras, no están supeditadas a una estructura de datos en forma de tablas y relaciones entre ellas, permitiendo a los usuarios almacenar información en formatos diferentes a los tradicionales. Algunas aplicaciones de estas bases de datos pueden observarse en [18], [19] y [20].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los repositorios digitales se estructuran mediante un modelo de contenidos. Las actuales búsquedas de información en estos repositorios, en términos generales, permiten escribir la consulta por medio de un campo de texto en el que se pueden seleccionar palabras claves o bien discriminadas por autor o institución académica. Estas búsquedas proporcionan resultados que satisfacen las consultas de los usuarios de acuerdo a los valores ingresados, sin embargo, es posible inferir más información de la obtenida hasta el momento.

Por este motivo, este trabajo pretende mostrar la mayor cantidad de información que puede estar representada explícita e implícitamente en el repositorio. Es por ello que la principal línea de investigación involucrada en el proyecto es la Ingeniería de software, mediante la implementación de un repositorio digital junto a una herramienta de búsqueda que permita mejorar los resultados a través de la integración de tecnologías de web semántica y sistemas NoSQL.

El proyecto se adecúa a las líneas prioritarias expuestas por la Facultad de Ingeniería de la UNJu en la Resolución FI Nº 071/98, la cual incluye el área temática "Ingeniería de Software", en la cual se consideran las siguientes líneas de acción:

- Repositorios digitales.
- Gestión de la información y el conocimiento.
- Sistemas de información web y bases de datos.
- Recuperación de la información.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Este proyecto tiene estipulados cuatro años de duración (2020-2023) y su principal objetivo es desarrollar una herramienta de búsqueda que facilite el análisis y comprensión de los datos almacenados en un repositorio digital de trabajos finales de grado de la Facultad de Ingeniería de la UNJu. La herramienta propuesta combinará para su desarrollo, conceptos de web semántica y sistemas NoSQL. La información extraída de estos repositorios será utilizada como apoyo para la toma de decisiones, tanto a nivel administrativo y operativo de los estudiantes de grado ya que les proporciona el conocimiento necesario para llevar a cabo la selección del tema de trabajo final. Además, esta información permitirá que otros usuarios como egresados, docentes, investigadores y agentes externos conozcan las diferentes líneas de investigación de los trabajos desarrollados, logrando de esta forma la transferencia de la UNJu hacia la comunidad.

El desarrollo de este proyecto tiene la intención de reformular el funcionamiento de los repositorios de trabajos finales de la Facultad de Ingeniería de la UNJU, de forma tal que el resultado del proceso de búsqueda información sea utilizado como apoyo en la toma de decisiones en el ámbito académico y social.

En el ámbito académico se observó que los alumnos antes de encarar el inicio de su trabajo final, realizan búsquedas exhaustivas de antecedentes sobre trabajos concluidos en la unidad académica a la cual pertenece para dotar a su trabajo final de originalidad, en muchas ocasiones no lo logran por la falta de información, provocando esta situación demoras en el inicio de la etapa final de su

carrera. En este contexto se pretende que la propuesta facilite al alumno en la elección de un tema, un director o la línea de investigación en el cual desea desarrollar su trabajo final. En el caso de los docentes e investigadores, disminuye la incertidumbre respecto a ciertos temas disciplinares consultados, por ejemplo, aplicación de nuevas tecnologías, nuevos métodos entre otros.

En el ámbito social/laboral, no se identifican a las personas especializadas en ciertos temas que sean de utilidad, por ejemplo, se ignoran los recursos humanos que tienen conocimiento sobre algunos temas como litio, Big Data o desarrollo de sistemas. Este desconocimiento implica que la universidad pierda la oportunidad de realizar la transferencia a la sociedad, y que recursos formados por la UNJu tengan oportunidades laborales. En este sentido agentes externos a la UNJu pueden captar mano de obra regional en áreas que les sean de interés, de esta forma se potenciará los recursos humanos formados en la UNJU.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto está siendo desarrollado por un equipo conformado por docentes investigadores del Grupo de Investigación y Desarrollo en Ingeniería de Software (GIDIS) de la Facultad de Ingeniería de la Universidad Nacional de Jujuy. La estructura del equipo de investigación es la siguiente:

- Directora: Mg. María del Pilar Gálvez. Categoría de Investigación III.
- Codirector: Mg. Ing. Sergio Luis Martínez. Categoría de Investigación III.

Investigadores:

- Ing. Nélida Raquel Cáceres. Categoría de Investigación IV. Actualmente realizando tesis de maestría vinculada al área de bases de datos.
- Ing. Ana Carolina Tolaba. Categoría de Investigación V. Actualmente realizando tesis de doctorado vinculada al área de modelado conceptual de datos a través de modelos semánticos.
- Esp. Ing. Laura Rita Villarrubia. Categoría de Investigación IV.
- Lic. Felipe Fernando Mullicundo. Categoría de Investigación V.
- Mg. Ing. José Rolando Quispe.
- APU Marcelo Raúl Sanguezo Ballon.

Participan del proyecto alumnos avanzados de la carrera de Ingeniería Informática:

- Jairo Joel Maximiliano Quispe
- Daniel Alberto Lamas
- Iván Leandro Sandoval.

Con la realización de este proyecto de investigación se espera la consolidación de los miembros del grupo en especial de los alumnos como jóvenes investigadores. Además, el proyecto brindará un marco propicio para la iniciación de trabajos finalesde grado de la carrera Ingeniería Informática.

5. BIBLIOGRAFÍA

[1] Maenza, R., & Darin, S. (2016). Universidades abiertas trabajando en la innovación tecnológica y la transparencia. Revista Internacional Transparencia e Integridad, RITI nro, 2.

[2] Peña, K. I. C. (2014). Modelos de acceso abierto en educación y ciencia. Educación y educadores, 17(2), 8. DOI: 10.5294/edu.2014.17.2.7

[3] Texier, J. (2013). Los repositorios institucionales y las bibliotecas digitales: una somera revisión bibliográfica y su relación en la educación superior. 11th Latin American and Caribbean Conference for Engineering and Technology. Cancún, México. p. 9

[4] Ramírez, M. R., Soto, M. D. C. S., Moreno, H. B. R., Rojas, E. M., Millán, N. D. C. O., &Cisneros, R. F. R. (2019). Metodología SCRUM y desarrollo de Repositorio Digital.

Revista Ibérica de Sistemas e Tecnologías de Informação, (E17), 1062-1072.

[5] Senso, J. A., y De la Rosa, A. (2003). El concepto de metadato: algo más que descripción de recursos electrónicos. Ciência da Informação, 32(2), 95-106. doi:10.1590/S0100-19652003000200011

[6] Testa, P. (2013). Esquemas de metadatos para los repositorios institucionales de las universidades nacionales argentinas.

[7] Delgado, J. C. S., & Alvarado, M. A. C. (2017). Repositorios institucionales digitales: Análisis comparativo entre SEDICI (Argentina) y Kérwá (Costa Rica). e-Ciencias de la Información, 1-32.

[8] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific american, 284(5), 28-37.

[9] Sulé, A., Centelles, M., Franganillo, J., & Gascón, J. (2016). Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos de España. Revista española de documentación científica, 39(1), 121.

[10] McBride, B. (2004). Theresource description framework (RDF) and its vocabulary description language RDFS. In Handbook on ontologies (pp. 51-65). Springer, Berlin, Heidelberg.

[11] RDF, Resource Description Framework. W3C recommendation. Disponible en: http://www.w3.org/RDF Acceso: Octubre, 2019.

[12] Gruber, T., Ontology, I. L. L., & Özsu, M. T. (2009). Encyclopedia of database systems. Springer-Verlag, ISBN 978-0-387-49616-0.

[13] Smine, B., Faiz, R., & Desclés, J.-P. (2012). Extracting relevant learning objects using a semantic annotation method. In International Conference on Education and e-Learning Innovations, (pp. 1–6). IEEE.

[14] Portolés Sánchez, M. J. (2010). Búsqueda semántica en repositorios de conceptos biomédicos estandarizados: CT Hunter (Doctoral dissertation).

[15] Venkatraman, S., Fahd, K., Kaspi, S., & Venkatraman, R. (2016). SQL Versus NoSQL movement with big data analytics. Int. J. Inform. Technol. Comput. Sci, 8, 59-66.

[16] "NOSQL Databases," Disponible en: http://nosql-database.org Acceso: Octubre, 2019.

[17] Arévalo, H. H. R., & Cubides, J. F. H. (2013). Un viaje a través de bases de datos espaciales NoSQL. Redes de ingeniería, 4(2), 57-69.

[18] Rodríguez Pérez, A., Rodríguez Hernández, D., & Díaz Martínez, E. (2016). Selección de Base de Datos No SQL para almacenamiento de Históricos en Sistemas de Supervisión. Revista Cubana de Ciencias Informáticas, 10(3), 159-170.

[19] Martín, A. E., Chávez, S. B., Rodríguez, N. R., Valenzuela, A., & Murazzo, M. A. (2013, June). Bases de datos NoSQL en cloud computing. In XV Workshop de Investigadores en Ciencias de la Computación.

[20] Valenzo, M. R., Valencia, R. E. C., & Castro, J. M. M. (2013). Integración de búsquedas de texto completo en Bases de Datos NoSQL. Revista Vínculos, 8(1), 80-92.

METODOLOGÍA PARA EL DESARROLLO DE APLICACIONES CON RA EN EDUCACIÓN (MEDUC_AR)

Nélida Raquel Cáceres, Ana Carolina Tolaba, María del Pilar Gálvez, Natalia María del Huerto Flores, Eduardo Elías Hinojosa

Facultad de Ingeniería - Universidad Nacional de Jujuy
Ítalo Palanca 20 San Salvador de Jujuy – 0388 4221576
nrcaceres@fi.unju.edu.ar

RESUMEN

El Grupo de Investigación y Desarrollo en Ingeniería de Software (GIDIS) de la Facultad de Ingeniería de la UNJu, diseñó la metodología Meduc_AR para el desarrollo de aplicaciones de Realidad Aumentada (RA) destinada al ámbito educativo. La metodología consiste en tres fases: Análisis del Problema, Elección de la Solución y Evaluación de la Aplicación. Entre algunas de sus características se destaca el desarrollo iterativo de las fases que la comprenden y el trabajo en equipo entre los desarrolladores y docentes, éstos últimos son los que solicitan el software a desarrollar según requerimientos específicos que se relacionan con la disciplina, el nivel de conocimiento a enseñar, edad de los estudiantes, entre otras cosas.

El propósito de este proyecto es realizar el refinamiento de la metodología mencionada, mediante un estudio comparativo e identificación de vacancias en metodologías para el desarrollo de aplicaciones educativas usando RA, de modo de ampliar los conceptos relacionados a la elicitación de requerimientos, la usabilidad de aplicaciones, los nuevos estándares para el desarrollo de software, aspectos de calidad, entre otros.

Las aplicaciones educativas con RA creadas con otras metodologías, representan los casos de estudio en los cuales se usará Meduc_AR para destacar las ventajas de su utilización.

Palabras claves: *Metodología de desarrollo, Software de RA, Realidad Aumentada, Ingeniería de Software, Calidad de software.*

CONTEXTO

Este trabajo de investigación está comprendido en el proyecto "*Estudio comparativo e identificación de vacancias en metodologías para el desarrollo de aplicaciones con RA en educación (Meduc_AR)*". El mismo fue aprobado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de Jujuy como proyecto categoría B (código D/B041) y se encuentra bajo incentivo.

1. INTRODUCCIÓN

Metodología de Desarrollo

Una metodología de desarrollo es una colección de procedimientos, técnicas, herramientas y documentos auxiliares que ayudan a los desarrolladores de software en sus esfuerzos por implementar nuevos sistemas de información. Una metodología está formada por fases, cada una de las cuales se puede dividir en sub-fases, que guiarán a los desarrolladores de sistemas a elegir las técnicas más apropiadas en cada momento del proyecto de desarrollo y también a planificarlo, gestionarlo, controlarlo y evaluarlo [1].

Una técnica, según la RAE (Real Academia Española), es un conjunto de

procedimientos y recursos de los cuáles se sirve una ciencia o un arte. Los recursos se refieren a herramientas y procedimientos que forman parte del proceso de desarrollo de software.

Una herramienta es un conjunto de instrumentos que se utilizan para desempeñar un oficio o un trabajo determinado. En el desarrollo de software, una herramienta es el instrumento mediante el cual pueden llevarse a cabo técnicas determinadas por una metodología.

Realidad Aumentada

Una de las definiciones más comúnmente aceptada es la propuesta por Azuma [2] que manifiesta que la Realidad Aumentada (RA) es una tecnología que combina contenido real y virtual, es interactivo en tiempo real y se registra en 3D. La RA propone superponer o añadir, en tiempo real, elementos virtuales al campo de visión de una persona, de modo tal que la información requerida esté presente de una manera natural para el usuario.

El funcionamiento de aplicaciones de RA, tienen tres subsistemas fundamentales [3]: visualización (salida), ubicación de objetos virtuales en el mundo real (registro) y métodos de interacción (entrada).

Los elementos virtuales consisten en objetos, sonidos, imágenes y textos visuales generados por computadora, llevando a que el usuario no se sumerja completamente en un mundo virtual sino en una mezcla de éste con el mundo real, manteniéndolo en contacto con la realidad y en ocasiones permitiéndole interactuar con objetos virtuales.

RA en el Ámbito Educativo

El ámbito educativo ha adoptado la RA de manera permanente entre sus recursos tecnológicos. Son numerosas las aplicaciones educativas que tienen como base la realidad aumentada y que son utilizadas tanto en las aulas como fuera de ellas. Se utilizan en proyectos de clase, como complemento educativo hasta incluso como protagonistas, siendo alumnos y profesores los creadores de la propia información [4]. Existen diferentes propuestas de aplicaciones educativas que emplean RA, ejemplo de ellas se describen a continuación.

En [5], los libros de texto con RA permiten al alumno explorar objetos desde diversas perspectivas debido a que éstos pueden visualizarse en 3D. En el trabajo presentado por Ardila Pérez *et al*. [6], se desarrollan prototipos rápidos que permiten la geolocalización del salón de clases mediante un dispositivo móvil para una escuela de ingeniería.

La propuesta presentada en [7] permite la construcción de objetos virtuales de aprendizaje basados en RA. Estos objetos fueron aplicados en la descripción de órganos dentales para una facultad de odontología. En [8] se desarrolla una aplicación móvil para la enseñanza de la clasificación de los seres vivos a niños de tercer grado utilizando RA.

El desarrollo de un libro interactivo de geografía para niños de 5 a 10 años es descrito en [9]. En [10] se muestra un libro aumentado que cambia la forma tradicional de interactuar con un libro mediante modelos 3D a las páginas y reproduce imágenes en el movimiento dentro del impreso, constituyendo un recurso didáctico para estimular la lectura en niños de 6 y 7 años.

De los trabajos mencionados anteriormente se realizó un análisis que permitió concluir que la RA provee herramientas que posibilitan mejorar el entendimiento de los conceptos logrando mayor participación de los alumnos. Además, aumentan su motivación y enriquecen la experiencia de aprendizaje en general.

En el grupo GIDIS se realizaron distintas aplicaciones educativas con RA, entre las que se destacan:

- Aplicación de RA destinada a alumnos de escuelas primarias, cuyo objetivo es que los estudiantes identifiquen las provincias y capitales de la República Argentina de una manera interactiva, fácil y a la vez divertida, la cual puede ser instalada en un dispositivo móvil o computadora de escritorio [11].
- Aplicación de geografía que utiliza la RA en el ámbito de la educación primaria, permite a los estudiantes identificar los departamentos

que componen una provincia, su capital, la región y cantidad de habitantes [12].

- Aplicación para la enseñanza de los sistemas óseo, muscular y digestivo del cuerpo humano, que son contenidos curriculares obligatorios de la asignatura Ciencias Naturales correspondiente al quinto grado del nivel primario en el ámbito educativo de la Provincia de Jujuy [13].

Metodología de Desarrollo para Aplicaciones con Realidad Aumentada en la Educación

En [14] se realizó una comparación de metodologías utilizadas en los trabajos: [6], [7], [8], [9]. Como resultado, se observaron que en las mismas se aplicaron metodologías ágiles, combinación de metodologías de desarrollo de software, pero se observó que no existe una metodología diseñada específicamente para el desarrollo de aplicaciones con RA destinada a la educación.

Es por ello que en [14] se propuso una metodología para el desarrollo de aplicaciones educativas usando RA que es iterativa, en la cual se fomenta el trabajo en equipo entre los desarrolladores y el plantel educativo. La metodología propuesta posee 3 fases (figura 1):

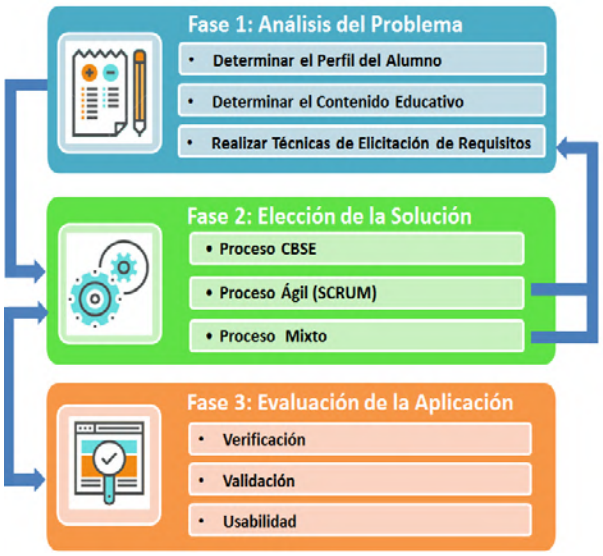


Figura 1. Metodología Meduc_AR[14].

- *Análisis del Problema*: en donde se identifican los usuarios destinatarios y receptores de la aplicación con RA.
- *Elección de la Solución:* se determina la viabilidad de la aplicación solicitada y la forma de realizarla, fundamentándose en los requerimientos del usuario.
- *Evaluación de la Aplicación*: se realiza la verificación y validación de la aplicación desarrollada.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En la actualidad, la incorporación de tecnologías como una forma de apoyo al dictado de clases en instituciones educativas, sumada a las características actuales de los alumnos, a los cuales les interesa aprender lo más posible en el menor tiempo, de manera práctica antes que teórica y en ambientes amenos. Estos en general no quieren leer y sus destrezas de escritura suelen ser deficientes, debido a que la tecnología se ha convertido en el instrumento interactivo de comunicación que permite velocidad de navegación y estimula, mediante la hipertextualidad, la asociación entre contenidos no alcanzados anteriormente por la mente humana. Siempre han vivido rodeados de computadoras, celulares, Internet, video juegos y tecnología.

En este escenario la educación compite con video juegos, internet y la televisión y es ideal para que la Realidad Aumentada sea utilizada y aceptada como una herramienta tecnológica usada en el proceso de enseñanza-aprendizaje. Esta tecnología brinda a los alumnos una motivación que les permite ampliar el conocimiento y observar facetas de alguna temática que puede resultar compleja cuando los docentes enseñan los contenidos de su asignatura.

La metodología propuesta por GIDIS, Meduc_AR, en cuyo primer desarrollo se plantearon los conceptos fundamentales que

describen un marco de trabajo donde se exponen fases, recursos materiales y humanos para la construcción de aplicaciones de RA para educación, constituye una guía para desarrollar las aplicaciones a medida de los docentes, las cuales pueden ser entregadas en un breve lapso de tiempo, además de constituir un soporte educativo para los contenidos impartidos en el dictado de clases.

La ampliación de esta metodología, contempla agregar conceptos relacionados a la elicitación de requerimientos, usabilidad de la aplicación, actualización de los estándares de calidad en el desarrollo de software, reutilización de componentes y ampliar su aplicación a otros niveles educativos, entre otros. Esta metodología está orientada a las aplicaciones educativas de RA utilizadas en computadoras de escritorio, y es necesario analizar su expansión hacia los dispositivos móviles.

Las aplicaciones educativas con RA que fueron creadas utilizando otras metodologías o modelos de proceso constituyen los casos de estudio a partir de los cuales se podrá comparar y refinar la metodología Meduc_AR.

El proyecto se adecúa a las líneas prioritarias expuestas por la Facultad de Ingeniería de la UNJu en la Resolución FI N° 071/98, la cual incluye el área temática "Ingeniería de Software" en la cual se consideran las siguientes líneas de acción:

- Metodología de desarrollo de Software
- Desarrollo ágil de software
- Calidad de software
- Verificación y Validación de software
- Estándares para el desarrollo de software.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Este proyecto tiene estipulados dos años de duración (2020-2021) y su principal objetivo es realizar el refinamiento de la metodología Meduc_AR para el desarrollo de aplicaciones de Realidad Aumentada destinada al ámbito educativo.

Con el desarrollo de este proyecto se espera:

- Estudiar conceptos vinculados a las fases de la metodología, tales como elicitación de requerimientos, usabilidad, estándares de calidad para el desarrollo de software, verificación y validación de aplicaciones, entre otros.
- Realizar estudios comparativos entre distintas metodologías o modelos de proceso utilizados en el desarrollo de aplicaciones de RA educativas, para determinar ausencias que permitan refinar la metodología Meduc_AR.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto está siendo desarrollado por un equipo conformado por docentes investigadores de la Facultad de Ingeniería de la UNJu. La estructura del equipo de investigación es la siguiente:

- Directora: Ing. Nélida Raquel Cáceres. Categoría de Investigación IV. Actualmente realizando tesis de maestría vinculada al área de bases de datos.

Investigadores:

- Ing. Ana Carolina Tolaba. Categoría de Investigación V. Actualmente realizando tesis de doctorado vinculada al área de modelado conceptual de datos a través de modelos semánticos.
- Mg. María del Pilar Gálvez Díaz. Categoría de Investigación III.

Alumnos avanzados de la carrera de Ingeniería Informática:

- Natalia María del Huerto Flores.
- Eduardo Elías Hinojosa. Becario de Iniciación a la Investigación: "Estímulo a las Vocaciones Científicas".

5. BIBLIOGRAFÍA

[1] Avison D. E. and Fitzgerald, G. Information Systems "Development: Methodologies, Techniques, and Tools", McGraw-Hill. 1995.

[2] Azuma, R. T. "A survey of augmented reality". Presence: Teleoperators and virtual environments, 6(4), 1997, pp. 355-385.

[3] Gil, G. D., Arias Figueroa, D., Gimson Saravia, L. E., Sánchez, E., & Silvera, J. A. "Uso de realidad aumentada como complemento a los prácticos tradicionales de enseñanza-aprendizaje". XVII Workshop de Investigadores en Ciencias de la Computación. Salta. 2015.

[4] Blázquez Sevilla, A. "Realidad Aumentada en Educación". Universidad Politécnica de Madrid. España. 2017.

[5] Fracchia, C. C., Alonso de Armiño, A. C., & Martins, A. "Realidad Aumentada aplicada a la enseñanza de Ciencias Naturales" TE& ET. pp 7-15. 2015.

[6] Ardila-Pérez, A. E. "Prototipo funcional para el uso de realidad aumentada en la EIA usando dispositivos móviles" Doctoral dissertation, Administrativa, Financiera, Sistemas y Computación. 2014.

[7] Tovar, L. C., Bohórquez, J. A., & Puello, P. "Propuesta metodológica para la construcción de objetos virtuales de aprendizaje basados en realidad aumentada". Formación universitaria, 7(2), pp. 11-20. 2014.

[8] Solano, C. A., Díaz, J. F. C., & Bolaños, J. C. G. "Aplicación móvil de realidad aumentada para la enseñanza de la clasificación de los seres vivos a niños de tercer grado". Ingeniería, 20(1). 2015.

[9] Abdulmushli, M. "Análisis de sistemas de realidad aumentada y metodología para el desarrollo de aplicaciones educativas". LSI1-Proyectos Fin de Máster. Universidad Rey Juan Carlos. 2012.

[10] Calisaya, M. A., Paredes, M. del M., Gálvez, M. del P. & V. Quincoces. "Realidad aumentada aplicada a la lectura en Educación Primaria", IX Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del Noa, Santiago del Estero, 3 4 de octubre de 2013.

[11] Quispe, J. "Provincias y capitales de Argentina con Realidad Aumentada". 6to Congreso Nacional de Ingeniería Informática /Sistemas de Información. Mar del Plata, 2018.

[12] Quispe, J. & Gálvez, P. "Realidad Aumentada aplicada al aprendizaje de Geografía en escuelas primarias". VI Seminario Argentina – Brasil de Tecnologías de la Información y la comunicación. Concepción del Uruguay, 2018.

[13] Veramendi, B., Cámara, A. & Calisaya, M. "Realidad Aumentada Innovación en la Enseñanza del Cuerpo Humano en la Educación Primaria de Jujuy". IV Jornadas de Jóvenes Investigadores. San Salvador de Jujuy. 2018.

[14] Cáceres, N. & Tolaba, A. "Metodología de Desarrollo para Aplicaciones con Realidad Aumentada en la Educación". V Congreso Nacional de Ingeniería en Informática/Sistemas de Información (CoNaIISI). 2017.

Adecuación de la forma de la construcción de glosarios al estilo del discurso

Renata S. Guatelli[1], Gladys N. Kaplan[1], Jorge H. Doorn[2,3]

[1]Departamento de Ingeniería e Investigaciones Tecnológicas – Universidad Nacional de La Matanza
[2]Escuela de Informática, Universidad Nacional del Oeste
[3]Departamento de Ingeniería, Universidad Nacional de Tres de Febrero
{rguatelli; gkaplan}@unlam.edu.ar; jdoorn@uno.edu.ar;

RESUMEN

Es conocida la importancia que tiene la construcción de un glosario en los procesos de requisitos. La experiencia recogida a través de numerosos casos reales ha mostrado que la construcción del mismo es simultáneamente laboriosa y que sus resultados son pocos confiables. Es razonable suponer que el uso de estrategias de procesamiento de lenguaje natural puede contribuir a atemperar ambas dificultades. La construcción de glosarios difiere de la minería de textos clásica en el sentido que se tiene un cierto conocimiento previo de aquello que se busca. Sin embargo, tanto debido a las peculiaridades de la fuente de información como a las características del Universo de Discurso, los estilos de las narrativas ofrecen una sensible dispersión. En el presente proyecto se planifica utilizar un enfoque de ingeniería inversa, utilizando fuentes de información confiables y glosarios ya construidos por seres humanos para estudiar el contexto concreto de uso de los símbolos incluidos en el glosario en la fuente de información. De este estudio se espera deducir reglas que permitan detectar símbolos no descubiertos a partir de la combinación del contexto de uso y del estilo del discurso.

Palabras clave: *Lenguaje natural, entrevista, transcripción, LEL, dispersión de información*.

CONTEXTO

La propuesta que se presenta es parte del proyecto de investigación “Aspectos No Funcionales de los Procesos de Requisitos” del Departamento de Ingeniería de Investigaciones Tecnológicas (DIIT) de la Universidad Nacional de La Matanza (UNLaM).

1.0 INTRODUCCIÓN

No todos los proyectos de software concluyen con éxito [1] [2] [3] [4]. Diversos estudios llevados a cabo para analizar este tema argumentan que el problema principal encontrado son requisitos inadecuados, mal comprendidos, incompletos y volátiles. Para poder definir los requisitos primero se debe adquirir conocimiento sobre el Universo de Discurso (UdeD), capturar las demandas, necesidades y problemas presentes en él. Esta información es analizada por la ingeniería de requisitos (IR), con el fin de generar la Especificación de Requisitos de Software (ERS) que contiene los servicios que el sistema de software debe satisfacer para mejorar los procesos del negocio. Para Loucopoulos [5] la IR se compone de actividades que permiten comprender las necesidades de los clientes-usuarios y traducir dichas necesidades en declaraciones precisas y sin ambigüedades que posteriormente se pueden utilizar en el proceso de desarrollo del nuevo sistema de software. Una ERS adecuada, es la base de las actividades de Gestión de Proyecto relacionadas con el presupuesto y el cronograma, influyendo sobre la calidad del mismo.

La baja calidad de los productos de software desarrollados, ERS deficientes, elevados costos de corrección y mantenimiento, o el completo fracaso del proyecto, podrían evitarse o al menos mitigarse si hubiera una mayor preocupación por la rigurosidad durante las actividades relacionadas con la IR [6].

Los problemas mencionados, en su mayoría, tienen su origen en la utilización de modelos que los clientes-usuarios no pueden comprender. Mejorar las estrategias de comunicación entre los involucrados es de gran importancia para identificar, validar y verificar de manera correcta y oportuna dichos modelos [7].

La etapa inicial de la IR, implica la obtención o transferencia de la mayor parte del conocimiento desde el UdeD. Plasmar ese conocimiento en modelos basados en lenguaje natural (LN), además de mejorar la comunicación entre los participantes, asegura que su validación sea más segura y acertada. Los clientes-usuarios comprenden mucho mejor las descripciones basadas en LN en lugar de aquellas basadas en esquemas técnicos. En especial si este lenguaje es lo más cercano posible a la jerga utilizada por ellos.

El LN tiene un gran poder expresivo, pero es polisémico ya que contiene ambigüedad, ironías, expresiones típicas y muchas otras características, que generan problemas de interpretación, dado que su significado depende de los puntos de vista y del modelo mental de los interlocutores. A pesar de sus inconvenientes, el uso del LN mejora significativamente la transferencia de conocimiento, asegurando una mejor comprensión de los modelos [8].

En el campo la IR existen diferentes modelos basadas en LN, algunos son:

- Casos de Uso [9], son descripciones narrativas de la interacción entre un actor y el sistema.
- Glosarios [10] [11] contienen palabras y/o expresiones, comentadas o explicadas. Se los crea con diferentes fines, por ej. aclarar el significado de conceptos del dominio de la aplicación, unificar la terminología empleada en los diferentes modelos, mejorar la comunicación entre los involucrados; se pueden centrar en la terminología de los clientes - usuarios, o en la de los documentos.
- El LEL [12], es un glosario que describe el vocabulario de la aplicación, sin necesidad de comprender la funcionalidad del proceso del negocio. Las palabras o frases que contiene son llamados símbolos. Cada símbolo se identifica con un nombre (o más de uno en caso de sinónimos). Se detalla indicando la noción (denotación, define su significado) e impacto (connotación, identifica la relación del símbolo que se está describiendo con los demás símbolos del léxico). Noción e impacto se deben describir teniendo en cuenta el "principio de circularidad" (maximizar el uso de símbolos pertenecientes al LEL) y el uso de "vocabulario mínimo" (acotar el uso de lenguaje externo al dominio de la aplicación). Generalmente los símbolos se clasifican en sujeto, objeto, verbo y estado, de acuerdo a su uso en el dominio. Pueden crearse clasificaciones especiales.
- Escenarios [13], son descripciones de las situaciones que ocurren en el contexto. Pueden representar situaciones actuales o la planificación de situaciones futuras.
- Historias de Usuarios [14], consisten en funcionalidades descriptas por el

propio usuario. Usualmente responden al siguiente formato: i) Quién requiere la funcionalidad, ii) Cuál es la funcionalidad y iii) Por qué esa funcionalidad es necesaria (opcional).

Sin importar el modelo que se utilice, todos están afectados por los beneficios e inconvenientes del LN. Algunos utilizan el LN sin restricciones, otros intentan atemperar los posibles inconvenientes.

Dentro de las distintas fuentes de información, las personas son las más apropiadas, pero también son las que requieren un tratamiento más elaborado. Por lo tanto, la entrevista es la técnica de elicitación más utilizada [15] [16]. Estos datos han sido avalados por varios estudios [17] [18] los cuales confirman que se suelen privilegiar la entrevista por sobre otras técnicas de elicitación. A partir de las entrevistas se pueden generar una serie de productos intermedios, tales como minutas, anotaciones y transcripciones. Estos productos intermedios son algunas de las fuentes de información que se utilizan en los procesos de la IR.

Se debe prestar una debida atención a que cada persona consultada tiene su propio estilo de discurso y este puede ser narrativo, expositivo o argumentativo. Además, este discurso puede estar orientado a describir el proceso que se está analizando, el producto final o hacer hincapié en los componentes necesarios y sus interrelaciones.

Las narrativas de los entrevistados, especialmente en las entrevistas no estructuradas, suelen diferir mucho unas de otras, tanto por las propiedades del UdeD como por el punto de vista del entrevistado. Por ejemplo, en un sistema de naturaleza hospitalaria las narrativas estarán fuertemente influenciadas por los protocolos médicos por lo que serán orientados a los procesos. En cambio, es frecuente encontrar narrativas que describen entidades y relaciones entre las mismas, especialmente en organizaciones administrativas.

2.0 LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este proyecto sigue la línea de investigación correspondiente al proceso de requisitos [20] basado en modelos en LN, orientado a atender las necesidades del cliente-usuario.

Se utilizarán técnicas de procesamiento de lenguaje natural, aplicadas sobre el modelo LEL, para mejorar la calidad del mismo, procurando reducir la notoria incompletitud observada en estudios previos [19] [20] [21] [22]. Usar estrategias de procesamiento de LN implica el uso de textos, por lo que se utilizará en todos los casos las transcripciones de las entrevistas en total concordancia con la enorme experiencia de las ciencias sociales que ampliamente las promueven [23] [24] [25]. Fundamentalmente en las transcripciones de las entrevistas es donde mejor se puede analizar el estilo discursivo para deducir nuevas reglas.

La técnica planificada consiste en una suerte de ingeniería inversa ya que se tomará un caso ya estudiado, y se analizarán todas las ocurrencias de los símbolos del LEL detectados por otros autores, y se estudiarán los conectores o frases que ligan los pares de símbolos en el texto.

En este sentido, se supone que los conectores lingüísticos o palabras auxiliares que ligan los símbolos del LEL en el discurso son fuertemente dependientes del estilo de la narrativa. Por ejemplo, es esperable que la frase “es un”, “es el”, “es parte de”, “está compuesto por” o similares estén presentes en un discurso centrado en entidades y ausentes en un discurso centrado en procesos.

Este estudio se planifica en el contexto de una calificación previa del discurso e intenta detectar las frases que vinculan los símbolos del

LEL. Estas frases serán luego útiles para mejorar la detección de símbolos en futuros proyectos. Obviamente, se panifica replicar el mismo estudio en varios casos para intentar determinar el grado de dependencia de estos conectores con el estilo del discurso utilizado y la variabilidad dentro de un mismo estilo.

Esta estrategia se basa en la hipótesis que en todo proyecto siempre existen unos pocos términos que son muy evidentes para el ingeniero de requisitos, los que serán incluidos en el LEL sin lugar a dudas. Es así que se planifica que la estrategia final resultante, utilice estos pocos símbolos iniciales como núcleo en la búsqueda semiautomática de nuevos símbolos. Es justamente en esa búsqueda de nuevos símbolos en la que el conocimiento de los conectores más usuales, posiblemente enriquezca y facilite la misma. Se espera que esto ayude a disminuir la incompletitud del LEL, permitiendo mejorar las heurísticas para colaborar en la detección de símbolos no triviales.

3.0 RESULTADOS ESPERADOS

En experiencias preliminares se han estudiados casos reales utilizando el mismo patrón de trabajo que se planifica utilizar sistemáticamente en el presente proyecto. Los resultados obtenidos son promisorios en el sentido que se han descubierto regularidades que de ser confirmadas permitirían guiar el procesamiento del LN con reglas basadas en el estilo del discurso.

En estos trabajos preliminares se transcribieron audios de entrevistas. En estos documentos se marcaron los símbolos que figuraban en el LEL, estudiando las palabras o frases que actuaban como conectores entre pares de símbolos, encontrándose que el número de conectores utilizados en la narrativa es relativamente reducido.

Como resultado principal del presente proyecto se espera detectar en los diferentes estilos discursivos cuáles son los conectores más frecuentes entre pares de símbolos del LEL. Estos se utilizarán luego para construir una herramienta que utilice las ocurrencias de los mismos para sugerir posibles nuevos términos a ser incluidos en el LEL.

4.0 FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación presentada colabora en la tesis doctoral de la Mg. Gladys Kaplan y es parte directa de la tesis de maestría de la Lic. Renata Guatelli.

5.0 BIBLIOGRAFÍA

[1] Gibbs, W. W. (1994). Software's chronic crisis. Scientific American, 271(3), 86-95.

[2] Finkelstein, A., & Dowell, J. (1996). A comedy of errors: the London Ambulance Service case study. In Proceedings of the 8th International Workshop on Software Specification and Design, pp. 2-4.

[3] Lindstrom, D. R. (1993). Five ways to destroy a development project (software development). IEEE Software, 10(5), pp.55-58.

[4] El Emam, K., & Koru, A. G. (2008). A replicated survey of IT software project failures. IEEE software, 25(5), pp.84-90.

[5] Loucopoulos, P., & Karakostas, V. (1995). System requirements engineering. McGraw-Hill, Inc.

[6] de Almeida Ferreira, D., & da Silva, A. R. (2009). A controlled natural language approach for integrating requirements and model-driven engineering. In Fourth International Conference on Software Engineering Advances, pp. 518-523.

[7] Boehm, B. W. (1984). Verifying and validating software requirements and design specifications. IEEE software, (1), pp. 75-88.

[8] Jackson, M. (1995). Software requirements & specifications: a lexicon of practice, principles and prejudices. ACM Press/Addison-Wesley Publishing Co.

[9] Jacobson, I. (1993). Object-oriented software engineering: a use case driven approach. Pearson Education.

[10] Weidenhaupt K., Pohl K., Jarke M., Haumer, P (1998) Scenarios in System Development: Current Practice, IEEE Software, pp 34-45.

[11] Robertson S. and Robertson J. (2006) Mastering the Requirements Process, 2nd Ed, AddisonWesley.

[12] Leite, J. D. P., & Franco, A. P. M. (1993, January). A strategy for conceptual model acquisition. In Proceedings of the IEEE International Symposium on Requirements Engineering, pp. 243-246.

[13] Carroll, J. M. (Ed.). (1995). Scenario-based design: envisioning work and technology in system development. John Wiley & Sons.

[14] Beck, K. (2000). Extreme programming explained: embrace change. Addison-Wesley.

[15] Pan, D., Zhu, D., & Johnson, K. (2001). Requirements Engineering Techniques. Internal Report. Department of Computer Science. University of Calgary. Canada.

[16] Bourque, P., & Fairley, R. E. (2014). Guide to the software engineering body of knowledge (SWEBOK (R)): Version 3.0. IEEE Computer Society Press.

[17] Antonelli, L., & Oliveros, A. (2002). Fuentes Utilizadas por desarrolladores de Software en Argentina para Elicitar Requerimientos. In fifth Workshop on Requirements Engineering, pp. 106-116.

[18] Oliveros, A., & Antonelli, R. L. (2015). Técnicas de elicitación de requerimientos. In XXI Congreso Argentino de Ciencias de la Computación, pp.546-555

[19] Ridao, M., & Doorn, J. H. (2006). Estimación de Completitud en Modelos de Requisitos Basados en LN. In 9th Workshop on Requirements Engineering, pp.146-152.

[20] Litvak, C. S., Hadad, G. D., & Doorn, J. H. (2012). Un abordaje al problema de completitud en requisitos de software. In XVIII Congreso Argentino de Ciencias de la Computación, pp. 827-836.

[21] Litvak, C. S., Hadad, G. D., & Doorn, J. H. (2013). Mejoras semánticas para estimar la Completitud de Modelos en Lenguaje Natural. In 1er Congreso Nacional de Ingeniería Informática / Sistemas de Información, (p.9) http://conaiisi.frc.utn.edu.ar/PDFsParaPublicar/1/schedConfs/7/7-477-1-DR.pdf (consultado el 26/03/2020)

[22] Litvak, C. S., Hadad, G. D. S., & Doorn, J. H. (2016). Procesamiento de lenguaje natural para estudiar completitud de requisitos. In XVIII Workshop de Investigadores en Ciencias de la Computación, pp. 498-502

[23] Valles, M. S. (2007). Entrevistas cualitativas (Vol. 32). CIS.

[24] Kvale, S. (2011). Las entrevistas en investigación cualitativa. Ediciones Morata.

[25] Díaz-Bravo, L., Torruco-García, U., Martínez-Hernández, M., & Varela-Ruiz, M. (2013). La entrevista, recurso flexible y dinámico. Investigación en educación médica, 2(7), pp. 162-167.

Análisis de calidad y recuperación de información en foros de discusión

Nadina Martinez Carod, Valeria Zoratto, Gabriela Aranda, Alejandra Cechich,
Agustín Chiarotto, Carina Noda, Mauro Sagripanti

Grupo de Investigación en Ingeniería de Software del Comahue (GIISCo) http://giisco.uncoma.edu.ar
Facultad de Informática. Universidad Nacional del Comahue -Buenos Aires 1400, (8300) Neuquén
Contacto: {nadina.martinez,vzoratto, gabriela.aranda, valeria.zoratto}@fi.uncoma.edu.ar

RESUMEN

Actualmente las organizaciones dedican mucho esfuerzo a resolver problemas utilizando estrategias ya probadas, lo que requiere brindar especial importancia a los sistemas de Information Retrieval (IR). Ante un problema que surge, lo primero que se analiza es si ha sido resuelto antes, bajo qué situaciones y en qué contexto se dieron problemas similares. Es aquí donde cobran gran importancia los sitios web donde se propician discusiones e intercambio de ideas sobre problemas comunes. Herramientas colaborativas como los blogs, wikis y foros de discusión hacen posible el reúso de conocimiento disponible en la Web. Entre ellas, nuestro proyecto se enfoca en los foros de discusión como herramienta fundamental, ya que contiene una base de conocimiento lo suficientemente completa para ser reutilizada. La enorme cantidad de información existente en los foros, sumado a la simplicidad de uso de los mismos, hacen necesario definir estrategias para clasificar las soluciones ya probadas, fundamentales en los sistemas IR.
De esta manera el principal objetivo de este proyecto es la creación de una herramienta que realice una clasificación automática de la información contenida en foros de discusión, utilizando la información existente en los hilos de discusión, que luego de ser analizada, será procesada, reutilizada y clasificada para solucionar problemas específicos recurrentes que fueran surgiendo.

CONTEXTO

Nuestro proyecto se enmarca en el Programa "Desarrollo de Software Basado en Reúso – Parte II", de la Universidad Nacional del Comahue, a realizarse en el período 2017-2020, el cual extiende al Programa "Desarrollo de Software Basado en Reúso" realizado en el período 2013-2016. Dicho Programa está basado en el proceso sistemático de detectar, seleccionar, organizar, presentar y utilizar la información del capital intelectual de las organizaciones, con el fin de brindar mayor competitividad a las mismas, aplicando el conocimiento en el problema objeto para resolver. Dicho Programa está compuesto por tres subproyectos los cuales coinciden en el tratamiento del desarrollo de software basado en reúso, pero desde aspectos diferentes: orientado a dominios, orientado a servicios y orientado a foros de discusión. El proyecto actual, denominado "Reúso de Conocimientos en Foros de Discusión – Parte II", continúa la línea de investigación enfocada en la recuperación de información y de conocimiento disponible en foros de discusión técnicos.

1. INTRODUCCIÓN

Los foros de discusión son utilizados para consultar, discutir e intercambiar conocimiento en la Web, pero su principal característica es que facilita la comunicación de personas distribuidas geográficamente. Se puede describir un foro como un espacio virtual de una comunidad de usuarios con temas de interés común. De esta manera

encontramos foros con temas variados como viajes, cocina, historia, software entre otros, sin embargo, en nuestro proyecto nos interesan en particular los foros sobre temas aplicados a las ciencias de la computación.

Dentro de un foro, los participantes no necesitan encontrarse conectados al mismo tiempo (herramienta asíncrona), incluso en la gran mayoría de las veces los participantes no se conocen personalmente, pero sí a través de sus nombres, alias o avatares (representaciones gráficas que se asocian a usuarios para identificarse). Algunos de estos foros, como Stackoverflow[1], se destacan por ofrecer entre sus soluciones fragmentos de código, así como información valiosa de los usuarios (por ejemplo, la reputación que se construye a partir de su participación y de votos de los otros usuarios). Otros, como lawebdelprogramador[2] arman una valoración del usuario de manera automática a partir de información del usuario que van obteniendo a lo largo de los mensajes que estos escriben, como preguntas iniciales, respuestas dadas y puntuación de sus miembros. Esto implica que, al trabajar con foros, como cada uno tiene una estructura diferente, se complejiza la tarea de recopilación de información y del análisis a realizar sobre ella. Estas características hacen de los foros de discusión no sólo una herramienta fundamental que facilita el trabajo colaborativo y distribuido, sino también una fuente esencial de conocimiento que puede ser reutilizado.

De acuerdo al permiso dado a los participantes para ver o comentar dentro del foro, se pueden distinguir diferentes tipos de foros: en los públicos todos los participantes pueden comunicarse o leer mensajes escritos por el resto sin necesidad de registrarse; en cambio en los foros protegidos es necesario registrarse para luego poder enviar mensajes. Por último, en los foros privados se exigen ciertas restricciones para participar y utilizar la información.

[1] https://stackoverflow.com/

[2] https://www.lawebdelprogramador.com/

Como el proyecto se enfoca en trabajar con la información existente en foros de discusión disponibles para cualquier usuario, sin necesidad de tener una registración en los mismos; los foros elegidos para nuestras evaluaciones son públicos, donde la totalidad de los mensajes queda disponible para ser consultada en problemas similares.

Respecto a la estructura de un foro de discusión, el mismo contiene una colección de hilos. Un hilo se genera a partir de una pregunta inicial sobre un problema específico. A partir de ese momento, los usuarios de la comunidad podrán debatir en función del tema y del problema enunciado. Luego, para obtener conocimiento proveniente de los hilos de discusión se utilizan diferentes técnicas y estrategias para establecer cuáles de las posibles soluciones obtenidas de los foros pueden ser relevantes para consultas sobre problemas similares.

El proyecto realiza por un lado el tratamiento del texto contenido en los hilos dentro del foro. Luego, la búsqueda transita por caminos independientes, pero con un objetivo en común: complementar el conjunto de estrategias de clasificación de hilos para definir un orden de prioridad. Para ello se ha continuado con el enfoque de Elsas & Carbonelli [6], el cual utiliza la estructura de los hilos separando la pregunta inicial del resto del hilo, tratándolo con la dupla <pregunta, hilo>. Una estructura similar se logra también mediante el uso de patrones Feng & Shaw [7], y de otros tratamientos como el de Cong & Wang [5] que separan preguntas de respuestas con un robot reconocedor de campos condicionales. Otra orientación que se ha analizado es la de clasificar o estructurar temas mediante jerarquías, como el enfoque de Nicoletti [17] o el de Helic et al. [11]. También se analizó la concordancia con Liu et al. [12], que busca la satisfacción de las respuestas sobre una pregunta realizada, o el enfoque que presenta Bathia [4] a partir del análisis de expertitud de

los usuarios, detectando niveles de conocimiento de los comentarios, para clasificar con mayor valoración los hilos en los cuales intervienen personas expertas o con altos conocimientos.

Otro camino analizado es mediante algoritmos de lenguaje natural para clasificar tipos de fragmentos dentro de los hilos de discusión utilizando técnicas de aprendizaje automático, como la propuesta por Tigelar et al. [19]. Dentro de este camino se encuentran también técnicas de minería de datos, para la búsqueda de patrones y reglas significativas, haciendo énfasis en la minería web (minería de datos sobre la web), tanto en procesamiento de lenguaje natural como opinion mining, en conjunto con la aplicación de *sentiment analysis*. Este estudio está focalizado en la búsqueda de patrones reconocedores de una determinada respuesta sobre la pregunta en un hilo, para determinar si tuvo aceptación o rechazo, y en qué escala. De esta manera se puede detectar aquellos hilos donde la respuesta fue satisfactoria y el grado de satisfacción del mismo.
Teniendo en cuenta las distintas direcciones mencionadas, se realizaron actividades, con resultados favorables en su mayoría, permitiendo la extensión de algunas líneas de avance y la elaboración de nuevas líneas a favor del objetivo del proyecto.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación principal está enmarcada en un Programa de investigación de la Universidad Nacional del Comahue, llamado "Desarrollo de Software Basado en Reúso – Parte II", dentro del período comprendido entre los años 2017 y 2020. El programa tiene 3 líneas de investigación, relacionadas; reúso orientado a dominios, reúso orientado a servicios y la nuestra, cuya denominación es "Reúso de Conocimientos en Foros de Discusión – Parte II". El programa está desarrollado por el Grupo GIISCo de la Facultad de Informática, y su objetivo es continuar con las investigaciones iniciadas en el programa inicial, período 2013-2016 denominado "Desarrollo de Software Basado en Reúso", perteneciente a la misma universidad.
El Grupo GIISCo está conformado por docentes de la Facultad de Informática de la Universidad Nacional del Comahue, no sólo del área de Ingeniería de Software, sino también incluye docentes con otras ópticas establecidas por las áreas de Programación y Teoría de la Computación. Esto, sumado a la colaboración de la Facultad de Ciencias Exactas de la Universidad Nacional del Centro de la Provincia de Buenos Aires, le brinda un marco heterogéneo propicio para la investigación de una amplia gama de soluciones.

3. RESULTADOS OBTENIDOS/ESPERADOS

A partir del objetivo de nuestro proyecto, el cual es brindar asistencia inteligente en la recomendación de hilos de discusión para buscar soluciones a problemas recurrentes presentados en foros de discusión técnicos, podemos citar resultados obtenidos hasta el momento y mencionar lo que se espera a partir del presente.
Los resultados obtenidos sobre nuestra línea de investigación parten de un modelo de calidad para foros de discusión en base a modelos de calidad de datos e información en la Web y estándares para la calidad de datos software, En base a de dicho modelo se determinaron métricas, que fueron validadas mediante encuestas [1], posteriormente se avanzó en el diseño e implementación de una herramienta para la recuperación de información [14], aplicando un grupo de métricas de calidad [21]. Para poder manipular la información de foros, se trabajó en el análisis y tratamiento de textos, con herramientas como Lucene[3], lo que dio lugar a tratamientos de recuperación de información para lenguaje específico de Java, [2, 22]. En esta línea se continuó con la utilización de

[3] https://lucene.apache.org

sinónimos, mediante el uso de la base de datos léxica en inglés WordNet[4] [22], mediante un analizador morfológico como Stanford POS Tagger[5] [3]. La tercera línea que se siguió fue la clasificación del conocimiento de los usuarios participantes de los hilos de discusión. Esta línea se basó en las ópticas de Lui & Baldwin [13], la de Bathia & Mitra [4] y la de Hecking et al. [10], bajo las cuales se diseñó una estrategia para determinar una jerarquía de roles determinados por el nivel de conocimientos de los participantes en los hilos de acuerdo a los posts realizados [15].

Considerando los resultados obtenidos hasta el momento como punto de partida, los resultados que se esperan obtener continúan en distintas direcciones, pero con nuevas alternativas. Por un lado, se continúa trabajando en la tarea de clasificación de tipos de mensajes para poder destacar las preguntas de las respuestas, informando la retroalimentación positiva y negativa, mediante la utilización de técnicas de procesamiento de lenguaje natural (PNL), utilizando técnicas de Data Mining, a partir de la propuesta de Liu [12], tanto en modelos de aprendizaje supervisados como no supervisados presentadas por Witten [20], además, siguiendo la línea de Shoji et al.[18], se está analizando al usuario que realiza la pregunta, teniendo en cuenta su historia dentro de la comunidad, para detectar posibles patrones de comportamientos Esta línea de investigación se sigue desarrollando en una tesis de doctorado en la cual se evalúa en conjunto: el rol del post, respuestas de calidad, expertitud del usuario que responde y satisfacción del usuario que pregunta.
En otra dirección se continúa trabajando en estrategias para la obtención de roles, aplicando un análisis de redes sociales en patrones de respuestas de los participantes que han escrito un post dentro del mismo hilo, propuesto por Fisher et al. [8]. Siguiendo con la línea de análisis de los usuarios participantes, se está realizando una tesis para buscar una estrategia para predecir la satisfacción de los usuarios que preguntan respecto a las soluciones propuestas por otros participantes.

[4] https://wordnet.princeton.edu/

[5] https://nlp.stanford.edu/software/tagger.shtml

4. FORMACIÓN DE RECURSOS HUMANOS

Debido a que la naturaleza del proyecto es multidisciplinaria, dentro de sus integrantes hay docentes de diferentes áreas. En particular en las de Ingeniería en Sistemas, Programación, y Teoría de la Computación. Las personas que forman parte del proyecto, tanto como colaboradores, asesores o integrantes son:

- 2 profesores adjuntos con dedicación exclusiva.
- 1 docente investigador con beca del CONICET.
- 2 docentes con dedicación simple.
- 1 profesora adjunta, asesora de la UNCo, con dedicación exclusiva.
- 1 profesora adjunta, asesora externa de la UNCPBA.
- 5 estudiantes de la carrera de Licenciatura en Ciencias de la Computación.

Los integrantes del proyecto que pertenecen al Departamento de Programación son dos docentes con dedicación exclusiva, doctores en Informática; el otro docente es asistente simple con beca de CONICET para hacer el doctorado, un docente investigador con dedicación simple. El otro docente investigador con dedicación simple pertenece al Departamento de Ingeniería de Sistemas.
La docente colaboradora de la UNCo que pertenece al área de Teoría de la Computación, posee un Diploma de Estudios Avanzados y se encuentra realizando la Maestría en Ciencias de la Computación de la Universidad Nacional del Comahue.
También integra el proyecto una asesora externa que pertenece a la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA), con experiencia en

Sistemas de Recomendación y Recuperación de Información, y pertenece al Instituto Superior de Ingeniería de Software (ISISTAN). Dicha docente posee el título de Doctora en Ciencias de la Computación.

Tres estudiantes de Licenciatura en Ciencias de la Computación desarrollaron sus tesis dentro del proyecto, uno de cuales está en etapa de defensa de la misma. Además, se han sumado cuatro estudiantes, los cuales están realizando sus tesis. De esta manera, se van incorporando actividades para extender líneas de investigación al proyecto inicial con nuevos enfoques.

5. BIBLIOGRAFÍA

[1] G. Aranda, N. Martínez Carod, P. Faraci, A. Cechich. Hacia un framework de evaluación de calidad de información en foros de discusión técnicos. ASSE 2013,

[2] G. Aranda, N. Martínez-Carod, S. Roger, P. Faraci, and A. Cechich. Una herramienta para el análisis de hilos de discusión técnicos. In CACIC 2014, pages 803 - 812, 2014.

[3] G. Aranda, V. Zoratto, N. Martínez Carod, Sandra Roger, F. Otermín, A. Cechich. Clasificación de contenido de hilos de discusión mediante análisis sintáctico y morfológico. CICCSI 2018.

[4] S. Bhatia and P. Mitra. Classifying user messages for managing web forum data. In Z. G. Ives and Y. Velegrakis, editors, WebDB , pages 13-18, 2012

[5] G.Cong, L. Wang, C. Lin, Y. Song, and Yueheng (2008). Finding question-answer pairs from online forums. In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '08). Association for Computing Machinery, 467–474.

[6] J.L Elsas and J. G Carbonell, "It pays to be picky: an evaluation of thread retrieval in online forums", in Proceedings of the 32nd international ACM SIGIR (2009), pp. 714—715.

[7] D. Feng, E. Shaw, J.Kim & E.Hovy (2006). An intelligent discussion-bot for answering student queries in threaded discussions. In Proceedings of the 11th international conference on Intelligent user interfaces (IUI '06). Association for Computing Machinery, 171–177.

[8] D. Fisher, M. Smith, and H. T. Welser, You are who you talk to: Detecting roles in usenet newsgroups, in Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), vol. 3, pp. 59b59b, IEEE, 2006.

[9] A. Gangemi, R. Navigli, P. Velardi. The OntoWordNet Project: Extension and Axiomatization of Conceptual Relations in WordNet, In Proc. of ODBASE 2003, Catania, Sicily (Italy), 2003, pp. 820–838.

[10] T. Hecking, I. Chounta, and H. U. Hoppe. Investigating social and semantic user roles in MOOC discussifon forums. In LAK, pages 198-207. ACM, 2016

[11] D. Helic, N. Scerbakov (2003), "Reusing Discussion Forums as Learning Resources in WBT Systems".

[12] B. Liu. Web Data Mining. Exloring Hyperlinks, Contents, and Usage Data. Springer. 2008

[13] M. Lui and T. Baldwin. Classifying user forum participants: Separating the gurus from the hacks, and other tales of the internet. In Proceedings of Australasian Language Technology Association Workshop, pages 49-57, 2010.

[14] N.Martínez Carod, G. Aranda. Análisis de la información presente en foros de discusión técnicos. In CACIC 2013, pp. 847- 856, 2013.

[15] N.Martínez Carod, G. Aranda. Valeria Zoratto, Christian Murray. Una propuesta para clasificación de roles de usuarios en foros de discusión técnicos. In CACIC 2019, pp. 836- 845, 2019.

[16] G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. 1990. WordNet: An online lexical database. Int. J. Lexicograph.pp. 235–244.

[17] M. Nicoletti, S. Schiafino, and D. Godoy. Mining interests for user profiling in electronic conversations. Expert Syst. Appl., Feb. 2013.

[18] Y. Shoji, S. Fujita, A. Tajima, and K. Tanaka. Who stays longer in community qa media?-user behavior analysis in cqa. In International Conference on Social Informatics, pages 33 48. Springer, 2015. .

[19] A. Tigelaar, R. Op Den Akker and D. Hiemstra, Automatic summarisation of discussion fora, Natural Language Engineering, ISSN 1469-8110, Vol 16, Issue 02, pp. 161-192, 2010.

[20] I. Witten, E. Frank and M. Hall. Data Mining: Practical Machine Learning Tools and Techniques. Elsevier. 2011

[21] V. Zoratto, G. Aranda, S. Roger, A. Cechich, Análisis de estrategias para clasificar contenidos en foros de discusión: Un caso de estudio ASSE 2015, pp. 176-190.

[22] V. Zoratto, G. Aranda, S. Roger, A. Cechich, Analyzing Discussion Forums ThreadsAbout Java Programming Language Usage, Electronic Journal of SADIO, 2016.

VALIDACIÓN Y VERIFICACIÓN CON ALLOY DE ONTOLOGÍAS EN OWL QUE UNIFICAN MODELOS DE DATOS HETEROGÉNEOS

Mario Luis Accattoli [1], Ana Garis [1], Daniel Riesco [1],

[1] Departamento de Informática - Facultad de Ciencias Físico-Matemáticas y Naturales (FCFMyN) Universidad Nacional de San Luis

accattolimario@gmail.com, agaris@unsl.edu.ar, driesco@unsl.edu.ar

RESUMEN

Ante el requerimiento de unificar sistemas de información preexistentes, las ontologías se presentan frecuentemente como la mejor opción para resolver interoperabilidad semántica e integrar modelos de datos de fuentes heterogéneas. Entre los lenguajes más populares para especificar dichas ontologías se encuentra OWL, un lenguaje basado en lógica descriptiva y soportado por un razonador que permite chequear la consistencia lógica. Sin embargo, tiene ciertas limitaciones para verificar y validar algunos aspectos relevantes. Por ejemplo, carece de características que posibiliten chequear si la ontología es coherente con el dominio del conocimiento, si es compatible para recuperar la información deseada, o si es completa y consistente de acuerdo a los requerimientos. Alloy es un lenguaje formal soportado por un analizador automático SAT, que habilita la Verificación y Validación (V&V) de modelos.

La presente línea de investigación tiene como objetivo estudiar el uso de Alloy para la V&V de ontologías en OWL desarrolladas para unificar modelos de datos heterogéneos. Como caso de estudio se plantea la V&V con Alloy, de una ontología genérica a nivel nacional especificada para unificar el modelo de datos utilizado en de las distintas jurisdicciones dentro los sistemas de información de los Poderes Judiciales de Argentina.

Palabras Claves: Ontologías, OWL, Alloy, Verificación y Validación de Modelos, Unificación Modelo de Datos Poder Judicial.

CONTEXTO

La línea de investigación se enmarca en la continuación del Proyecto de Investigación "Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad" – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Dicho proyecto fue acreditado con evaluación externa y contó con el financiamiento de la UNSL.

1. INTRODUCCIÓN

Dentro del ámbito de la Ingeniería de Software las ontologías son utilizadas para múltiples propósitos, tales como el modelado conceptual, la interoperabilidad semántica, y también para facilitar el intercambio e integración de los datos [1]. Los lenguajes ontológicos permiten definir la semántica de los conceptos, habilitando no solo el almacenamiento de la información sino también la búsqueda y recuperación desde modelos de datos que provienen de fuentes heterogéneas. Sin embargo, los beneficios estarían limitados si la definición de la ontología se llevara a cabo de manera informal, generando un modelo ambiguo e impreciso.

Los métodos formales, proveen técnicas, métodos y herramientas para asegurar la correctitud de los modelos. Típicamente, incluyen un lenguaje de especificación soportado por herramientas de verificación y validación utilizados para eliminar inconsistencias y chequear propiedades deseadas. OWL [2] es un lenguaje formal para ontologías, aplicado principalmente

para representar la semántica Web. Si bien, el lenguaje está basado en lógica descriptiva y es soportado por un razonador que permite chequear la consistencia lógica de la ontología, no es posible verificar y validar algunos aspectos relevantes, tales como si es coherente con el dominio del conocimiento, o compatible para recuperar la información deseada.

Alloy [3] es un lenguaje formal soportado por un analizador automático SAT, que habilita la Verificación y Validación (V&V) de modelos, y permite realizar especificaciones no ambiguas, así como identificar y demostrar propiedades del sistema. El potencial de Alloy podría ser aprovechado para la V&V de ontologías OWL.

La presente línea de investigación se enfoca en el estudio de Alloy para V&V de ontologías en OWL desarrolladas para unificar modelos de datos heterogéneos. En particular, la propuesta procura resolver un problema que se plantea en los sistemas de los Poderes Judiciales de Argentina relacionado a la dificultad de unificar información y homogeneizar conceptos, de tal manera que también se consideren las particularidades de los distintos distritos provinciales.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los ejes de la investigación se basan en los siguientes puntos:

- La evaluación de los mecanismos para llevar a cabo la unificación de modelos de datos de fuentes heterogéneas mediante OWL y la identificación de problemas de inconsistencias que deberían ser luego verificados y validados para chequear la correctitud del modelo.
- El análisis de antecedentes en la adopción de Alloy para la V&V de ontologías OWL y enfoques de transformación entre OWL y Alloy.
- La especificación en OWL de la ontología que unifica modelos de datos heterogéneos en el dominio de la Justicia. En este sentido, es necesario estudiar la existencia de ontologías ya definidas en dicho dominio y evaluar los elementos que deben ser verificados y validados.

3. RESULTADOS OBTENIDOS/ESPERADOS

Con respecto a los mecanismos para realizar la unificación de modelos de datos de fuentes heterogéneas, se han analizado enfoques generales presentados en la literatura, esto es, el de ontología global, múltiple e híbrida [1,4,5,6]. El estudio preliminar indica que se debería partir de un proceso de ingeniería inversa de los modelos actuales [7], para luego armar una ontología unificada del modelo de datos. Es necesario evaluar aún si el proceso de ingeniería inversa puede ser ejecutado en forma automático o semiautomático [8,9].

Resta analizar cómo se lleva a cabo la unificación con OWL en el caso de estudio planteado en el trabajo: Modelos de Datos de Sistemas de Gestión Judicial en Argentina. Se espera poder identificar problemas de inconsistencia mientras se desarrolla esta actividad, que permitan definir reglas de V&V generales a chequear con Alloy.

La adopción de Alloy para la V&V de estas ontologías OWL requiere investigar propuestas similares y enfoques de transformación entre OWL y Alloy, tales como el presentado en [10].

Para chequear la propuesta se utilizará un caso de estudio real. Es decir se planteará la definición de un modelo de datos unificado de los Sistemas de Gestión Judicial en Argentina a través de ontologías validado y verificado formalmente. La unificación no solo mejoraría la comparación entre provincias, sino que se convertiría en un base de datos de consultas de antecedentes, facilitando la recopilación de casos similares para una mejor resolución de los

actuales, así como el registro de la trazabilidad de comportamientos de personas relacionadas con causas judiciales.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de la línea de investigación está integrado por investigadores de la UNSL, asesores externos y alumnos de grado y posgrado. En el grupo se llevan a cabo numerosas tesis de especialización, maestría y doctorado, así como trabajos finales de carreras de grado. En particular, el presente trabajo se desarrolla en el marco de la tesis de uno de los autores para optar al grado de Doctor en Ingeniería Informática en la UNSL.

Los avances logrados en esta línea de investigación, en el marco del Proyecto de Investigación mencionado en la Sección Contexto, establecen una base para el inicio de otras tesis de posgrado, tanto en el doctorado como en la Maestría en Ingeniería de Software, la Maestría en Calidad del Software, la Especialización en Ingeniería de Software de la UNSL. De igual forma, el presente trabajo, brinda un aporte para desarrollar trabajos finales en las carreras de grado de la UNSL, específicamente en la Licenciatura en Ciencias de la Computación, Ingeniería en Informática e Ingeniería en Computación.

5. BIBLIOGRAFÍA

[1] I.F. Cruz, H. Xiao, “The role of ontologies in data integration”. Engineering Intelligent Systems for Electrical Engineering and Communications, 13 (4), p. 245. 2005.

[2] W3C, OWL Web Ontology Language Reference. Disponible en https://www.w3.org/TR/owl-ref/ (ultima visita marzo de 2020)

[3] D. Jackson. Software Abstractions: Logic, Language, and Analysis. MIT Press, revised edition, 2012.

[4] D. Skoutas and A. Simitsis, "Ontology-based Conceptual Design of ETL Processes for both Structured and Semi-structured Data”, International Journal on Semantic Web and Information Systems, Special Issue on Semantic Web and Data Warehousing, 2007.

[5] H. Wache, T. Vogele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann and S. Hubner, "Ontology-Based Integration of Infonnation. A Survey of Existing Approaches," In Proc. ofthe IJCAI workshop on Ontologies and Information Sharing, 2001.

[6] A. Buccella, A. Cechich, N. Brisaboa, “Ontology-Based Data Integration Methods: A Framework for Comparison”, Revista Colombiana de Computación, 2005.

[7] Justas Trinkunas, Olegas Vasilecas: Building Ontologies from Relational Databases Using Reverse Engineering Methods (2007)(CompSysTech’07)

[8] H. Afzal, M. Waqas, T. Naz, “OWLMap: Fully Automatic Mapping of Ontology into Relational Database Schema”, (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 7, No. 11, 2016.

[9] L. Zemmouchi-Ghomari, “Cohabitation of Relational Databases and Domain Ontologies in the Semantic Web Context”, Journal of Systems Integration, 2018.

[10] Y. Song, R. Chen, Y. Liu, “A Non-Standard Approach for the OWL Ontologies Checking and Reasoning”, Journal of Computers, Vol.7(10): 2454-2461, 2012.

CREACIÓN DE HERRAMIENTAS DE SOFTWARE PARA LA CONSTRUCCIÓN DE SISTEMAS DE IOT ESCALABLES Y SEGUROS

Sebastián U. Flores, Mario Berón, Daniel Riesco, Pedro Rangel Henriques
Departamento de Informática - Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis - Ejército de Los Andes 950 - San Luis - Argentina
Universidade do Minho Braga - Portugal
s.flores@outlook.com.ar, { mberon, driesco }@unsl.edu.ar
pedrorangelhenriques@gmail.com

RESUMEN

En la actualidad, los avances en el dominio del Internet of Things están transformando las formas en que las personas interactúan con el entorno y están dotando de inteligencia al mismo. Estos progresos buscan mejorar la calidad de vida de las personas, optimizar los procesos industriales e incrementar el cuidado del medio ambiente y de sus recursos. Grandes empresas han implementado modelos de producción innovadores incluyendo dispositivos IoT en sus sistemas, con el fin de incrementar su competitividad de cara a la industria 4.0. Sin embargo, el correcto desarrollo de un sistema de IoT podría ser un desafío si no se toman ciertos recaudos. Uno de los problemas que se presentan, desde la planificación del sistema, es el diseño de una arquitectura que favorezca la escalabilidad en cuanto a cantidad y variedad de dispositivos y usuarios conectados al mismo tiempo, publicando y recibiendo información, garantizando estabilidad, fluidez y seguridad en sus comunicaciones.

En este artículo, se presenta una línea de investigación que aborda la *creación de herramientas de software para la construcción de sistemas de IoT escalables y seguros*, buscando superar las dificultades mencionadas.

Palabras clave: *Sistema de IoT, Dispositivo, Cosa, Escalabilidad, Seguridad, Privacidad, Arquitectura, Internet.*

CONTEXTO

La presente línea de investigación se enmarca en dos Proyectos de Investigación. El primero: "*Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad*" – Facultad de Ciencias Físico- Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto N.º P-031516. Tal proyecto es la continuación de diferentes proyectos de investigación, a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el Programa de Incentivos. El segundo proyecto: "*Fortalecimiento de la Seguridad de los Sistemas de Software mediante el uso de Métodos, Técnicas y Herramientas de Ingeniería Reversa*" realizado en conjunto con la Universidade do Minho Braga, Portugal, fue aprobado por el Ministerio de Ciencia, Tecnología e Innovación Productiva (MinCyT), y su código es PO/16/93.

1. INTRODUCCIÓN

El término IoT es la abreviación de la frase en inglés "Internet of Things" (i.e. Internet de las Cosas). El mismo corresponde a un dominio de aplicación que integra diferentes campos tecnológicos y sociales. De acuerdo

con IEEE et. al. [1], aún no se ha alcanzado un consenso en cuanto a una definición del mismo, que contenga todas sus características y pueda facilitar una mejor comprensión de su alcance.
Existen varios puntos importantes a destacar en el Software de los sistemas de IoT [1]:

- Los sistemas operativos usados en dispositivos IoT deben estar diseñados para ejecutarse de forma eficiente (ya que trabajan en componentes de pequeña escala), sin dejar de proporcionar funcionalidades básicas que den soporte a los objetivos y propósitos de las aplicaciones ejecutadas en ellos [2][3][4].

- Al contar con escaso almacenamiento y con una capacidad de procesamiento reducida, es muy importante el desarrollo de interfaces de programación de aplicaciones (APIs) que favorezcan la reutilización de componentes y una adecuada gestión de los datos a almacenar/procesar[2][5][6][7].

- Los sistemas de IoT, potencialmente pueden crecer y llegar a componerse por millones de dispositivos diferentes, cada uno ubicado en lugares remotos del planeta y con usuarios ubicados también remotamente. En este contexto, la autogestión y auto-optimización de cada dispositivo y/o subsistema individual podría convertirse en la norma [2][3][8].

- La privacidad y seguridad deben ser garantizadas en cada proceso de los sistemas de IoT, ya que los mismos producen y manejan información reservada [7][9][10][11].

- Para garantizar la escalabilidad de un sistema de IoT, este debe poseer un diseño arquitectónico adecuado. Por la característica previamente mencionada, se entiende la creación de un sistema flexible que permita interconectar tantos dispositivos IoT como sea necesario. Tal interconexión se realiza sin que importe el medio de conexión físico o el sistema operativo que posea cada uno de ellos, siempre y cuando utilicen las interfaces de software y protocolos de comunicación adecuados [12][13][2][8].

- Finalmente, dado los sistemas de IoT manipulan información sensible de personas y organizaciones, deben ser concebidos y conducidos dentro de las restricciones y regulaciones de cada país.

Más allá de las características particulares de cada uno, los sistemas de IoT pueden ser utilizados en los siguientes ámbitos:

- Sistemas de domótica.

- Extracción y análisis automáticos o semi-automáticos de datos en líneas de producción industrial, que permitan optimizar los procesos de negocio [14].

- Monitoreo del estado físico de personas o animales, o extensión de sus capacidades físicas, a través de dispositivos que puedan ser vestidos.

- Monitoreo de cultivos, riego inteligente y control remoto de maquinaria utilizada en la industria agropecuaria (e.g. cosechadoras automáticas).

- Monitoreo del estado del medio ambiente, para la creación de ciudades inteligentes, el control del impacto en la naturaleza de la extracción de recursos naturales y la construcción de una industria más ecológica en general.

Los sistemas de IoT no necesariamente requieren de interacción humana en sus procesos. ETSI [12] trata los sitemas de IoT autónomos y semi-autónomos al definir las comunicaciones "*Machine-to-Machine (M2M)*" (en español, comunicaciones Máquina a Máquina), como aquellas realizadas directamente entre dispositivos y/o subsistemas de IoT,

con interacción humana escasa o nula. Esta clase de comunicaciones posee varias aplicaciones:

- Automatización de actividades repetitivas y/o de carácter determinístico, en las cuales los avances tecnológicos nos permiten prescindir de intervención humana.
- Aprendizaje de patrones de comportamiento humano o patrones en el funcionamiento de procesos propios del sistema. En este caso, la única interacción humana que podría presentarse es la de supervisión de los resultados del aprendizaje realizado por el sistema.
- Autonomización de un sistema aislado, localizado en una ubicación de difícil acceso para las personas.
- Edificios inteligentes que se adapten automáticamente a las condiciones del entorno. En este caso, existiría una constante comunicación entre subsistemas de IoT distribuidos en todo el edificio y sus alrededores, cada uno evaluando las condiciones de su propio entorno y comunicando los cambios relevantes al resto de los subsistemas, para que cada uno se adapte de la mejor forma.

Por lo mencionado en los párrafos anteriores, se puede notar que es difícil la creación de un marco general de trabajo que permita desarrollar e integrar sistemas de IoT de diversa escala y con ámbitos de aplicación heterogéneos. Este marco debe ser lo suficientemente flexible para permitir modificaciones estructurales de los sistemas, de una forma comprensible y segura. También, debe permitir la obtención de múltiples vistas que representen diferentes propiedades de los mismos. Para crear este marco de trabajo, es necesario detectar y comprender los problemas a los que se enfrentan los arquitectos y desarrolladores de sistemas de IoT y que, de acuerdo con Kranz et al. [14], llevan a que un 60 % de las iniciativas IoT se estanquen en la fase de prueba de concepto y, del 40 % restante, sólo el 26 % sean consideradas un completo éxito. Para concluir esta introducción, es importante mencionar que se prevé un amplio crecimiento en la cantidad de objetos conectados a internet a través de sistemas de IoT. De hecho, Kranz et al. estima que, durante el año 2020, se alcanzará una cifra de 50.000 millones de objetos conectados en todo el mundo [14].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En la actualidad, existe una gran variedad de sistemas operativos destinados a dispositivos IoT (e.g. TinyOS, Raspbian, RIOT, Ubuntu, Windows 10 IoT Core) [1][4]. También, existe una gran cantidad de librerías y aplicaciones implementadas con el fin de brindar soporte a desarrolladores de sistemas de IoT.
De acuerdo con IEEE et. al. [1], frente al crecimiento en la cantidad de *cosas conectadas a internet*, cada una perteneciendo posiblemente a diferentes *dominios de administración* (i.e. diferentes sistemas/subsistemas de IoT), es necesario repensar por completo los enfoques tradicionales en el desarrollo de aplicaciones web, siendo prioritarios elementos como la escalabilidad y la lógica distribuida.
Por otro lado, mientras que el IoT promete una mejor vida a través de dispositivos conectados y de la información que ellos generen, también marca el comienzo de una nueva era en cuanto a la privacidad y la seguridad. OWASP lista los 10 principales problemas de seguridad que suelen presentarse en sistemas de IoT de toda índole [15].
Frente a lo mencionado en los párrafos anteriores, está claro que los desarrolladores de sistemas de IoT se enfrentan a un problema principal: *la creación de herramientas de software para la construcción de sistemas de IoT escalables y seguros.*
Esta línea de investigación propone:

- Investigar en profundidad las vulnerabilidades de seguridad que afectan a los

sistemas de IoT.

- Investigar el estado del arte de los patrones/estilos arquitectónicos de sistemas de IoT, y de los protocolos utilizados en su desarrollo.

- Integrar las investigaciones realizadas con el fin de crear un conjunto de herramientas que faciliten al desarrollador la creación de sistemas de IoT escalables y seguros, brindando soporte en la integración de subsistemas que posiblemente estén ubicados en diferentes posiciones geográficas, posean diferentes sistemas operativos y utilicen diferentes protocolos de software y hardware.

3. RESULTADOS OBTENIDOS/ESPERADOS

Con el objetivo de evaluar los conceptos investigados, se implementaron dos herramientas de software. La primera de ellas es F-IoT Core [16], una aplicación backend desarrollada en el framework Grails [17] que nuclea toda la lógica de negocio de F-IoT y expone APIs REST para la configuración y monitoreo de sistemas de IoT de diferentes formas. Con el fin de desarrollar un Frontend para acceder de forma más amigable a las configuraciones de los sistemas de IoT gestionados por F-IoT Core, se incorporó la segunda herramienta llamada F-IoT Frontend. La misma consiste en una aplicación web implementada en el framework Angular 8 [18], que brinda al usuario una vista de todos los sistemas de IoT registrados y le permite tanto modificarlos estructuralmente, como visualizar el último estado reportado de los dispositivos que lo comprenden.
A futuro se poseen los siguientes objetivos:

- Implementar en F-IoT Frontend un sistema de visualización en tiempo real del estado de los dispositivos IoT, a través de sistemas del estado del arte. Actualmente se está incursionando en el uso de sockets Web, a través de la librería Socket.IO [19], una de las más usadas por los desarrolladores Web en el mundo, dado que posee una gran flexibilidad y robustez, mucha documentación, es gratuita y de código abierto.

- Realizar una investigación profunda acerca de las amenazas de seguridad que afectan a los sistemas de IoT, y crear una capa de seguridad que abarque todas las comunicaciones realizadas en F-IoT.

- Crear patrones de diseño arquitectónico que puedan ser usados en diferentes ámbitos de aplicación de sistemas de IoT e indicar, para cada patrón, aquellos ámbitos en los que sea más viable su implementación.

- Analizar y comparar las diferentes tecnologías IoT disponibles para ejecución en la Nube, a través de servicios provistos por terceros (e.g. Microsoft, Google, Amazon, IBM, Oracle).

4. FORMACIÓN DE RECURSOS HUMANOS

Los progresos obtenidos en esta línea de investigación sirven como base para el desarrollo de tesis de posgrado, ya sea de doctorado o maestrías en Ingeniería de Software y desarrollo de trabajos finales de las carreras Licenciatura en Ciencias de la Computación, Ingeniería en Informática e Ingeniería en Computación de la Universidad Nacional de San Luis, en el marco de los Proyectos de Investigación mencionados en la Sección *Contexto*.

5. BIBLIOGRAFÍA

[1] IEEE. (mayo de 2015). Towards a Definition of the Internet of Things(IoT), dirección: `https://iot.ieee.org/images/files/pdf/IEEE_IoT_Towards_Definition_Internet_of_Things_Revision1_27MAY15.pdf`. Último acceso: 01 03 2020.

[2] IETF. (2010). The Internet of Things - Concept and Problem Statement, dirección: https://www.ietf.org/archive/id/draft-lee-iot-problem-statement-05.txt.

[3] CERP-IoT. (2010). Visions and Challenges for Realising the Internet of Things, dirección: http://www.internet-of-things-research.eu/pdf/IoT_Clusterbook_March_2010.pdf.

[4] S. Pérez, P. Gaur y M. P. Tahiliani, «Operating Systemas for IoT Devices: A Criical Survey», *IEEE Region 10 Symposium*, mayo de 2015.

[5] A. Bassi, M. Bauer, M. Fiedler, T. Kramp, R. van Kranenburg, S. Lange y S. Messner, «Enabling Things to Talk: Designing IoT Solutions with the IoT Architectural Reference Model», *SpringerOpen*, 2013.

[6] H. Ren, H. Li, Y. Dai, K. Yang y X. Lin. (mar. de 2018). Querying in Internet of Things with Privacy Preserving: Challenges, Solutions and Opportunities, dirección: http://ieeexplore.ieee.org/document/8315210/.

[7] T. Choudhury, A. Gupta, S. Pradhan, P. Kumar e Y. S. Rathore. (oct. de 2017). Privacy and Security of Cloud-Based Internet of Things (IoT), dirección: http://ieeexplore.ieee.org/document/8307328/.

[8] A. Zanella, N. Buis, A. Castellani, L.Vangelista y M. Zorzi, «Internet of Things for Smart Cities», *IEEE Internet of Things Journal*, feb. de 2014.

[9] M. El-hajj, M. Chamoun y F. S. Ahmed. (dic. de 2017). Taxonomy of authentication techniques in Internet of Things (IoT), dirección: http://ieeexplore.ieee.org/document/8305419/.

[10] J. Singh, T. Pasquier, J. Bacon, H. Ko y D. Eyers, «Twenty Security Considerations for Cloud-Supported Internet of Things», *IEEE Internet of Things Journal*, 2015.

[11] S. Pérez, J. L. Hernández-Ramos, S. N. Matheu-García, D. Rotondi, A. F. Skarmeta, L. Straniero y D. Pedone. (feb. de 2018). A lightweight and flexible encryption scheme to protect sensitive data in Smart Building scenarios, dirección: http://ieeexplore.ieee.org/document/8279412/.

[12] ETSI. (sep. de 2010). ETSI Technical Specification, Machine-to-Machine Communications (M2M); M2M Service Requirements. Technical Specification. ver. 1.1.1, dirección: www.etsi.org/deliver/etsi_ts/102600_102699/102689/01.01.01_60/ts_102689v010101p.pdf.

[13] D. Zeng, S. Guo y Z. Cheng. (2011). Journal of Communicaitons, dirección: http://www.jocm.us/index.php?m=content&c=index&a=show&catid=50&id=134.

[14] M. Kranz, *Internet Of Things. Construye nuevos modelos de negocio*. LID Editorial, 2017, ISBN: 9788416894895.

[15] OWASP. (2018). Project Internet of Things, dirección: https://owasp.org/www-project-internet-of-things/.

[16] S. Flores, M. Berón, D. Riesco, P. Henriques y M. Bustos, «F-IoT Core, una Herramienta para el Desarrollo de Soluciones IoT Escalables y Flexibles», *CoNaIISI*, 2018.

[17] I. Object Computing. (2020). Grails Framework, dirección: https://grails.org/.

[18] Google. (2020). Angular Framework, dirección: https://angular.io/.

[19] S. IO. (2020). Socket.IO, dirección: https://socket.io/.

Técnicas, Estrategias y Herramientas de Comprensión de Programas para facilitar el entendimiento de Sistemas Multiparadigmas

Enrique Miranda[(1)], Corina Abdelahad[(1)], Mario Berón[(1)], Daniel Riesco[(1)],

[(1)]Departamento de Informática / Facultad Ciencias Físico Matemáticas y Naturales/
Universidad Nacional de San Luis
Ejército de los Andes 950 – San Luis – Argentina
{eamiranda, cabdelah, mberon,driesco}@unsl.edu.ar

Resumen

Dentro del contexto del ciclo de vida del producto de software, una de las tareas que más tiempo y dedicación consume es la de Mantenimiento y Evolución de Software (MES). A partir de la necesidad de asistir al arduo proceso de comprensión requerido en la etapa mencionada anteriormente, surge una disciplina de la Ingeniería de Software denominada Comprensión de Programas (CP). La CP se presenta como un área de investigación interesante para impulsar el trabajo de MES a través de técnicas y herramientas que asistan al ingeniero de software en la difícil tarea de comprender sistemas.

Por otra parte, los desafíos recientes en la industria de software requieren cada vez más de lenguajes y frameworks de programación con características multiparadigmas, los cuales a su vez presentan desafíos a las estrategias de comprensión.

En esta línea de investigación se propone estudiar conceptos, definir estrategias e implementar herramientas de Comprensión de Programas para analizar sistemas desarrollados con lenguajes con soporte multiparadigma.

Palabras clave: Ingeniería Reversa, Comprensión de Programas, Lenguajes con soporte Multiparadigma, Sistemas de Software de gran Envergadura.

Contexto

La presente línea de investigación se enmarca en el Proyecto de Investigación "Ingeniería de Software: Conceptos, Prácticas y Herramientas para el Desarrollo de Software con Calidad" de la Facultad de Ciencias Físico Matemáticas y Naturales de la Universidad Nacional de San Luis. Proyecto No P-031516 que es la continuación de diferentes proyectos de investigación, a través de los cuáles se ha logrado vínculos con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el Programa de Incentivos.

Introducción

Siempre que se realice un cambio a una pieza de software, es importante que el ingeniero de software obtenga un completo entendimiento de la estructura, comportamiento y funcionamiento de la parte del sistema que se está modificando. Es sobre la base de este entendimiento, que se pueden generar las propuestas de modificación para lograr los objetivos en la etapa de mantenimiento.

Los ingenieros de software encargados del mantenimiento, empeñan mucho de su tiempo leyendo el código y la documentación complementaria para comprender su lógica, propósito y estructura. En este contexto, el entendimiento de un sistema no es tarea fácil, ya que generalmente, la persona que realiza el mantenimiento no es la misma que escribió el código o bien ha pasado un período de tiempo considerable desde que esta persona finalizó las tareas de desarrollo. En este sentido, Eagleson refleja claramente este aspecto: *"Cualquier pieza de código que no se haya visto desde hace seis meses o más, bien podría haber sido escrita por alguien más"*.

La ley expone una visión de lo difícil que puede resultar mantener un sistema de software, al punto tal que la misma persona que ha desarrollado un sistema particular puede resultarle desconocido luego de un determinado tiempo. A todo lo anterior se le debe sumar que los sistemas en su gran mayoría no cuentan con documentación que cumpla ciertos requisitos básicos como por ejemplo que la misma se encuentre actualizada, que sea consistente, que esté completa, entre otros.

Es evidente que todos los aspectos mencionados anteriormente han sido los motivos que han impulsado el surgimiento de distintos recursos para asistir al ingeniero de software en la difícil tarea de comprender un sistema. En este contexto, surge una disciplina denominada Comprensión de Programas (CP), la cual está dirigida a proveer modelos, métodos, técnicas y herramientas, basada en un proceso de aprendizaje específico y procesos de ingeniería, a fin de encontrar un conocimiento más profundo acerca de un sistema de software [1].

A través de un extenso estudio y análisis de herramientas de comprensión [1,2,3], se pudo comprobar que para lograr una comprensión efectiva del sistema bajo estudio, el sujeto cognoscente debe interrelacionar el Dominio del Problema con el Dominio del Programa. El primer dominio, hace referencia a la salida del sistema. El segundo, a las componentes de software usadas para producir dicha salida. Claramente, existe una relación (o relaciones) concreta y robusta entre ambos dominios. Durante la implementación del sistema el desarrollador va construyendo (desde su perspectiva) la relación entre los dominios y de alguna manera la misma se encuentra activa durante el ciclo de desarrollo. Sin embargo, en la etapa de mantenimiento y evolución del software es muy posible que dicha relación pierda fuerza y en determinados casos llegue a desaparecer. La reconstrucción de este tipo de relación es compleja y requiere de estrategias de comprensión efectivas que sean elaboradas tomando como base diferentes temáticas como lo son las Técnicas de Extracción de la Información, Administración de la Información, Estrategias de Vinculación de Dominios, Visualización de Software, entre otras [1,2,3,4].

Actualmente, existen numerosas herramientas de CP con sofisticadas técnicas de exploración de código. Muchas de estas funcionan de manera adecuada, sin embargo, ciertas tareas de comprensión son todavía muy complejas. Ciertas herramientas proveen diferentes vistas de la información extraída del sistema. No obstante, el uso de este tipo de técnicas no asiste al programador en las complejas tareas de abstracción y vinculación de los Dominios del Problema y del Programa, ya que las vistas provistas brindan solo un aspecto del código fuente [4]. Una forma de salvar este inconveniente es a través de la elaboración de estrategias que interconecten los Dominios del Problema y Programa [1,4,5]. La característica mencionada previamente simplifica la exploración porque el ingeniero de software solamente inspecciona las partes del sistema relacionadas con la funcionalidad de estudio. De esta manera, se alcanza una considerable reducción de esfuerzos humanos, lo que se traduce en reducción de costos del proyecto.

Tomando en cuenta este contexto, a través del estudio del estado del arte de CP es posible inferir los siguientes aspectos:

- El proceso de elaboración de estrategias de vinculación de Dominios es arduo y complejo. El mismo necesita de un estudio profundo de las distintas disciplinas que se mencionaron previamente; es por esto que es difícil encontrar numerosas estrategias de vinculación de Dominios con estas características.
- La mayoría de las técnicas y herramientas de CP abordan exclusivamente el Dominio del Programa. Es decir, realizan análisis de los artefactos relativos al código fuente de un programa sin tener en cuenta componentes del Dominio del Problema. Ciertos trabajos sólo proponen teorías y técnicas novedosas, pero no presentan herramientas, casos de estudios o Dominios de aplicación interesantes en donde se puedan poner en práctica efectivamente las mismas.
- Entre las herramientas y estrategias propuestas es difícil encontrar algunas que

abarquen lenguajes multiparadigma como por ejemplo Python. Este estilo de programación ha tomado cada vez más relevancia ya que es ampliamente utilizado en la actualidad.

Teniendo en cuenta los aspectos destacados en los ítems anteriores, es posible determinar que la mayoría de las técnicas y herramientas existentes de CP no proponen estrategias robustas y eficaces para interrelacionar los Dominios del Problema y del Programa. Además, del limitado conjunto de estrategias disponibles que abordan dicho enfoque, algunas poseen ciertas desventajas que impide su aplicación a sistemas de gran tamaño escritos en lenguajes ampliamente utilizados en el contexto de desarrollo de software a gran escala.

Como objetivo general de esta línea de investigación se pretende mejorar la comprensión de sistemas multiparadigmas, por medio de la definición de estrategias de CP que permitan vincular los Dominios del Problema y del Programa. Como medio para alcanzar dicho objetivo se propone el estudio de distintas temáticas relacionadas a modo de líneas de investigación comprendidas dentro de la temática principal.

Líneas de Investigación y Desarrollo

A continuación, se describe brevemente un conjunto no exhaustivo de líneas de investigación desprendidas de la línea principal.

Extracción de Información

Una de las actividades iniciales de cualquier enfoque que se enmarque dentro del contexto de Comprensión de Programas, es la extracción de la información del sistema bajo estudio. Por esta razón, se utilizan técnicas que permiten obtener distintos tipos de datos relacionados con determinados aspectos de la estructura y/o el comportamiento del programa. Dichas técnicas se conocen en el contexto de CP como Técnicas de Extracción de Información (TEI). Se puede decir que las TEI se subdividen en dos categorías dependiendo del tipo de información que se desee extraer: TEI estática y TEI dinámica. Las primeras permiten recuperar todos los atributos de los objetos definidos en el código fuente del programa [6]. Las segundas posibilitan conocer los componentes del programa utilizados para una ejecución específica del sistema [6,7].

Es difícil determinar cual es la técnica más eficaz, ya que se podrían encontrar numerosos trabajos que presentan resultados interesantes usando alguna de las técnicas. Incluso diversos enfoques combinan ambos tipos de técnicas, logrando así destacar distintos aspectos del sistema mediante la información obtenida con técnicas dinámicas en conjunto con la información estática [1,6].

Esta temática se torna más determinante cuando los sistemas analizados están desarrollados en lenguajes con soporte multiparadigma. Este tipo de lenguajes presentan numerosos desafíos en el contexto de esta temática [5,8].

Reducción de información

En el contexto de análisis de software de gran envergadura, es necesario contar con técnicas de reducción de información para poder generar abstracciones. Por lo general, en dichos sistemas se extrae cierta cantidad de información que no está vinculada con aspectos relativos a la solución del problema para el cual fue desarrollado el sistema. Para que las representaciones de base contengan información precisa referente al problema que el sistema intenta solucionar, es necesario filtrar aquellos artefactos de software que no están relacionados con la lógica subyacente del programa. Para llevar a cabo la abstracción es posible valerse de distintas técnicas y herramientas, como por ejemplo, métricas de software, técnicas de *clustering*, estrategias de identificación de artefactos irrelevantes, entre otras [9,10].

Transformaciones entre Modelos

Un enfoque que permite lidiar con las características y dificultades que presentan los sistemas multiparadigmas es la Arquitectura Dirigida por Modelos (MDA) [11,12]. La idea principal de MDA, es separar la funcionalidad del sistema de su implementación sobre plataformas específicas, considerando la evolución desde modelos abstractos hasta su implementación, acrecentando el grado de automatización y logrando interoperabilidad con

múltiples plataformas, lenguajes formales y lenguajes de programación.
Usando MDA, cualquier especificación puede ser expresada con modelos, en consecuencia, una estrategia de comprensión puede comprender un proceso de refinamiento y transformación entre modelos de manera tal que el nivel de abstracción vaya aumentando hasta llegar a una representación independiente de ciertos aspectos específicos tales como una plataforma, paradigmas de programación, lenguajes de programación, entre otros [11,13].

Modelos del Dominio del Problema
A la hora de interconectar los Dominios del Problema y del Programa es necesario contar con representaciones de cada uno de estos [1,3,4,5]. En el contexto de CP existen diversas estrategias que realizan extracción y representación del Dominio del Programa. Sin embargo, se han encontrado relativamente escasas propuestas de CP que realicen extracción de información del Dominio del Problema para posteriormente construir la representación del mismo. Uno de los principales factores que justifican este hecho es que el Dominio del Problema no cuenta con la cantidad de artefactos que pueden relevarse en el Dominio del Programa. Por lo general, el primero se encuentra plasmado en narrativas escritas en lenguaje natural, especificaciones formales u ontologías. También pueden considerarse la salida del sistema o su comportamiento como un artefacto del Dominio del Problema [1,3,6]. Como puede observarse, los artefactos de este dominio no son fáciles de analizar y por consiguiente no es sencillo extraer información de los mismos. La mayoría de las estrategias de comprensión que consideran el Dominio del Problema, lo hacen teniendo en cuenta representaciones del mismo ya definidas. Por esto motivo se considera relevante ahondar en el estudio de esta temática para lograr obtener otras representaciones de este dominio y así contar con mayores recursos a la hora de definir estrategias efectivas de CP.

Estrategias de Interconexión entre Dominios
Una vez construidas (o definidas) las representaciones de los Dominios del Problema y del Programa, es necesario definir la estrategia de vinculación que permitirá al ingeniero de software relacionar aquellos artefactos del Dominio del Programa que implementan funcionalidades del Dominio del Problema. Este aspecto es fundamental y el mismo presenta numerosos desafíos dentro del contexto en que se plantea esta línea, es decir, sistemas de software con soporte multiparadigma.

Estrategias de Visualización de Información
Uno de los aspectos fundamentales a tener en cuenta en cualquier estrategia efectiva de comprensión es la manera en que se mostrará la información que ha sido procesada.
La Visualización de Software (VS) es una disciplina de la Ingeniería de Software cuyo propósito es visualizar la información relacionada con los sistemas de software con el objetivo de simplificar el análisis y la comprensión de los mismos [1,14,15]. En el contexto de CP, este aspecto se torna trascendental debido a que, es el vehículo por medio del cual el ingeniero de software puede efectivamente reconstruir, de diversas maneras, la relación entre los dominios [1,4,5].

Resultados Obtenidos/Esperados

Algunos de los resultados destacados obtenidos por esta investigación son:

- Se desarrolló e implementó una estrategia de vinculación de Dominios que fue aplicada a sistemas en Python [5].
- Se utilizaron técnicas de transformación entre modelos para desarrollar una estrategia de Comprensión de Programas [13].
- Se establecieron temáticas de investigación que posibilitarán un abordaje más abarcativo del objetivo principal de la línea.

Entre los objetivos planteados a corto y largo plazo se pueden mencionar:

- Extender los tipos de representaciones del Dominio del Problema utilizados en estrategias de CP.
- Ampliar los tipos de plataformas y lenguajes abarcados por las estrategias vigentes.

- Establecer un conjunto de métricas orientadas a determinar el índice de comprensibilidad de sistemas multiparadigmas.
- Ampliar las estrategias de Ingeniería Reversa de sistemas multiparadigmas que toman como base MDA.
- Definir nuevos mecanismos de abstracción para las representaciones intermedias de las estrategias de CP.

Formación de Recursos Humanos

Las investigaciones realizadas, así como los resultados obtenidos en este trabajo contribuyen al desarrollo de tesis de posgrado, ya sea de doctorado o maestrías en Ingeniería de Software y desarrollo de trabajos finales de las carreras Licenciatura en Ciencias de la Computación, Ingeniería en Informática e Ingeniería en Computación de la Universidad Nacional de San Luis, en el marco del proyecto de investigación mencionado.

Bibliografía

[1] Mario Berón. "Program Inspection to interconnect Behavioral and Operational Views for Program Comprehension". Ph.D. Thesis Dissertation at University of Minho. Braga. Portugal, 2010.

[2] Margaret-Anne Storey. "Theories, Methods and Tools in Program Comprehension: Past, Present and Future". Proceedings of the 13th International Workshop on Program Comprehension, p. 181–191, 2005.

[3] Rugaber S. The Use of Domain Knowledge in Program Understanding. Annals of Software Engineering, 9 (1-2), p. 143–192, 2000.

[4] Carvalho, N. R., Almeida, J. J., Henriques, P. R., y Varanda, M. J. From source code identifiers to natural language terms. Journal of Systems and Software, 100, p. 117–128, 2015.

[5] Miranda, E. et al. Using reverse engineering techniques to infer a system use case model. Journal of Software: Evolution and Process, vol. 31, no 2, p. e2121, 2019.

[6] Eisenbarth, T., Koschke, R., y Simon, D. Aiding Program Comprehension by Static and Dynamic Feature Analysis. En Proceedings of the IEEE Intern. Conference on Software Maintenance (ICSM'01) p. 602–611. IEEE Computer Society, 2001.

[7] Cornelissen, B., Zaidman, A., Van Deursen, A., Moonen, L., y Koschke, R. A systematic survey of program comprehension through dynamic analysis. IEEE Transactions on Software Engineering, 35 (5), p. 684–702, 2009.

[8] Wampler, D. et al. Multiparadigm programming in industry: A discussion with Neal Ford and Brian Goetz. IEEE software, vol. 27, no 5, p. 61-64, 2010.

[9] Abdelwahab Hamou-Lhadj and Timothy Lethbridge. Summarizing the Content of Large Traces to Facilitate the Understanding of the Behaviour of a Software System. In 14th International Conf. on Program Comprehension, ICPC '06. p. 181–190. IEEE, 2006.

[10] Jones, K. Automatic summarising: The state of the art. Information Processing & Management, vol. 43, no 6, p. 1449-1481, 2007.

[11] Roxana Giandini, Claudia Pons, and Gabriela Pérez. A two-level Formal Semantics for the QVT Language. In Proceeding of Conferencia Iberoamericana de Software Engineering, CIbSE '09, p. 73–86, 2009.

[12] Pereira, C., Martinez, L., y Favre, L. Recovering Use Case Diagrams from Object Oriented Code: an MDA-based Approach. En 8th International Conference on Information technology: New generations (ITNG), p. 737–742, 2011.

[13] Miranda, Enrique Alfredo, et al. Inferring Use-cases from GUI Analysis. IEEE Latin America Transactions, vol. 13, no 12, p. 3942-3952, 2015.

[14] Storey, M.-A., Fracchia, D., y Müller, H. Cognitive Design Elements to Support the Construction of a Mental Model During Software Visualization. En Proceedings of 5th Iternational Workshop Program Comprehension (IWPC'97), p. 17–28, 1997.

[15] Lanza, M., Ducasse, S., Gall, H., y Pinzger, M. Codecrawler- an information visualization tool for program comprehension. En proceeding on 27th

International Conference on Software Engineering, ICSE p. 672–673, 2005.

Evaluación de técnicas para la validación de requerimientos en entornos de trabajo para el desarrollo de software

Sonia R. Santana[1], Lucrecia R. Perero[1], Amalia G. Delduca[1,]
Gladys N. Dapozo[2]
(1) Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos
(2) Facultad de Ciencias Exactas y Naturales y Agrimensura - Universidad Nacional del Nordeste
e-mail: sr.santana675@gmail.com

Resumen

La validación de requerimientos es un proceso continuo en el proyecto de desarrollo de software con el fin de asegurar que los requerimientos elicitados sean representaciones exactas de las necesidades y expectativas de los usuarios. Esta actividad contribuye a mejorar la calidad de los requerimientos, a reducir costos, tiempos y riesgos en el desarrollo de software. En este proyecto se aborda el estudio de técnicas y herramientas libres para la validación de requerimientos de software con el objetivo de contribuir con información que apoye la actividad de los profesionales del sector de desarrollo de software, en cuanto a lograr la integridad de la especificación de los requerimientos de software. Se presentan los avances obtenidos sobre el aporte de las herramientas libres en los distintos procesos de la ingeniería de requerimientos. A partir de los resultados obtenidos se propone conseguir información sobre la performance de técnicas y herramientas libres para validar requerimientos en base a las restricciones de los entornos de trabajo para el desarrollo de software.

Palabras clave: Ingeniería de Requerimientos, Técnicas para la ingeniería de requerimientos. Validación de requerimientos.

Contexto

El presente PID 7057 se encuadra en la línea de investigación "Ingeniería de Software", establecida como prioritaria desde la carrera Licenciatura en Sistemas de la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos (UNER). Se adecua, además, a una de las prioridades de la UNER considerando que es un proyecto aplicado a la investigación sobre Tecnologías de la Información y la Comunicación.

Introducción

El desarrollo de software no es solo un desafío tecnológico sino también un proceso social complejo. En este proceso, la comunicación efectiva entre los interesados desempeña un papel vital en el desarrollo del proyecto de software que conduce a recopilar requerimientos de manera correcta y consistente.

La tecnología cambia rápidamente y la creciente competencia está poniendo cada vez mayor presión sobre el proceso de desarrollo. El desarrollo de software en un entorno de trabajo global es una tendencia emergente de la Ingeniería de Software. Busca disminuir costos y acceder al fondo común de recursos mundiales, son los principales factores de la globalización en el desarrollo de software. Presenta ciertas características como la distancia, la zona horaria

y las diferencias interculturales que se diferencia del desarrollo de software en un entorno de trabajo compartido. Estas transformaciones llevan asociadas la aparición de nuevas restricciones, no contempladas en un entorno de trabajo compartido, y que, por lo tanto, obligan a adoptar nuevas estrategias para validar los requerimientos.

El principal objetivo de la validación de los requerimientos es certificar que los requerimientos elicitados son representaciones exactas de las necesidades y expectativas de los usuarios [1] [2] [3] y deben ser completos, correctos y consistentes [4]. La validación de los requerimientos en el desarrollo de software de un entorno de trabajo compartido, se produce en un solo lugar con los usuarios. Esto facilita la comunicación y la coordinación entre ellos, es decir, requerimientos que faltan, problemas, conflictos, errores y contradicciones pueden resolverse fácilmente. Debido a desafíos impuestos por los avances tecnológicos en un entorno de trabajo global, esta actividad se vuelve crítica.

Las técnicas para validación de requerimientos desempeñan un papel fundamental para detectar y prevenir posibles defectos y errores en los requerimientos, pero se comportan en forma diferente en un entorno de trabajo compartido que en un entorno de trabajo global.

Trabajar en el contexto de la validación de requerimientos se está convirtiendo en un desafío para los equipos, clientes y usuarios, existen diferentes causas que imponen problemas de comunicación, control, intercambio de conocimientos, confianza y retrasos en el desarrollo del software [5].

Muchos investigadores han desarrollado y estudiado las técnicas para la validación de requerimientos. Algunos de los estudios se dan de la siguiente manera:

Moketar [6] realizó una revisión de la literatura donde muestra la tendencia del enfoque de la validación de requerimientos de software, desde el año 2007 hasta 2016. Según el estudio muestra que el enfoque / método semi-formal y el Lenguaje Unificado de Modelado (UML) son las más utilizados para la validación de requerimientos. En cuanto a las técnicas, la creación de prototipos es el favorito seguido de la simulación y la validación de requerimientos basados en pruebas y en modelos. El estudio también publicó que los criterios de calidad más importantes de los requerimientos son la consistencia, exactitud e integridad.

Sourour [7] analizó la evolución de las técnicas para validación y su estado actual en Ingeniería de Requerimientos. La literatura tiende a considerar la validación de requerimientos, como un proceso heterogéneo basado en la aplicación de una variedad de técnicas independientes; sin ser capaz de especificar: qué, por qué, quién, cuándo y cómo validar los requerimientos. Las técnicas que desempeñan un papel importante son: la revisión, inspección, técnicas de lectura, creación de prototipos, validación basada en modelos, validación basada en pruebas y validación basada modelos.

Sulehri [8] presentó en su tesis los pros y los contras de las técnicas para validación de requerimientos practicada en diferentes empresas de software en Suecia y disponible en el ámbito académico. El informe del estudio muestra claramente que la creación de prototipos de requerimientos se practica tanto en pequeñas como en grandes empresas de software. Por otra

parte, el informe indica que los prototipos y las pruebas son muy efectivos para detectar defectos, pero son semanales en términos de tiempo y costo con respecto a otros. Las revisiones tienen control de la semana sobre la detección de defectos pero muy excelente en términos de tiempo y costo.

Hoy, la demanda de flexibilidad, desarrollo ágil, colaboración mundial y el software avanzado están cambiando la manera en que se gestionan los requerimientos. Por otro lado, el desarrollo en entorno de trabajo global los equipos necesitan acceder fácil y ampliamente a los requerimientos y las especificaciones con trazabilidad a lo largo de todo el ciclo de vida. Uno de los principales problemas con la fase de validación de requerimientos es la falta de aplicación de una estrategia o marco para validar los requerimientos. Las empresas practican la validación de requerimientos sobre una base ad hoc, debido a la falta de personal capacitado o falta de formación o la exposición a la validación de requerimientos. Se presta más atención a la fase de pruebas de software que se lleva a cabo por lo general en el final, cuando se integran en todos los módulos de proyectos de software [9].

En función de lo señalado, nuestro abordaje busca profundizar en las técnicas y herramientas para validar requerimientos en entornos de trabajo para el desarrollo de software.

Líneas de Investigación, Desarrollo e Innovación

En el marco de la investigación se pretende:

- Estudiar en profundidad los conceptos y técnicas para la validación de requerimientos en el marco de la Ingeniería de Requerimientos.
- Estudiar los principales problemas causados según las restricciones en los entornos de trabajo para el desarrollo de software.
- Evaluar las técnicas y herramientas para la validación de requerimientos según las restricciones preestablecidas y obtener un cuadro resumen de la performance en los entornos de trabajo para el desarrollo de software.

Resultados obtenidos/esperados

En el marco de este proyecto y respecto de la línea de gestión de requerimientos se lograron los siguientes resultados [10]:

- El aporte de cada herramienta libre en los distintos procesos de la ingeniería de requerimientos, se muestran en la Fig. 1.

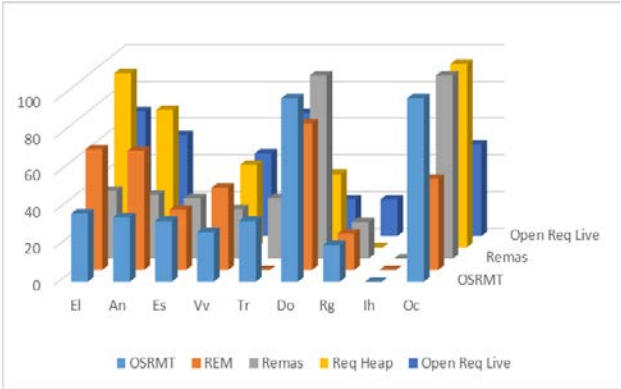


Fig.1. Evaluación de las herramientas libres en la gestión de requerimientos.

- El grado de cumplimiento de las funcionalidades de las herramientas libres requeridas para una completa gestión de requerimientos, se muestra en la Fig.2.

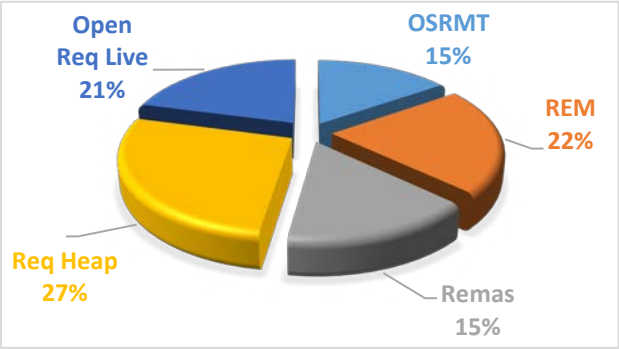


Fig.2. Grado de cumplimiento de las funcionalidades requeridas

- El grado de capacidad de las herramientas libres en base a determinadas funcionalidades y escenarios, se muestran en la Tabla 1.

Tabla 1. Capacidades de las herramientas libres para los escenarios.

Escenarios	*Capacidad Herramienta*	*Nombre Herramienta*
Escenario 1 *(Elicitación **y** V&V de requerimientos)*	79%	ReqHeap
Escenario 2 *(Modelado y especificación de requerimientos)*	41%	OSRMT
Escenario 3 *(Gestión de requerimientos)*	68%	ReqHeap

En cuanto a los resultados esperados, el procedimiento metodológico adoptado para esta investigación se basa en tres etapas distintas. En la primera etapa se seleccionarán diversas técnicas para la validación de requerimientos teniendo en cuenta su objetivo y alcance, con el propósito de establecer una contribución efectiva del presente proyecto frente al escenario actual. Identificada las técnicas, en la segunda etapa se analizarán los diferentes problemas causados por las restricciones en los entornos de trabajo para el desarrollo de software. Finalmente, como última etapa las técnicas serán analizadas con los siguientes criterios: Comunicación – Control - Intercambio de conocimientos – Demora – Confianza – Costo - Tamaño de la empresa y equipo de desarrollo - Reutilización.

Para formalizar la comparación entre los criterios de evaluación se orientará a responder cuestiones tales como:

a) Puntos fuertes y débiles de las técnicas y herramientas para la validación de requerimientos en los entornos de trabajo.
b) Capacidad de las técnicas y herramientas para detectar posibles defectos y errores en los requerimientos.
c) Disponibilidad de datos sobre la performance de las técnicas y herramientas para validación en los entornos de trabajo.

Con esta información sobre técnicas y herramientas para la validación de requerimientos aplicadas a entornos de trabajo para el desarrollo se espera contribuir con la actividad de los profesionales del sector de desarrollo de software. Además, la comparación presentada debería ayudar en la aplicación de las técnicas y herramientas en la ejecución de proyectos dentro de la línea de tiempo, presupuesto, y de acuerdo con la funcionalidad deseada.

Formación de Recursos Humanos

Este proyecto prevé la iniciación en actividades de investigación de tres docentes de la carrera Licenciatura en Sistemas, y la realización de una tesis de maestría correspondiente a la Maestría de Ingeniería de Software de la Facultad de Informática de la UNLP. La línea cuenta con la

colaboración de una docente investigadora de la UNNE.

Referencias

1. P. A. Laplante: Requirements Engineering for Software and Systems, CRC Press (2019).
2. B. H. C. Cheng and J. M. Atlee: Current and Future Research Directions in Requirements Engineering, Design Requirements Engineering A Ten-Year Perspective, Lecture Notes in Business Information Processing, vol. 14, pp. 11–43 (2019).
3. G. Kotonya and I. Sommerville: Requirements Engineering Processes and Techniques, Wiley (1998).
4. S. L. Pfleeger: Software Engineering – Theory and Practice, Prentice Hall (1998).
5. Loucopoulos, P., Karakostas, V.: System Requirements Engineering, McGraw-Hill, London, ISBN 0-07-707843-8 (1995).
6. Moketar, Nor Aiza & Kamalrudin, Massila & Mohd-Yusof, Mokhtar & Sidek, Safiah, "A review on requirements validation for software development", Journal of Theoretical and Applied Information Technology. 96. 3182-3193 (2018).
7. Sourour, Maalema Nacereddine, Zarourb: Challenge of validation in requirements engineering, Journal of Innovation in Digital Ecosystems, volume 3, Issue 1, Pages 15-21, (2016).
8. Sulehri Latif Hussain: Comparative Selection of Requirements Validation Techniques Based on Industrial Survey, Blekinge Institute of Technology, School of Computing, (2010).
9. U. A. Raja: Empirical studies of requirements validation techniques, 2nd International Conference on Computer, Control and Communication, Karachi, pp. 1-9 (2009).
10. Santana, S., Perero, L., Delduca, A., Dapozo, G.: Evaluación de las herramientas libres para la gestión de requerimientos. CACIC 2019. ISBN 978-987-688-377-1.

Construcción de Interfaces Gráficas de Usuario aplicando técnicas de la Ingeniería del Software y utilizando Metodologías Centradas en el Usuario

M. Claudia Albornoz – Corina N. Abdelahad
Departamento de Informática/Universidad Nacional de San Luis-U.N.S.L./San Luis/Argentina
Ejército de los Andes 950, Tel: +54 (0266) 4520300 (int 2102)
{albornoz, cabdelah }@unsl.edu.ar

Resumen

El continuo avance de la digitalización de contenidos y sistematización de la información hace que todas las personas se vean obligados a interactuar con productos de software.

Al comenzar a trabajar con cualquier dispositivo (computadora, tablet, celular, etc.) el usuario comienza a interactuar con la interfaz, ya sea la del sistema operativo, la de un software/aplicación en particular o la de cualquier sitio web.

La Interfaz Gráfica de Usuario (GUI por su nombre en inglés, Graphical User Interface) es parte fundamental de cualquier producto de software; allí comienza la interacción hombre-computadora. El diseño y construcción de la GUI no se lo debe considerar como una tarea secundaria y sin importancia; por el contrario, el equipo de desarrollo debe contar con integrantes especializados en el tema.

Para esto se deben aplicar técnicas de diseño e incluir al usuario en cada etapa de la construcción.

El objetivo del presente trabajo es mostrar la importancia del diseño de la GUI utilizando técnicas para su construcción en sistemas interactivos y mostrar la importancia de realizar un diseño centrado en el usuario.

***Palabras Claves**:*

Interfaz Gráfica de Usuario (GUI), Interacción Persona-Computadora (IPC), Diseño Centrado en el Usuario (DCU)

Contexto

Este trabajo de investigación se encuentra enmarcado dentro de una de las líneas de investigación del Proyecto de Incentivos cuyo título es: "Ingeniería de Software: conceptos, prácticas y herramientas para el desarrollo de software de calidad", código P-031516, de la Facultad de Ciencias Físico, Matemáticas y Naturales, de la Universidad Nacional de San Luis.

Este proyecto es la continuación de diferentes proyectos de investigación, los cuales han logrado importantes vínculos con diversas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el Programa de Incentivos.

1. Introducción

Todo producto de software tiene como principal objetivo permitir a los usuarios concretar sus tareas de manera eficiente y efectiva. Para que esto sea posible, no solo se debe poner énfasis en la funcionalidad, sino también en la interfaz.

Es necesario incorporar técnicas que ayuden a captar las necesidades del usuario, y así poder lograr interfaces intuitivas, fáciles de usar y satisfactorias. Algunas de estas técnicas son el análisis de casos de uso y de tareas [1].

El desarrollo de software a través de la aplicación estructurada de técnicas científicas, se lo considera una ingeniería. Así mismo, el desarrollo y la construcción de interfaces gráfica de usuario se convierte en una ingeniería al utilizar técnicas específicas de ingeniería [2]. La utilización

de estas técnicas asegura un producto de software de calidad.

Desde hace ya un tiempo, para el desarrollo de interfaces se aplican metodologías que involucra al usuario en todo el ciclo de vida del diseño, y se las conoce como metodologías centradas en el usuario.

Es necesario tener en claro que los sistemas interactivos deben mostrar un fluido diálogo con el usuario; pero el usuario no está interesado en la estructura interna del producto, lo único que le importa es cómo usarlo, cómo comunicarse [3].

2. Diseño de sistemas interactivos

Los sistemas interactivos son aquellos sistemas de software que se interrelacionan con los usuarios; el sistema podrá reaccionar según las acciones del usuario.

La interacción se realiza a través de la interfaz del sistema; el usuario se comunica/interactúa con el sistema por medio de la parte visible del sistema, su interfaz.

Con el objetivo de lograr una interacción fluida y satisfactoria, surge una nueva disciplina denominada Interacción Persona Ordenador (IPO), mas conocida por sus siglas en inglés HCI (Human Computer Interaction). La IPO se centra en estudiar la interacción entre usuarios y sistemas informáticos; y se ha convertido en un área muy importante dentro de la Ingeniería de Software.

A modo de definición, se puede decir que: la interacción persona-computadora es el intercambio observable de información, datos y acciones entre un humano (usuario) y la computadora, y viceversa [9].

Es de gran importancia que el sistema cuente con una interfaz que ayude, guíe al usuario a concretar sus tareas.

Lo que antes se tomaba como tarea secundaria y complementaria, hoy cuenta con una metodología y una planificación para lograr la mayor fluidez en la interacción. Actualmente se le ha dado protagonismo al diseño y construcción de la interfaz de usuario. Es por esto que se aplican técnicas de la ingeniería del software para lograr un producto de calidad, las mismas son utilizadas con la participación del usuario.

3. Análisis de técnicas

La técnica de Casos de uso es muy utilizada en ingeniería del software, dado que permite describir los límites de un producto de software y también describir los requisitos del mismo.

La Ingeniería de Software la define como una técnica que [4]:

- Describe el comportamiento de un sistema desde el punto de vista del usuario en términos de los requisitos funcionales y por tanto se pueden considerar el punto de partida para poder especificar las necesidades de los usuarios finales.
- Permite definir los límites del sistema.
- El usuario debería poder entender cada caso de uso para poder validarlo.
- Guía todo el proceso de desarrollo del sistema.

Estos cuatro ítems se los puede considerar como los puntos fuertes de la técnica. Sin embargo, es importante mencionar sus debilidades, y es que no capturan requisitos no funcionales, tal como fiabilidad, portabilidad, mantenibilidad, eficiencia y sobre todo usabilidad; y tampoco sirven para capturar las necesidades de los usuarios, en cuanto a la interacción (aspecto esencial a la hora de desarrollar interfaces de usuario).

Es por esto que surge la necesidad de encontrar la manera de revertir esa debilidad. Para esto se utiliza el Análisis de Tareas, técnica que se enfoca en capturar los requisitos de la Interfaz del Usuario. Analizar las tareas permite obtener información respecto a cómo los usuarios

realizan su trabajo y así conocer la funcionalidad que el sistema debe mostrar a través de la interfaz.

Los objetivos fundamentales de esta técnica son [5]:

- Identificar usuarios (características principales: experiencia, conocimientos previos, grupo etario, etc).
- Identificar qué debe hacer el usuario (qué quiere hacer el usuario).
- Comprender el dominio del problema.
- Describir el mecanismo de interacción.
- Considerar las diferentes respuestas del usuario.

De manera sintética se puede decir que el análisis de tareas es desarrollar modelos genéricos, abstractos de las tareas del usuario.

Como diferencia entre las dos, se puede mencionar que la técnica de Casos de Uso es utilizada por ingenieros de software que se ocupan del desarrollo del sistema; en tanto que el Análisis de Tareas se utiliza en el área de Interacción Humano-Computadora, donde se hace mayor énfasis en los requisitos para el desarrollo de la interfaz de usuario.

Ambas técnicas son utilizadas en las etapas iniciales del diseño y construcción de un sistema de software.

4. Diseño centrado en el usuario

El Diseño Centrado en el Usuario (DCU) es el término que se utiliza para describir el diseño en el que el usuario influye en el resultado final. Se considera que es una filosofía y un proceso. Una filosofía porque sitúa al usuario en el centro con la intención de desarrollar un producto adecuado a sus requerimientos, necesidades y un proceso de diseño porque se centra en los factores cognitivos de las personas y como éstos intervienen en sus interacciones con los productos [7].

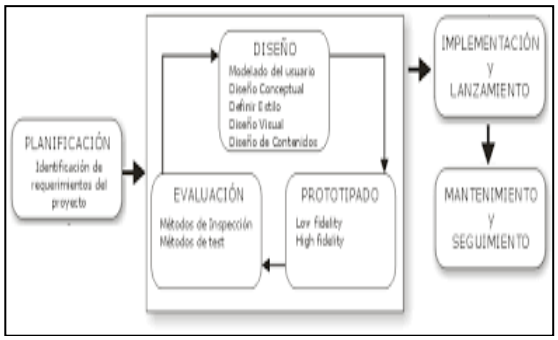


Fig. 1: Etapas del DCU

Esta metodología consta de diferentes etapas (Fig. 1) [8]. Si bien estas etapas corresponden a todo el proceso de desarrollo, es obvio que el diseño y desarrollo de la Interfaz se ven involucrados en ellas. Como es evidente, el usuario se ve involucrado en cada etapa o fase.

El concepto Diseño Centrado en el Usuario nació tras la investigación de Donald A. Norman en la Universidad de California San Diego (UCSD); autor del libro *User centered system design: new perspectives on human-computer interaction,* y posteriormente se consolidó con el libro de Norman *The design of everyday things*. Obra referente en el ámbito del DCU y las disciplinas afines como la interacción persona-computadora, el diseño y la experiencia de usuario.

La metodología de diseño centrado en el usuario, según la ISO/DIS 13407 (ISO-98) [6], implica:

- Que el usuario participe activamente, para que el equipo de desarrollo logre comprender: el contexto de uso y de las tareas; y los requisitos.
- Una interacción continua entre las soluciones propuestas por el equipo y el usuario, y así poder hacer una retroalimentación.
- Un diseño multidisciplinar que incluya varias especialidades: desarrolladores de software, expertos en usabilidad, en ergonomía, etc.

El usuario final se ve involucrado en cada etapa o fase del proceso de desarrollo, con esto se garantiza que el producto se ajuste a sus necesidades. Ésta forma de diseño le otorga al usuario el rol central del desarrollo. En ocasiones, el usuario es un miembro más del equipo de diseño. El objetivo es obtener productos fáciles de usar, efectivos y eficientes.

En la siguiente figura se muestra cómo el usuario participa activamente del ciclo de vida de un sistema interactivo.

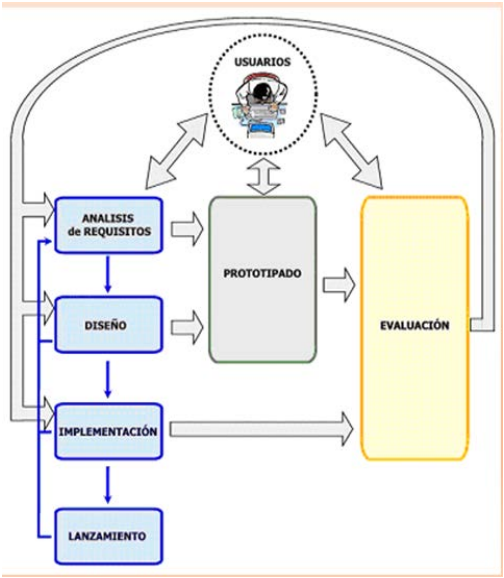


Fig. 2: Ciclo de vida de sistemas interactivos.

La Fig. 2 muestra como característica fundamental del ciclo de vida la presencia constante del usuario, desde la fase de análisis hasta en la evaluación; como también la naturaleza cíclica, dado que hay un a continua retroalimentación a partir de la fase o etapa de evaluación.

5. Actividades Llevadas a Cabo en la Línea de Investigación

Hasta el momento, en la línea de investigación descripta previamente, se están llevando a cabo las siguientes tareas:

- Revisión Sistemática de la literatura referente al tema de diseño y construcción de Interfaces Gráficas de Usuario; con el fin de evidenciar la importancia de su diseño y cómo lograr una óptima y satisfactoria interacción con los usuarios. Haciendo una mención especial respecto al rol del usuario en el diseño de la Interfaz.
- Estudio de Métodos de Evaluación Multicriterios. Esta tarea tuvo como objetivo seleccionar un método de evaluación para evaluar interfaces la calidad de las interfaces gráficas de usuario. Hasta el momento se ha encontrado que el método LSP (Logic Scoring of Preference) se adapta muy bien para la evaluación de las GUI.
- Elaboración de Criterios de Evaluación. Esta tarea se lleva a cabo con el objetivo de establecer cuáles son las características que las interfaces gráficas de usuario deben tener para que sean consideradas adecuadas para su utilización por parte del usuario. Estos criterios serán la entrada al método de evaluación multicriterio mencionado en el ítem precedente.
- Estudio de Reglas, Técnicas y Principios de Diseño de GUI. Esta tarea tiene como finalidad proveer la información necesaria para que los equipos de desarrollo de GUI puedan desarrollar GUIs adecuadas para el usuario.
- Análisis de la Influencia Emocional, el objetivo es determinar cómo influye el estado emocional de los usuarios en la interacción con las computadoras y, a su vez, cómo el diseño de la Interfaz influye en el estado emocional del usuario.

6. Conclusiones y Trabajos Futuros

Luego de la lectura de diferentes trabajos respecto de cómo mejorar la interacción usuarios-computadoras se puede observar que el diseño de una Interfaz Gráfica de Usuario no es una tarea secundaria. Además, no sólo involucra un

grupo de expertos en programación; sino que es fundamental que el equipo integre a los usuarios. Diversos estudios demuestran que cada vez son más las investigaciones que se realizan con el objetivo de lograr una óptima interacción.

Como resultado, esta investigación pretende determinar cómo lograr una GUI 'usable', comprensible, satisfactoria; con el análisis de los diferentes tópicos que se involucran en la Interacción Persona-Computadora.

Futuros trabajos: i) Enriquecer el estudio sistemático de la literatura; ii) Implementar un método de evaluación multicriterio para evaluar GUIs, iii) Definir criterios de evaluación de interfaces gráficas de usuario, iv) Elaborar un manual con las Reglas, Técnicas y Principios de Diseño de GUI y v) Profundizar el estudio sobre la influencia emocional.

7. Formación de Recursos Humanos

En el contexto de esta línea de investigación se están llevando a cabo diferentes tesis de grado, trabajos finales integradores de ingeniería y trabajos de especialización.

En lo que respecta a tesis de grado una de ellas finalizó exitosamente y con repercusión internacional. En dicho trabajo se implementó el método de evaluación Logic Scoring of Preference (LSP). Por otra parte, se está realizando un trabajo final integrador de ingeniería en donde se implementa el método de evaluación AHP (Analytic Hierarchy Process). Ambos trabajos se van a utilizar para evaluar GUIs y también para analizar las ventajas y desventajas de cada método para evaluar el tipo de sistemas estudiados en esta línea de investigación.

En Ingeniería de Software, a nivel de Especialización, se puede decir que se está desarrollando un trabajo final en donde se pretende integrar todos los conceptos abordados en la línea de investigación.

Referencias

[1] Lozano, María & González, Pascual & Ramos, Isidro & Montero, Francisco & Molina Masso, Jose Pascual. (2002). Desarrollo y generación de interfaces de usuario a partir de técnicas de de Análisis de Tareas y Casos de Uso. Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. 6. 83-92. 10.4114/ia.v6i16.745.

[2] Losada, Begoña, David López, and Javier Martínez. "Guía de actuación en el desarrollo de interfaces de usuario según la metodología centrada en el usuario INTERGRAM." *VIII Congreso Internacional de Interacción Persona-Ordenador Interacción*. Vol. 7. 2007.

[3] http://www.lsi.us.es/docencia/get.php?id=4330
Material de estudio del Departamento de Lenguajes y sistemas Informáticos – Universidad de Sevilla.

[4] https://www.ctr.unican.es/asignaturas/is1/is1-t08-trans.pdf
Material de studio INGENIERÍA DEL SOFTWARE -Contexto y Requisitos del Sistema (en desarrollo OO) Univ. Cantabria – Fac. de Ciencias

[5] Gea, İ. Miguel, and Fco Luis Gutiérrez. "Diseño centrado en el usuario." (2011). http://lsi.ugr.es/~mgea/docencia/diu/Temario/Diu_Tema3.pdf

[6] ISO (International Organization for Standardisation), 13407, p.7, HumanCentered Design Processes for Interactive Systems, 1997

[7] Domingo, Muriel Garreta, and Enric Mor Pera. *Diseño centrado en el usuario*, 2010.

[8] Yusef Hassan & Francisco J. Martín Fernández & Ghzala Iazza. *Diseño Web Centrado en el Usuario: Usabilidad y Arquitectura de la Información.* "Hipertext.net", núm. 2, 2004. <http://www.hipertext.net>

[9] Díaz, F. J., Harari, I., and Amadeo, A. P. (2013). Guía de recomendaciones para diseño de software centrado en el usuario. Editorial de la Universidad Nacional de La Plata (EDULP).

Experiencia de Usuario en APPs Móviles basado en emociones.

Iván Balmaceda Castro[1,2], Mariela Fernandez[1], Sandra Rodriguez[1]
Carlos Vera[1], Luis Fuentes[1], Juan Magaquian[1]

[1] Laboratorio de Ingeniería de Software - Instituto de Investigación para el desarrollo socioeconómico de los Llanos Riojanos (INDELLAR)
Sede Regional Chamical. Universidad Nacional de La Rioja
e-mail: {ibalmaceda89, mariel.lucia, carloseduardonvn, jmagaquian, neluis97}@gmail.com, sandraorona27@yahoo.com.ar

[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Resumen

Hoy en día, las aplicaciones móviles son un elemento clave de los dispositivos inteligentes, ya que transforman el estilo de vida de los usuarios; así también, revolucionan los negocios e impulsan a distintas industrias.
La Experiencia de Usuario (UX), se convierte en un punto de partida para los programadores y desarrolladores de APPs Móviles, donde se debe diseñar aplicaciones que busquen consolidarse en la mente de los usuarios. Estas deben contar con pocas funcionalidades, útiles y fáciles de usar. Por ello, es importante que las organizaciones aprendan cómo aprovechar el gran potencial de las APPs para atraer a los usuarios con eficacia, así como optimizar sus procesos.
El presente trabajo, se orienta a medir la experiencia de usuario a través de emociones, que permitirá mejorar la usabilidad de aplicaciones móviles.

Palabras clave: UX | Experiencia de Usuario | Emociones | Evaluación | Aplicaciones Moviles.

Contexto

El presente trabajo se enmarca dentro del Proyecto de Investigación *"MÉTODO DE INSPECCIÓN DE LA EXPERIENCIA DE USUARIO BASADO EN EMOCIONES PARA APLICACIONES MÓVILES"* del Laboratorio de Ingeniería de Software del Instituto de Investigación para el desarrollo socioeconómico de los Llanos Riojanos (INDELLAR) - Sede Regional Chamical Universidad Nacional de La Rioja.

Introducción

El estudio de la UX y la Usabilidad de un sistema software, nos permite generar valor agregado a un producto, el cual afecta directamente en la aceptación del mismo involucrando factores sociales, culturales y de contexto. [1,2,3].
El incremento y la diversificación del uso de dispositivos móviles han generado un gran auge los últimos años. Hoy en día, no existe ninguna actividad que se realiza en computadoras, que no pueda ser realizada en un dispositivo móvil, convirtiéndolos en elementos vitales en la cotidianeidad.
La UX, es el conjunto de múltiples factores y elementos relativos, que inciden en el uso de un producto o servicio, cuyo resultado genera una percepción y respuesta por parte del usuario [4]

Los dispositivos móviles cuentan con una capacidad especial de procesamiento, conexión permanente o intermitente de red, memoria limitada, diseñados específicamente para una función, pero tienen la versatilidad de desarrollar otras funciones más genéricas [5]. En la mayoría de los casos, un dispositivo móvil puede definirse con cuatro características que lo diferencian de otros dispositivos [6]:

- Movilidad: Es la cualidad de un dispositivo para ser transportado o movido con frecuencia y facilidad.

- Tamaño Reducido: Ser fácilmente usado con una o dos manos sin necesidad de ninguna ayuda o soporte externo.
- Comunicación Inalámbrica: la capacidad que tiene un dispositivo de enviar o recibir datos sin la necesidad de un enlace cableado.
- Interacción con Usuarios: el proceso de uso que establece un usuario con un dispositivo.

Es de gran importancia que los usuarios puedan tener una placentera experiencia con las APPs Móviles y que de esta forma puedan ser utilizadas de forma fácil y sin ningún tipo de problemas, es por ello que se debe estudiar y evaluar dicha experiencia en estas aplicaciones, pero ¿Cómo se evalúa la UX en las diversas aplicaciones de dispositivos móviles? ¿Cómo se determina el grado de satisfacción de los usuarios?

La investigación, en la disciplina de Interacción Humano Computadora (HCI, por su sigla del inglés Human-Computer Interaction**),** estudia los productos informáticos que ayudan a realizar tareas con facilidad de uso, bajo tiempos de ejecución, evitando los posibles errores y con un alto grado de satisfacción [7]. De ello deviene, que la UX, es uno de los principales factores en una aplicación móvil, que depende de una buena estrategia de diseño y comunicación hacia el usuario [8], constituyéndose una herramienta muy poderosa a la hora de acercar propuestas de información, productos y servicios a los usuarios.
Es de gran importancia considerar cómo van a interactuar los usuarios a través de las APPs y ayudarles a encontrar el camino más fácil para encontrar la información/acción que necesitan.

Las emociones determinan la calidad de las interacciones con un producto en el entorno del usuario y están directamente relacionadas con la evaluación de la experiencia de éste. Los usuarios generan emoción como un modo de minimizar los errores, interpretar la funcionalidad o aligerar la complejidad de un trabajo [9], con esto podemos entender mejor el nivel de implicación y de motivación del usuario.
El análisis de esta dimensión en los métodos de diseño centrados en el usuario, nos ayuda a estar seguros de que nuestros usuarios se sienten interesados y motivados mientras utilizan nuestros sistemas y/o aplicaciones.
Por otro lado se comenzaron a utilizar modelos motivacionales para explicar de manera más amplia la aceptación tecnológica del usuario en términos de "calidad de experiencia", considerando que las emociones y las experiencias estéticas son valores económicos que determinan la calidad de un producto, bien o servicio.
Construir aplicaciones móviles que ofrezcan una buena UX significa especificar productos enfocados en las personas que los usan y las tareas que esas personas llevan a cabo [10].
UX es una forma de medir si un usuario puede predecir satisfactoriamente cuál es el resultado de una acción. Si el resultado es muy diferente a lo que el usuario está esperando, la experiencia será frustrante y negativa [4]. Por esto, es importante evaluar experiencias a corto plazo, considerando los cambios dinámicos en los objetivos y necesidades de los usuarios relacionados con el contexto de uso, y cómo y por qué las experiencias evolucionan con el tiempo [11,12,13]. Esto no significa que la usabilidad ya no sea relevante; por el contrario, es un elemento importante de la UX, la usabilidad se vuelve aún más relevante, porque un mayor grado de usabilidad de un producto de software también genera mejor experiencia del usuario. [12,14,15,16].
Las revisiones de diseño y los recorridos cognitivos generalmente se realizan a través de los ojos del usuario, aunque muchas veces esto nunca se hace explícito o específico. Los usuarios pueden hacer una revisión experta o un recorrido cognitivo.[17]

Para poder medir la UX, se deben utilizar métodos de evaluación, pero ¿cuál de los

métodos es apropiado para evaluar la UX en aplicaciones móviles?
La UX puede ser evaluada con diferentes métodos que permiten una apreciación emocional del usuario frente al producto. Con el estudio de antecedentes, podemos definir que la UX es evaluada cuando el producto ya está desarrollado. Ésta permite a los diseñadores comprender y tener una idea de cómo los usuarios perciben y valoran las aplicaciones. El propósito de este trabajo, es el de analizar, definir e incluir criterios de experiencias de usuarios teniendo en cuenta las emociones del usuario, a través de métodos de inspección basados en heurísticas [18], que permiten evaluar la forma en cómo es la interacción, considerando que es el más apropiado para la línea de investigación presentada.

Líneas de Investigación, Desarrollo e Innovación

Ésta línea de investigación se centra en el análisis de las pautas de diseño de interfaces en aplicaciones móviles, especificando los tipos y estrategias de los factores que condicionan las emociones a través de la experiencia de usuario, acorde a los objetivos planteados en el Proyecto de Investigación "MÉTODO DE INSPECCIÓN DE LA EXPERIENCIA DE USUARIO BASADO EN EMOCIONES PARA APLICACIONES MÓVILES" perteneciente a la Sede Regional Chamical de la Universidad Nacional de La Rioja.

Resultados Obtenidos y Objetivos

Como objetivo general de la presente línea de investigación, se plantea en definir un método que sirva de instrumento de Evaluación de la UX para aplicaciones móviles que permitan cuantificar el grado de satisfacción de los usuarios de este tipo de herramientas.
Para el mismo, se plantean los siguientes objetivos particulares:

1. Analizar las pautas de Diseño de Interfaces en aplicaciones móviles.
2. Analizar y definir los requerimientos funcionales y no funcionales basada en las pautas de experiencia de usuario.
3. Especificar los tipos y estrategias de los factores que condicionan las emociones tanto en lo cognitivo como en lo afectivo.
4. Definir métricas de experiencia de usuario.
5. Integrar y definir una heurística de evaluación que incorpore requisitos de experiencia de usuario y diseño de interfaces.
6. Validar el modelo en aplicaciones móviles.

Formación de Recursos Humanos

El equipo de investigación está conformado por 3 Docentes Investigadores, 1 graduado, 2 alumnos, uno de ellos becado por el Consejo de Investigación Científicas y Tecnológicas (CICyT) de la UNLaR.
A su vez, se está trabajando en: 1 tesis de Doctorado en Ciencias de la Informática de la UNSJ. 2 tesis de Maestría, uno de la Maestría en Informática de la UNSJ y otro de la Maestría en Ingeniería de Software de la UNSL.
Trabajos Finales de Grado de la Licenciatura en Sistemas de Información y Analista Universitario en Sistemas de Información de la UNLaR.

Referencias

1. J. Nielsen, "Usability inspection methods," in Conf. Human factors in computing systems, Boston, Massachusetts, United States, 1994.
2. J. Nielsen, R. Molich, "Heuristic Evaluation of User Interfaces", in Proceeding SIGCHI '90 Conference on Human factors in Computing Systems, pp.249-256, 1990.
3. N. Bevan, "International Standards for HCI and Usability", International journal of human-computer studies, vol. 55, no. 4, pp. 533-552, 2001.

4. Garrett, J. J. (1900). The elements of user experience. Pearson Education Incorporated.
5. Guevara Soriano, A. (2010). Dispositivos móviles.
6. Rivera, Y. J. M., Cardona, J. S., & Franco, S. A. T. (2012). Sistema Operativo Android: caracteristicas y funcionalidad para dispositivos moviles (Doctoral dissertation, Universidad Tecnológica de Pereira. Facultad de Ingenierías Eléctrica, Electrónica, Física y Ciencias de la Computación. Ingeniería de Sistemas y Computación).
7. C. Cobo Romani (2005). "Organización de la información y su impacto en la usabilidad de las tecnologías interactivas". Recuperado: www.tdx.cbuc.es/TESIS_UAB/AVAILABLE/TDX-0403106215231//ccr1de1.pdf
8. B. BALLARD (2007). Designing the mobile user experience. Recuperado: http://www.blablacar.es/
9. P. Branco, P. Firth, L. Encarnao, P. Bonato, "Faces of emotion in human-computer interaction". En Ext. Abstracts CHI 2005 (Portland, OR, 2-7 de abril de 2005), ACM Press, 1236-1239.
10. ISO 9241-11, Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts, International Organization for Standardization, 2018.
11. A. Allam, A. Razak and H. Mohamed, "User Experience: Challenges and Opportunities", Journal of Information Systems Research and Innovation, vol. 3,pp. 28-36, 2013.
12. Usability Body of Knowledge, Usability Evaluation Methods. Recuperado: http://www.usabilitybok.org/usability-evaluation-methods
13. J. Nielsen, "Ten Usability Heuristics". Recuperado: https://www.nngroup.com/articles/ten-usability-heuristics/.
14. J. Nielsen, "Usability engineering", 1 edition ed., San Diego, CA 92101, Estados Unidos: Academic Press, 1993.
15. ISO 9241-210. Ergonomics of human-system interaction — Part 210: Humancentred design for interactive systems. International Organization for Standardization, 2010.
16. S. Joshi, "Customer experience management: An exploratory study on the parameters affecting customer experience for cellular mobile services of a telecom company", in Procedia - Social and Behavioral Sciences, vol. 133, pp. 392-399, 2014.
17. P. Desmet Measuring emotions: Development of an instrument to measure emotional responses to products. En Blythe, M. A.; Overbeeke, K.; Monk, A. F.; Wright, P. C. (ed.), Funology: from usability to enjoyment. Kluwer Academic Publishers. Dordrecht, Boston, Londres, 2003.
18. M. Mascheroni,C. Greiner, R. Petris, G. Dapozo, M. Estayno "Calidad de Software e Ingenieria de Usabilidad" Workshop de Investigadores en Ciencias de la Computación (WICC 2012, UNM, Misiones), 2012

ARQUITECTURA TECNOLÓGICA PARA LA IMPLEMENTACIÓN DE GESTIÓN DEL CONOCIMIENTO EN PEQUEÑAS Y MEDIANAS FÁBRICAS DE SOFTWARE

Straccia, Luciano[1]; Pollo-Cattaneo, Ma. Florencia[1]

[1]Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS)
Universidad Tecnológica Nacional. Facultad Regional Buenos Aires.
Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires Tel +54 11 4867-7511

lstraccia@frba.utn.edu.ar, flo.pollo@gmail.com

Resumen

La aplicación de un Sistema de Gestión de Conocimiento (SGC) implica considerar cuatro conductores base: roles y responsabilidades, procesos, gobernanza y tecnología. La mayoría de los trabajos conocidos tratan sobre los primeros dos conductores, hallándose menor cantidad de investigaciones vinculas a los restantes.

La identificación de una apropiada arquitectura tecnológica permite dotar de factibilidad y aplicabilidad a cualquiera de los modelos.

En este contexto, el objetivo del proyecto es definir una arquitectura tecnológica y su metodología de aplicación en las pequeñas y medianas fábricas de software, dando continuidad a trabajos previos donde se generan modelos de procesos para un SGC para ese tipo de organizaciones.

Palabras clave: Gestión del Conocimiento, arquitectura tecnológica, PyMEs, fábricas de software, Ingeniería de Software.

Contexto

La Universidad Tecnológica Nacional ha definido diversos Programas de Investigación, Desarrollo e Innovación, entre los cuales se encuentra el Programa de Sistemas de Información e Informática que tiene como objetivo "intensificar y focalizar las acciones tendientes a fortalecer y promover el crecimiento de temáticas de investigación en Sistemas de Información e Informática, y promover la interacción con la industria informática en general y de desarrollo de sistemas de información y de software en particular" [1].

Introducción

La Jerarquía DIKW y el conocimiento

Shannon [2] y Weaver [3] introdujeron tempranamente los conceptos de dato e información, aunque con algunas diferencias terminológicas, incorporación de las ideas de lo "físico" (lo captable con el sentido) y el "significado". Timpson [4-7] retoma el debate respecto a estos conceptos. En este trabajo se considerará al dato, como representación de un hecho aislado, fuera de contexto; e información, como colección de datos estructurados en un contexto (datos con significado) [8-9].

Las nociones de dato e información fueron incorporadas en la denominada

Jerarquía DIKW, presentada por Ackoff [10], que sostiene que se trata de una cadena jerárquica en la que cada concepto añade valor al previo. Esta Jerarquía tiene como elemento base al dato (*data*); un nivel superior, a la información (*information*); posteriormente, al conocimiento (*knowledge*) y en el nivel superior, la sabiduría (*wisdom*).

Cornella [11] sostiene que "la 'metabolización' de la información, su conversión en estructuras mentales, generalmente permanentes, lleva a la creación de conocimiento en nuestras mentes", siendo hechos con sentido y con estructura [12]. El conocimiento resulta una "respuesta apropiada y persistente a una entrada dada" [13] y su valor está asociado a las ocasiones en las cuales se requiere su aplicación. Finalmente, "si al conocimiento se lo somete a juicios de valor y se lo dota de una ética, deviene sabiduría" [14].

Gestión del conocimiento

La Gestión del conocimiento (GC) es un proceso a través del cual las organizaciones logran descubrir, utilizar y mantener el conocimiento, con la idea de alinearlo con las estrategias de negocio para la obtención de ventajas competitivas [15]. El conocimiento "se constituye en el recurso estratégico más importante, y la habilidad para generarlo, adquirirlo, codificarlo, transferirlo, aplicarlo y reutilizarlo, se ha convertido en la competencia sustancial para la obtención de una ventaja competitiva sostenible" [16].

Nonaka y Takeuchi [17] introdujeron la tipología del conocimiento: tácito (aquel difícil de expresar, formalizar y compartir, muy personal y subjetivo, derivado de la experiencia) y explícito (aquel que se puede expresar y formalizar fácilmente, por lo cual, se adquiere mediante métodos formales de estudio) y la posibilidad de conversión del conocimiento tácito en explicito.

En [18;19] se realiza una revisión de modelos de gestión del conocimiento. Los modelos relevados y analizados fueron Nonaka y Takeuchi, Wigg, Sveiby, Earl, Kerschberg, Bustelo y Amarilla, Mc. Elroy, CEN, KMC, K-TSACA y, destinados especialmente a la industria del software: Bovea y García e ISECO. Además, en dichos trabajos se propone un modelo específicamente orientado a las pequeñas y medianas fábricas de software, analizando las dificultades que enfrenta la industria y tipología de organizaciones específica y las características de los diferentes modelos de GC.

En [20] se presentan los cuatro conductores base para la Gestión del Conocimiento: roles y responsabilidades; elementos del proceso; elementos de gobernanza (política de GC, métricas e incentivos) y los elementos tecnológicos.

En trabajos previos del Grupo [18;19] se abordan los primeros dos conductores. En el presente trabajo se aborda el cuarto conductor.

Elementos tecnológicos

Entre los modelos de gestión de conocimiento relevados y ya mencionados, sólo los modelos de Kerchsberg [21] y Borghoff et al. [22] proponen una estructura tecnológica para su implementación.

Otros autores hacen referencia a las tecnologías o herramientas requeridas para la implementación de gestión del conocimiento tales como Peter Tyndale [23] y Carvalho et al. [24]. En [25;26] se propone integrar al sistema de gestión de conocimiento las tecnologías ERP. Sin embargo, estos autores no integran estas tecnologías informáticas en un modelo

arquitectónico que favorezca su análisis e implementación. En [27] se presenta una arquitectura, pero la misma se limita a considerar sistemas multiagentes.

Arquitectura tecnológica

La arquitectura tecnológica de un sistema de información y un sistema de gestión del conocimiento es "el conjunto de estructuras necesarias para razonar sobre el sistema. Comprende elementos de software, relaciones entre ellos, y propiedades de ambos" [28]. La arquitectura involucra "conceptos fundamentales o propiedades de un sistema en un entorno definido, encarnado en elementos, las relaciones que existen entre ellos; y los principios que guían su diseño y evolución" [29].

Kerchsberg [21] presenta una arquitectura tecnológica basada en el Modelo 3 Capas: Vista (o Interacción), Gestión y Almacenamiento, mientras que Borghoff et al. [22] proponen una estructura de cuatro componentes: depósitos y almacenes de conocimiento; comunidades de trabajadores del conocimiento; flujo del conocimiento; cartografía del conocimiento. Tiwana [30] presenta una arquitectura de siete capas: interfaz, acceso y autenticación, inteligencia colaborativa y filtrado, aplicación, transporte, integración de *middleware* y *legacy* y los repositorios [31]. En [32] se presenta una arquitectura de cuatro capas: datos, integración de la información, procesamiento del conocimiento y presentación. Además, otros autores presentan diferentes arquitecturas o integración tecnológica [33-35].

Líneas de Investigación, Desarrollo e Innovación

En la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional (UTN.BA) se ha conformado en el año 2009 el Grupo GEMIS, con dependencia del Departamento de Ingeniería en Sistemas de Información, integrado por un equipo de docentes, alumnos y graduados con interés en la sistematización de conocimientos y su promoción sobre el campo de la Ingeniería en Sistemas de Información y la Ingeniería de Software incluyendo sus aplicaciones y abordajes metodológicos en todo tipo de escenarios.

En el año 2018 se ha iniciado un Proyecto de Investigación y Desarrollo (PID) denominado "La gestión del conocimiento en pequeñas y medianas fábricas de software en el Área Metropolitana de Buenos Aires" que indaga en los modelos de gestión del conocimiento.

A partir de las experiencias y resultados obtenidos en dicho PID que se encuentra en etapa de finalización y considerando la detección de dificultad de aplicabilidad de los modelos teóricos a las características de las pequeñas y medianas fábricas de software, se comienza una nueva línea de trabajo vinculada específicamente a los elementos tecnológicos de la gestión del conocimiento, cuya denominación es "Arquitectura tecnológica para la implementación de Gestión del Conocimiento en pequeñas y medianas fábricas de software".

De esta manera, se articula dentro de los objetivos de GEMIS en el campo de la Informática, la generación de nuevos conocimientos en el área de la Ingeniería de Software.

Resultados esperados

Objetivo General

Definir una arquitectura tecnológica y su metodología de aplicación en la gestión de conocimiento en pequeñas y medianas fábricas de software.

Objetivos Específicos

Los objetivos específicos son: a) identificar elementos tecnológicos que permitan implementar un modelo de gestión de conocimiento; b) identificar arquitecturas tecnológicas para la implementación de GC; c) identificar las características, recursos y posibilidades de las pequeñas y medianas fábricas de software; d) proponer una arquitectura tecnológica para la implementación de GC para este tipo de organizaciones.

Resultados esperados

Como resultado de las tareas a desarrollar se espera:

- contar con una arquitectura tecnológica para la gestión del conocimiento en las pequeñas y medianas fábricas de software;
- validar en una organización real la arquitectura generada;
- incorporar los resultados parciales y finales de esta investigación a la enseñanza de grado y posgrado en diversas regionales de la UTN, donde los investigadores del proyecto desarrollan su actividad docente;
- vincular con otros PID del Grupo, relacionados con Sistemas Expertos y Sistemas Inteligentes.

Formación de Recursos Humanos

El equipo se encuentra conformado por investigadores formados, investigadores de apoyo, graduados de grado y alumnos de la carrera de Ingeniería en Sistemas de Información. Esta nueva línea de trabajo permitirá:

- la generación nuevas oportunidades y experiencias para la formación de los investigadores;
- la incorporación de nuevos becarios graduados a través de las becas BINID;
- el desarrollo de por lo menos 2 tesis de la Maestría en Ingeniería en Sistemas de Información. Además, se prevé el inicio de por lo menos 2 trabajos finales de la Especialidad en Ingeniería en Sistemas de Información

Referencias

[1] Universidad Tecnológica Nacional (2016). Resolución Nro. 2508/16

[2] C. E. Shannon, “A Mathematical Theory of Communication,” The Bell System Technical Journal, vol. 37, pp. 379-423, 1948.

[3] C. Shannon and W. Weaver, The mathematical Theory of Communication, University of |Illinois Press, 1949.

[4] Timpson, C. (2004). Quantum Information Theory and the Foundations of Quantum Mechanics. PhD diss., University of Oxford (quant-ph/0412063).

[5] Timpson, C. (2006). “The Grammar of Teleportation.” The British Journal for the Philosophy of Science, 57: 587-621.

[6] Timpson, C. (2008). “Philosophical Aspects of Quantum Information Theory.” Pp. 197-261, in Dean Rickles (ed.), The Ashgate Companion to the New Philosophy of Physics. Aldershot: Ashgate Publishing.

[7] Timpson, C. (2013). Quantum Information Theory and the Foundations of Quantum Mechanics. Oxford: Oxford University Press.

[8] Parra Mesa (2004). Los modernos alquimistas. Epistemología corporativa y gestión de conocimiento. Editorial EAFIT. Medellín.

[9] Lombardi, Olimpia; Holik, Federico; Vanni, Leonardo (2014). What is Shannon information? En Synthese, 193(7). Springer.

[10] Ackoff RL. From data to wisdom. Journal of Applied Systems Analysis. 1989;16:3-9

[11] Cornella, Alfons (2000). Infonomia!. La empresa es información. Deusto. Bilbao, España.

[12] Chekland, P. y Holwell, S. (1998) Information, systems and information systems.Chichester, UK: John Wiley & Sons

[13] Sanders, John (2016). Defining Terms: Data, Information and Knowledge. SAI Computing Conference. Londres, UK.

[14] Pollo-Cattáneo, María Florencia (2012). Resolviendo problemas en los sistemas de información. Enfoque para informáticos. Editorial CEIT. Buenos Aires.

[15] Bueno, E. (2000). Dirección del Conocimiento y Aprendizaje: Creación, distribución y mediación de Intangibles.

[16] Gelaf, Graciela (2010). Abordajes creativos en situaciones de crisis organizacionales. Contaduría General de la Nación. Tucumán, Argentina.

[17] Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford university press

[18] Maulini, A.; Straccia, L; Pollo-Cattaneo, M.F. (2018). Una aproximación a un modelo de gestión de conocimiento aplicable a las pequeñas y medianas fábricas de software. En Conferencia Internacional de Ingeniería InNGENIO. Medellín, Colombia.

[19] Maulini, A.; Straccia, L; Pollo-Cattaneo, M.F. (2019) Un modelo de gestión de conocimiento aplicable a las pequeñas y medianas fábricas de software. En "Desarrollo e Innovación en Ingeniería - Cuarta Edición" (Ed. Prof. Edgar Serna M). Editorial Instituto Antioqueño de Investigación

[20] Milton, Nick (2014). The 4 legs on the Knowledge Management table

[21] Kerschberg, L y Weishar, J (2000). Conceptual Models and Architectures for Advanced Information Systems

[22] Borghoff, U y Parechi, R (1998) Information Technology for Knowledge Management

[23] Tyndale, Peter (2002). A taxonomy of knowledge management tools: origins and applications. Evaluation and Program Planning 25 (183-190).

[24] Carvalho y Ferreira (2001). A Proposal of Taxonomy for Knowledge Management Tools.

[25] Lopes de Almeida, Tiago (2018). A Knowledge Management Architecture for Information Technology Services Delivery. Universidad Europeia.

[26] M. F. Acar, M. Tarim, H. Zaim, S. Zaim, and D. Delen (2017) "Knowledge management and ERP: Complementary or contradictory?," Int. J. Inf. Manage., vol. 37, no. 6, pp. 703–712.

[27] Thitiya Manakitsirisuthi, Yacine Ouzrout, Abdelaziz Bouras. A PLM/KMS integration for Sustainable Reverse Logistics. 8th IFIP Inernational Conference Product Lifecycle Management PLM 11, Jul 2011, Eindhoven, Netherlands. pp.1-11. ffhal-01550197

[28] Bass, L.; Clements, P.; Kazman, R. (2012). Software Architecture in Practice. Addison-Wesley Professional.

[29] ISO/IEC/IEEE 42010, ISO, 2011.

[30] Tiwana, A. (2000). The Knowledge Management Toolkit: Practical Techniques for Building a Knowledge Management System. Upper Saddle River, NJ: Prentice Hall.

[31] Halawi, L; McCarthy, R; Aronson, J. (2017). Success Stories in Knowledge Management Systems. Issues in Information Systems, 18(1).

[32] S. Rance, ITIL service Transition, 2011 Editi. Norwich: TSO, 2011

[33] Meroño Cerdan, A. (2004). Tecnologías de información y gestión del conocimiento: integración de un sistema. En Economía industrial, Nº 357. España.

[34] Perez, D.; Dressler, M. (2007). Tecnologías de la información para la gestión del conocimiento. En Intangible Capital - Nº 15 - Vol. 3- pp. 31-59

[35] Quintanilla Juarez, N. (2014). Herramientas TICs y la gestión del conocimiento.

Enfoques y Tendencias en el Desarrollo de Aplicaciones para Dispositivos Móviles.

Pablo Thomas, Lisandro Delia, Leonardo Corbalán, Germán Cáseres, Juan Fernández Sosa, Fernando Tesone, Verónica Aguirre, Verena Olsowy, Patricia Pesado

Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

50 y 120 La Plata Buenos Aires

Centro Asociado a la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

526 e/ 10 y 11 La Plata Buenos Aires

{pthomas, ldelia, corbalan, gcaseres, jfernandez, ftesone, vaguirre, volsowy, ppesado}@lidi.info.unlp.edu.ar

RESUMEN

Se presenta una línea de investigación y desarrollo centrada en el estudio de aspectos de Ingeniería de Software aplicados a las diferentes tendencias en el desarrollo de Aplicaciones para Dispositivos Móviles.

Palabras claves: Plataformas para Dispositivos Móviles - Aplicaciones Nativas- Aplicaciones Híbridas – Aplicaciones Interpretadas – Aplicaciones por Compilación Cruzada – Aplicaciones Web Progresivas – Offline First. – Instant App

CONTEXTO

Esta línea de Investigación forma parte del Proyecto (2018-2021) "*Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso*", en particular del subproyecto *Ingeniería de Software para escenarios híbridos* del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación de la Nación.

Hay cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa y Latinoamérica.

1. INTRODUCCIÓN

Los avances tecnológicos en la comunicación móvil, junto con los avances en el campo de la computación han dado lugar a la computación móvil, que puede definirse como un entorno de cómputo con movilidad física que brinda al usuario la capacidad de acceder a datos, información u otros objetos lógicos desde cualquier dispositivo en cualquier red, al mismo tiempo que va moviéndose, cambiando su locación geográfica [1].

Las características específicas de la computación móvil plantean nuevos desafíos en el desarrollo de software para este tipo de dispositivos. La necesidad de tratar con diversos estándares, protocolos y tecnologías de red; las capacidades limitadas, no sólo en cuanto a procesamiento, sino también, en cuanto a consumo, tamaño físico y capacidad de almacenamiento de los dispositivos; las restricciones de tiempo impuestas por un mercado altamente dinámico y la existencia de distintas plataformas de hardware y software son sólo algunas de las dificultades a las que se enfrentan los desarrolladores en esta área.

Las aplicaciones se generan en un entorno dinámico e incierto. En su mayoría se trata de aplicaciones pequeñas, no

críticas, destinadas a un gran número de usuarios finales que son liberadas en versiones rápidas para poder satisfacer las demandas del mercado. En otros casos las aplicaciones son de mayor tamaño, a veces para brindar movilidad a una parte de la funcionalidad de un sistema más grande, mientras que otras veces son el único punto de interacción del sistema.

Algunas iniciativas como *offline first* proponen nuevas maneras de abordar el desarrollo de aplicaciones móviles tomando como requerimiento no funcional la posibilidad de seguir brindando un servicio incluso cuando el dispositivo pierde conectividad con la red móvil.

Todas las particularidades previamente mencionadas hacen que el desarrollo de software para dispositivos móviles difiera considerablemente del tradicional. Ello conduce a nuevas prácticas y metodologías que promueven el crecimiento de la Ingeniería de Software como disciplina, acompañando este proceso de desarrollo tecnológico.

La existencia de una enorme competencia en el mercado de las aplicaciones móviles, promovida en parte por la proliferación de dispositivos móviles y el número de negocios que están migrando sus servicios a Internet, revela la importancia que implica para las empresas el posicionamiento en este mercado. Para conseguirlo, es necesario reducir al máximo el tiempo de desarrollo de las aplicaciones y al mismo tiempo hacer que éstas se ejecuten en el mayor número de dispositivos posible.

Este propósito se ve obstaculizado por la excesiva fragmentación de hardware y software existente, originada por el alto número de dispositivos distintos, con sus propios sistemas operativos y plataformas de desarrollo.

La alternativa más costosa para hacer frente a esta problemática consiste en el desarrollo específico de aplicaciones de manera nativa en cada una de las plataformas existentes, utilizando entornos de desarrollo, lenguajes de programación y tecnologías propias de cada plataforma.

El desarrollo nativo de aplicaciones móviles posee ciertas ventajas como la posibilidad de acceder sin limitaciones a todas las características del dispositivo (cámara, GPS, acelerómetro y agenda, entre otras), el alto rendimiento, interfaces gráficas consistentes con el resto de la plataforma, la posibilidad de trabajar sin acceso a Internet y de correr procesos en segundo plano. Estas aplicaciones pueden distribuirse a través de las tiendas en línea correspondientes. Sin embargo, el precio de todas estas ventajas es alto: no es posible reutilizar el código fuente entre plataformas diferentes, el esfuerzo se multiplica y se elevan los costos de desarrollo, actualización y distribución de nuevas versiones.

Como alternativa al desarrollo nativo en múltiples plataformas, se presentan diversos enfoques de desarrollo que procuran optimizar la relación costo/beneficio compartiendo la misma base de código fuente entre las versiones para las distintas plataformas [2]. La construcción de aplicaciones Web Móviles constituye un ejemplo representativo de este enfoque. Estas aplicaciones se diseñan para correr dentro de un navegador, se desarrollan con tecnologías web bien conocidas (HTML, CSS y JavaScript), no necesitan adecuarse a ningún entorno operativo móvil específico; su puesta en marcha es rápida y sencilla.

Las desventajas de las aplicaciones Web Móviles recaen sobre su rendimiento. Los tiempos de respuesta se dilatan afectados por la interacción cliente-servidor y las restricciones de seguridad impuestas a la ejecución de código por medio del navegador limitan el acceso a todas las capacidades del dispositivo. Además, la experiencia de usuario dista de ser similar a la de las aplicaciones nativas, resultando menos atractiva para el usuario final.

El enfoque híbrido es una alternativa que posee también la ventaja de estar basado en tecnologías web estándar (HTML, Javascript y CSS) pero, a diferencia del anterior no funciona dentro de un navegador, sino en un contenedor web especial con mayor acceso a las características del dispositivo a través de una API específica.

Las aplicaciones híbridas permiten la reutilización de código en las distintas plataformas, el acceso al hardware del dispositivo, y la distribución a través de las tiendas de aplicaciones. Sin embargo, conservan algunas de las desventajas de las aplicaciones Web Móviles: la utilización de componentes no nativos en la interfaz perjudica la experiencia de usuario, y la ejecución se ve ralentizada por la carga asociada al contenedor web.

El enfoque interpretado se presenta como una alternativa en donde las aplicaciones son traducidas en su mayor parte a código nativo, mientras que un resto se interpreta en ejecución. Se implementan de forma independiente de las plataformas utilizando diversas tecnologías y lenguajes, tales como Javascript, Typescript y XML, entre otros.

La obtención de interfaces nativas constituye una de las principales ventajas de este tipo de aplicaciones, mientras que la definición de nuevas componentes suele tener un alto grado de complejidad debido a la necesidad de definir abstracciones compatibles con diferentes plataformas.

Finalmente, las aplicaciones generadas por compilación cruzada también constituyen un tipo de desarrollo multiplataforma. Estas aplicaciones se compilan de manera nativa creando una versión específica de alto rendimiento para cada plataforma destino.

Un nuevo concepto ha surgido en los últimos años denominado Aplicaciones Web Progresivas (PWA por sus siglas en inglés). Una PWA es una aplicación web que utiliza las últimas tecnologías disponibles en los navegadores para ofrecer en dispositivos móviles una experiencia lo más parecida posible a la de una aplicación nativa.

Los objetivos que persiguen las PWA son: lograr el mayor rendimiento posible en dispositivos móviles, que la aplicación cargue de manera casi instantánea, que la interfaz de usuario se parezca lo máximo posible a una nativa, que se pueda trabajar sin conexión (*offline first*) y que se puedan enviar notificaciones a los usuarios, como en una aplicación nativa.

Una nueva tecnología, desarrollada recientemente por Google, denominada TWA (Trusted Web Activities) o Actividades Web de Confianza, permite integrar una PWA con una aplicación Android. Las TWA son ejecutadas desde un APK y distribuidas desde Google Play Store. Muestran en pantalla completa un navegador web dentro de una aplicación Android sin mostrar la interfaz del navegador.

Las TWA representan un punto de inflexión para los desarrolladores que ahora pueden distribuir sus PWA en Google Play Store que ha dejado de ser una tienda exclusiva de aplicaciones nativas.

Desde 2017 los desarrolladores de Android tienen una nueva opción para hacer llegar sus apps a los usuarios finales. El concepto de *Instant App* permite ejecutar una funcionalidad específica de una aplicación sin necesidad de instalar la App completa.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

- Enfoques de desarrollo de Aplicaciones para Dispositivos Móviles.
- Metodologías y Técnicas de la Ingeniería de Software y su aplicación en el desarrollo de software para dispositivos móviles.
- Aplicaciones Nativas en Android [3].

- Aplicaciones Nativas en iOS [3].
- Aplicaciones Web Móviles.
- Aplicaciones Móviles Híbridas (PhoneGap [4], Ionic [5]).
- Aplicaciones Móviles Interpretadas (Appcelerator Titanium [6], NativeScript [7]).
- Aplicaciones Móviles generadas por compilación cruzada (Xamarin [8], Corona [9]).
- PWA, Instant App y Offline First.
- TWA (Trusted Web Activities),
- Análisis y estudio comparativo de requerimientos no funcionales, tales como rendimiento, consumo de energía, tamaño de software, entre otros, en los distintos enfoques de Aplicaciones Móviles.
- Experiencia de usuario en Aplicaciones Móviles generadas con distintos enfoques de desarrollo.
- Internet de las Cosas (IoT), Internet de Todo (IoE) y aplicaciones de sensado móvil o sensado urbano.
- Aplicaciones Móviles y su utilización en el proyecto de *Smart City.*
- El rol de los dispositivos móviles en aplicaciones de *crowdsourcing* o colaboración abierta distribuida.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados obtenidos/esperados se pueden resumir en:

- Se ha realizado un estudio [10] sobre el impacto que distintos frameworks de desarrollo multiplataforma tienen sobre el tamaño de la aplicación construida. Los resultados obtenidos en este estudio son relevantes dado que el espacio disponible se ha convertido en un recurso crítico para muchos usuarios.
- Se ha elaborado un detallado estudio comparativo [11] que contempla una lista de 23 características de interés para la Ingeniería de Software y su impacto en 9 tecnologías para el desarrollo de aplicaciones móviles.
- Se ha presentado un minucioso estudio sobre la eficiencia energética y su relación con los enfoques de desarrollo [12]. Se estudió la autonomía de las baterías de los dispositivos móviles en aplicaciones con funciones multimedia (acceso a imágenes y reproducción de video) y con alta carga de procesamiento, generadas con diversos enfoques de desarrollo.
- Se ha estudiado [14] [15] la manera en que el enfoque de desarrollo utilizado afecta el rendimiento de aplicaciones que realizan procesamiento intensivo. Se extrajeron conclusiones considerando las plataformas iOS y Android utilizando varios frameworks de desarrollo multiplataforma.
- Se estudiaron de manera integral y conjunta dos tópicos importantes en el área de las aplicaciones móviles: los enfoques de desarrollo y los requerimientos no funcionales.[26] El objetivo planteado fue cuantificar el impacto que tiene la elección de un enfoque de desarrollo, y en particular un determinado *framework*, sobre tres de los requerimientos no funcionales más demandados en el área de las aplicaciones para dispositivos móviles: rendimiento, consumo de energía y uso de espacio de almacenamiento.
  Se analizaron enfoques nativos y multiplataformas. El estudio incluyó a 6 de los *frameworks* de desarrollo más conocidos y utilizados en el mercado: 1) Android SDK (enfoque nativo), 2) Cordova (enfoque híbrido), 3) Titanium (enfoque interpretado), 4) NativeScript (enfoque interpretado), 5) Xamarin (enfoque compilación cruzada) y 6) Corona (enfoque compilación cruzada).
  Se obtuvieron conclusiones que sirven de apoyo a las decisiones que los

desarrolladores de software deben realizar al momento de elegir el enfoque/*framework* más adecuado de acuerdo con el nivel de expectativas creadas sobre el rendimiento, consumo de energía y uso de espacio de almacenamiento.

- En [27] se analizaron las principales características de otro enfoque de desarrollo para aplicaciones móviles denominado Aplicaciones Web Progresivas o PWA, por sus siglas en inglés. Dado su estrecha relación con las aplicaciones web móviles resultó de interés realizar un detallado estudio comparativo entre ambos enfoques explicitando diferencias y similitudes. Se concluyeron una serie de ventajas y desventajas de un enfoque respecto del otro.
- Se realizó el lanzamiento de la versión para iOS de "*Informática UNLP*", una aplicación móvil presentada en [13] desarrollada para la comunidad de la Facultad de Informática de la Universidad Nacional de La Plata. Las primeras versiones de esta aplicación fueron liberadas para la plataforma Android. En la actualidad también está disponible en la tienda App Store para iPhone y iPad. "*Informática UNLP*" mejora la comunicación entre alumnos y docentes. Incluye una cartelera virtual donde los alumnos reciben información de interés sobre sus cursos. También permite consultar información de la ocupación de las aulas en tiempo real utilizando tecnología de realidad aumentada
- Se continúa con el desarrollo de nuevas características de "*Informática UNLP*, una aplicación multiplataforma en continua evolución y crecimiento.
- Se continúa con el estudio de PWA al que se agrega el análisis de los alcances de la tecnología TWA (Trusted Web Activities), recientemente aparecida. TWA es la estrategia implementada por Google para integrar una PWA con una aplicación Android que puede ser entonces distribuida desde Google Play Store. Actualmente se está realizando un análisis detallado de la combinación PWA+TWA, teniendo en cuenta características no funcionales, de desarrollo, y de la gestión de proyectos de software.
- El Centro de Innovación y Transferencia Tecnológica (CIyTT) de la Facultad de Informática de la UNLP fue sede de una muestra de Ciencia y Tecnología 2019 en la que se exhibieron diferentes desarrollos tecnológicos innovadores. La muestra consistió de 10 estaciones, distribuidas a lo largo del centro, las cuales contenían distintos proyectos que podían ser visitados por el público en general. Para realizar una visita guiada, se desarrolló una PWA denominada “InnovApp”. Su funcionalidad principal consistió en describir, por medio de texto y contenido multimedia (imágenes, audio y video), cada uno de los proyectos dentro de las diferentes estaciones.
- Se esperan resultados concretos al estudiar las posibilidades que surgen de la utilización de los sensores de los dispositivos móviles en aplicaciones de Internet de las Cosas (IoT), Internet de Todo (IoE) y particularmente aplicaciones de sensado móvil, también conocido como sensado urbano.
- Se plantea analizar la utilidad de las Aplicaciones Móviles en el contexto de las ciudades inteligentes, haciendo foco especialmente en aquellas que implementan la colaboración abierta distribuida (*crowdsourcing*).
- Analizar las ventajas y desventajas de las *"Instant App"* de Android. Resulta de interés su estudio y posterior comparación con las aplicaciones nativas tradicionales.
- Promover el avance sostenido y continuo de la formación de los miembros involucrados en ella.

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con este proyecto. Además, participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

5. BIBLIOGRAFÍA

1. Hongxing Li, Guochu Shou, Yihong Hu, Zhigang Guo. Mobile Edge Computing: Progress and Challenges. 2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud). Oxford UK.
2. Spyros Xanthopoulos, Stelios Xinogalos, ***A Comparative Analysis of Cross-platform Development Approaches for Mobile Applications***, BCI' 2013, Greece
3. Tracy, K.W., ***Mobile Application Development Experiences on Apple's iOS and Android OS***, Potentials, IEEE, 2012.
4. http://phonegap.com/
5. https://ionicframework.com/
6. http://www.appcelerator.com/
7. https://www.nativescript.org/
8. http://xamarin.com/
9. https://coronalabs.com/
10. J. Fernández Sosa, P. Thomas, L. Delía, G. Cáseres, L. Corbalán, F. Tesone, A. Cuitiño, P. Pesado. ***"Mobile Application Development Approaches: A Comparative Analysis on the Use of Storage Space"***. XV Workshop de Ingeniería de Software - XXIV Congreso Argentino de Ciencias de la Computación CACIC 2018. Tandil, Argentina. 8 al 12 de octubre de 2018.
11. L. Delia, P. Thomas, L. Corbalan, J. Fernandez Sosa, A. Cuitiño, G. Cáseres, P. Pesado. ***"Development Approaches for Mobile Applications: Comparative Analysis of Features"*** SAI - Computing Conference 2018. London, United Kingdom. 10 al 12 de Julio de 2018.
12. L. Corbalan, J. Fernandez Sosa, A. Cuitiño, L. Delía, G. Cáseres, P. Thomas, P. Pesado. ***"Development Frameworks for Mobile Devices: A Comparative Study about Energy Consumption".*** 5th IEEE/ACM International Conference on Mobile Software Engineering and Systems. Gothenburg, Sweden. 27 y 28 de Mayo del 2018.
13. Fernández Sosa J., Cuitiño A., Thomas P., Delia L., Caseres G., Corbalán L., Pesado P. ***"Informática UNLP" la App de la Facultad de Informática***. CACIC 2017. La Plata, del 9 al 13 de Octubre de 2017.
14. Delia L., Galdamez N., Corbalan L., Thomas P., Pesado P. ***Approaches to Mobile Application Development: Comparative Performance Analysis*** SAI Computing Conference (SAI), 2017. Londres, del 18 al 20 de Julio de 2017.
15. Delia L., Galdamez N., Corbalan L., Thomas P, Pesado P., ***Un Análisis comparativo de rendimiento en Aplicaciones Móviles Multiplataforma***, CACIC 2015, UNNOBA Junín, Octubre 2015.
16. Delia L., Galdamez N., Corbalan L., Thomas P, Pesado P., ***Multi-Platform Mobile Application Development Analysis***, IEEE Ninth International Conference on Research Challenges in Information Science, May 2015, Athens, Greece, ISBN 978-1-4673-6630-4
17. Delia L., Galdamez N., Thomas P, Pesado P., ***Un Análisis Experimental de Tipo de Aplicaciones para Dispositivos Móviles***, CACIC 2013, CAECE Mar del Plata, Octubre 2013
18. Digital Possibilities. Mobile Development Frameworks Overview http://digital-possibilities.com/mobile-development-frameworks-overview/

19. Markus Falk. Mobile Frameworks Comparison Chart, http://www.markus-falk.com/mobile-frameworks-comparison-chart/
20. Anup Kumar y Bin Xie, Handbook of Mobile Systems Applications and Services. Editorial CRS Press, ISBN 978-1-4398-0152-9, Año 2012.
21. Sambasivan, D.; John, N.; Udayakumar, S.; Gupta, R., Generic framework for mobile application development, Internet (AH-ICI), 2011 Second Asian Himalayas International Conference on Computing & Processing (Hardware/Software).
22. Choi, Y.; Yang, J.-S.; Jeong, J., Application framework for multi platform mobile application software development, Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on Computing & Processing (Hardware/Software)
23. Anthony Wasserman, Carnegie Mellon Silicon Valley, Software Engineering Issues for Mobile Application Development, 2º Workshop on Software Engineering for Mobile Application Development, MobiCASE '11, Santa Monica, California, USA, October 2011.
24. Jemel, M., Serhrouchni, A. Content protection and secure synchronization of HTML5 local storage data. Consumer Communications and Networking Conference (CCNC), 2014 IEEE 11th, Las Vegas, NV, USA.
25. Tim A. Majchrzak, Andreas Biørn-Hansen, Tor-Morten Grønli, ***Progressive Web Apps: the Definite Approach to Cross-Platform Development ?***, Proceedings of the 51st Hawaii International Conference on System Sciences j, 2018, ISBN: 978-0-9981331-1-9
26. Corbalán, L., Thomas, P., Delía, L., Cáseres, G., Sosa, J. F., Tesone, F., & Pesado, P. (2019, June). A Study of Non-Functional Requirements in Apps for Mobile Devices. In *Conference on Cloud Computing and Big Data* (pp. 125-136). Springer, Cham
27. Aguirre, V., Ortu, A., Delía, L. N., Thomas, P. J., Corbalán, L. C., Cáseres, G., & Pesado, P. M. (2019). PWA para unificar el desarrollo Desktop, Web y Mobile. In *XXV Congreso Argentino de Ciencias de la Computación (CACIC)(Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019)*.

Reuso, Composición y Refactorización de Servicios Heterogéneos

Andres Flores, Martin Garriga, Alejandra Cechich,
Marcelo Moyano, Alan De Renzis, Diego Anabalon, Franco Corgatelli
Grupo de Investigación en Ingeniería de Software del Comahue (GIISCo)
http://giisco.uncoma.edu.ar
Departamento de Ingeniería de Sistemas – Facultad de Informática
Universidad Nacional del Comahue
Buenos Aires 1400, (8300) Neuquén
Contacto: [andres.flores, martin.garriga]@fi.uncoma.edu.ar

Resumen

El paradigma de Computación Orientada a Servicios (SOC), promueve el desarrollo de aplicaciones distribuidas en ambientes heterogéneos, que son construidas ensamblando o componiendo servicios reusables, que se publican a través de una red y se acceden mediante protocolos específicos. SOC ha sido ampliamente adoptado con la tecnología de Servicios Web. Existen diferentes estilos de Servicios Web que amplían las oportunidades de selección de soluciones, pero generan un desafío de evaluación y ajuste de servicios heterogéneos. Entre los estilos se encuentran los servicios SOAP (con descripciones WSDL) y los servicios RESTful (con múltiples lenguajes de descripción tal como WADL, OpenAPI, etc.). Para afrontar estos desafíos se definió un Metamodelo de Servicios Heterogéneos que permite la evaluación y composición de servicios. Además, el desarrollo de servicios para reuso afronta la necesidad de reducir la complejidad de los servicios que afecta su comprensión e interoperabilidad. Para ello, se pueden utilizar métricas de complejidad de servicios y realizar refactorizaciones hasta alcanzar la complejidad deseada. Estos desafíos también son posibles por medio del Metamodelo de Servicios, para que un proveedor reajuste sus servicios y ofrezca nuevas soluciones en base a sus desarrollos previos.

Palabras Clave: Ingeniería de Software basada en Reuso – Software Orientado a Servicios – Servicios Web – WSDL – REST.

Contexto

La línea presentada se inserta en el contexto de los siguientes proyectos y acuerdos de cooperación:

- 04/F009-2: "Reuso Orientado a Servicios – Parte II". Financiado por UNCo. (2017-2020).
- PIP GI 11220170100951CO: "Construcción de Líneas de Productos Software guiada por Estándares de Dominio". Financiado por CONICET. (2018-2020).
- PICT-2017-1725: "Extensión de Metamodelos de Aplicaciones Orientadas a Servicios para Descripción, evaluación y despliegue de Servicios en la Nube". Financiado por ANPCyT. (12/2018-12/2020).
  - Investigaciones conjuntas con ISISTAN-UNICEN, Tandil.
- Acuerdo de Cooperación con el Grupo Alarcos, Escuela Superior de Informática, Universidad de Castilla-La Mancha, España.

Introducción

Actualmente la industria de software observa cómo mediante el reuso de software se puede alcanzar un proceso de desarrollo de software acelerado y confiable al basarse en artefactos software que ya han sido probados en diferentes contextos de aplicación. Para ello se

adopta el concepto denominado "tercerización", por medio del cual se acuerdan contratos para adquisición y provisión de artefactos software reusables y se establecen relaciones comerciales entre vendedores y clientes. Por lo tanto, desde el punto de vista de un cliente implica la posibilidad de acelerar el desarrollo de un producto software para reducir el lanzamiento al mercado, y desde el punto de vista de un proveedor implica la posibilidad de observar sus productos con una perspectiva nueva que los coloque dentro del mercado de artefactos reusables.

Un paradigma que promueve altamente el reuso de software se denomina Computación Orientada a Servicios (SOC), donde la funcionalidad a ser reusada adopta la forma de servicios, o unidades lógicas que presentan entornos heterogéneos de ejecución y pueden ser ensambladas para formar otras unidades lógicas de mayor nivel de abstracción que resuelvan (directamente o en parte) los procesos de negocios para un contexto de aplicación [SH05,PTDL07]. El paradigma SOC encontró una plataforma potencial de aprovechamiento mediante la Web, desde donde se desarrolló la tecnología de Servicios Web [NSS03, Wetal05], con notaciones formales específicas para la descripción las interfaces de servicios. Así el paradigma SOC bajo la implementación con Servicios Web ha logrado su amplia adopción en la industria, principalmente bajo la flexibilidad de ejecución remota que permite a las compañías descentralizar aún más sus procesos de negocios y la ventaja de que las plataformas específicas de ejecución se encuentran ocultas, por lo cual no se requiere de inversiones adicionales en tecnología (incluyendo costos y esfuerzo de aprendizaje) al adquirir funcionalidad de terceras partes. El beneficio que la tecnología de Servicios Web provee al paradigma SOC se ha denominado "relación sin responsabilidad", donde una aplicación cliente no requiere asumir cómo se ha implementado el servicio con el que se comunica. Sin embargo, los proveedores de servicios tienen la responsabilidad de evaluar la calidad de los productos ofrecidos como servicios y los consumidores de servicios a su vez deben ser capaces de identificar tal calidad que influirá sobre las aplicaciones en desarrollo.

El funcionamiento concreto del paradigma SOC se basa en la Arquitectura orientada a Servicios (SOA) [SH05] que se encuentra compuesta por tres actores principales: un proveedor, un consumidor y un registro de servicios; donde el proveedor desarrolla y publica servicios en el registro, para que luego el consumidor busque servicios y establezca una comunicación con el proveedor. Sin embargo, la búsqueda de servicios publicados en un registro UDDI (según la tecnología de servicios Web) [OASIS04], en general requiere invertir un esfuerzo considerable para distinguir servicios candidatos que satisfagan los requerimientos de la aplicación cliente [NSS03, Wetal05]. En particular, cuando varios candidatos ofrecen funcionalidades similares se requieren métodos eficientes de selección de servicios que discriminen tanto aspectos funcionales como no-funcionales, considerando además las interacciones válidas para un servicio candidato en función de los procesos de negocio que implementará la aplicación cliente. En particular el ensamblaje de servicios considerando procesos de negocio e interoperabilidad de servicios plantea el uso de dos conceptos de reciente investigación: Orquestación y Coreografía de servicios [P03, Wetal05]. El primero relacionado a una aplicación particular que describe un proceso de negocios específico, y el segundo relacionado a las interacciones válidas que pueden ocurrir entre distintos servicios predestinados a intervenir en una colaboración.

Los Servicios Web pueden ser desarrollados mediante diferentes estilos

distintivos, entre los cuales se encuentran los servicios SOAP y los servicios RESTful. Los servicios SOAP se describen mediante el lenguaje WSDL (Web Service Description Language) que permite especificar las operaciones e intercambio de mensajes para un Servicio Web. Existen diferentes versiones del lenguaje WSDL (1.0, 1.1, 2.0), siendo la versión WSDL 2.0 definido como estándar OMG. Los servicios SOAP se consideran servicios "grandes", debido a que el intercambio de mensajes se realiza por medio de documentos XML, que también es la base para el lenguaje WSDL. Esto hace que se deba afrontar una sobrecarga de transferencia de archivos. Los servicios RESTful se pueden describir mediante diferentes lenguajes que intentan solventar la sobrecarga mencionada, mediante el intercambio de mensajes livianos, generalmente usando JSON o YAML. El primer intento de notación para REST fue WADL (Web Application Description Language) como analogía de WSDL, y también basado en XML. La industria ha propuesto diferentes lenguajes, entre los que se encuentra OpenAPI/Swagger, REST API Modeling Language (RAML), o MIT's API Blueprint, entre otros.

Evaluar y consumir tal variedad de servicios heterogéneos podría reducir el espacio de solución de los servicios a consumir, de acuerdo con la tecnología que se pueda soportar. De este modo, la amplia gama de servicios probablemente reusables, se vuelve seriamente limitada.

Para ello, hemos desarrollado un Metamodelo de Servicios Heterogéneos basado en estándares para ampliar la gama de compatibilidad de servicios de diversas tecnologías, que pudieran ser potenciales candidatos para aplicaciones consumidoras. Los principales estándares para la descripción del servicio incluyen: WSDL 2.0, WADL y OpenAPI, considerando así ambos estilos de servicios, basados en SOAP y REST.

Por otro lado, los proveedores de servicios se ven en la necesidad de ajustar sus servicios de acuerdo a los requerimientos de los clientes. En particular, la complejidad del servicio puede afectar la comprensión de sus aspectos funcionales, y reducir su descubrimiento e interoperabilidad. Para ello, se ha investigado sobre métricas de complejidad de servicios y la posibilidad de realizar refactorizaciones de servicios, hasta alcanzar servicios con una complejidad deseada. Esto también permitiría que los servicios SOAP pudieran ser migrados a servicios RESTful, que son más livianos en su consumo desde una aplicación cliente. Aunque incluso los servicios REST podrían superar los niveles deseables de complejidad, afectando seriamente su reuso, por lo cual también son susceptibles de ser refactorizados con los mismos propósitos.

Líneas de Investigación, Desarrollo e Innovación

El perfil de esta línea puede definirse en base a las actividades de investigación y transferencia, a las que da soporte el grupo GIISCo. Los temas específicos consideran los desafíos diferentes asociados al crecimiento de la Tecnología de la Información y las Comunicaciones. Actualmente, abordamos los siguientes aspectos:

- Compatibilidad y selección de servicios.
- Adaptación y Composición de servicios.
- Complejidad y legibilidad de servicios
- Refactorización de servicios
- Testing de servicios.
- Herramientas para evaluación, selección, composición y testing de servicios.
- Definición de aplicaciones en dominios específicos.

Resultados y Objetivos

En [FCGMRAC19] hemos enumerado una serie de contribuciones anteriores. Durante el año 2019, hemos profundizado la investigación en aspectos de *evaluación de compatibilidad y complejidad de servicios*, generando métodos y herramientas enfocados en las interfaces y comportamiento dinámico de los servicios [GF19]. Parte de este avance se ha efectuado en colaboración con investigadores de ISISTAN (UNICEN) [MZFM19], y del Politecnico di Milano (Italia) [BMGGQ19], [TNLGDG19].
Las líneas de investigación convergen en el tratamiento del desarrollo de software basado en el reuso de servicios desde la perspectiva de las aplicaciones orientadas a servicios. Una aplicación orientada a servicios implica una solución de negocio que consume servicios de uno o más proveedores y los integra en un proceso de negocio [SW04]. Además puede verse como una aplicación basada en componentes que integra dos tipos de componentes: internos localmente empotrados en la aplicación, y externos estática o dinámicamente enlazados a algún servicio [CMZC14]. El Metamodelo de Servicios Hetorgéneos habilita la posibilidad de reusar Servicios Web de diferentes estilos y tecnologías, extendiendo las oportunidades para la solución de aplicaciones cliente. Las capacidades que están desarrollando para la composición de servicios heterogéneos abren otro abanico de posibilidades cuando los servicios candidatos recuperados no proveen las funcionalidades requeridas. El Metamodelo de Servicios ha sido complementado con nuestro framework para evaluación y selección de servicios desarrollado en trabajo previos [GRLFMCZ18]. Por otro lado la refactorización de servicios basado en métricas de complejidad permite no solamente la adecuación de Servicios Web que un proveedor ofrece, sino también la posibilidad de estudiar escenarios de composición de servicios, para validar nuestra propuesta. Se prevee la aplicación de estos modelos y las herramientas de soporte a dominios específicos, con particular énfasis en aquellos que requieran rigurosidad como aporte de validación efectiva. La visión de esta línea de investigación se resume en:

"Definir técnicas y herramientas para la mejora del desarrollo de software, en función del reuso de servicios web. La definición de modelos de evaluación, selección y refactorización de servicios, y la posibilidad de composición de servicios".

Formación de Recursos Humanos

Este proyecto se compone de 11 investigadores, entre los que se cuentan docentes y estudiantes del Grupo GIISCo de UNComa y asesores externos. Algunos de los docentes–investigadores se encuentran realizando carreras de postgrado. Se cuenta actualmente con 3 doctores (1 investigador adjunto y 1 investigador asistente CONICET), 2 doctorandos y 1 maestrando entre los miembros del proyecto. Dirección de Tesis de Grado durante 2019: 5 tesis.

Referencias

[BMGGQ19] Baresi, L., Mendonça, DF., Garriga, M., Guinea, S. Quattrocchi, G. (2019). *A Unified Model for the Mobile-Edge-Cloud Continuum*. ACM Transactions on Internet Technology (TOIT) 19 (2): 1-21. (**Indexed SCI IF JCR2018: 2.382**).

[CMZC14] Crasso, M., Mateos, C., Zunino, A., Campo, M. (2014). *EasySOC: Making Web Service Outsourcing Easier*. International Journal on Information Sciences, 259: 452–473.

[GRLFMCZ18] Garriga, M., De Renzis, A., Lizarralde, I., Flores, A., Mateos, C., Cechich, A., Zunino, A. (2018). A Structural-Semantic Web Service Selection Approach to Improve Retrievability of Web Services. Information Systems Frontiers, 20(6): 1319-1344 **(Indexed SCI IF JCR2017: 3.232)**.

[GF19] Garriga, M., Flores, A. (2019). *Standards-driven Metamodel to Increase Retrievability of Heterogeneous Services*. 34th ACM/SIGAPP Symposium on Applied Computing, pp. 2507-2514. (**Qualis A1**).

[FCGMRAC19] Flores, A., Garriga, M., Cechich, A., Moyano, M., De Renzis, A., Anabalon, D., Corgatelli, F. (2019). *Reuso de Servicios basado en Análisis de Compatibilidad y Complejidad*. WICC'19, XXI Workshop de Investigadores en Ciencias de la Computación. Abril, San Juan.

[MZFM19] Mateos, C., Zunino, A, Flores, A., Misra, S. (2019). *COBOL Systems Migration to SOA: Assessing Antipatterns and Complexity*. Information Technology and Control 48 (1): 71-89 (**Indexed SCI IF JCR2018: 0.707**).

[NSS03] Nagappan, R.; Skoczylas, R.; Sriganesh, R. (2003). *Developing Java™ Web Services: Architecting and Developing Secure Web Services Using Java*. Wiley Publishing Inc.

[OASIS04] OASIS Consortium (2004). *UDDI Version 3.0.2*. UDDI Spec Technical Committee Draft, October.

[P03] Peltz, C. (2003). *Web Services Orchestration and Choreography*. IEEE Computer, 36(10): 46–52.

[PTDL07] Papazoglou, M.; Traverso, P.; Dustdar, S.; Leymann, F. (2007). *Service-Oriented Computing: State of the Art and Research Challenges*. IEEE Computer, 40(11): 38–45.

[SH05] Singh M.; Huhns, M. (2005). *Service-oriented computing: Key concepts and principles*. IEEE Internet Computing, 9(1): 75–81.

[SW04] Sprott, D.; Wilkes, L. (2004). *Understanding Service-Oriented Architecture*. The Architecture Journal. MSDN Library. Microsoft Corporation, http://msdn.microsoft.com/en-us/library/aa480021.aspx

[TNLGDG19] Tsigkanos, C. Nenzi, L. Loreti, M. Garriga, M. Dustdar, S. Ghezzi, C. (2019). *Inferring Analyzable Models from Trajectories of Spatially-Distributed Internet of Things*. IEEE/ACM 14th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), pp. 100-106.

[W3C05] W3C Candidate Recommendation (2005) *Web Services Choreography Description Language Version 1.0*. http://www.w3.org/TR/ws-cdl-10/

[Wetal05] Weerawarana, S.; Curbera, F.; Leymann, F.; Storey, T.; Ferguson, D. (2005). *Web Services Platform Architecture: SOAP, WSDL, WS-Policy, WS-Addressing, WS-BPEL, WS-Reliable Messaging, and More*. Prentice Hall PTR.

Aplicaciones Móviles 3D y Realidad Virtual

Pablo Thomas, Federico Cristina, Sebastián Dapoto, Patricia Pesado

Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata

50 y 120 La Plata Buenos Aires

Centro Asociado a la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

{pthomas, fcristina, sdapoto, ppesado}@lidi.info.unlp.edu.ar

Resumen

Se presenta una línea de investigación y desarrollo, que tiene por objeto estudiar temas relacionados con aspectos de Ingeniería de Software, orientados al desarrollo de aplicaciones móviles tridimensionales (3D) sobre diversas plataformas. En particular, se pone el foco en el análisis de performance y consumo de energía de las aplicaciones 3D, como así también en el desarrollo de aplicaciones 3D inmersivas, a través del uso de Realidad Virtual (RV).

Palabras claves: Dispositivos Móviles - Aplicaciones 3D - Aplicaciones Multiplataforma - M-Learning – Performance – Consumo de energía – Realidad Virtual – Internet de la Cosas

Contexto

Esta línea de Investigación forma parte del proyecto (2018-2021) "*Metodologías, técnicas y herramientas de ingeniería de software en escenarios híbridos. Mejora de proceso.*", en particular del subproyecto "*Ingeniería de Software para escenarios híbridos"*, del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación de la Nación.

Existe una importante cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa en proyectos financiados por el Ministerio de Ciencia y Tecnología de España y la AECID.

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Por otra parte, se tiene financiamiento de Telefónica de Argentina en Becas de grado y posgrado.

Introducción

Actualmente los dispositivos móviles son sofisticados y su evolución tecnológica permite ejecutar aplicaciones complejas y con exigentes requerimientos de hardware. Así, es cada vez más común el uso de frameworks que permiten desarrollar aplicaciones tridimensionales para dispositivos móviles.

La RV es una simulación interactiva por computadora en la cual se sustituye el mundo real con información sensorial que recibe el usuario. La RV permite generar entornos inmersivos donde el usuario puede interactuar con representaciones virtuales de objetos, que de otro modo sería difícil o imposible de acceder [1].

El ámbito educativo debe adaptarse a los cambios y nuevas formas de aprendizaje. M-learning (mobile learning) plantea métodos modernos de apoyo al proceso de aprendizaje mediante el uso de dispositivos móviles. Las aplicaciones móviles 3D son una herramienta ideal para acercar a los alumnos [2] [3] [4].

Sin embargo, muchos de los potenciales usuarios de estas aplicaciones educativas pueden no disponer de dispositivos de última generación. Por esta razón, resulta de vital importancia realizar un análisis en profundidad de los parámetros que inciden en la performance final de una aplicación 3D [5] y en el consumo de energía que dichas aplicaciones generan, el cual es generalmente alto [6] [7].

Los sistemas de domótica comprenden un conjunto de módulos cuyo objetivo en común es la automatización de las viviendas y de sus funciones. Los objetos cotidianos, tales como electrodomésticos, cuentan con conexión a Internet mediante la integración de sensores y/o dispositivos (Internet de las Cosas, o IoT por sus siglas en inglés) [8] [9].

Una aplicación móvil 3D de domótica facilita la interacción con los objetos conectados a una vivienda, mediante controles visuales, cómodos e intuitivos para el usuario.

Líneas de Investigación y Desarrollo

- Metodologías y Técnicas de la Ingeniería de Software y su aplicación en el desarrollo de software para dispositivos móviles.
- Aplicaciones Móviles 3D Multiplataforma
- Mobile Learning
- Frameworks para el desarrollo de Aplicaciones Móviles 3D
- Performance de aplicaciones móviles 3D
- Consumo de energía en aplicaciones móviles 3D
- Realidad Virtual en aplicaciones móviles 3D
- Domótica en aplicaciones móviles 3D

Resultados esperados/obtenidos

Los resultados esperados/obtenidos se pueden resumir en:

- Avanzar en la capacitación continua de los miembros de la línea de investigación.
- Avanzar en el aprendizaje de frameworks que permiten desarrollar aplicaciones 3D multiplataforma, particularmente para dispositivos móviles [10] [11] [12].
- Avanzar en el desarrollo de aplicaciones educativas, teniendo como finalidad enriquecer las experiencias interactivas y motivar el aprendizaje mediante su uso.
- Avanzar en el análisis de performance de ejecución de aplicaciones generadas con diferentes frameworks 3D.
- Avanzar en el análisis de consumo de energía de aplicaciones generadas con diferentes frameworks 3D.
- Se ha ampliado el prototipo móvil R-Info3D [2], una herramienta de aprendizaje sencilla de los conceptos básicos para la construcción de algoritmos en la Facultad de Informática. El prototipo ahora puede ser utilizado con lentes de RV, permitiendo una inmersión completa en el escenario virtual. Figura 1.

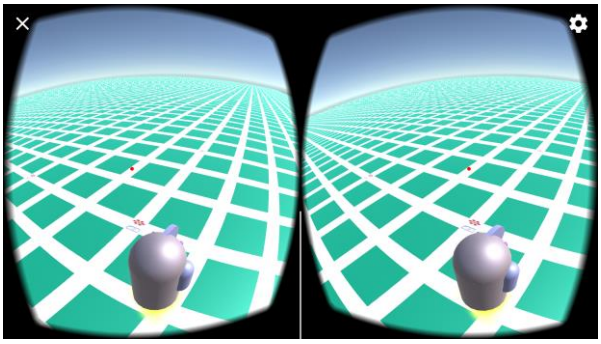


Figura 1. R-Info3D, controlado mediante las lentes de RV.

- Se ha desarrollado mediante los motores de juego multiplataforma Unity y Unreal Engine dos prototipos de análisis de performance de ejecución de aplicaciones 3D [13]. Figura 2.
- Se ha desarrollado mediante los motores de juego multiplataforma Unity y Unreal Engine dos prototipos de análisis de consumo de energía de las aplicaciones 3D [14]. Figura 3.
- Se ha desarrollado una aplicación móvil 3D inmersiva que permite recorrer el HMS Beagle, barco en el que Charles Darwin viajó a la Provincia de Buenos Aires para realizar diversas exploraciones. La aplicación permite movilizarse por la embarcación y visualizar sus descubrimientos mediante lentes de RV [15].

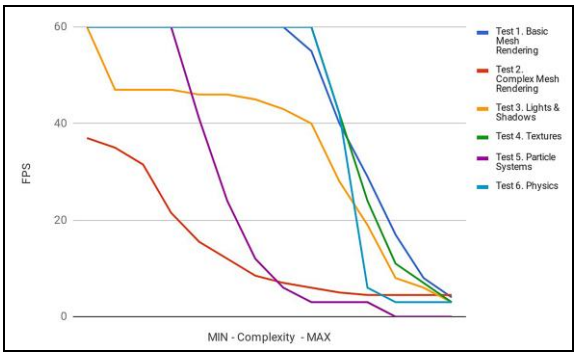


Figura 2. Prototipo de análisis de performance de ejecución. Resultados obtenidos.

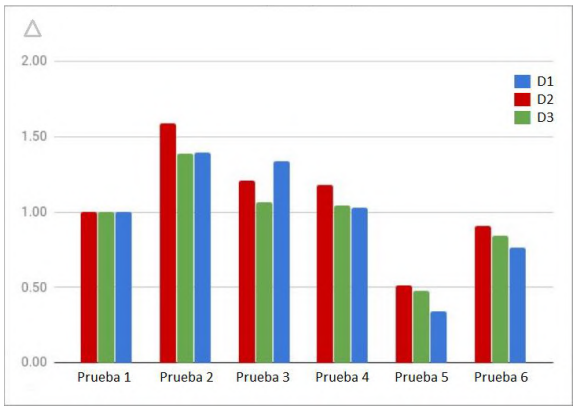


Figura 3. Prototipo de análisis de consumo de energía. Resultados obtenidos.

- Se está avanzando en el desarrollo de un prototipo móvil 3D de domótica. La aplicación facilitará la interacción con los objetos conectados a una vivienda, mediante controles visuales, cómodos e intuitivos. Figura 4.

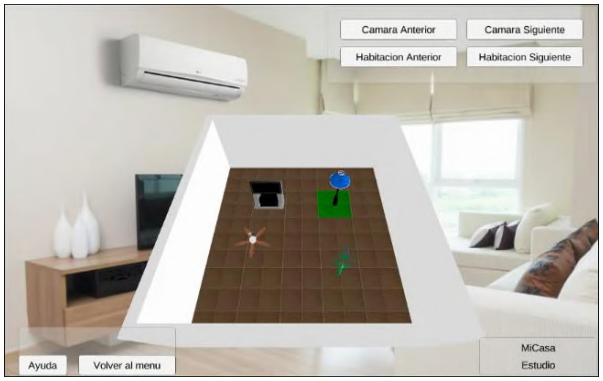


Figura 4. Prototipo móvil 3D de domótica.

- Se está avanzando en el estudio comparativo de los motores de juego multiplataforma Unity y Unreal Engine con el fin contrastar las ventajas y desventajas de cada motor.

Formación de Recursos Humanos

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con el proyecto. Además, participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

Referencias

1. Linowes J. “Unity Virtual Reality Projects”. 2015. ISBN-13: 978-1783988556.
2. Federico Cristina, Sebastián Dapoto, Pablo Thomas, Patricia Pesado. Capítulo de Libro: “3D Mobile Prototype for Basic Algorithms Learning”. Libro: “Computer Science & Technology Series - XXI Argentine Congress Of Computer Science. Selected Papers” (300 páginas). EDULP. ISBN: 978-987-4127-00-6, páginas 239-247. Año 2016.
3. Kantel E., Tovar G., Serrano A.”Diseño de un Entorno Colaborativo Móvil para Apoyo al Aprendizaje a través de Dispositivos Móviles de Tercera Generación.” IEEE-RITA 5, no. 4 (2010): 146-151.
4. Rosa Paredes, J. Alfredo Sánchez, Liliana Rojas, Daniel Strazzulla, Ronel Martínez-Teutle. "Interacting with 3D Learning Objects". 2009 LA Web Congress. ISBN: 978-0-7695-3856-3/09.
5. Akekarat Pattrasitidecha. "Comparison and evaluation of 3D mobile game engines". Chalmers University of Technology. University of Gothenburg. 2014.
6. Riaz, M. "Energy consumption in hand-held mobile communication devices: A comparative study". International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). 2018. ISBN: 978-1-5386-1370-2/18.
7. Aaron Carroll, Gernot Heiser. "An Analysis of Power Consumption in a Smartphone". USENIXATC'10 Proceedings of the 2010 USENIX conference on USENIX annual technical conference. 2010.

8. J. C. Montesdeoca Contreras, R. S. Avila Campoverde, J. C. Cabrera Hidalgo and P. E. Vintimilla Tapia, "Mobile applications using TCP/IP-GSM protocols applied to domotic," 2015 XVI Workshop on Information Processing and Control (RPIC), Cordoba, 2015, pp. 1-4.
9. Protocolo de red abierto MQTT. http://mqtt.org/
10. Unity 3D Homepage: https://unity3d.com/.
11. Unreal Engine Homepage: https://www.unrealengine.com/.
12. CryEngine. Homepage: https://www.cryengine.com/
13. Federico Cristina, Sebastián Dapoto, Pablo Thomas, Patricia Pesado. Capítulo de Libro: “Performance evaluation of a 3D engine for mobile devices”. Libro: “Computer Science – CACIC 2017. Communications in Computer and Information Science, vol 790”. Springer. ISBN: 978-3-319-75213-6, 978-3-319-75214-3, páginas 155-163. Año 2018.
14. Federico Cristina, Sebastián Dapoto, Pablo Thomas, Patricia Pesado. "Análisis de consumo de energía en aplicaciones 3D sobre dispositivos móviles". CACIC 2018. Tandil, Argentina. ISBN: 978-950-658-472-6, páginas 622-630.
15. Federico Cristina, Sebastián Dapoto, Pablo Thomas, Patricia Pesado. "HMSBeagleVR: Una experiencia inmersiva a bordo del HMS Beagle". Aceptado para el XXIV Congreso Argentino de Ciencias de la Computación CACIC 2019. Río Cuarto, Argentina. Páginas 1011-1019.

Aspectos de Ingeniería de Software, Bases de Datos Relacionales y Bases de Datos No Relacionales para el Desarrollo de Sistemas de Software en Escenarios Híbridos

Luciano Marrero, Pablo Thomas, Ariel Pasini, Rodolfo Bertone, Eduardo Ibáñez, Verónica Aguirre, Verena Olsowy, Fernando Tesone, Patricia Pesado.

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 La Plata Buenos Aires
Centro Asociado CIC
{lmarrero, pthomas, apasini, pbertone, eibanez, vaguirre, volsowy, ftesone, ppesado}@lidi.info.unlp.edu.ar

RESUMEN

Se presenta una línea de investigación que tiene por objeto estudiar las problemáticas actuales que afronta los procesos de Ingeniería de Software y Bases de Datos ante una gran variedad, de nuevas aplicaciones Web, que exigen tiempos de respuestas inmediatos y generan grandes volúmenes de información. Bajo esta demanda, las bases de datos no relacionales (NoSQL), aparecen como un complemento y/o alternativa a las bases de datos relacionales en el proceso de diseño de la información. Además, se tiene como objetivo principal generar metodologías y prácticas de Ingeniería de Software considerando las características de este nuevo contexto. Esto implica, tener en cuenta aspectos que hace algunos años no eran considerados, tales como, movilidad, geolocalización, generación de grandes volúmenes de información y la diversidad de los dispositivos electrónicos involucrados.

Palabras claves: Ingeniería de Software, Almacenamiento no estructurado, Bases de Datos NoSQL, Aplicaciones Móviles.

CONTEXTO

Esta línea de Investigación forma parte del proyecto (2018-2021) “Metodologías, técnicas y herramientas de ingeniería de software en escenarios híbridos. Mejora de proceso.”, en particular del subproyecto “Ingeniería de Software para escenarios híbridos”, del Instituto de Investigación en Informática LIDI (III-LIDI) de la Facultad de Informática UNLP, acreditado por el Ministerio de Educación de la Nación.

Hay cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa en proyectos financiados por el Ministerio de Ciencia y Tecnología de España y la AECID.

Se utilizan los recursos de Hardware y Software disponibles en el III-LIDI para diseñar, desarrollar y probar diferentes soluciones a problemáticas relacionadas con escenarios a investigar. Como resultado de esto, se espera obtener métricas reales que sirvan como

referencia para los investigadores en la comparación de resultados.

1. INTRODUCCIÓN

"Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso", 2018-2021 del Programa de Incentivos, es un proyecto que propone profundizar las investigaciones que se vienen realizando en el III-LIDI y extender la mirada a nuevos desafíos y cambios que están en gestación. Se organiza en tres subproyectos que permiten atender de manera ordenada el objetivo general propuesto:

"SP1 - Ingeniería de Software para escenarios híbridos", SP2 - Gobernanza Digital. Mejora de Procesos." Y "SP3 - Metodologías y herramientas para la apropiación de tecnologías digitales en escenarios educativos híbridos [15].

Este artículo se centra en el subproyecto "SP1 -Ingeniería de Software para escenarios híbridos". Se orienta a la investigación de metodologías y técnicas de la Ingeniería de Software, Bases de Datos Relacionales y Bases de Datos NoSQL, con énfasis en los escenarios híbridos y las nuevas aplicaciones Web adaptables a cualquier dispositivo informático.

La disponibilidad de la información en todo momento, debido a la expansión del acceso a internet y el uso de los dispositivos móviles, han cambiado los aspectos en que la información debe ser analizada y diseñada. Se presentan así, nuevas alternativas y desafíos en la Ingeniería de Software y las arquitecturas tradicionales para el almacenamiento de la información. En este contexto, las bases de datos no relacionales representan la evolución en el almacenamiento de los datos en una nueva generación de sistemas Web que deben responder eficientemente a las exigencias del usuario [1, 2, 8, 9 y 11]. El aumento en el acceso y procesamiento de la información digital a través de una variedad de dispositivos electrónicos con diferentes características ha alterado el concepto de diferentes ámbitos, por ejemplo, ciudades inteligentes, enseñanza virtual, domótica, etc. [10 y 11].

A razón de esto, los sistemas de software han aumentado sus exigencias, no solo el volumen de datos a procesar y almacenar, sino que también, deben responder de manera óptima al aumento de usuarios que lo acceden de forma simultánea. Esto significa que la escalabilidad y el rendimiento se han convertido en auténticos retos para la representación en el almacenamiento de la información. Las tecnologías de bases de datos más comúnmente utilizadas, las relacionales (SQL), pueden no ser suficientes y requieren de un complemento o alternativa para satisfacer las demandas actuales.

En respuesta a esta problemática, la industria del software ha incorporado a las bases de datos no relacionales (NoSQL). NoSQL no es tipo de base de datos, sino que es un conjunto de tipos de bases de datos que brinda diferentes formas para el almacenamiento no estructurado de información, es decir, operan sin un esquema predefinido. Las bases de datos NoSQL, surgen como respuesta para la administración de grandes volúmenes de información. Son altamente escalables y no respetan estrictamente las propiedades ACID (Atomicidad, Consistencia, Aislamiento y Durabilidad) [5, 6 y 7]. NoSQL propone un sistema llamado BASE (Básicamente Disponible, Estado Suave y Consistencia Eventual) [8, 9 y 11].

El desarrollo de software, año tras año, actualiza sus prácticas tradicionales de Ingeniería de Software. Los modelos y la forma en que se gestionan sus requerimientos se vuelven cada vez más ágiles. En ciertas ocasiones, la implantación de sus datos se combina en distintas categorías de almacenamiento (relacional, documental, orientado a columnas, clave-valor o grafos). Es importante el estudio de nuevas formas de almacenamiento no estructurado y comprender el funcionamiento de las herramientas que utilizan estos tipos de almacenamiento para comprender en qué tipo de problemas son aplicables. Herramientas NoSQL como MongoDB (documental), Cassandra (orientada a columnas), Redis (clave-valor) y Neo4j (grafos) son ejemplos de una nueva generación de motores de bases de datos que surgen en respuesta a diversas problemáticas existentes en las nuevas aplicaciones Web. En el almacenamiento documental, se almacena, recupera y gestiona documentos, en donde se codifican datos bajo un formato estándar (XML, JSON, BSON, etc.). En el almacenamiento orientado a columnas, los datos se organizan por columnas y no por filas, suelen ser un híbrido con el almacenamiento relacional. En el almacenamiento clave-valor, almacena un dato bajo una clave, similar a una tabla hash y en el almacenamiento de grafos, los datos son almacenados en objetos denominados nodos y son ideales para representar las relaciones o vínculos que hay entre ellos [9].

La comunicación y la sincronización del trabajo continúa siendo un pilar fundamental para el éxito de un proyecto. La utilización de repositorios de información, por ejemplo, GIT, permiten realizar un control de versiones distribuido, trabajando en modo offline o en modo online, con la facilidad de disponer herramientas específicas para la resolución de conflictos entre versiones [4, 5, 7, 8, 13 y 14].

La calidad en las prestaciones de los dispositivos móviles ha logrado que el acceso de miles de usuarios a ciertas aplicaciones sea constante. En esta realidad, es fundamental la combinación de distintas tecnologías para lograr el mejor rendimiento posible.

El desarrollo de nuevas aplicaciones Web para escenarios híbridos y con características de movilidad, plantea nuevos desafíos en diversas áreas de la Ciencia Informática.

Todas las particularidades previamente mencionadas conducen a nuevas prácticas y metodologías que promueven el crecimiento de la Ingeniería de Software y la gestión de Bases de Datos como disciplina [2, 3, 7, 8, 9 y 12].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

- Metodologías y Técnicas de la Ingeniería de Software y su aplicación en el desarrollo de software para escenarios híbridos.
- Investigar los distintos tipos de almacenamiento no estructurado de información (documental, orientado a columnas, clave-valor y grafos).
- Investigar Bases de Datos no relacionales (NoSQL). MongoDB, Cassandra, Redis, Neo4j, entre otras.
- Desarrollo de casos de estudio, pruebas de comparación y rendimiento de bases de datos no relacionales y relacionales [9].
- Repositorios GIT.

- Trabajo colaborativo en Proyectos de Software.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados esperados/obtenidos se pueden resumir en:

- Capacitación continua de los miembros de las líneas de investigación.
- Analizar, comparar y adaptar las nuevas metodologías y herramientas de la Ingeniería de Software para el desarrollo del software.
- Definición de técnicas, atributos y métricas para el análisis de usabilidad de sistemas en escenarios híbridos.
- Estudio y análisis de Bases de Datos no relacionales [9].
- Analizar, comparar y determinar escenarios para los distintos tipos de almacenamiento no estructurado de información.
- Comparación y análisis de resultados para diversos casos de estudio entre Bases de Datos Relacionales y Bases de Datos No Relacionales [9].
- Definición de procesos de Gestión de Incidencias utilizando repositorios GIT.
- Análisis de metodologías para la interoperabilidad de sistemas web y aplicaciones móviles [12].

El III-LIDI realiza transferencias de Aplicaciones Web y Aplicaciones Móviles en el marco de este proyecto a distintas entidades (Universidad Nacional de La Plata, Facultad de Ciencias Exactas de la UNLP, Albergue Universitario de la UNLP, entre otras).

4. FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con el proyecto. Además, participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

5. BIBLIOGRAFÍA

1. Ingeniería de Software. Un Enfoque Práctico. Séptima Edición. Roger S. Pressman. McGraw-Hill Interamericana Editores, S.A. 2010. ISBN: 978-607-15-0314-5
2. Ingeniería de Software Teoría y Práctica. Pfleegger Shari Lawrence. Pearson / Prentice Hall. 2002. ISBN: 9789879460719
3. Ingeniería de Software. Novena Edición. Ian Sommerville. Addison Wesley / Pearson. 2011. ISBN: 978-607-32-0603-7
4. Ingeniería de Software Clásica y Orientada a Objetos. Sexta Edición. Stephen R. Schach. Mc Graw Hill Interamericana Editores S.A. 2006. ISBN: 970-10-5636-1
5. Administración de Proyectos. Guía para el Aprendizaje. Francisco Rivera Martínez, Gisel Hernández Chávez. Prentice Hall / Pearson. 2010. ISBN: 978-607-442-620-5.
6. Diseño Conceptual de Bases de Datos, un enfoque de entidades-interrelaciones. Carlo Batini, Stefano Ceri, Shamkant B. Navathe. Addison-Wesley / Díaz de Santos. ISBN 0-201-60120-6 (1994).
7. Utilización de NoSQL para resolución de problemas al trabajar con cantidades masivas de datos. Róttoli, Giovanni, López Nocera,

Marcelo, Pollo Cattaneo María Florencia. 2015. XVII Workshop de Investigadores en Ciencias de la Computación (Salta, 2015). http://sedici.unlp.edu.ar/handle/10915/45514.

8. Aspectos de Ingeniería de Software y Bases de Datos para el Desarrollo de Sistemas de Software en Escenarios Híbridos. Patricia Mabel Pesado, Rodolfo Bertone, Pablo Thomas, Ariel Pasini, Luciano Marrero, Eduardo Ibáñez, Alejandra Rípodas, Verónica Aguirre, Verena Olsowy, Fernando Tesone. XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019). Universidad Nacional de San Juan (Abril 2019). http://sedici.unlp.edu.ar/handle/10915/77088. ISBN: 978-987-3984-85-3
9. Un estudio comparativo de bases de datos relacionales y bases de datos NoSQL. Pesado Patricia Mabel, Thomas Pablo, Delía Lisandro, Marrero Luciano, Olsowy Verena, Tesone Fernando, Fernandez Juan Sosa. XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019). Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019. ISBN 978-987-688-377-1. http://sedici.unlp.edu.ar/handle/10915/91403.
10. NoSQL A Brief Guide To The Emerging World of Polyglot Persistence. Pramod J. Sadalage y Martin Fowler. Pearson Education. 2013. ISBN: 978-0-321-82662-6.
11. Data Modeling with NoSQL Database. Ajit Singh, Sultan Ahmad. ISBN 978-1072978374 (2019).
12. Aplicaciones para Dispositivos Móviles. Estrategias y enfoques de desarrollo. Thomas Pablo Javier, et. al.. XXI Workshop de Investigadores en Ciencias de la Computación (San Juan 2019). ISBN: 978-987-3984-85-3
13. Which Change Sets in Git Repositories Are Related?. Ramadani, J., Wagner, S. EEE 2016. International Conference on Software Quality, Reliability and Security (Viena, Austria.)
14. Synchronization and replication in the context of mobile applications. STAGE, A. (2005 Joint Advanced Student School Course 6: Next-Generation User-Centered).
15. III-LIDI: http://weblidi.info.unlp.edu.ar/wp/proyectos/investigacion/

TOMA DE DECISIONES MULTICRITERIO EN PROBLEMAS DE LA INGENIERÍA DE SOFTWARE UTILIZANDO COMPUTACIÓN BLANDA

Carlos Casanova, Manuel Chichi, María Luján Gabioud, Fernando Pereyra Rausch, Lucas Prado, Giovanni Daián Rottoli, Esteban Schab, Anabella De Battista

Grupo de Investigación sobre Inteligencia Computacional e Ingeniería de Software, Departamento Ingeniería en Sistemas de Información, Fac. Reg. Concepción del Uruguay, Universidad Tecnológica Nacional

Entre Ríos, Argentina

{casanovac, chichim, gabioudl, pereyraf, pradol, rottolig, schabe, debattistaa}@frcu.utn.edu.ar

RESUMEN

La Ingeniería de Software (IS) como la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento de software. Sin embargo, la adopción de herramientas formales que complementen la experiencia y el buen juicio en las distintas actividades de un proceso de desarrollo de software todavía es un pendiente dentro la industria del software, particularmente en la Argentina. Dos carencias que pueden explicar esto son, por un lado, la falta de conocimientos respecto de enfoques realistas para resolver problemas relativos a la IS, y por otro, la falta de herramientas software que auxilien a los tomadores de decisiones utilizando tales enfoques. Las líneas de investigación aquí propuestas tienden a suplir ambas carencias. Para esta tarea se propone la utilización de técnicas comprendidas en lo que se conoce como computación blanda (*soft computing*), dentro de las cuales se encuentran la Teoría de Conjuntos Difusos, las Redenes Neuronales y los métodos de Búsqueda Heurística y Metaheurística. La *soft computing* es capaz de brindar la flexibilidad necesaria para crear métodos y modelos que sean tolerantes a la imprecisión, la falta de información y la aproximación, características que le son propias a los contextos de decisión en la IS.

Palabras clave: Ingeniería de Software, Optimización, Soft Computing, Teoría de Conjuntos Difusos, Metaheurísticas, Preferencias.

CONTEXTO

El presente trabajo se desarrolla en el ámbito del Grupo de Investigación sobre Inteligencia Computacional e Ingeniería de Software (GIICIS), perteneciente al Departamento Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay.

1. INTRODUCCIÓN

El Vocabulario de la Ingeniería de Sistemas y Software de ISO/IEC/IEEE (SEVOCAB) define a la ingeniería de software como la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento de software; es decir, la aplicación de la ingeniería al software [1]. El Software es un bien intangible que cumple un doble rol. Es un producto y al mismo tiempo es el vehículo para entregar el producto más importante de nuestro tiempo: la información.

El software se desarrolla o modifica con intelecto: no se manufactura en el sentido clásico. Los costos del software se concentran en la ingeniería. Esto significa que los proyectos de software no pueden administrarse como si fueran proyectos de manufactura.

En este contexto, un proceso de software es un conjunto de actividades relacionadas que conducen a la producción de un producto de software [2]. Sin embargo, no es una prescripción rígida de cómo elaborarlo. Por el contrario, debe ser ágil y adaptable (al problema, al proyecto, al equipo y a la cultura organizacional). Por tanto, un proceso adoptado para un proyecto puede ser significativamente distinto de otro adoptado

para otro proyecto [2]. El proceso de software es complejo y, como todos los procesos intelectuales y creativos, confía en que las personas tomen decisiones con criterio.

2. ESTADO DE LA INDUSTRIA EN ARGENTINA

Aún así, en la República Argentina la industria del software, de un tiempo a esta parte, ha ido ganando progresivamente visibilidad e importancia para la economía del país. Un estímulo importante que ha incentivado el crecimiento de esta industria es el Régimen de Promoción de Software, instituido por la Ley 25.922 y con vigencia hasta el 31 de diciembre de 2019, el cual se reconoce como un instrumento que resultó eficiente para la promoción de dicha industria. Las actividades de la Economía del Conocimiento son las más dinámicas de la economía argentina: entre 2007 y 2017 generaron 65% más empleo que en el resto de las actividades productivas y sus operaciones crecieron en el mismo periodo un 70%, frente a un crecimiento general del 12%. Además, los salarios son un 35% más elevados que en el resto de los sectores. Actualmente, emplean a 215 mil personas y según el informe [3] del año 2017 de Argenconomics (entidad conformada por empresas prestadoras de servicios basados en el conocimiento, relacionadas con los mercados externos) las exportaciones de sectores vinculados a la Economía del Conocimiento ascendieron a USD 6.168 millones de dólares en ese año, convirtiéndose en el tercer rubro exportador del país, sólo detrás de la soja y sus derivados, y los cereales, y en el segundo exportador de Servicios Basados en Conocimiento (SBC) a nivel regional. En el primer semestre de 2019, según datos de la Secretaría de Transformación Productiva sobre la base del INDEC [4], los tres principales complejos exportadores fueron el cerealero/oleaginoso (US$ 13.596 millones), el automotriz (US$ 3802 millones) y el de los SBC (US$ 3385 millones).

3. PROBLEMAS QUE SURGEN EN EL CONTEXTO DE LA IS

La comunidad científica viene realizando esfuerzos en la definición realista de problemas propios de la ingeniería de software. El primer paso en cualquier metodología científica es la definición del problema. Sin una definición apropiada del problema, se corre el riesgo de encontrar una solución elegante para un problema que no existe. Características comunes que pueden identificarse son las siguientes:

• Hay una necesidad de **balancear objetivos que compiten** entre sí, es decir, se clasifican como problemas de optimización multi-objetivo.

• Existe la necesidad de tratar con **información incompleta, imprecisa o inconsistente** desde la formulación de los problemas. Este escenario es propio de entornos donde se desenvuelven los seres humanos. Por caso, puede darse que no se cuente con suficiente información para estimar con un grado de certeza aceptable ciertos parámetros. También puede darse la falta de acuerdo entre varios expertos al realizar una estimación, o que la misma sea establecida de forma imprecisa, por ejemplo, puede requerirse que el acoplamiento entre dos módulos sea “bajo” o que ciertos *stakeholders* resulten “muy favorecidos” en la próxima versión.

• Existen muchas soluciones potenciales, usualmente caracterizadas por una **explosión combinatoria** de las variables de decisión.

4. INVESTIGACIÓN OPERATIVA

La Investigación de Operaciones (IO) es un enfoque científico en la toma de decisiones que busca el mejor diseño y operación de un sistema, por lo regular en condiciones que requieren la asignación de recursos escasos [5]. Como su nombre lo indica, el objetivo de esta disciplina implica “investigar sobre las operaciones”. En consecuencia, esta disciplina se aplica a la problemática relacionada con la conducción y la coordinación de actividades en una organización.

La IO incluye el término investigación porque utiliza un enfoque similar al que se aplica en las áreas científicas establecidas. El proceso comienza por la observación cuidadosa y la formulación del problema, lo cual incluye la recolección de los datos pertinentes. El

siguiente paso es la construcción de un modelo científico —generalmente matemático— con el cual se intenta abstraer la esencia del problema real. En esta etapa se propone la hipótesis de que el modelo será una representación tan precisa de las características esenciales de la situación, que permitirá que las conclusiones —soluciones— que se obtengan sean válidas también para el problema real. Después se llevan a cabo los experimentos adecuados para probar esta hipótesis, para modificarla si es necesario y para verificarla en determinado momento. En cierto sentido, la IO involucra la investigación científica creativa de las propiedades fundamentales de las operaciones. Sin embargo, es más que esto. La IO se ocupa también de la administración práctica de la organización. Por lo tanto, para tener éxito, también debe proporcionar conclusiones claras que el tomador de decisiones pueda usar cuando sea necesario [5].

5. SOFT COMPUTING

La *soft computing* difiere de la "*hard computing*" o computación convencional en que, a diferencia de esta última, es tolerante a la imprecisión, la incertidumbre, la verdad parcial y la aproximación. En efecto, el modelo a seguir para la *soft computing* es la mente humana. El principio guía de la *soft computing* es: aprovechar la tolerancia a la imprecisión, la incertidumbre, la verdad parcial y la aproximación para conseguir tratabilidad, robustez y soluciones de bajo costo. Las técnicas de la *soft computing* incluyen la lógica difusa, algoritmos genéticos, redes neuronales artificiales, aprendizaje de máquinas y sistemas expertos [6]. Estos múltiples métodos no son competitivos entre sí, sino que son complementarios y pueden ser utilizados juntos para resolver un problema dado.

La precisión, el rigor y la certeza son atributos que implícitamente establecen un compromiso: se está en posesión de toda la información necesaria para poder encontrar la mejor solución posible para el problema dado. En efecto, la complejidad de ciertos problemas produce que no sea posible utilizar técnicas tradicionales para abordarlos si lo que se pretende son buenas soluciones en lapsos razonables. Más aún, para problemáticas cercanas a la realidad, las hipótesis y supuestos de las mismas pueden hacer imposible la aplicación de técnicas de computación convencional, generalmente debido a la presencia de formas específicas de incertidumbre en los parámetros. La *soft computing*, en tal caso, apunta a resolver los problemas aprovechando la imprecisión y la incertidumbre presente en el proceso de toma de decisiones, brindando las "soluciones más satisfactorias" en tal contexto de toma de decisiones.

6. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

6.1. Formulación realista de problemas de la IS como problemas de búsqueda/optimización

A nivel de industria de software mundial, ninguna técnica formal ha tenido un impacto significativo [7]. Una de las razones que puede explicar este fenómeno es que los modelos y algoritmos que se utilizan no tienen en cuenta desde su formulación, en su gran mayoría, la incertidumbre inherente al proceso de desarrollo de software (y por lo tanto, las soluciones obtenidas resultan de una vida útil casi efímera), o bien, si la contemplan, los modelos y resultados son difíciles de entender para una persona no experta en optimización o inteligencia artificial [8]. En este sentido, se busca formular modelos matemáticos utilizando mecanismos formales de captación de incertidumbre (como la teoría de conjuntos difusos, o la de posibilidades) que resulten adecuados para la realidad de un proceso de desarrollo de software. El primer paso, por tanto, consiste en formular el problema como uno de búsqueda/optimización. En este sentido, se trabaja junto a expertos de la industria del software en la especificación de problemas de la IS, con el objetivo de darle utilidad a las soluciones que se pueden obtener con los distintos algoritmos disponibles.

6.2. Mecanismos de captación de preferencias

Existen múltiples enfoques de optimización multiobjetivo, entre ellos los basados en preferencias. Estos suponen que puede incluirse información del tomador de decisiones para guiar la búsqueda hacia regiones prometedoras del espacio de soluciones. Idear mecanismos adecuados para captar esta información no es una tarea trivial, y se trabaja en el estudio de los mecanismos disponibles, y el diseño e implementación de nuevos mecanismos para mejorar principalmente la usabilidad de los métodos. Los componentes analizables de estos mecanismos son principalmente tres: la cantidad de interacciones con el usuario, qué información se le solicita, y en qué momento del proceso de optimización [9].

6.3. Integración con herramientas existentes de gestión de proyecto / proceso / producto

Como ya se ha dicho, los modelos y resultados de esta línea de investigación pueden resultar difíciles de entender para una persona no experta en optimización o inteligencia artificial. Se pretende, por lo tanto, construir una herramienta software que soporte la toma de decisiones que sea usable por administradores no expertos en *soft computing*. Es razonable suponer que si esta herramienta se integra a herramientas que la industria ya utiliza tendrá mayores posibilidades de ser incorporada.

7. RESULTADOS OBTENIDOS/ESPERADOS

Hasta el momento se han resuelto satisfactoriamente varios problemas, aunque ninguno de ellos está agotado. Entre ellos se encuentran el problema del próximo lanzamiento (*Next Release Problem*, NRP), resuelto mediante múltiples algoritmos evolutivos como NSGA-II, IBEA y una versión novedosa de PSO denominada FMOPSO [10], desarrollada por los autores, escrita en C++. Otro problema es el de la planificación de lanzamientos (*Release Planning Problem*, RPP), una extensión del NRP, en este caso utilizando programación matemática difusa, incorporando restricciones blandas [11]. También se ha resuelto el problema de priorización de requerimientos de software utilizando relaciones de preferencia difusa [12]. Para esto se desarrolló un prototipo en Octave que todavía debe ser mejorado.

Lograr un primer prototipo de interfaz usable para usuarios no expertos en este tipo de técnicas es un objetivo esperado. Este prototipo debe incorporar la posibilidad de realizar la carga de los datos necesarios de los problemas y la inducción de las preferencias de los usuarios. Además se proyecta poder construir *plugins* o mecanismos de interoperabilidad con *suites* existentes usadas en la industria.

Finalmente, se espera contar al final del proyecto con una colección de modelos altamente cohesivos que brinden información para facilitar la toma de decisiones relativas a distintos problemas de la Ingeniería de Software.

8. FORMACIÓN DE RECURSOS HUMANOS

La investigación presentada constituye las líneas fundacionales de un nuevo grupo de investigación dentro de la UTN-FRCU, el GIICIS. Un investigador ha sido invitado a realizar una investigación posdoctoral en el Grupo de Investigación en Modelos de Decisión y Optimización, Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada, España. Dos investigadores se encuentran realizando su tesis de doctorado. Además participan en el proyecto dos becarios alumnos de la carrera Ingeniería en Sistemas de Información que inician su formación en la investigación, dos alumnos realizan su práctica supervisada y uno su trabajo de fin de carrera en el contexto de esta investigación. En el marco de este proyecto ya se ha defendido exitosamente una práctica profesional supervisada y hay otra en inminente presentación.

9. BIBLIOGRAFÍA

[1] Sweb. E. Board, *Guide to the software engineering body of knowledge (SWEBOK (R)): Version 3.0*. IEEE Computer Society, 2014.

[2] I. Sommerville, *Software Engineering*, 9.ª ed. Pearson Education, Inc., 2011.

[3] Argenconomics, «Informe Economía del Conocimiento 2017», 2018. [En línea]. Disponible en: http://www.argencon.org/nota628-Argencon-realizo-su-Asamblea-Anual-y-presento-su-nuevo-informe-Argenconomics. [Accedido: 17-feb-2020].

[4] S. de la T. P. Dirección Nacional de Análisis y Estadísticas Productivas, Subsecretaría de Desarrollo y Planeamiento Productivo, «Argentina productiva: economía del conocimiento», 2019.

[5] F. S. Hillier y G. J. Lieberman, *Introducción a la Investigación de operaciones*. McGRAW-HILL/INTERAMERICANA EDITORES, S.A. DE C.V., 2010.

[6] L. A. Zadeh, «Fuzzy Logic, Neural Networks, and Soft Computing», *Fuzzy Systems*, vol. 37, n.º 3, pp. 77-84, 1994.

[7] D. C. C. Peixoto, G. R. Mateus, y R. F. Resende, «The issues of solving staffing and scheduling problems in software development projects», en *Proceedings - International Computer Software and Applications Conference*, 2014.

[8] J. M. Rojas y G. Fraser, «Is search-based unit test generation research stuck in a local optimum?», en *Proceedings - 2017 IEEE/ACM 10th International Workshop on Search-Based Software Testing, SBST 2017*, 2017, n.º c, pp. 51-52.

[9] T. N. Ferreira, S. R. Vergilio, y J. T. de Souza, «Incorporating user preferences in search-based software engineering: A systematic mapping study», *Inf. Softw. Technol.*, vol. 90, pp. 55-69, 2017.

[10] C. Casanova *et al.*, «Aproximación del Frente Pareto-Óptimo de un Problema NRP Bi-Objetivo mediante un Algoritmo basado en Enjambres de Partículas», en *Anales VI Seminario Argentina-Brasil de Tecnologías de la Información y la Comunicación*, 2018.

[11] C. Casanova, F. Pereyra Rausch, y L. Prado, «Modelo de Planificación de Lanzamientos de Software utilizando Restricciones Blandas», en *VII Congreso Nacional de Ingeniería Informática / Sistemas de Información*, 2019.

[12] M. L. Gabioud y C. Casanova, «Priorización en Ingeniería de Requerimientos con Preferencias Difusas», en *VII Congreso Nacional de Ingeniería Informática / Sistemas de Información*, 2019.

Ingeniería de Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube

Mario Beron[1], Norma Perez[1], Daniel Riesco[1], Germán Montejano[1]
{mberon|nbperez|driesco|gmonte}@ unsl.edu.ar
Maria J. Pereira [2], Paulo Novais [3], Pedro Henriques [3]
mjoao@ipb.pt - {pjon|prh}@ di.uminho.pt
[1] Universidad Nacional de San Luis - Departamento de Informática
[2] Instituto Politécnico de Bragança - Escola Superior de Tecnología e Gestão
[3] Universidad de Minho - Departamento de Informática

1. Resumen

Los avances de las tecnologías de la información y de la comunicación requieren que las organizaciones se adapten al contexto. Esta tarea implica la construcción de nuevos sistemas y el ajuste de otros preexistentes para que los procesos se lleven a cabo de manera eficiente y confiable. La primer tarea se debe llevar a cabo siguiendo pasos bien definidos con el propósito de desarrollar software de calidad y seguro. La segunda implica la utilización de métodos y técnicas de mantenimiento y migración sofisticadas que permitan transformar un sistema que utiliza tecnología antigua en otro que usa tecnología de punta y que a su vez mantenga / mejore los requisitos de calidad y seguridad. Dicha tarea es extremadamente complicada porque no solo necesita de un conocimiento acabado de estrategias de mantenimiento, de conceptos de calidad y de seguridad sino que también exige entrelazar creativamente los saberes para elaborar estrategias eficientes y robustas de forma tal que las organizaciones puedan utilizar sus sistemas lo antes posible. En este artículo se describe un proyecto de investigación de reciente creación cuyo principal objetivo es el desarrollo de estrategias de mantenimiento y migración de sistemas de software que transformen sistemas que usan tecnología obsoleta en otros que empelan tecnología de punta y hacen uso de la nube preservando / mejorando las características de calidad y seguridad.

2. Contexto

El presente proyecto de investigación ha sido presentado para su evaluación en la categoría Proyectos de Investigación Consolidados (PROICO) en la Universidad Nacional de San Luis y se espera que su ejecución comience durante el período 2020-2023. El mismo es la continuación de otro

proyecto consolidado denominado "Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad " – Facultad de Ciencias Físico- Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto NRO. P-031516. Es importante mencionar que el proyecto antes mencionado ha tenido muy buen impacto a nivel nacional e internacional lo cual se refleja en la producción realizada y en los vínculos establecidos con diferentes universidades de nivel nacional e internacional.

3. Introducción

El software es una de las componentes fundamentales para la adaptación de los sistemas a los nuevos requerimientos que demandan los progresos tecnológicos como así también para todos aquellos productos que tienen como base a las computadoras. El paso del tiempo y el advenimiento de grandes cambios científicos-tecnológicos han conducido a que el software haya dejado de ser una herramienta de asistencia a los usuarios y a las organizaciones para convertirse en una industria pujante y vanguardista, y por lo tanto adolece, como cualquier otra industria, de problemas para desarrollar productos de alta calidad, dentro de los márgenes de tiempo apropiados y con los costos indicados. Si bien muchas de las tareas complejas del proceso de desarrollo han sido estudiadas con el propósito de mejorar las formas de llevarlas a cabo y de hecho se han logrado muchos avances en el área (basta con observar la gran evolución que han experimentado los diferentes enfoques para el desarrollo de software), los dominios de aplicación siguen ofreciendo problemáticas que aún demandan la realización de investigaciones que subsanen falencias o bien propongan soluciones a problemas no abordados o escasamente estudiados. Tenga presente que la gran mayoría de los sistemas son web o han sido desarrollados para operar en la nube y que por lo general hacen un fuerte uso de almacenamiento que se encuentra en lugares remotos [3, 9, 11]. En este contexto, la tarea de desarrollo, mantenimiento y migración de sistemas de software presentan problemas que requieren de estudios concienzudos para que se puedan elaborar las soluciones correspondientes [7, 6, 8]. Uno de tales inconvenientes es: Llevar adelante tareas de mantenimiento de los sistemas de software que ejecutan (o hacen uso) de la nube preservando y/o mejorando la calidad/seguridad de los mismos [11, 10, 14]. Está demás de decir que un problema similar se puede enunciar para el caso de la migración de sistemas de software [13, 10, 1]. Dado la relevancia y lo desafiante de la temática en este proyecto se pretende investigar: El desarrollo de métodos, técnicas y herramientas de mantenimiento y migración de sistemas de software que ejecutan en la nube preservando/mejorando la calidad y seguridad de los mismos.

4. Líneas de Investigación y Desrrollo

El proyecto de investigación brevemente presentado en este artículo consta de dos líneas de investigación las cuales se describen a continuación:

Desarrollo de Estrategias de Mantenimiento que Preserven la Calidad y Seguridad los Sistemas de Software que Ejecutan en la Nube: Esta línea de investigación se encarga del estudio de los procesos de mantenimiento de software orientados a preservar / mejorar la calidad y seguridad de los sistemas. Para llevar adelante esta actividad los estudios se centrarán en el análisis de las principales fallas en la calidad y seguridad de los sistemas de software tradicionales y de aquellos que ejecutan en la nube. Una vez que se hayan descrito los inconvenientes antes mencionados se construirán diferentes herramientas que permitan identificar y visualizar las fallas en la calidad y seguridad [15, 12]. Esto le permitirá al ingeniero comprender la causa de la falla y seleccionar la estrategia adecuada para subsanarla para lo cual se tendrá que desarrollar herramientas de re-ingeniería que contribuyan a facilitar el proceso. Las herramientas desarrolladas serán utilizadas para analizar sistemas de software de importante envergadura y que sean utilizado por las organizaciones o bien por la comunidad científica [4, 5, 2]. A partir de la experiencia adquirida en el uso de las herramientas se elaborarán mejoras a los procesos de mantenimiento utilizados con el fin de producir otra contribución en el campo de la Ingeniería de Software. Es importante mencionar que el sistema producto del proceso de mantenimiento también será sometido a pruebas de chequeo de calidad y seguridad con el fin de poder analizar las bondades de los enfoques propuestos por el equipo de investigación.

Desarrollo de Métodos Estrategias y Herramientas de Migración Preventivo de Sistemas de Software que Ejecutan en la Nube: En esta línea de investigación se procederá utilizando una enfoque similar al utilizado para las tareas de mantenimiento, es decir habrá una etapa en donde se realicen estudios sistemáticos de la literatura que describan los principales problemas por los que atraviesa el proceso de migración de sistemas de software y también se analizarán herramientas y enfoques existentes para migrar software convencional a otro que se ejecute o utilice la nube [9, 11]. A partir de esos estudios se desarrollarán herramientas de inspección, comprensión, re-ingeniería desde la óptica de las actividades de migración es decir están orientadas no solo a comprender el sistema y detectar fallas de calidad y seguridad sino también a identificar las partes del sistema que pueden provocar problemas de migración. También se considerarán la elaboración de herramientas que aborden aspectos de la migración de software vinculadas con el cambio de sistemas operativos, base de datos, ambientes, entre otras tantas posibilidades. Claramente, las herramientas desarrolladas tienen que poder analizar software de gran tamaño y que se use en la vida cotidiana.

5. Resultados Obtenidos / Esperados

Es importante destacar que el proyecto descrito comenzará su ejecución durante el año 2020. No obstante, los resultados con los que se cuentan actualmente consisten de un amplio estado del arte sobre temáticas de seguridad y calidad, varios proyectos finales integradores que abordan la temática

de seguridad, como así también varias tesis de maestría que hacen lo mismo con la temática de calidad. En cuanto a las tesis doctorales es posible decir que se ha finalizado una que describe métodos y estrategias para comprender sistemas de software y también otras se encuentran en desarrollo en donde el foco de estudio es calidad y seguridad.

Respecto de los resultados esperados para la ejecución de este proyecto los mismos se encuentran enmarcados varios ejes para los cuales en los párrafos siguientes se hace un listado no exhaustivo de contribuciones que se espera la ejecución del proyecto.

Eje: Contribución al avance del conocimiento científico y / o tecnológico:

a) Mantenimiento de Software: i) Desarrollo de diferentes estrategias de análisis de programas que tengan como objetivo identificar diferentes problemas de mantenimiento (perfectivo, adaptativo, correctivo, etc.); ii) Desarrollo de técnicas de transformación de código que faciliten las subsanar los problemas de mantenimiento; iii) Etc; b) Migración de Software: i) Elaboración de procesos de migración que preserven la calidad / seguridad del sistema migrado; ii) Detección de introducción de problemas de calidad y vulnerabilidades en el sistema migrado; iii) Migración de sistemas legado; iv) Etc;

c) Calidad de Sistemas de Software: i) Especificación de fallas en la calidad de los sistemas de software que ejecutan en la nube; ii) Elaboración de técnicas de re-ingeniería que subsanen problemas de calidad de los sistemas de software ejecutan en la nube; iii) Etc;

d) Seguridad Informática: i) Especificación de las vulnerabilidades de los sistemas de software que ejecutan en la nube; ii) Elaboración de técnicas de re-ingeniería que subsanen las vulnerabilidades de los sistemas de software que ejecutan en la nube. iii) Etc.

Eje: Contribución a la formación de RRHH: i) La elaboración de tesis doctorales, maestría, licenciatura, proyectos finales integradores de ingeniería en las temáticas investigadas en el proyecto; ii) La introducción de becarios para desarrollar tesis en diferentes niveles académicos y pasantes para apoyar las tareas que requieran las tesis de diferentes niveles académicos; iii) La realización de publicaciones en conferencias y revistas nacionales e internacionales; iv) Etc.

Eje: Contribución al desarrollo socio económico: i) Transferencia de conocimiento y prácticas a organizaciones / instituciones educativas de la región / otras provincias / países sobre temas de mantenimiento y migración de sistemas de software; ii) Transferencia de conocimiento y prácticas a organizaciones / instituciones educativas de la región / otras provincias / países sobre temas de calidad y seguridad de sistemas de software; iii) Etc.

6. Formación de Recursos Humanos

Los progresos obtenidos en este proyecto de investigación, tal como ha sucedido en los proyectos anteriores, servirán de base para el desarrollo de tesis de posgrado, ya sea de doctorado o maestrías en Ingeniería de Software y desarrollo de trabajos finales de las carreras Licenciatura en Ciencias de

la Computación, Ingeniería en Informática e Ingeniería en Computación de la Universidad Nacional de San Luis.

Referencias

[1] Allen, J. H., Barnum, S., Ellison, R. J., McGraw, G., and Mead, N. R. *Software Security Engineering: A Guide for Project Managers (The SEI Series in Software Engineering)*, 1 ed. Addison-Wesley Professional, 2008.

[2] April, A., and Abran, A. *Software Maintenance Management : Evaluation and Continuous Improvement*. 05 2008.

[3] Bass, L., Clements, P., and Kazman, R. *Software Architecture in Practice*. Addison-Wesley Longman Publishing Co., Inc., USA, 1998.

[4] Eilam, E. *Reversing: Secrets of Reverse Engineering*. John Wiley & Sons, Inc., USA, 2005.

[5] Erdogmus, H., and Tanir, O. *Advances in Software Engineering: Comprehension, Evaluation, and Evolution*. 01 2002.

[6] Laszewski, T., and Nauduri, P. *Migrating to the Cloud: Oracle Client/Server Modernization*, 1st ed. Syngress Publishing, 2011.

[7] Lomita, A., Litoiu, M., and Lewis, G. *Migrating Legacy Applications: Challenges in Service Oriented Architecture and Cloud Computing Environments*, 1st ed. IGI Global, 11 2012.

[8] Lomita, A., Litoiu, M., and Lewis, G. *Big Data Analytics Methods: Modern Analytics Techniques for the 21st Century: The Data Scientist's Manual to Data Mining, Deep Learning & Natural Language Processing*, 1st ed. 1530414830, 3 2016.

[9] Marinescu, D. Cloud computing: Theory and practice. *Cloud Computing: Theory and Practice* (06 2013), 1–396.

[10] Mead, N. R., and Woody, C. *Cyber Security Engineering: A Practical Approach for Systems and Software Assurance*, 1st ed. Addison-Wesley Professional, 2016.

[11] Rhoton, J. *Cloud Computing Protected: Security Assessment Handbook*.

[12] Singh, A. *Identifying Malicious Code Through Reverse Engineering*, 1 ed. Springer Publishing Company, Incorporated, 2009.

[13] Tilley, S., and Parveen, T. *Software Testing in the Cloud: Migration and Execution*. Springer Publishing Company, Incorporated, 2012.

[14] Tripathy, P., and Naik, K. *A Practitioner's Approach, Software Evolution and Maintenance*. John Wiley & Sons, Inc., USA, 2014.

[15] Wagner, C. *Model-Driven Software Migration: A Methodology: Reengineering, Recovery and Modernization of Legacy Systems*. 03 2014.

Desarrollo de Recursos Humanos para la Gobernanza de Ciudades Inteligentes Sostenibles – Proyecto CAP4CITY

Armando De Giusti, Patricia Pesado, Ariel Pasini, Pablo Thomas, Rocío Muñoz[1] Juan Santiago Preisegger[1] Elsa Estevez, Pablo Fillottrani, Sonia Rueda, Karina Cenci

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
[1] Becario postgrado UNLP
50 y 120 - La Plata, Buenos Aires
Centro Asociado CIC
526 e/ 10 y 11 - La Plata, Buenos Aires
(degiusti, ppesado, apasini, pthomas, rmunoz,jspreisegger) @lidi.info.unlp.edu.ar

Laboratorio de Ingeniería de Software y Sistemas de Información (LISSI)
Departamento de Ciencias e Ingeniería de la Computación – Universidad Nacional del Sur
Av. San Andrés 800 – Campus de Palihue - Bahía Blanca, Buenos Aires
Centro Asociado CIC
526 e/ 10 y 11 - La Plata, Buenos Aires
(ece, prf, svr, kmc) @cs.uns.edu.ar

RESUMEN

El proyecto CAP4CITY tiene como objetivo fortalecer y desarrollar la capacidad de instituciones académicas en América Latina y Europa para mejorar la calidad de la educación superior en el campo de ciudades inteligentes sostenibles. El Proyecto es co-financiado en el marco del Programa Erasmus+ de la Unión Europea, para la construcción de capacidades en la educación superior. Se presenta el avance del primer año del proyecto y las actividades a realizar en el siguiente periodo.

.

Palabras Claves
Ciudades Inteligentes - Capacidades de Gobernanza - Gobernanza Digital - Colaboración Internacional

CONTEXTO

La línea de investigación y desarrollo aquí descripta es parte de las tareas planificadas por el proyecto "Strengthening Governance Capacity for Smart Sustainable Cities (CAP4CITY)". El Proyecto, financiado como parte del programa Erasmus+ de la Unión Europea es ejecutado por un consorcio integrado por 12 universidades; cuatro de ellas europeas – Donau Universität für Weiterbildung (DUK) en Austria, Tallinn University of Technology (TUT) en Estonia, Delft University of Technology (TU Delft) en los Países Bajos, y Gdańsk University of Technology (GUT) en Polonia; y ocho universidades en la región de América Latina – Universidad Nacional de La Plata (UNLP) y Universidad Nacional del Sur (UNS) en Argentina; Ponitificia Universidade Catolica do Rio Grande do Soul (PUCRS) y Faculdade Meridional (IMED) en Brazil; Universidad Técnica Federico Santa María (UTFSM) y Universidad Católica del Norte (UCN) en Chile; y Universidad Externado de Colombia (UEC) y Escuela Colombiana de Ingeniería (ECI) en Colombia. El Proyecto lleva el número 598273 y se ejecuta bajo el acuerdo 598273-EPP-1-2018-1-AT-EPPKA2-CBHE-JP.

1. INTRODUCCION

El proyecto CAP4CITY tiene como objetivo fortalecer y desarrollar la capacidad de instituciones académicas en América Latina y Europa para mejorar la

calidad de la educación superior en el campo de ciudades inteligentes sostenibles (SSC).

Se define una SSC como una ciudad innovadora que utiliza Tecnologías de Información y Comunicación (TIC) y otros medios para mejorar la calidad de vida, la eficiencia de la operación y los servicios urbanos, y la competitividad, al tiempo que se garantiza que satisfaga las necesidades de las generaciones presentes y futuras con respecto a los aspectos económicos, sociales, ambientales y culturales.

Uno de los modelos de SSC propone cinco dimensiones para su desarrollo: 1) Social, 2) Económica, 3) Ambiental, 4) Gobernanza, y 5) Infraestructura Urbana. La Figura 1 muestra el modelo.

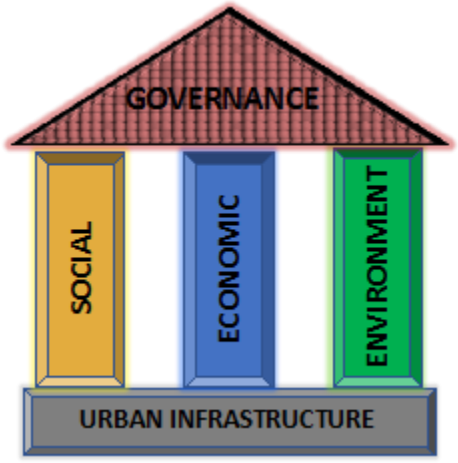


Figura 1. Dimensiones de una SSC

La dimensión Social cubre los aspectos relacionados con las personas y comunidades para garantizar la calidad de vida, como, por ejemplo: salud, seguridad, educación, entre otros. La Económica cubre los aspectos relacionados con el crecimiento económico responsable y sustentable y la generación de oportunidades laborales. Por su parte, la Ambiental se relaciona con la utilización de prácticas ecológicas, la protección y restauración del medio ambiente. La dimensión de Gobernanza se refiere a la capacidad de administrar recursos, políticas e involucrar a diferentes partes interesadas, proponiendo mecanismos y procesos regulatorios y de cumplimiento bien equilibrados de manera estandarizada y continua. Por último, la dimensión de Infraestructura Urbana se refiere a la infraestructura física (carreteras, transporte, etc.) y a la infraestructura digital (tecnología de la información y comunicación), que son herramientas esenciales para permitir ciudades inteligentes y sostenibles

Dadas estas cinco dimensiones, la complejidad de las mismas, así como la característica multidisciplinar de los problemas a resolver en SSC, hace que la construcción de capacidades humanas, particularmente, la formación de líderes que se dediquen al liderazgo y gobernanza de estas iniciativas sea un gran desafío. Justamente, el Proyecto CAP4CITY busca proponer soluciones al mismo, realizando diseño de cursos y materiales de enseñanza y aprendizaje que permitan crear las capacidades humanas necesarias.

En base al estudio del arte realizado que incluyó un relevamiento de oferta – programas de estudio ofrecidos; y demanda – puestos de trabajo requeridos, el creciente número de competencias necesarias y su característica interdisciplinaria, los cursos diseñados abarcan varias áreas temáticas, tales como: Tecnologías de la Información y la Comunicación (TICs), Administración de Empresas, Informática, Ingeniería, Arquitectura y Urbanismo, Planificación Urbana, y Ciencias Políticas, entre otras.

A fin de definir las competencias necesarias para el desarrollo de SSC en América Latina, se hizo un relevamiento de datos en la región. A tal efecto, se realizaron dos talleres en las ciudades de Bogotá y Medellín, Colombia, como experiencia piloto. Luego cada una de las universidades de Latino América replico el "*Taller de Relevamiento de Competencias para el Desarrollo de Ciudades*

Inteligentes y Sostenibles" con representante de su comunidad, incluyendo ciudadanos y representantes de la industria, la academia y el gobierno.

Con los resultados obtenidos de todas las universidades se estableció una lista de competencias en base a las cuales se definieron 31 cursos. La validación de los mismos realizará con los interesados locales durante el primer trimestre de 2020.

Los beneficiarios del Proyecto CAP4CITY incluyen instituciones de educación superior a nivel local, provincial, nacional e internacional. Estas podrán utilizar los cursos diseñados, compartir conocimientos e intercambiar experiencias sobre el desarrollo y la entrega de programas educativos relacionados con SSC.

El resto de este artículo se organiza de la siguiente manera. La Sección 2 explica los objetivos de investigación, la Sección 3 los resultados esperados/obtenidos y, por último, la Sección 4 discute la formación de recursos humanos.

1. OBJETIVOS DE INVESTIGACIÓN Y ACTIVIDADES

- Mejorar la calidad de la educación superior en el campo de SSC, aumentando su relevancia para el mercado laboral y la sociedad en general.
- Aumentar las competencias de los recursos humanos en los países en desarrollo de América Latina para enfrentar los desafíos del mundo digital.
- Proveer el desarrollo continuo de competencias en SSC a través de programas de capacitación y educación.
- Asegurar una enseñanza multidisciplinaria orientada a la resolución de problemas en SSC.
- Desarrollar una red de cooperación entre los socios de diferentes regiones del mundo.
- Promover la colaboración entre los socios del Consorcio, las entidades públicas, las empresas y otros interesados en SSC.
- Facilitar el intercambio de conocimientos, experiencias y buenas prácticas e iniciativas conjuntas en materia de SSC entre socios académicos ubicados en diferentes ciudades y países.

La Universidad Nacional de La Plata y la Universidad Nacional del Sur colaborarán con las siguientes actividades:

- Relevamiento de iniciativas de SSC que puedan identificarse como buenas prácticas y de programas de postgrado relacionados con temas de SSC.
- Capacitación de formadores en SSC.
- Implementación de nuevos programas educacionales en SSC.
- Difusión y explotación del concepto de SSC en Argentina.
- Fortalecimiento de los vínculos laborales con sectores públicos y privados de la región.

2. RESULTADOS ESPERADOS/OBTENIDOS

Al primer trimestre de 2020, los resultados obtenidos son:

- Se ha desarrollado un relevamiento relevamiento de programas de postgrado relacionados con temas de SSC.
- Se ha desarrollado un curso a distancia sobre la construcción de MOOC que serán utilizado en el dictado de los cursos en proyecto.
- Se diseñaron 31 cursos de diferentes niveles – pregrado, posgrado y formación continua, relacionados a

SSC que se encuentran pendientes de validación
- Organización de dos tipos de talleres en América Latina para: 1) identificar el tipo de competencias que se necesita desarrollar y 2) validar de los módulos de formación propuestos.

Como trabajo a futuro, los planes incluyen:

- Implementación de la "capacitación de formadores".
- Implementación de nuevos planes de estudio de postgrado relacionados con SSC en conjunto con otras universidades socias.
- Incremento de la matrícula de alumnos interesados en los postgrados de SSC.
- Incremento de publicaciones relacionadas con SSC por parte de las universidades involucradas.
- Generación de nuevas oportunidades laborales en empresas de la región.

3. FORMACION DE RECURSOS HUMANOS

- Capacitación de los miembros del proyecto en SSC.
- Desarrollo de tesis de postgrado y tesinas de grado en el área.
- Participación de los integrantes de esta línea de investigación en el dictado de asignaturas/cursos de grado/postgrado en la Facultad de Informática de la UNLP y en el Departamento de Ciencias e Ingeniería de la Computación de la UNS.

4. AGRADECIMIENTOS

Project co-funded by the Erasmus+ Programme of the European Union. Grant no: 598273-EPP-1-2018-1-AT-EPPKA2-CBHE-JP.

5. BIBLIOGRAFIA

[1] Akande, A., Cabral, P., Gomes, P., Casteleyn, S. 2019. "The Lisbon ranking for smart sustainable cities in Europe." Sustainable Cities and Society, 44, pp. 475-487.

[2] Caragliu, Andrea, Chiara Del Bo, and Peter Nijkamp. 2011. "Smart Cities in Europe." Journal of Urban Technology 18(2):65–82.

[3] Craglia, Massimo, Lila Leontidou, Giampaolo Nuvolati, and Jürgen Schweikart. 2004. "Towards the Development of Quality of Life Indicators in the 'Digital' City." Environment and Planning B: Planning and Design 31(1):51–64.

[4] Elsa Estevez, Nuno Lopes, Tomasz Janowski. 2016. "Smart sustainable cities: Reconnaissance Study", 1-330.

[5] Huovila, A., Bosch, P., Airaksinen, M. 2019. "Comparative analysis of standardized indicators for Smart sustainable cities: What indicators and standards to use and when?" Cities, 89, pp. 141-153.

[6] Keshvardoost, S., Renukappa, S., Suresh, S. 2019. "Developments of policies related to smart cities: A critical review". Proceedings - 11th IEEE/ACM International Conference on Utility and Cloud Computing Companion, UCC Companion 2018, art. no. 8605807, pp. 365-369.

[7] Kurebayashi, Toshihiko, Yoshihiro Masuyama, Kiyonori Morita, Naoyuki Taniguchi, and Fumio Mizuki. 2011. "Global Initiatives for Smart Urban Development." Hitachi Review 60(2):89–93.

[8] Martin, C., Evans, J., Karvonen, A., Paskaleva, K., Yang, D., Linjordet, T. 2019. "Smart-sustainability: A new urban fix?" Sustainable Cities and Society, 45, pp. 640-648.

[9] Paroutis, Sotirios, Mark Bennett, and Loizos Heracleous. 2014. "A Strategic View on Smart City Technology: The Case of IBM Smarter Cities During a Recession." Technological Forecasting and Social Change 89:262– 72.

Evaluación de la reusabilidad de los conjuntos de datos abiertos de un portal de Infraestructura de Datos Espaciales. Caso de Estudio: IDERA

María A. Barrera[1], Carlos Salgado[2]

[1]Departamento de Informática
Facultad de Tecnología y Ciencias Aplicadas
Universidad Nacional de Catamarca
Maximio Victoria Nº 55 – (4700) Catamarca - Argentina
Tel.: +54-383- 4435112 – Int. 168
marita.latingeo@gmail.com

[2]Departamento de Informática
Universidad Nacional de San Luis
Ejército de los Andes 950 – (5700) San Luis – Argentina
Tel.: +54-266-4520300– Int. 2101
csalgado@unsl.edu.ar

Resumen

Cada vez más organismos públicos ponen a disposición de los ciudadanos datos geoespaciales a través de sus Infraestructuras de Datos Espaciales (IDE), denominados también geoportales. Estos, son aplicaciones web que se utilizan para buscar y acceder a información geográfica (IG) y servicios geográficos asociados a través de Internet.

Sin embargo, estudios a nivel mundial sobre la utilización de las aplicaciones, muestran que aquellas creadas con datos abiertos (DA) públicos no se encuentran ni siquiera entre las 100 aplicaciones más descargadas por los usuarios y no se cuenta con estándares que evalúen los recursos que componen los conjuntos de datos de una IDE, por lo que pueden existir recursos difíciles de reutilizar debido a diversas características de los geoportales, y los consumidores de los mismos tienen que realizar un gran esfuerzo para encontrar datos de calidad y que se ajusten a sus necesidades.

Esta investigación realiza un análisis de la información publicada en los geoportales, con el objeto de medir el grado de reutilización de sus conjuntos de datos, para esto se propone una adaptación de la métrica MELODA, teniendo en cuenta la naturaleza específica de los geodatos y tomando como caso de prueba la IDE de la República Argentina.

Palabras Claves: Datos Abiertos, Datos Geoespaciales, Información Pública, Infraestructura de Datos Espaciales, Métricas e Indicadores.

Contexto

Se presenta aquí la investigación que se viene realizando en el marco de una de las líneas de investigación correspondiente al proyecto "Las TIC al Servicio del Dato Abierto: Situación actual, conceptualización e iniciativas de apertura de información Pública", aprobado y subsidiado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Catamarca y que se desarrolla en ámbitos del Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa.

La labor se realiza en forma colaborativa con el proyecto de investigación: Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto Nº P-031516. Además, ambos proyectos se encuentran reconocidos por el

programa de Incentivos.

Introducción

"Todos los datos producidos por las administraciones públicas, son datos públicos" (CEPAL, 2012). Muchas ciudades publican sus datos en formato abierto, pero el gran desafío en la actualidad es demostrar la capacidad y calidad de la información publicada, la reutilización que hace el ciudadano con esos datos y la forma en la que encuentran la información para que los datos puedan ser consultados, descargados o incluidos en aplicaciones externas.

Por otra parte, la disponibilidad, accesibilidad y el intercambio de información georeferenciada son requerimientos que cobran paulatinamente mayor importancia para una sociedad interesada en incorporar este tipo de información como apoyo fundamental en los procesos de planificación y toma de decisiones, tanto a nivel local como regional e inclusive global (Rix et al., 2011). Esta utilización de conjuntos de datos espaciales conlleva múltiples abordajes para el diseño y construcción de herramientas tecnológicas dirigidas hacia este fin (Rajabifard & Williamson, 2001). Sin embargo, cualquiera que sea el punto de partida, es primordial que los datos georeferenciados se brinden de forma precisa, sean mantenidos adecuadamente y garanticen su veracidad (Williamson, Rajabifard, & Binns, 2006).

Ante dicha coyuntura surge la noción de Infraestructura de Datos Espaciales (IDE) como alternativa para suplir información georeferenciada y, a la vez, cumplir con las demandas mencionadas anteriormente en relación a su integridad y validez. Una IDE consiste en una plataforma tecnológica distribuida que permite a los proveedores publicar sus datos geoespaciales y a los usuarios acceder a dicha información. Cabe resaltar que las IDE incluyen un conjunto de políticas, estándares y buenas prácticas que guían su uso (Rix et. al., 2011).

Estas infraestructuras de los países, se hacen visibles a través de los denominados geoportales. Estos son aplicaciones web que actúan como un punto de entrada a la información geográfica distribuida y disponible para ser compartida (European Commission, 2005).

Cada vez hay más ciudades que cuentan con portales de IDE que ponen a disposición de los ciudadanos datos, productos y servicios geográficos, publicados bajo estándares y normas, asegurando su interoperabilidad y uso, la propiedad sobre la información por parte de los organismos que la publican y su responsabilidad de actualización, que han sido publicados para que los ciudadanos puedan consultarlos, descargarlos o incluirlos en aplicaciones externas abriendo nuevas posibilidades para generar valor, tanto en el ámbito social como en el económico. Ahora, el desafío es demostrar las capacidades de esa información y la reutilización que se hace de esos datos. Sin embargo, los estudios a nivel mundial sobre la usabilidad de las aplicaciones, muestran que aquellas creadas con datos abiertos públicos no se encuentran ni siquiera entre las 100 aplicaciones más descargadas por los usuarios. Igualmente, el Barómetro de Datos Abiertos (Open Data Barometer en inglés, es un proyecto de la ONG World Wide Web Foundation que consiste en una revisión del estado de la apertura de los datos públicos en múltiples países del mundo. https://opendatabarometer.org/), indica que menos del 7% de los datos publicados son reutilizables adecuadamente. Esta es una de las mayores preocupaciones que se presentan en las discusiones de DA y servicios geográficos, probablemente debido a la complejidad de los metadatos y de los estándares, la tecnología usada para la publicación, la confianza de los usuarios, la baja integración y la actualización de los datos disponibles, entre otros.

Comienzan a observarse iniciativas que han tomado este nuevo reto y han planteado algunas estrategias para mejorar el nivel de reuso de la información geográfica, sin embargo hay interrogantes que siguen en la mira: ¿Cuál es el

nivel de utilidad de la información geográfica publicada en los portales de DA de las ciudades?, ¿los ciudadanos están aprovechando efectivamente la publicación de estos servicios para promover la transparencia, la colaboración y el emprendimiento?, ¿o son solo iniciativas que corresponden a intereses de las administraciones actuales en cuanto a los mandatos de transparencia?. Es por esto que la aplicación de métricas se presenta como una necesidad estratégica para calificar la información disponible y evaluar su grado de reutilización.

En esta línea de investigación se está realizando un análisis general de la información publicada en los portales de IDE, con el objeto de medir el grado de reutilización efectiva de sus conjuntos de datos, teniendo en cuenta estándares técnicos, acceso legal, formatos de datos, visualización, su utilización en la recuperación de información, y las exigencias que las mismas tienen por parte de organismos internacionales de normalización para asegurar su calidad e interoperabilidad, para que las IDE sean herramientas ciertas que mejoren el acceso y reutilización de los datos abiertos geoespaciales a toda la ciudadanía.

En este sido, según las encuestas de la Cámara de la Industria Argentina de Software, http://www.cessi.org.ar/opssi (2016) las métricas son un buen medio para entender, monitorizar, controlar, predecir y probar el desarrollo de software y los proyectos de mantenimiento, y para poder asegurar que un proceso o sus productos resultantes son de calidad, o poder compararlos, es necesario asignar valores, descriptores, indicadores o algún otro mecanismo mediante el cual se pueda llevar a cabo dicha comparación. Para ello, y según lo establecido en la nueva Ley de Promoción de la Industria del Software, (http://servicios.infoleg.gob.ar/infolegInternet/anexos/185000-189999/185701/norma.htm - 2011) es necesario llevar a cabo un proceso de medición del software que, en general, persigue tres objetivos fundamentales: (i) ayudarnos a entender qué ocurre durante el desarrollo y el mantenimiento, (ii) permitirnos controlar qué es lo que ocurre en los proyectos y (iii) poder mejorar los procesos y productos.

Abella, Ortiz-de-Urbina-Criado y De Pablos-Heredero (2017) han elaborado la métrica MELODA que permite evaluar el grado de reutilización de los datos abiertos. MELODA analiza seis dimensiones: Marco Legal, Estándares Técnicos, Acceso a la Información, Modelo de Datos, Geolocalización de la Información y Actualización en Tiempo Real. Cada una de estas dimensiones consta de cinco niveles mediante los cuales se obtiene un porcentaje de calidad de las mismas, dicho porcentaje se utiliza posteriormente en una fórmula que proporcionará el nivel de reusabilidad de los conjuntos de datos disponibles en los portales abiertos.

Esta última métrica se ha utilizado por otros autores como Vicente-Paños y Jordán-Alfonso (2017) para estudiar el acceso a la información del sector público y su reutilización en las comunidades autónomas de España. Estos autores observaron que cuando se calificaban a través de MELODA los portales de datos abiertos de esas comunidades, las puntuaciones más bajas se obtenían en las dimensiones de información geolocalizada e información en tiempo real.

Es por esto que se propone una adaptación de la métrica MELODA, con el propósito de evaluar el grado de reutilización de los conjuntos de datos publicados en una IDE, teniendo en cuenta la naturaleza específica de esos datos y tomando como caso de prueba la Infraestructura de Datos Espaciales de la República Argentina (IDERA).

Líneas de investigación y desarrollo

La presente es una investigación que se desprende del Proyecto: "Las TIC al Servicio del Dato Abierto: Situación actual, conceptualización e iniciativas de apertura de información Pública".

Para la metodología de trabajo se establecieron métodos comparativos y descriptivos y análisis documental y de contenidos, que abordan el estudio de los conceptos fundamentales sobre DA e IDE, su importancia y las diversas metodologías, técnicas, herramientas y propuestas que se encuentran disponibles para evaluar y diagnosticar su implementación. La revisión bibliográfica y el análisis del estado del arte en DA son la base para determinar las capacidades de los mismos en cualquier organismo y en especial en la Infraestructura de Datos Espaciales de la República Argentina (IDERA: http://www.idera.gob.ar/). Con respecto a la obtención de información se prevé recolección de muestras mediante examen de características de las administraciones públicas que ofrecen a la ciudadanía sus DA. La fuente de información se centra en la observación indirecta, a través de documentación de las tecnologías informáticas utilizadas en la administración pública nacional y provincial.
En cuanto a la métrica utilizada para el análisis de los datos se ha seleccionado a MELODA (MEtric for reLeasing Open DAta) para el análisis exploratorio de los principios de DA, y sobre la situación actual de iniciativas de DA, con el fin de que el proyecto sea de máxima aplicabilidad, se realizan pruebas de la métrica MELODA sobre conjuntos de datos de IDERA y de los portales de otras IDE regionales.

Resultados y Objetivos

Esta línea de I-D-I se encuentra en una etapa inicial. El objetivo principal planteado para la ejecución del trabajo fue: Evaluar la reusabilidad de los conjuntos de datos disponibles en las Infraestructura de Datos Espaciales, ejemplificando los pasos seguidos por medio del geoportal de la Infraestructura de Datos Espaciales de la República Argentina (IDERA).
En este sentido, se está realizando una adaptación de la métrica MELODA para la evaluación de la reusabilidad de los conjuntos de datos de los geoportales que requieren un análisis especial dado sus características y requerimientos exigidos por organismos internacionales de estandarización e interoperabilidad.

Formación de Recursos Humanos

El primer autor se desempeña como docente de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca. Mientras que el segundo autor es docente de la Facultad de Ciencias Físico-Matemáticas y Naturales de la Universidad Nacional de San Luis. La temática abordada en este proyecto de investigación en desarrollo, fue concebida a priori en el marco de investigación de la tesis para optar al título de Magister en Ingeniería de Software. El mismo se desarrolla en la Facultad de Ciencias Físico-Matemáticas y Naturales de la Universidad Nacional de San Luis, en donde el Mgter. Carlos Salgado, se desempeña como director de dicha Tesis.
También se desarrollan dentro de los trabajos de investigación becas CIN y trabajos finales de carreras de grado con relación a la temática por alumnos de la carrera de Ingeniería Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca y de la carrera Licenciatura en Computación, y trabajos finales de carrera de la Ingeniería en Informática e Ingeniería en Computación de la Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis.

Bibliografía

- Barómetro de los Datos Abiertos, 4ª Edición. Informe Global, 2017 - World Wide Web Foundation.
http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport-ES.pdf

-Conferencia Regional de Datos Abiertos para América Latina y el Caribe. *Datos* https://www.cepal.org/es/discursos/conferencia-regional-datos-abiertos-para-el-desarrollo-de-america-latina

- Albella A., Ortiz-De-Urbina-Criado, M Y De-Pablos-Heredero C. "*Meloda, métrica para evaluar la reutilización de datos abiertos*". El profesional de la información, 2014, noviembre-diciembre, v. 23, n. 6. ISSN: 1386-6710

- "meloda 40v1 – MELODA," 2016. [Online]. Available: http://www.meloda.org/meloda40v1/.

- Iniesto M., Núñez et all. (2014). "*Introducción a las Infraestructuras de Datos Espaciales"*. Edición digital. Catálogo general de publicaciones oficiales http://publicacionesoficiales.boe.es. NIPO: 162-14-022-0 - DOI: 10.7419/162.12.

- Sommerville, I. (2011) *Ingeniería de Software*. PEARSON EDUCACIÓN, México, ISBN2011:978-607-32-0 603-7 – 9na Edición

gubernamentales en formato abierto?, 2014.

- Infraestructura de Datos Espaciales de la República Argentina (IDERA) (2020) https://www.idera.gob.ar/

- Vicente-Paños, A.; Jordán-Alfonso, A. (2017). "*Acceso a la información pública y su reutilización en las comunidades autónomas: evaluación de la reutilización de datos abiertos"*. El profesional de la información, v. 26, n. 3, pp. 381-391. https://doi.org/10.3145/epi.2017.may.04

- Nueva Ley de Promoción de la Industria del Software, (2011) http://servicios.infoleg.gob.ar/infolegInternet/anexos/185000-189999/185701/norma.htm.

- CESSI, "Encuestas. Cámara de la Industria Argentina de Software" (2016) http://www.cessi.org.ar/opssi.

Ingeniería de Software al Servicio de la Informática Forense y la Evidencia Digital

Enrique Miranda[(1)], Hernán Bernardis[(1)], Daniel Riesco[(1)],

[(1)]Departamento de Informática / Facultad Ciencias Físico Matemáticas y Naturales/
Universidad Nacional de San Luis
Ejército de los Andes 950 – San Luis – Argentina
{eamiranda,hbernardis,driesco}@unsl.edu.ar

Resumen

En los últimos años la Informática Forense, ha demostrado ser una disciplina imprescindible para la aplicación de la ley. Sin embargo, actualmente enfrenta diversos desafíos que requieren del desarrollo de nuevas técnicas y herramientas que posibiliten hacer frente a los mismos. El vertiginoso avance de Internet y la proliferación de dispositivos tecnológicos de uso cotidiano se han convertido en la principal preocupación de los profesionales de la disciplina.

En este contexto de constante demanda, ha surgido una gran cantidad de técnicas y herramientas de software forenses para facilitar la tarea de los profesionales. Sin embargo, es posible distinguir un contexto de crecimiento con cierto déficit en el uso de métodos ingenieriles que faciliten y certifiquen de alguna forma el correcto desarrollo de dichos recursos. Por otra parte, los constantes desafíos requieren de profesionales con cierto grado de formación específica en la temática; aspecto que actualmente no es considerado seriamente por las instituciones de educación superior en nuestro país.

Teniendo en cuenta este contexto se propone una línea de investigación que tome como base los conceptos, técnicas y herramientas de la Ingeniería de Software para asistir a la Informática Forense y al tratamiento de Evidencia Digital.

Palabras clave: Informática Forense, Evidencia Digital, Ingeniería de Software, Herramientas Forenses, Educación.

Contexto

La presente línea de investigación se enmarca en el Proyecto de Investigación "Ingeniería de Software: Conceptos, Prácticas y Herramientas para el Desarrollo de Software con Calidad" de la Facultad de Ciencias Físico Matemáticas y Naturales de la Universidad Nacional de San Luis. Proyecto N° P-031516, continuación de diferentes proyectos de investigación a través de los cuáles se ha logrado vínculos con distintas universidades a nivel nacional e internacional. Además, el mismo se encuentra reconocido por el Programa de Incentivos de Ciencia y Técnica.

Introducción

Con el avance tecnológico actual estamos viviendo verdaderamente en una era digital. Sin duda alguna este aspecto mejora muchos quehaceres de nuestra vida diaria, al punto tal que no somos verdaderamente conscientes de la cantidad de tareas que realizamos utilizando recursos digitales. Este amplio uso de la tecnología informática en casi todos los ámbitos, sumado al avance revolucionario de las telecomunicaciones, ha favorecido la generación de un "ambiente virtual" en el que se llevan a cabo un sinnúmero de actividades y en donde se almacena un gran volumen de información relacionado a dichas actividades [1,2]. Este tipo de información no ha sido ajena en los procesos penales que se han tenido que adaptarse a este nuevo paradigma conforme fueron evolucionando las distintas conductas llevadas a cabo por medios digitales. En pocas palabras, estamos haciendo referencia a la evidencia digital (o electrónica), que se torna cada vez más esencial y se encamina a convertirse en la prueba fundamental de los procesos penales y de auditorías, posiblemente desplazando con el tiempo en gran medida a la evidencia física [1,4]. Incluso en investigaciones que no son

principalmente de naturaleza digital, en algún momento de la misma pueden surgir elementos de interés en base a archivos o datos informáticos que requieran de un análisis adicional. Es decir, en la actualidad este tipo de evidencia está comenzando a ser un elemento de prueba fundamental en la investigación de cualquier tipo de delito.

La evidencia digital posee ciertas características que la diferencian sobre otras formas de evidencia física. Por ejemplo, la evidencia digital es frágil y volátil, es decir, en ciertos casos la evidencia está almacenada en dispositivo donde cualquier acción puede provocar la alteración de la misma. Además, la evidencia digital se puede copiar *sin límites* de manera relativamente sencilla en la mayoría de los casos. Incluso, el término utilizado es aún más fuerte ya que se habla de "clonación" en vez de copiado ya que el objeto resultante de dicha operación mantiene todas las características del original [1,4]. Otro aspecto no menos relevante acerca de este tipo de evidencia es el volumen de información que normalmente debe ser procesada para la obtención de la misma. Por ejemplo, para lograr obtener una imagen determinada almacenada en un disco rígido de una CPU, es necesario procesar un sinnúmero de archivos que no poseen relevancia. Otra característica es la diversidad de dispositivos en donde se puede almacenar información que puede ser contener evidencia digital. Por ejemplo, discos rígidos, smartphones, drones, dispositivos de IoT, etc.

En este contexto se torna trascendental una disciplina denominada *Informática Forense*, la cual posibilita la detección y recuperación de la información digital que sirva de evidencia a la hora de reconstruir un hecho o sucesión de ellos. La actuación forense en informática permite recuperar y/o reconstruir rastros digitales garantizando su valor probatorio [4].

La Informática Forense requiere de profesionales que cuenten con los conocimientos y la capacitación adecuada para realizar el análisis. En este sentido, los mismos deben demostrar el rigor de las técnicas y herramientas y a su vez deben poder comunicar claramente los resultados a quien corresponda (juez, tribunal, jurado, etc.). Principalmente es por este motivo que los equipos, herramientas y técnicas forenses deben tener su correspondiente validación científica y al mismo tiempo producir un resultado que sea preciso, demostrable y reproducible [1,3].

Para asistir a los profesionales de la Informática Forense existe una amplia variedad de herramientas de software en el mercado, las cuales facilitan, entre otras cosas, la recuperación, análisis y posterior reporte de resultados sobre diversos recursos digitales. Sin embargo, existe una necesidad crítica tanto para las agencias de aplicación de la ley (comúnmente llamadas *law enforcement*) como para las corporaciones de hacer uso estrictamente de herramientas validadas con cierto rigor científico. Frecuentemente los profesionales en Informáticos Forenses hacen uso de herramientas, procedimientos y técnicas que no son accesibles o de público conocimiento para las personas ajenas a la disciplina, por lo tanto, estos no son comprendidos y/o aceptados fácilmente. Para que los hallazgos de un profesional forense sean aceptados, deben ser reconocidos por otros expertos en el campo y cumplir con los estándares de práctica nacionales e internacionales. Los investigadores forenses informáticos corren el riesgo de perder credibilidad si es posible introducir dudas sobre la idoneidad de las herramientas y/o procedimientos desarrollados durante el tratamiento de la evidencia digital presentada [2,4].

Teniendo en cuenta los aspectos relevados anteriormente relacionados con la fiabilidad y la admisibilidad de la evidencia digital, es necesario de común acuerdo entre todos los interesados dentro del proceso, responder algunas cuestiones como por ejemplo qué tipo de herramientas forense de software son las indicadas, cuándo una herramienta de este tipo puede ser utilizada y para qué clase tareas, qué criterios de calidad deben cumplir dichas herramientas, bajo qué requerimientos se debe llevar a cabo una selección, cómo se validan dichas herramientas para que sean "confiables"; por otra parte también es relevante responder a cuestiones como qué conocimientos debe tener ese profesional en Informática Forense para

hacer frente a los desafíos que presenta constantemente la disciplina.

Para dar tratamiento a los cuestionamientos planteados en el párrafo precedente, en esta línea de investigación se propone un abordaje desde la Ingeniería de Software como disciplina de base para facilitar el trabajo forense. Es decir, utilizar conceptos, técnicas, estrategias y herramientas de Ingeniería de Software para afrontar los constantes desafíos que presenta el uso de herramientas de software en Informática Forense para el tratamiento de la Evidencia Digital.

Líneas de Investigación y Desarrollo

A continuación se describe un conjunto no exhaustivo de temáticas de investigación que se enmarcan dentro de la línea principal.

Evaluación y Selección de Herramientas de Software Forense

Los planteos más frecuentes en el contexto de la actividad forense están relacionados con la confiabilidad y admisibilidad de la evidencia digital recuperada usando herramientas de software forense [2,3]. Esto ha resaltado la necesidad de que los actores del sistema de justicia respondan y justifiquen de manera precisa preguntas tales como ¿Qué hace que una herramienta de software forense sea aceptada? y ¿Cómo hacer una buena selección de herramientas de software forense? En este contexto, se considera fundamental proporcionar una comprensión más integral de los factores que los profesionales y las empresas de la disciplina consideran importantes para la selección de una herramienta de software forense adecuada [4]. En este sentido es relevante dar respuesta a la siguiente pregunta: ¿Cuáles son los criterios que debe cumplimentar una herramienta de software forense para cada tipo de investigación? La importancia de esta línea de investigación radica en ampliar la comprensión de la selección de herramientas de software forense tomando como base conceptos y prácticas inherentes a la Ingeniería de Software [5,6].

Validación y Verificación de Herramientas de Software Forense

Uno de los desafíos que enfrentan los profesionales de la Informática Forense es cómo asegurar la confiabilidad (o "solidez forense") de la evidencia digital adquirida por las herramientas utilizadas [7,8]. Actualmente las investigaciones dependen en gran medida de herramientas de software automatizadas. La confiabilidad de los resultados de dichas investigaciones está determinada principalmente por la validación y consistencia de dichas herramientas y su funcionamiento interno. Todos los actores del sistema han planteado una demanda insistente para validar y verificar las herramientas de software forense y así asegurar más la confiabilidad de la evidencia digital [7,8,9].

En esta línea de investigación se propone el estudio y especificación de métodos de verificación y validación de software aplicadas a técnicas y herramientas de Informática Forense.

Estrategias de Búsqueda y Reducción de Información Digital

Con el paso de los años los dispositivos han adquirido cada vez más capacidad de almacenamiento. A esto se le debe sumar la evolución de internet en una web más compleja donde se pueden encontrar más sitios de almacenamiento en la nube, mayor cantidad de sistemas web de información, base de datos, proliferación de gestores de contenido, etc. Por último, el avance tecnológico ha posibilitado el acceso a una gran variedad de dispositivo a cualquier usuario. De esta manera, cualquier causa a investigar puede involucrar diversos dispositivos con gran capacidad de almacenamiento de datos [4]. Por lo general, a partir del análisis forense de los mismos se extrae gran cantidad de información superflua que no está vinculada con los hechos que se investigan. Este volumen hace más ardua la tarea del profesional forense informático ya que el mismo debe concentrarse en determinar sobre qué tipo de información tiene que hacer foco [10].

En el contexto de análisis de evidencia digital, es necesario contar con técnicas de búsqueda eficiente y reducción de información para poder tratar con el volumen de datos masivo que generalmente resulta de la recuperación forense [10,11,12].

Modelado e Implementación de Procesos y Herramientas Forenses

En general, las soluciones existentes en el campo de la Informática Forense son en gran medida *ad-hoc* [5,9]. Dada la importancia de la disciplina y los desafíos y demanda constante en los últimos años, es necesario la especificación de modelos rigurosos de análisis forense y también un desarrollo de procesos y herramientas tomando en cuenta conceptos y técnicas ingenieriles [11,13]. En este sentido, dentro de las ventajas de adoptar este tipo de enfoques, se pueden mencionar: i) un proceso de análisis forense de más amplia aceptación por parte de los actores de la justicia; ii) procesos y herramientas mejores documentados; iii) herramientas más fáciles de mantener y evolucionar; iv) procesos de investigación mejores estructurados; entre otras.

Técnicas y Herramientas para nuevos Escenarios

Actualmente la Informática Forense enfrenta muchos desafíos que obligan a buscar y desarrollar nuevas técnicas de análisis de investigación. El desarrollo de dispositivos electrónicos es cada vez más estrepitoso y tiene como objetivo principal llevar los mismos al consumidor de la manera más rápida posible [4].
Dentro de este entorno, la Informática Forense se enfrenta cada vez más a desafíos innumerables, dirigidos principalmente por los nuevos escenarios y tipos de dispositivos que se presentan día a día en el mercado [9,14,15]. Es por esto que se considera fundamental el estudio e implementación de nuevas técnicas y herramientas que permitan hacer frente a este contexto cambiante.

Educación en Informática Forense y Evidencia Digital

La demanda de profesionales forenses calificados aumenta constantemente en todo el mundo y Argentina no es la excepción [3,4]. Desde la educación superior se intenta dar solución a los nuevos desafíos que se presentan en la disciplina a nivel científico. Sin embargo desde el aspecto de formación es posible observar que no existe una oferta académica consistente con la amplia demanda. Por otra parte, dentro del grupo de actores relacionados con la disciplina, existen diversas ideas sobre lo que constituye una educación forense adecuada, pero esto no se termina de plasmar en carreras de grado y posgrado; ni siquiera en asignaturas dentro de carreras informáticas. Se puede observar que la oferta de formación está principalmente vinculada a cursos dictados por personas idóneas con buena experiencia en trabajo de campo pero limitada en el ambiente científico-académico.
Por este motivo se considera fundamental abarcar el análisis de un aspecto importante dentro de la disciplina relacionado con la oferta educativa y nuevas estrategias de proliferación de las mismas [16,17]. Esto implica no sólo relevar las opciones de formación para la disciplina en el país, sino también analizar las temáticas excluyentes que debe abarcar dicha formación y finalmente, promover la mejora de la oferta académica tomando como base las instituciones nacionales de educación superior.

Resultados y Objetivos

Se consolidará un grupo de investigación, desarrollo y transferencia a la comunidad con capacidad para abordar los nuevos desafíos que presentan la Informática Forense y la Evidencia Digital. De esta manera, se plantean como principales objetivos la generación de oportunidades de formación de recursos humanos, desarrollo de oferta académica relativa a la temática, proporción de servicios a la comunidad en el área de estudio, promoción de los vínculos de cooperación con otras instituciones y actores relacionados a la disciplina, entre otros. Con respecto a las líneas

propuestas, se espera: i) estudiar y definir criterios y procesos de evaluación de herramientas de Informática Forense y tratamiento de Evidencia Digital; ii) implementar técnicas de verificación y validación de las herramientas de la disciplina; iii) investigar y unificar las propuestas de la comunidad académica y científica para resolver los distintos desafíos que se plantean en el contexto; iv) definir los criterios de diseño de software que conduzcan a una mejora de las herramientas en general que se utilizan en el entorno v) fomentar el aumento y la mejora de la oferta académica de grado y posgrado relacionado con la temática.

Formación de Recursos Humanos

Las investigaciones realizadas así como los resultados obtenidos en este trabajo contribuyen al desarrollo de trabajos finales de grado y tesis de posgrado, ya sea de doctorado o maestrías en Ingeniería de Software y desarrollo de trabajos finales de carreras informáticas de la Universidad Nacional de San Luis, en el marco del proyecto de investigación mencionado. Así mismo se pretende que en el corto plazo alumnos de grado y posgrado desarrollen pasantías de investigación y prácticas profesionales en instituciones de aplicación de la ley.

Bibliografía

[1] DI IORIO, Ana Haydée, et al. El rastro digital del delito: aspectos técnicos, legales y estratégicos de la Informática Forense. 2017.

[2] ARMILLA, Nicolás, et al. Buenas prácticas para la recolección de la evidencia digital en la Argentina. En *XXIII Congreso Argentino de Ciencias de la Computación, La Plata, 2017*.

[3] CANO, Jeimy. Computación forense descubriendo los rastros informáticos. Editorial: AlfaOmega, México, 2009.

[4] SALT, Marcos. Nuevos desafíos de la evidencia Digital. Acceso transfronterizo y técnicas de acceso remoto a datos informáticos. Editorial: Ad-Hoc, Argentina, 2017.

[5] ARMSTRONG, Colin. Developing a framework for evaluating computer forensic tools. En Evaluation in Crime Trends and justice: Trends and Methods Conference, Canberra Australia. 2003.

[6] RAGHAVAN, Sriram; RAGHAVAN, S. V. A study of forensic & analysis tools. En 8th International Workshop on Systematic Approaches to Digital Forensics Engineering (SADFE). IEEE, 2013.

[7] GUO, Yinghua; SLAY, Jill; BECKETT, Jason. Validation and verification of computer forensic software tools - Searching Function. digital investigation, 2009, vol. 6, p. S12-S22.

[8] BECKETT, Jason; SLAY, Jill. Digital forensics: Validation and verification in a dynamic work environment. En 40th Annual Hawaii International Conference on System Sciences, 2007, p. 266a-266a.

[9] WILSDON, Tom; SLAY, Jill. Digital forensics: exploring validation, verification & certification. En First International Workshop on Systematic Approaches to Digital Forensic Engineering, 2005, p. 48-55.

[10] ZOUBEK, Christian; SACK, Konstantin. Selective deletion of non-relevant data. Digital Investigation, 2017, vol. 20, p. S92-S98.

[11] BOGEN, A. Chris; DAMPIER, David A. Preparing for Large-Scale Investigations with Case Domain Modeling. En DFRWS. 2005.

[12] FREILING, Felix; GLANZMANN, Thomas; REISER, Hans P. Characterizing loss of digital evidence due to abstraction layers. Digital Investigation, 2017, vol. 20, p. S107-S115.

[13] BOGEN, A. Chris; DAMPIER, David A. Unifying computer forensics modeling approaches: a software engineering perspective. En First International Workshop on Systematic Approaches to Digital Forensic Engineering, 2005, p. 27-39.

[14] DALINS, Janis; WILSON, Campbell; CARMAN, Mark. Criminal motivation on the dark web: A categorisation model for law enforcement. Digital Investigation, 2018, vol. 24, p. 62-71.

[15] SERVIDA, Francesco; CASEY, Eoghan. IoT forensic challenges and opportunities for digital traces. Digital Investigation, 2019, vol. 28, p. S22-S29.

[16] NANCE, Kara; ARMSTRONG, Helen; ARMSTRONG, Colin. Digital forensics: Defining an education agenda. En 43rd Hawaii International Conference on System Sciences. IEEE, 2010. p. 1-10.

[17] CASEY, Eoghan. What constitutes a proper education? Digital Investigation: The International Journal of Digital Forensics & Incident Response, 2014, vol. 11, no 2, p. 79-80.

Diseño UX: Una Guía para Mejorar la Experiencia de los Usuarios haciendo un uso Eficiente de los Recursos Disponibles

Tatiana Hotimsky [1], Walter Molina [1], Carlos Salgado [2], Mario Peralta [2], Alberto Sánchez [2]

[1] Departamento de Ciencias Aplicadas;
Escuela de Gestión de Empresas y Economía;
Escuela de Ingeniería y Ciencias Ambientales
Universidad Nacional de Villa Mercedes
Las Heras 383, CP 5730, Villa Mercedes, San Luis, Argentina
email: {tatianahotimsky, waltermolina} @unvime.edu.ar

[2] Departamento de Informática de la Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950, CP 5700, San Luis, Argentina
email: {csalgado, mperalta, alfanego} @unsl.edu.ar

RESUMEN

Pensar en la experiencia que tendrá el usuario al interactuar con una pieza de software, es comenzar a entender que, en el proceso de interacción, influyen distintos factores: individuales, sociales, culturales contextuales y aquellos propios del producto en cuestión. Esto será determinante en la percepción -positiva o negativa- que tendrá el usuario de un producto, servicio y/o proceso.

Así se evidencia la necesidad de contar con alguna estrategia, técnica, modelo o método que permita organizar y/o sistematizar estas tareas a los desarrolladores de software y los profesionales de distintas disciplinas que trabajan en conjunto, de manera interdisciplinaria, en el desarrollo de interfaces de software. En este sentido, el objetivo de la presente propuesta es brindar a las empresas, organizaciones e instituciones de desarrollo de software una guía de calidad que sirva como instrumento o buenas prácticas que les permita posicionarse en un nivel altamente competitivo en el mercado actual, mediante la producción de software de calidad basado en la experiencia de usuario. Para ello se tuvo en cuenta la estructura de la Norma Internacional ISO 13407:1999, estándar de calidad e información recabada de la industria del software y servicios informáticos de la región.

Palabras Clave: Diseño de Experiencia de Usuario, Diseño Centrado en el Usuario, Necesidades de los Usuarios, Uso eficiente de Recursos UX.

CONTEXTO

Este trabajo se enmarca dentro del proyecto de investigación “Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad” del Departamento de Informática de la Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto Nº P-031516. Dicho proyecto se encuentra reconocido por el programa de Incentivos.

1. INTRODUCCIÓN

Cuando escuchamos hablar de *experiencia de usuario* (UX), nuestro pensamiento rápido evoca una persona interactuando con una pantalla, en un dispositivo electrónico [1]. Aquí es donde debemos detenernos: pensar que la UX se limita únicamente al desarrollo de software es limitar su potencial, es olvidarse que idear, diseñar y desarrollar cualquier producto o servicio pensando en el usuario, nos brinda una *ventaja competitiva* difícil de igualar.

Es decir, debemos reemplazar ese pensamiento rápido, por uno más analítico para comprender

que el diseño de *experiencia de usuario* es mucho más que sólo una técnica de diseño centrada en el usuario y aplicada al software: es el diferencial que en un contexto complejo nos ayudará a que nuestro producto/servicio tenga mayor éxito, sea más utilizado y aceptado, y brinde una *experiencia* más gratificante respecto a otros productos/servicios que no fueron ideados teniendo como centro al usuario.

Efectivamente, la experiencia de usuario es "*la impresión que un usuario tiene al relacionarse (...) con un producto o servicio*" [2]. El diseño de experiencia de usuario no es otra cosa que aplicar esta *filosofía* en la *búsqueda de soluciones inmediatas* a través de *experiencias únicas y satisfactorias* [3].

Pero, encontrar estas soluciones es a menudo una tarea compleja [4], que requiere del uso de modelos y métodos que permitan diseñar y evaluar la calidad en la experiencia de usuario.

Diseñar para usuarios es diseñar para humanos [5], tarea que consta de dos partes donde debe dividirse el esfuerzo de investigación [6]:

- La tecnología, *centrándose en el análisis de las pautas de Ingeniería de Usabilidad, análisis y definición de requerimientos funcionales y no funcionales* [7] [8] [9].
- Aspectos emocionales del usuario, *aprovechando los conocimientos humanos que complementen a los tecnológicos para un diseño de interfaces correcto, que sean capaces de ofrecer aspectos tangibles y no estresantes a sus usuarios* [10] [9].

Ahora bien, si aunamos ambas partes y comenzamos a pensarlo como un todo podríamos encontrarnos con una traducción del modelo de Maslow de las necesidades humanas: Jerarquía de las Necesidades de los Usuarios, también conocida como la Pirámide de UX (Figura 1) [5].

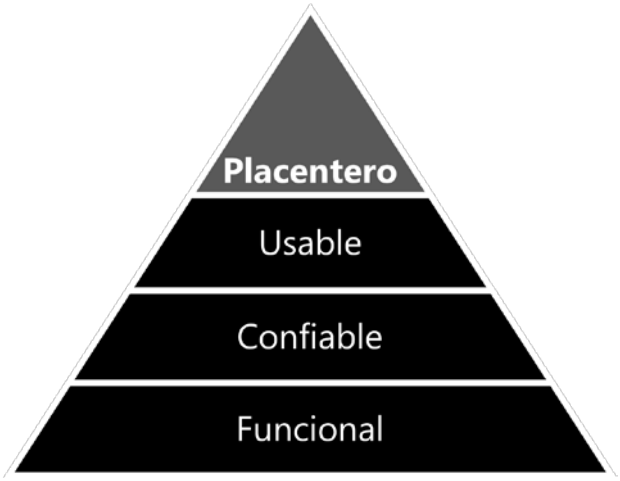


Figura 1. Pirámide de UX

Si analizamos la pirámide desde su base podemos ver que partimos de lo tangible (la tecnología) a lo subjetivo (aspectos emocionales del usuario).

Es probable que, al iniciar un desarrollo observando la Pirámide de UX, nos encontremos con cierta reticencia. Inmediatamente se pensará en los costos económicos y temporales asociados a la segmentación de los grupos de usuarios y el cumplimiento de cada etapa propuesta. Pero para disipar todas las dudas podemos plantear el siguiente interrogante:

¿Cómo podemos lograr que nuestro producto o servicio sea competitivo?

Simplemente poniendo el foco, no sólo en una experiencia de usuario memorable, sino que además es importante observar el uso eficiente de los recursos con los que se cuenta.

Atendiendo a esta reflexión se proponen los siguientes pasos para el proceso de diseño de una solución centrada en el usuario.

1. Analizar el problema y tomarnos el tiempo para preguntarnos en qué contexto será utilizado el producto y conocer a quiénes lo utilizarán. Plantear interrogantes para tener en cuenta, que nos permiten proponer ideas.

2. Conocer el contexto en profundidad, preguntarnos cuál de las propuestas que tenemos pensadas e ideadas será la más adecuada y óptima.
3. Comprender al usuario y sus características. Ponernos en sus zapatos. Hacer testeos con la opción seleccionada para desarrollo, y así verificar si la decisión fue la adecuada.
4. Conociendo el contexto y a los potenciales usuarios y habiendo realizado una prueba preliminar, ésta será una instancia para pulir las propuestas que fueron surgiendo a lo largo del proceso. Teniendo en claro que lograremos una excelente experiencia de usuario, mejorando así la calidad de nuestro producto digital.

Estos pasos no hacen más que sintetizar y dar un orden secuencial que conduzca a obtener un resultado que cumpla con lo que plantea Donald Norman en *La psicología de los objetos cotidianos* [11]*:*

> *Dicho, en otros términos, asegurar que:*
>
> *1) el usuario pueda imaginar lo que ha de hacer.*
>
> *2) el usuario pueda saber lo que está pasando.*
>
> *El diseño debe utilizar las propiedades naturales de la gente y del mundo: debe explotar las relaciones y las limitaciones naturales. En la medida de lo posible debe funcionar sin instrucciones ni etiquetas. No debería ser necesario recibir instrucción ni formación más que una vez; con cada explicación, la persona debe poder decir: «naturalmente» o «claro, ya entiendo». Bastará con una explicación sencilla si el diseño es razonable, si todo tiene su lugar y su función y si los resultados de los actos son visibles. Si la explicación lleva a la persona a pensar o decir: «¿cómo voy a recordar esto?», el diseño es malo.*

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación persigue el desarrollo de un esquema de validación de calidad para el diseño de experiencia de usuario, teniendo en cuenta los pasos propuestos en el artículo *Cambiando Percepciones: UX Design* [12]*,* los cuales están alineados con la Pirámide de UX en su totalidad, y que son descriptos en la introducción.

Para poner a prueba los pasos planteados anteriormente y en el marco de esta línea de investigación; se llevaron adelante dos Workshops sobre UX: el primero en el *XIII Encuentro Latinoamericano de Diseño de la Universidad de Palermo* y el segundo en la *Escuela Técnica N°20 "Antonio Berni", de la Ciudad de Villa Mercedes (San Luis)*, con orientación en servicios informáticos, con estudiantes del último año (con conocimiento en programación de software).

Como eje del workshop se eligió trabajar con una adaptación de la actividad propuesta en el libro Diseño de Experiencias de Usuario denominada Ciclo de Desarrollo Centrado en el Usuario, que *"introduce la idea de un proceso de diseño iterativo centrado en un grupo de usuarios y en un contexto específico*" [13] a través del diseño de una combinación de 3 platos de comida y en 4 etapas. Siguiendo lo que propone la ISO 9241-210:2019, se armaron grupos heterogéneos e interdisciplinares, con miembros con habilidades diversas [14].

3. RESULTADOS ESPERADOS/OBTENIDOS

Luego de analizar la experiencia de los Workshops se pudo observar que desarrollar aplicando los pasos propuestos y teniendo como centro al usuario nos permite hacer más eficiente el uso de los recursos, optimizar el tiempo y lograr un mayor grado de satisfacción, lo que conduce a un porcentaje elevado de fidelización.

Todo esto forma parte de nuestra *ventaja competitiva* logrando un mejor posicionamiento respecto a la elección de los usuarios de nuestro producto o servicio.

Entonces, cuando hablamos de *experiencia de usuario* no sólo hablamos de una técnica de diseño, sino que hablamos de una filosofía organizacional que nos aporta:

- Organización, simplificación de tareas.
- Control, sabemos qué desarrollar, cómo hacerlo y cómo medir los resultados obtenidos.
- Información, conocemos mejor a nuestros usuarios y podemos trabajar para mantenerlos cautivos.

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de investigación se trabaja en lo referente a mejorar la experiencia de los usuarios haciendo un uso eficiente de los recursos con los que se cuenta o se tiene disponible.

En este sentido, se está trabajando en forma conjunta entre la Universidad Nacional de San Luis y la Universidad Nacional de Villa Mercedes.

El equipo de trabajo está compuesto por 3 docentes investigadores categorizados a nivel nacional, 3 Tesistas de Posgrado y 1 becario de grado, todos pertenecientes a la carrera de Ingeniería en Informática. En este marco se están llevando a cabo 2 tesis de posgrado y algunos trabajos finales de carrera de la Ingeniería en Informática y tesinas de la Licenciatura en Ciencias de la Computación.

REFERENCIAS

[1] D. Kahneman, Pensar rápido, pensar despacio, 6 ed., Ciudad Autónoma de Buenos Aires: Debate, 2016.

[2] International Organization for Standardization, ISO 9241-210:2010 Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems, 2010.

[3] E. C. Sosa Bruchmann, C. Chayle, G. A. Montejano, A. G. Garis y L. E. Martín, «El diseño y las emociones en la Experiencia de Usuario,» de *XXI Workshop de Investigadores en Ciencias de la Computación - WICC 2019: Libro de Actas.*, N. R. Rodríguez, M. A. Murazzo, M. O. Ortega y M. I. Lund, Edits., San Juan, San Juan: Universidad Nacional de San Juan, 2019, pp. 428-432.

[4] C. Di Cicco, A. Jaszczyszyn, P. Luengo, B. Cicerchia, E. Álvarez, J. Charne, E. Goitea y M. Blanco, «Interfaz humano-máquina (diseño UX),» de *XXI Workshop de Investigadores en Ciencias de la Computación - WICC 2019: libro de actas*, N. R. Rodríguez, M. A. Murazzo, M. O. Ortega y M. I. Lund, Edits., San Juan, San Juan: Universidad Nacional de San Juan, 2019, pp. 298-301.

[5] W. Aarron, Designing for Emotion, K. Stevens, Ed., New York, New York: A Book Apart, 2011.

[6] M. Hassenzahl, "The hedonic/pragmatic model of user experience," in *Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK (HCI)*, Lancaster, 2007.

[7] I. Balmaceda Castro, C. H. Salgado, M. Peralta, A. Sánchez, M. Fernández, J. Magaquian y N. Fuentes, «Experiencia de Usuario en Plataforma virtual de Aprendizaje,» de *XXI Workshop de Investigadores en Ciencias de la Computación - WICC 2019: libro de actas*, N. R. Rodríguez, M. A. Murazzo, M. O. Ortega y M. I. Lund, Edits., San Juan, San Juan: Universidad Nacional de San Juan, 2019, pp. 444-448.

[8] J. Nielsen, Usability Engineering, Mountain View, California, 1993.

[9] M.-C. Marcos, «HCI (human computer interaction): concepto y desarrollo,» *El Profesional de la Información,* vol. 10, nº 6, pp. 4-16, Junio 2001.

[10] E. C. Sosa Bruchmann, G. A. Montejano y A. G. Garis, «Análisis de la Experiencia del Usuario: Relación entre el Comportamiento Emocional y la Satisfacción de Uso,» de *XVII Workshop de Investigadores en Ciencias de la Computación - WICC 2015: libro de actas*, 1 ed., A. De Giusti, G. D. Gil y L. E. Gilson Saravia, Edits., Salta, Salta: Universidad Nacional de Salta, 2015.

[11] D. Norman, La psicología de los objetos cotidianos, 4 ed., Madrid: Nerea, 1990.

[12] T. Hotimsky, W. Molina, C. H. Salgado, M. Peralta y A. Sánchez, «Cambiando Percepciones: UX Design,» de *XXI Workshop de Investigadores en Ciencias de la Computación - WICC 2019: libro de actas*, N. R. Rodríguez, M. A. Murazzo, M. O. Ortega y M. I. Lund, Edits., San Juan, San Juan: Universidad Nacional de San Juan, 2019, pp. 399-403.

[13] G. Allanwood and P. Beare, Diseño de Experiencias de Usuario, 1 ed., Badalona, Barcelona: Parramón Arts & Design, 2015, pp. 12-36.

[14] International Organization for Standardization, ISO 9241-210:2019 Ergonomics Of Human-System Interaction - Part 210: Human-Centred Design For Interactive Systems, 2 ed., 2019.

Heurística de Evaluación de la Experiencia de Usuario en Sistemas e-Learning

Iván Balmaceda Castro[1], Carlos Salgado[2], Mario Peralta[2], Alberto Sanchez[2]
[1]Departamento Académico de Ciencias Exactas, Físicas y Naturales
Sede Regional Chamical. Universidad Nacional de La Rioja
ibalmaceda89@gmail.com
[2]Departamento de Informática Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina
{csalgado, mperalta, alfanego}@unsl.edu.ar

RESUMEN

Los avances en las tecnologías de información, proponen un cambio en los paradigmas tradicionales de aprendizaje, permitiendo una constante evolución en la enseñanza desde lo presencial a lo virtual. El aprendizaje virtual o e-Learning (Electronic Learning), ha sido consolidado como una gran alternativa en los procesos de enseñanza-aprendizaje.

La Experiencia de Usuario (UX), resulta muy relevante a la hora de saber si el sistema cumple con lo esperado por el usuario. La misma busca guiar, controlar y mejorar las sensaciones durante el uso de estos sistemas. La presente línea de investigación, propone como enfoque principal la evaluación, desde la perspectiva de evaluadores expertos, la definición de métricas que permiten generar un modelo que sirva de instrumento de Evaluación de la UX en sistemas e-Learning, a partir de la identificación de algunos criterios que permiten cuantificar el grado de satisfacción de los usuarios.

Palabras clave: UX, Experiencia de Usuario, Métricas y Heuristica, e-Learning.

CONTEXTO

Este trabajo se enmarca dentro del proyecto de investigación "Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad" del Departamento de Informática de la Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto Nº P-031516. Dicho proyecto se encuentra reconocido por el programa de Incentivos.

1. INTRODUCCIÓN

En los últimos años, el e-Learning ha crecido rápidamente, como todo lo referente al campo de la tecnología al ofrecer nuevas formas de capacitaciones e innovaciones de aprendizaje, en términos de contenido educativo, recursos tecnológicos y posibilidades de interacción [1].
E-Learning es una modalidad de enseñanza y aprendizaje, que puede representar todo o parte del modelo educativo en el que se aplica, que explota medios y dispositivos electrónicos para facilitar el acceso, la evolución y la mejora de la calidad de la educación y la formación [2].
Estas posibilitan generar un mejor espacio de comunicación entre docentes y alumnos durante el desarrollo de un curso bajo diferentes modalidades, sea éste completamente a distancia, presencial, o semipresencial.
En estos, los estudiantes deberían poder interaccionar fácilmente con los contenidos educativos y con el entorno de aprendizaje y concentrarse en adquirir los conocimientos y competencias previstos en su formación. La relación entre los procesos epistemológicos y la informática toma en consideración el estudio de la Interacción Humano-Computadora (HCI por su sigla del inglés Human-Computer Interaction) y las teorías cognitivas en los que se basa el diseño, evaluación e implementación de interfaces interactivas [3].
Al evaluar la calidad del proceso de enseñanza-aprendizaje, la Experiencia del Usuario (UX) es de gran importancia, ya que

busca que los sistemas e-learning, sean más usables, accesibles y seguros con el fin de ofrecer a los usuarios una interfaz adecuada para lograr mayor efectividad en las tareas que se desarrollan durante la navegación [4,5].

Definimos que la investigación de la HCI y, en particular, la UX se centran en las interacciones que tienen lugar durante el ciclo de vida del usuario. De esta forma, la UX no es solo un estudio técnico de lo que realiza el usuario, sino que es necesario comprender por qué lo hace para así determinar su experiencia [6,7].

La UX desempeña el papel de identificar los sentimientos de los usuarios que pueden resultar positivos o negativos [8,9]. La norma ISO 9241 la define como "las percepciones y respuestas del usuario que resultan del uso y / o uso anticipado de un sistema, producto o servicio" [5], teniendo una visión completa de las emociones, actitudes y expectativas de una persona. El campo de estudio de la UX se centra principalmente en el diseño y la evaluación de las experiencias que tienen las personas a través del uso (o uso anticipado) de un sistema, producto o servicio. Este uso se lleva a cabo en un contexto específico, que tiene un impacto en la UX.

Por ello es que la UX, ha sido visto como algo evaluable después, antes y durante la interacción con un producto.

Las visiones que contempla experiencia del usuario, pueden clasificarse en:

- Útil: que tenga una utilidad práctica para el usuario que utiliza la aplicación o visita el sitio.
- Utilizable, usabilidad: Que sea fácil de usar, entendible y auto explicativo. Ésta es una característica necesaria pero no suficiente.
- Deseable: Se trata de conseguir con el diseño que la integración de los elementos sea equilibrada entre imágenes, identidad, marca, sonidos, animaciones, y el resto de componentes del diseño emocional.
- Encontrable: Los sitios web deben ser fácilmente navegables y permitir que, de una manera sencilla y práctica, los usuarios puedan encontrar lo que necesiten.
- Accesible: Una buena experiencia del usuario pasa por dar accesibilidad a las personas con discapacidad.
- Creíble: Existen elementos de diseño y de la organización de los contenidos que influyen en los usuarios y generan confianza y credibilidad en la web.
- Valioso: La página web debe poder ofrecer valor a los usuarios. En esta investigación no sólo se evaluará la usabilidad de la aplicación, también se busca integrar otros aspectos relacionados con la experiencia del usuario.

Los métodos de evaluación deben poder identificar los factores emocionales, funcionales y racionales que entran en juego en cada interacción y uso, en un contexto cultural específico. Las evaluaciones pueden ser formativas o aditivas, en las formativas se detectan problemas, mientras que en las aditivas se hace uso de mediciones mediante métricas.

Existen muchos métodos de evaluación de experiencia del usuario, muchos de ellos con un costo muy alto, para pensar en utilizarlos, pero otros pueden ser adaptados y utilizados. Dependiendo de lo que uno quiera evaluar, estos métodos de evaluación de UX se pueden clasificar en cuatro categorías [10]:

1) Estudios de campo: se llevan a cabo en entornos reales de uso;
2). Estudios de laboratorio: se llevan a cabo en lugares establecidos por los investigadores, en contraste con contextos de uso auténticos;
3) Estudios en línea: se realizan a través de Internet y los participantes pueden ser personas anónimas de todo el mundo;
4) Cuestionarios y encuestas.

Así mismo, los métodos de inspección basados en heurísticas [11], han permitido evaluar la forma en que el sistema interactúa con el usuario a fin de determinar cómo se realizan las tareas básicas.

La evaluación heurística es un método ampliamente adoptado de las pruebas de usabilidad, ya que es fácil de realizar y relativamente barato, pero se necesita tener expertos. Consta de una lista de preguntas que el experto o un usuario calificado responden al ir recorriendo la interfaz, sobre la base de las respuestas dadas, se califica dicha interfaz [12].

Desde esta perspectiva, analizando la situación actual en la Universidad Nacional de La Rioja (UNLaR), se pudo observar que no se realiza un proceso formal de evaluación de la UX en el Campus Virtual, dependiente de la Dirección de Educación a Distancia y Tecnología Educativa. En base a esta observación, se comenzó a trabajar en un modelo de evaluación heurística que permita evaluar la UX de sistemas e-Learning, como los sistemas de aulas virtuales como apoyo a los procesos pedagógicos. Cabe destacar, que, como un caso de estudio, y en las distintas etapas del diseño la definición del modelo, se utilizará al Campus Virtual de la UNLaR, para evaluar la UX del mismo y, de esa manera obtener una validación del modelo propuesto.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

A partir de las líneas de investigación presentadas en WICC 2019 [13], se continúa trabajando en el estudio de distintos modelos y métodos de evaluación de calidad en la experiencia de usuario.

Desarrollar un producto digital sin preguntarnos quiénes lo usarán, para qué, desde qué dispositivos, etc., no hace más que alejarnos de la premisa que reina hoy en día, en esta era digital: todo por el usuario.

A menudo en la industria del software nos encontramos con productos digitales que son ideados y desarrollados por personas expertas en el tema, a las que resulta fácil entender el uso de un *menú hamburguesa* o que conviven en lo cotidiano con términos como *scroll, input* o *form.* Pero ¿qué pasa si hablamos de usuarios que no son expertos en el tema? ¿Y si hablamos de los diferentes grupos etarios? ¿De comunidades vulnerables o que no poseen fácil acceso a la tecnología? ¿Cómo se sienten al encontrarse con un producto digital que no fue ideado o desarrollado para su situación o su contexto?

Ahora preguntémonos: ¿cómo podrían los equipos de desarrollo sacar al mercado un producto digital pensando en el usuario si generalmente trabajan contra reloj? Si, a menudo conocen al usuario en el transcurso del desarrollo, ¿podrían cubrir las modificaciones que van siendo solicitadas? A las interrogantes sumémosle que los equipos están conformados por personas con diferente visión, y que pueden pensar diversas soluciones según su experiencia ¿Cómo se podría lograr un consenso?

Y si tenemos en cuenta que, a menudo, los equipos de desarrollo deben producir software multiplataforma, impactando directamente sobre la UX por los riesgos propios de las tecnologías que se usan [12], ¿podríamos llegar a un producto digital con el que todos, equipo y usuarios, se sintieran satisfechos e identificados?

Para ello, se está trabajando en la definición de modelos, métodos, métricas, heurísticas, guías, estrategias que permitan evaluar la calidad de las interfaces con foco en las experiencias de usuario.

3. RESULTADOS ESPERADOS/OBTENIDOS

El objetivo general de la presente línea de investigación consiste en generar un instrumento de evaluación de la UX en sistemas e-Learning, que permita proporcionar un modelo útil y comprensible, estableciendo criterios para la satisfacción de los usuarios. En este contexto, se plantean los siguientes objetivos particulares:

- Investigar el estado del arte de los métodos, modelos y técnicas de evaluación de la UX en sistemas e-learning.
- Analizar en detalle los criterios, heurísticas, técnicas y herramientas

para la evaluación de experiencia de usuario.

- Desarrollar métricas de UX y usabilidad que permitan evaluar el modelo.
- Proponer un conjunto de heurísticas y/o checklist para Entornos Virtuales de Aprendizaje.
- Validación del modelo propuesto aplicado a casos de estudio de sistemas e-Learning, como, por ejemplo, el Campus Virtual de la UNLAR.

Para la definición del modelo, se tendrá en cuenta las guías de experiencias de usuario, las características sugeridas por la Norma ISO 25010.

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de investigación, se logró un trabajo conjunto entre la Universidad Nacional de San Luis (UNSL) y la Universidad Nacional de La Rioja (UNLaR) más puntualmente con su sede universitaria en Chamical.

Por parte de la UNSL, el equipo está compuesto por 4 docentes investigadores categorizados a nivel nacional, 1 tesista de posgrado de la Maestría en Ingeniería de Software, 2 tesistas de posgrado de la Maestría en Calidad de Software y 1 becario de la Ingeniería en Informática. En este marco se están llevando a cabo 3 tesis de posgrado y algunos trabajos finales de carrera de la Ingeniería en Informática y tesinas de la Licenciatura en Ciencias de la Computación.

Desde la UNLaR, la formación de recursos humanos incluye a 1 de los autores de este documento, docente de la universidad, quien es maestrando de la Maestría en Ingeniería de Software de la UNSL.

5. BIBLIOGRAFÍA

1. Rodrigues, H., Almeida, F., Figueiredo, V., Lopes, SL. Monitoring e-learning through published articles: a systematic review Computers and Education, 136, 87-98. (2019)
2. Adam, M. R., Vallés, R. S., & Rodríguez, G. I. M. E-learning: características y evaluación. Ensayos de economía, 23(43), 143-159. (2013)
3. Litwin, E. (2003). Los desafíos y los sinsentidos de las nuevas tecnologías en la educación. entrevista portal EDUC. AR.
4. Martin, G. Gaetaán, V. Saldaño, A. Pires, G. Miranda, S. Villagra, A. Carrizo, C. Cardozo, H. Sosa "Un enfoque integrador para diseñar y evaluar interfaces de usuario web" XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires), 2007
5. ISO 9241-11, Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts, International Organization for Standardization, 2018.
6. E. Sosa Bruchmann, G. Montejano, A. Garis "Análisis de la experiencia del usuario: relación entre el comportamiento emocional y la satisfacción de uso" XVII Workshop de Investigadores en Ciencias de la Computación, 2015.
7. Carreras, J. (2003). "Midiendo la experiencia de usuario".
8. Montuwy, A.; Cahour, B.; Dommes, A. Using Sensory Wearable Devices to Navigate the City: Effectiveness and User Experience in Older Pedestrians. Multimodal Technologies and Interaction 3, 17. doi:10.3390/mti3010017.
9. Hu, K.; Gui, Z.; Cheng, X.;Wu, H.; McClure, S. The Concept and Technologies of Quality of Geographic Information Service: Improving User Experience of GIServices in a Distributed Computing Environment. ISPRS International Journal of Geo-Information 2019, 8, 118. (2019)
10. All About UX, Information for user experience professionals. [Online]. Available: http://www.allaboutux.org/
11. A. Ferreira, C. Saenz "Usabilidad de los entornos virtuales de enseñanza y aprendizaje" VI Congreso de Tecnología

en Educación y Educación en Tecnología, 2011.
12. DAdamo, M. H., Baum, A., Luna, D., & Argibay, P. (2011). Interacción ser humano-computadora: usabilidad y universalidad en la era de la información. Rev. Hosp. Ital. B. Aires Vol, 31(4).
13. Balmaceda Castro, I., Salgado, C. H., Peralta, M., Sánchez, A., Fernández, M., Magaquian, J., & Fuentes, N. "Experiencia de usuario en plataforma virtual de aprendizaje." XXI Workshop de Investigadores en Ciencias de la Computación, San Juan 2019.

Fiabilidad en la Calidad del Software: Modelos, Métodos y Estrategias

Walter A. Lucero, Carlos Salgado, Alberto Sánchez, Mario Peralta
Departamento de Informática Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
Ejercito de los Andes 950 – C.P. 5700
San Luis - Argentina
walteradrianlucero@gmail.com - {csalgado, alfanego, mperalta}@unsl.edu.ar

RESUMEN

El software ha cobrado, en la mayoría de nuestros sistemas, un papel muy importante. El interés por la calidad del mismo ha aumentado de manera significativa puesto que uno de los aspectos más importante de la calidad del software es la ***fiabilidad***. Un aspecto particular de la calidad del software que ha recibido gran atención, es la modelización de la fiabilidad del software, que tiene por objeto dar una descripción precisa en términos probabilísticos de la fiabilidad de dichos sistemas en base a hipótesis sobre los factores que pueden afectarle y a datos empíricos. En este sentido, los modelos de calidad son referencias que las organizaciones utilizan para mejorar su gestión. Los modelos, a diferencia de las normas, no contienen requisitos que deben cumplir los sistemas de gestión de la calidad sino directrices para la mejora. La aplicación de modelos de calidad y métodos de evaluación busca aportar a las empresas una visión general de su situación actual, como de cada uno de los aspectos principales que hacen a su desarrollo, elementos clave para su desempeño. En este sentido, se ha propuesto un modelo de Gestión de Calidad en empresas SSI y un método de evaluación para la medición de la satisfacción de dicho modelo de calidad en las empresas del medio, y su utilización para evaluar la ***fiabilidad*** de los productos software de dichas empresas.

Palabras claves: Modelos, Métodos de Evaluación, Fiabilidad. Software de Calidad.

CONTEXTO

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto Nº P-031516. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el programa de incentivos.

1. INTRODUCCIÓN

En el ámbito de la gestión, diariamente se toman decisiones que pueden afectar a la economía de todo el país, el futuro de la organización/institución, el impacto en el mercado, etc. ¿Cómo se toman estas decisiones? ¿Son decisiones buenas o malas? La meta de una evaluación es informar sobre los efectos de las reglas de negocio y políticas actuales y potenciales. Existen varios métodos de evaluación con diferentes niveles de fiabilidad. La calidad de la evaluación es de suma importancia para poder entregar resultados correctos. Así, los métodos pueden seguir metodologías cuantitativas o cualitativas. En este sentido, algunos ejemplos de métodos que se pueden mencionar son: Matching y PSM (Propensity Score Matching) [1], Regresión Discontinua [2], Diferencias en Diferencias (DIFF-IN-DIFF) [3], Evaluación Aleatoria [4, 5], entre otros.

Desde el punto de vista de la calidad, en [6], se presenta un método para la evaluación de modelos de Procesos de Negocio (PN) sin importar su representación. Dicho método permite evaluar las principales características de calidad que se considera que todo modelo de PN debe satisfacer.

En base a estas propuestas, surge la necesidad de contar con un método de evaluación que permita la instanciación de un modelo de Gestión de Calidad en empresas SSI. Es decir, poder evaluar el grado en que una empresa u organización cumple con un modelo de calidad. El método propuesto tiene sus bases en modelos matemáticos y estadísticos que permiten que los procesos sean sistemáticos y repetibles. Para definirlo, se tuvieron en cuenta diversas herramientas, técnicas, modelos y métodos. Se parte eligiendo un modelo de gestión de calidad, en particular se tomó como punto de partida el modelo definido en [7]. Cabe aclarar que dicho modelo, se plantea como inicio para el método, pero no es de ninguna manera estático, es decir, se define de manera que sea aplicable a otros modelos de calidad. Se puede ajustar a las distintas situaciones, tecnologías o reglas de negocio/mercado a medida que vayan surgiendo.

El objetivo del método es poder validar/verificar si el modelo de la empresa se ajusta, y en qué medida, al modelo de gestión de la calidad. En base a dicho modelo de gestión, y usando el método Delphi [8], se definen las preguntas de la encuesta que va a servir para la recolección de los datos de cómo cada empresa estudiada se acerca o aleja del modelo de la gestión de la calidad. Para la ponderación de las preguntas que conforman la encuesta se utiliza una escala LIKERT [9]. Como paso siguiente, se procede a la toma de la encuesta sobre la población motivo de estudio. Con las respuestas de la encuesta se procede a evaluar la fiabilidad y validez del modelo en las empresas. Para ello, se utiliza el método de Cronbach [10] y el análisis factorial [11], ambos métodos ampliamente probados en distintos ámbitos. Por último, en base a estos resultados se procede a hacer un reporte de sugerencias y recomendaciones. Ese reporte, permite a la empresa acondicionar sus procesos/productos/servicios y luego, como un paso hacia la mejora continua, volver a evaluar la fiabilidad o validez del modelo de gestión.

Para nuestro caso en particular nos vamos a centrar en la confiabilidad, según la cual, de manera informal, se define que el software es confiable si el usuario puede tenerle confianza. Formalmente [12], la confiabilidad se define en términos del comportamiento estadístico: la probabilidad de que el software opere como es esperado en un intervalo de tiempo especificado. A diferencia de la corrección, la confiabilidad es una cualidad relativa. Cualquier desviación de los requerimientos hace que el sistema sea incorrecto, por otro lado, si la consecuencia de un error en el software no es seria, el software aún puede ser confiable.

En la práctica, como la especificación es un modelo de lo que quiere el usuario que puede ser o no adecuado para sus necesidades y requerimientos reales, lo máximo que puede hacer el software es cumplir los requerimientos especificados del modelo, sin asegurar la adecuación del mismo. Pueden tenerse aplicaciones correctas diseñadas para requerimientos "incorrectos", por lo que la corrección del software puede no ser suficiente para garantizar al usuario que el software se comporta como se espera.

La confiabilidad es un estado que comprende varios conceptos, como fiabilidad, disponibilidad, mantenimiento y seguridad. Por ello, no se mide directamente, sino por medio de sus atributos. Así, una medida de confiabilidad importante es la tasa de fallos (i.e. el número de fallos por unidad de tiempo) que sirve para evaluar la frecuencia de fallos de un sistema tal y como es percibida por el usuario. En hardware, la fluctuación normal de la tasa de fallos suele ser la conocida curva de la bañera (Fig. 1) [13].

Sin embargo, el comportamiento de la tasa de fallos para el software es diferente, ya que cuando el sistema se encuentra disponible para el usuario se puede considerar que la tasa de fallos permanece constante en el tiempo. Ahora bien, durante el periodo de desarrollo de un sistema de software este comportamiento no es así, ya que la tasa fallos no es constante en el tiempo. De hecho, durante la fase de pruebas la tasa de fallos debería ser decreciente y

aproximadamente constante durante su fase de producción (Fig. 2) [13].

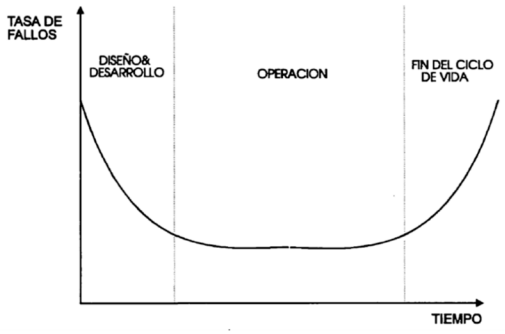


Fig.1. Tasa de fallo en Hardware.

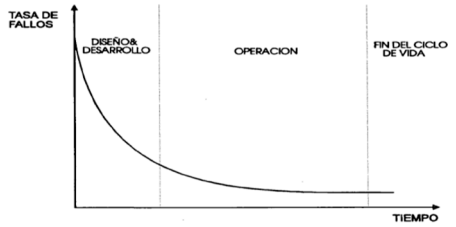


Fig.2. Posibles tasas de fallo en Software.

Así, las principales razones de estudio de la confiabilidad de sistemas de software son:

1. Determinar el tiempo "tp" hasta el cual se espera que falle una proporción "p" dada de los sistemas en operación.
2. Encontrar el tiempo "tp" al cual se espera que sobreviva una proporción "1-p" dada de los productos en operación. Es una estimación de la confiabilidad de los productos.
3. Determinar la propensión a fallar que tienen el producto en un tiempo dado. Para comparar dos o más diseños o procesos, o lo que se publicita por un proveedor.
4. Dado que un artículo ha sobrevivido un tiempo "T0", encontrar la probabilidad de que sobreviva un tiempo "t" adicional. Para planear el reemplazo de los equipos.
5. Los puntos anteriores se pueden hacer de manera comparativa para diferentes materiales, proveedores o modos de falla.

Como se mencionó, la confiabilidad se mide a través de sus atributos, siendo uno de ellos la ***fiabilidad***. En particular, la fiabilidad de un software está vinculada a la tasa de fallos del mismo. La necesidad de tener conocimiento de este tipo de medidas, se enmarca dentro del objetivo de cuantificar la confiabilidad de un sistema software, que hace referencia a la "calidad del servicio prestado" de forma que se pueda confiar en su servicio.

Con lo visto hasta ahora, el concepto más simple de ***fiabilidad***, es aquel que comprueba que el producto cumple ciertas especificaciones, y cuando esto ocurre, es enviado al cliente. El cliente por su parte acepta que el producto pueda fallar con el tiempo, y en algunos casos el período de garantía es una forma de prever esta posibilidad a corto plazo. Evidentemente fallos continuados, incluso durante el período de garantía, producen altos costos, tanto al proveedor como al comprador.

Todo esto conduce a la necesidad de considerar un control de calidad basado en el tiempo. El control de calidad habitual, o de inspección, no tiene continuidad temporal: el producto pasa un control o no lo pasa (testing). Pero nada garantiza que vaya a fallar pasado un cierto tiempo. El estudio de fallos de los productos en el dominio del tiempo es el campo de la ***fiabilidad***, que así definida, está relacionada con fallos durante la vida del producto. La definición de ***fiabilidad*** mediante conceptos probabilísticos, indica que cualquier intento de cuantificación pasa por la utilización de técnicas estadísticas que, según el problema, pueden llegar a ser muy sofisticadas.

En resumen, se puede definir que el problema fundamental en ***fiabilidad***, es estimar la vida de un sistema software y la probabilidad de que se produzca un fallo en cada momento. Así, La ***fiabilidad*** de un sistema se estudia para poder determinar si el sistema va a realizar su tarea cuando así se requiere, o fallará.

La evaluación de la fiabilidad de un sistema lógico o software se realiza mediante distribuciones de probabilidad, lo que se explica intuitivamente al considerar cómo fallan elementos de un mismo tipo, que han sido fabricados y que trabajan en las mismas condiciones. El tiempo de funcionamiento correcto es específico para cada uno, es decir,

que todos los elementos no fallan después del mismo tiempo de operación. Si se registran los tiempos hasta el fallo de cada uno de los elementos observados, se obtendrá la distribución de probabilidad de fallo de este tipo de elementos cuando se fabrican y operan en las condiciones definidas. En consecuencia, el tiempo hasta el fallo T es una variable aleatoria.

La falla de un sistema de software sucede cuando deja de operar, funcionar o no realiza satisfactoriamente la función para la que fue creado. El tiempo de falla es el tiempo que transcurre hasta que el producto deja de funcionar.

En función de todo lo explicado, dentro de la confiabilidad, la fiabilidad es uno de los conceptos más importantes a tener en cuenta a la hora de evaluar la calidad de un producto software. Por ello, y considerando las características del método propuesto, se lo aplicará para evaluar en qué medida los modelos de negocio de las empresas SSI, se ajustan a la característica.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En los últimos años se evidenció un significativo incremento en la cantidad de empresas del sector de Software y Servicios Informáticos (SSI) en la República Argentina, según datos oficiales, actualizados por el Observatorio de Empleo y Dinámica Empresarial del Ministerio de Trabajo, el número de empresas creció un 132% en 10 años consecutivos [14]. Por ello, para lograr la captación de nuevos clientes, las empresas SSI se ven en la necesidad de diferenciarse de la competencia, como así también mantener el crecimiento al ritmo de los avances tecnológicos y los requerimientos del mercado.

Muchas empresas del sector no se encuentran aplicando un modelo de gestión integrado que esté orientado a los clientes, el liderazgo, los procesos, los resultados y la mejora continua lo cual las aleja de la excelencia y, por ende, son poco productivas. Si bien un importante porcentaje de empresas de software, alrededor del 60% según los últimos resultados publicados [6], poseen algún tipo de certificación, no cuentan con un modelo a seguir para su éxito a largo plazo. Así, se ha propuesto un modelo de Gestión de la Calidad para el sector Software y Servicios Informáticos basado en el enfoque de Gestión por Procesos orientado a los clientes, el liderazgo directivo, el personal, los resultados y la mejora continua. Además, se diseñó un método basado en Técnicas Estadísticas como es el Alfa de Cronbach, que valide el Modelo de Gestión de la Calidad propuesto.

Se propone, continuando en esta línea de investigación, la aplicación del método propuesto en la evaluación de las distintas características de calidad que el modelo de referencia para la gestión de la Calidad en las empresas SSI propone. Esto tendrá como mayor beneficio la mejora continua en los procesos de las empresas, lo cual impactará directamente en la satisfacción de las partes interesadas, mayormente en los clientes.

3. RESULTADOS OBTENIDOS/ESPERADOS

La ***fiabilidad del software*** es un concepto importante en la calidad del software. Llevar a cabo su estudio debe ser abordado en tres partes, modelado, medición y mejora.

El modelado de la ***fiabilidad*** del software ha ido madurando durante los últimos años hasta el punto de que se puede obtener información muy útil a la hora de desarrollar una aplicación informática, mediante la aplicación de un modelo determinado a cada problema.

Son muchos los modelos que pueden ser empleados en el análisis de la ***fiabilidad*** de una aplicación informática, el principal problema radica en que no se puede encontrar un modelo que funcione bien en cualquier situación. Es tarea de los ingenieros adaptar un método o componer uno nuevo a partir de los ya existentes para conseguir una solución que se amolde al problema en particular que se está tratando.

La medición de la ***fiabilidad*** del software es

motivo de consenso ya que, a diferencia de otras ramas de la ingeniería, no existe una determinación convergente en el tema.

La mejora es el tercer ítem a tener presente, es arduo, ya que todavía existe un vacío importante en la materia.

Por todo ello, se espera, introducir nuevos aportes en estos tres ítems, en particular, en lo que respecta a la mejora continua, ya que, al poder establecer si el modelo de negocio de una empresa se ajusta o no a un modelo de calidad, y poder establecer los puntos más débiles en este sentido, ayudará a dicha empresa a buscar los medios y herramientas para adecuar su modelo de negocio de manera que el mismo satisfaga todos los requisitos de calidad, lo que redundará en una mayor calidad en sus productos y, por ende, en la satisfacción del cliente.

4. FORMACIÓN DE RECURSOS HUMANOS

Bajo esta línea, en el grupo de investigación, se han desarrollado dos tesis de maestría en Calidad de Software y una tesis en maestría en Ingeniería de Software. Desde el punto de vista de la fiabilidad/confiabilidad, en una de ella se definió un modelo de calidad para las empresas del sector SSI basado en un enfoque de gestión por procesos orientados a los clientes, el liderazgo directivo, el personal, los resultados y la mejora continua. La otra correspondió a la definición de un Método de Evaluación para el Modelo de Gestión de Calidad en Empresas de SSI. También se están llevando a cabo trabajos de grado con relación a la temática por alumnos de la carrera de Ingeniería Informática y Licenciatura en Ciencias de la Computación en la Universidad Nacional de San Luis.

5. BIBLIOGRAFÍA

[1] P. R. Rosenbaum, D. B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika, 70, 1,* pp. 41-55, 1983.

[2] G. Imbens, T. Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*. 2008.

[3] A. Abadie, "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies, Vol. 72, Issue 1,* 2005.

[4] E. Duflo, et al., "Using Randomization in Development Economics Research: A Toolkit," *Handbook of Development Economics,* pp. 3895-3962, 2007.

[5] J. Ludwig, J. Kling, S. Mullainathan, "Mechanism Experiments and Policy Evaluations," *Journal of Economic Perspectives,* 2011.

[6] N. Debnath, et al. "MEBPCM: A Method for Evaluating Business Process Conceptual Models. A Study Case.," presented at the 9th ITNG, Las Vegas, Nevada, USA, 2012.

[7] M. C. Gette, A. Sánchez, and C. Salgado, *Modelo de Gestión de la Calidad para Empresas SSI de Argentina*. Editorial Académica Española, 2017.

[8] J. Landeta, *El método Delphi. Una técnica de previsión para la incertidumbre.* . Barcelona: Ariel, 1999.

[9] G. J. Posadas Fernandez, "Elementos básicos de estadística descriptiva". 2016.

[10] J. A. González Alonso, M. Pazmiño, "Cálculo e interpretación del Alfa de Cronbach para el caso de validación de la consistencia interna de un cuestionario, con dos posibles escalas tipo Likert," *Publicando,* vol. 2(1), pp. 62-77, 2015.

[11] P. Morales, "El análisis factorial en la construcción e interpretación de tests, escalas & cuestionarios. ," *Madrid. http://web.upcomillas.es/personal/peter/investigacion/AnalisisFactorial.pdf* 2013.

[12] C. Ghezzi, et al. "Chapter 1 – Software: its nature and qualities.," in *Fundamentals of Software Engineering*: Prentice-Hall, 1991.

[13] R. Pressman, *Ingenieria del Software*, 6 ed. 2005.

[14] *Cámara de la Industria Argentina de Software*. Available: http://www.cessi.org.ar/opssi

LA EXPERIENCIA DE USUARIO EN EL CONTEXTO DE LOS MULTI-DISPOSITIVOS

Adriana MARTIN, Gabriela GAETAN, Viviana SALDAÑO, Claudia CARDOZO, Alejandra CARRIZO, Hernán SOSA, Rosalina PIÑERO, Silvia VILLAGRA

Grupo de Investigación y Formación en Ingeniería de Software (GIFIS)
Instituto de Tecnología Aplicada (ITA)
Universidad Nacional de la Patagonia Austral, Unidad Académica Caleta Olivia (UNPA-UACO)

adrianaelba.martin@gmail; *gabrielagaetan@yahoo.com.ar*; *vsaldanio@gmail.com*; *claudia_yoryi@yahoo.com.ar*; *acarrizo@uaco.unpa.edu.ar*; *hassio_09@hotmail.com*; *pinerorosy@gmail.com*; *svillagra@uaco.unpa.edu.ar*.

RESUMEN

A lo largo del día, las personas consumen contenido desde múltiples dispositivos. Estos consumidores han impuesto el desafío de desarrollar productos que les ofrezcan experiencias adaptadas a sus necesidades y preferencias. El desafío consiste en proporcionar desde el dispositivo elegido por el usuario, el contenido correcto, en el momento correcto. En este contexto, el diseño de productos multi-dispositivos es más que simplemente cambiar el tamaño del contenido para mostrarlo en diferentes pantallas.

El Proyecto de Investigación (PI) que se presenta en este trabajo, estuvo dirigido a desarrollar productos Web, aplicando y validando propuestas integradoras de técnicas y herramientas conceptuales y prácticas basadas en la eXperiencia de Usuario (UX), a los efectos de satisfacer las necesidades y preferencias de diferentes grupos de usuarios en el diseño y desarrollo de productos multi-dispositivos.

Palabras clave: *Experiencia de Usuario (UX)* | *Diseño y Evaluación* | *Grupos de Usuarios* | *Experiencia Multi-Dispositivos* | *Usabilidad* | *Accesibilidad Web.*

CONTEXTO

Nuestro Grupo de Investigación y Formación en Ingeniería de Software (GIFIS) perteneciente al Instituto de Tecnología Aplicada (ITA), Universidad Nacional de la Patagonia Austral (UNPA), Unidad Académica Caleta Olivia (UACO), ha ejecutado en 2019 el segundo y último año del Proyecto de Investigación (PI) Nº29/B222, Período: 2018-2020, denominado: "*Diseño y Evaluación de Experiencia de Usuario para Multi-Dispositivos*", dirigido por la Dra. Martín y codirigido por la Mg. Gaetán. Para desarrollar la problemática propuesta por este PI, GIFIS cuenta con sólidos antecedentes recabados a partir de la ejecución de los siguientes proyectos: PI Nº29/B194 (2016-2018), denominado: "*Un Enfoque Integrador para Diseñar y Evaluar Interfaces de Usuario Web*"; PI Nº29/B167 (2014-2016), denominado: "*Identificación, Desarrollo y Uso de Soluciones Web Centradas en el Usuario*" y, PI Nº29/B144 (2012-2014), denominado: "*Diseño y Evaluación de Portales Web*".

En este contexto, en el PI Nº 29/B222, ha dirigido sus esfuerzos en estos 2 años de ejecución, a desarrollar productos multi-dispositivos aplicando técnicas y herramientas basadas en la experiencia del usuario (UX), que satisfagan los requerimientos de usuarios de interés.

1. INTRODUCCIÓN

La expectativa mínima de los usuarios cuando deciden navegar en la Web desde algún tipo de dispositivo móvil, es que los sitios y aplicaciones a las que acceden, sean capaces de brindarles al menos una experiencia uniforme y sin problemas. Sin embargo, este requerimiento que parece tan básico, conlleva a algo para nada elemental que es conocer las necesidades y preferencias de nuestros usuarios para poder satisfacer sus expectativas.

Jakob Nielsen, una de las personas más respetadas en el ámbito mundial de la Usabilidad, nos enfrenta a un verdadero dilema cuando pronuncia su ley referida a la experiencia del usuario de Internet [1]: "*Los usuarios pasan la mayor parte de su tiempo en otros sitios…"* Muchas son las interpretaciones que se pueden hacer al respecto de este enunciado, pero sin lugar a dudas, y sin entrar a analizar en profundidad todas las implicancias de esta ley, no existen dudas de que este enunciado es un llamado a conocer a nuestros usuarios. Identificar las expectativas y patrones de necesidades y preferencias con los cuales están familiarizados los destinarios de nuestros productos Web, representa un factor clave al diseño UX.

Tal como ya señalamos, GIFIS ha estado trabajando fuertemente la problemática UX [5] en el marco de varios PI-UNPA y desde diferentes perspectivas, tales como la Usabilidad [2], Accesibilidad Web [3], Diseño Centrado en el Usuario [4] y hasta desde el desarrollo aplicando Metodologías Ágiles [6]. Este trabajo previo, constituye un sólido antecedente y motivación para desarrollar la problemáticas UX en el contexto de los multi-dispositivos. A continuación, se presentan los Objetivos y las Actividades (asociados a estos Objetivos) que definieron los 2 años de ejecución del PI 29/B222:

Objetivo General: "*Desarrollar productos multi-dispositivos aplicando técnicas y herramientas basadas en la experiencia del usuario (UX), que satisfagan los requerimientos de grupos de usuarios de interés.*"

Objetivos Específicos (para usuarios de interés):
Objetivo Específico 1. Seleccionar técnicas y herramientas para desarrollar productos multi-dispositivos.
Objetivo Específico 2. Definir requerimientos de experiencias multi-dispositivos.
Objetivo Específico 3. Diseñar experiencias multi-dispositivos.
Objetivo Específico 4. Validar experiencias multi-dispositivos.

Para alcanzar el Objetivo General, el PI Nº29/B222 ha ejecutado un Plan de 4 Actividades. Cada Actividad se corresponde con un Objetivo Específico y se compone de un conjunto de Tareas a ejecutar:

Actividad 1 → Objetivo Específico 1
Tarea1.1 Evaluar patrones y estrategias de diseño multi-dispositivo.
Tarea1.2 Analizar y seleccionar herramientas de conceptualización y visualización de UX.
Tarea1.3 Evaluar y seleccionar herramientas para prototipado y evaluación de UX.
Tarea1.4 Seleccionar criterios para establecer una estrategia de diseño.
Entregables: Guía de técnicas y herramientas recomendadas.

Actividad 2 → Objetivo Específico 2
Tarea2.1 Analizar y entender las necesidades de usuarios de interés.
Tarea2.2 Identificar patrones de comportamiento y tendencias de uso de usuarios de interés.
Tarea2.3 Identificar sinergias derivadas del uso combinado de varios dispositivos.
Entregables: Fichas de persona, Historias de usuario, Escenarios, Resultados de evaluación de Accesibilidad y Usabilidad.

Actividad 3 → Objetivo Específico 3
Tarea3.1 Analizar los diferentes modos de interacción disponibles en los dispositivos para los que se diseñarán las experiencias de usuario.
Tarea3.2 Desarrollar el diseño de IU a través de prototipos de baja y alta fidelidad.
Tarea3.3 Definir la estructura de navegación del producto.
Entregables: Prototipos, Mapas de sitio, Esquemas de página, Arquitectura de Información, Estrategia de contenidos

Actividad 4 → Objetivo Específico 4
Tarea4.1 Revisión de prototipos con usuarios de interés.
Tarea4.2 Evaluación de resultados e incorporación de mejoras.
Entregables: Informe de evaluación de UX.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Durante 2019, GIFIS ha estado trabajando en concretar el Objetivo general a través de sus Objetivos específicos/ Actividades (ver Sección

1: Introducción), planteados para el PI 29/B222. A continuación, se describen las Líneas de Investigación (LI) basadas en grupos de usuarios de interés, dado que son los usuarios el centro para la ejecución del Plan de Actividades:

LI.1: Usuarios Adultos Mayores
Entre 2013-2018, GIFIS estuvo vinculado a UPAMI y/o directamente a cargo de la elaboración y dictado de los cursos de capacitación a los abuelos de la región y, en este marco de trabajo, y durante 2018 se hicieron numerosos aportes [7][12][15]. En 2019, y enfocada en el uso de Facebook desde dispositivos del tipo *tablet*, la línea presenta un diseño UX concreto destinado a los usuarios Adultos Mayores (AM).

LI.2 Usuarios del Dominio Universitario
Entre 2013-2018, GIFIS ha estado trabajando sobre diferentes aspectos referidos al diseño UX de productos Web universitarios en Argentina, y durante 2018 se hicieron los siguientes aportes [8][11][14]. En 2019, y con bases en este trabajo previo, la línea refuerza su objetivo enfocándose en asistir con estrategias a la creación de contenidos para aplicaciones móviles, que satisfagan las necesidades de esta comunidad.

LI.3 Usuarios de Gobierno Electrónico
Entre 2017-2018, GIFIS se ha enfocado en los sitios Web gubernamentales, a los efectos de analizar la UX de los ciudadanos con las aplicaciones de gobierno móvil en Argentina [13]. En 2019, la línea concreta estudios de Usabilidad/ UX en gobierno móvil y tiendas de aplicaciones (*apps*).

LI.4 Usuarios del Ámbito de la Salud
En Agosto de 2019, GIFIS ha comenzado las gestiones para trabajar en un proyecto nuevo de desarrollo para el Hospital Zonal de la ciudad de Caleta Olivia. El proyecto consiste en proveer a esta entidad de una aplicación SAAC (Sistema Aumentativo y Alternativo de Comunicación), para dispositivo *tablet*. La misma deberá brindar soporte específico al proceso de comunicación entre los pacientes (personas que tienen gravemente limitado el lenguaje oral), y los agentes de salud, en las áreas de guardia e internación.

3. RESULTADOS OBTENIDOS/ ESPERADOS

En lo que se refiere a resultados obtenidos, y tan sólo considerando el período 2018-2019, podemos citar las siguientes contribuciones [7][8][9][10][11][12][13][14][15][16][17][18] [19][20][21][22], entre otras.

3.1. Resultados Esperados

Alineados al Objetivo general/ Objetivos específicos y las Líneas de Investigación (Secciones 1. y 2., respectivamente), se especificó alcanzar los siguientes resultados a la finalización del PI Nº29/B222:

- Enfoques que permitan el diseño y la evaluación de la UX y,
- Propuestas móviles concretas que permitan validar estos enfoques, al demostrar mejoras en la UX de usuarios de interés.

En este punto, cabe señalar, que habiendo finalizado en 2019 el PI Nº29/B222, GIFIS ya tiene aprobado para su ejecución el nuevo PI 29/B256, Período: 2020-2022, denominado: "*Contextos Digitales para Asistencia de los Ciudadanos: Enfoques de Experiencia de Usuario*".

3.2. Resultados Obtenidos

Durante 2019, segundo y último año de ejecución del PI 29/B222, las contribuciones científicas-técnicas por Línea de Investigación, son las siguientes:

En la LI.1: Usuarios Adultos Mayores, a partir de un profundo análisis de Facebook, se concreta el desarrollo de una propuesta completa de prototipo de red social destinada a mejorar la UX de los Adultos Mayores (AM) desde dispositivo del tipo *Tablet* [18][22].

En la LI.2: Usuarios del Dominio Universitario, con base en la estrategia de ecosistemas de contenido, se propone una guía paso a paso para redefinir contenido Web. El prototipo resultante, demuestra mejorar la UX de los usuarios, ya que resuelve los problemas identificados en el sitio Web universitario original.[16].

En la LI.3: Usuarios de Gobierno Electrónico, se concreta un estudio de Usabilidad sobre un

número relevante de aplicaciones de gobierno electrónico en Argentina [21] y, se ofrece un completo análisis de UX basado en comentarios de tiendas de aplicaciones (*apps*) [19].

En la LI.4 Usuarios del Ámbito de la Salud: se está explotando la trayectoria en Accesibilidad Web puesta de manifiesto por los integrantes de GIFIS; sin ninguna duda, esta nueva línea proveerá valiosos aportes al área y la UX de sus usuarios.

Además, los integrantes del GIFIS estuvieron presentes con contribuciones C-T en los siguientes eventos 2019:

- XXI WICC: se envió 1 artículo [20] y su respectivo poster.
- 48 JAIIO - Simposio de Informática en el Estado (SIE): se envió 1 artículo [21] y se viajó a exponerlo.

Otros resultados 2019 vinculados el PI 29/B222, son la ejecución de las siguientes actividades de capacitación/ extensión:

- Proyecto Especial UACO 2018-2019: "Entrenamiento a Usuarios Adultos Mayores en el Uso de Aplicaciones Móviles" Director y Ejecutor: Mg. Gabriela Gaetán. Disposición Nro: 272/18. Finalizado: 30-06-2019.
- Curso de Capacitación - XV Escuela de Informática UNPA-UACO: "Evaluación de Diseños de Experiencia de Usuario" Capacitadores: Mg. Viviana Saldaño, Mg. Gabriela Gaetán. Destinatarios: alumnos y profesionales interesados en diseño de UX. Duración: 15hs.
- Curso de Capacitación Inter-Institucional: "Manejo y Organización Digital de los Datos: Módulo Bases de Datos" Capacitador: Mg. Viviana Saldaño. Destinatarios: Bomberos de la Policía de la Provincia de Santa Cruz. Duración: 20hs.

4. FORMACIÓN DE RECURSOS HUMANOS

Es importante señalar que GIFIS está enfocado en brindar el marco adecuado para que cada integrante alcance sus objetivos particulares de investigación y formación al contribuir con el objetivo del PI en curso.

Durante 2019, los integrantes (actuales y próximos del equipo de trabajo 2020-2022):

- Mg. Claudia Cardozo defendió su Tesis MIS-UNPA, en Agosto de 2019 [22] y continuará como integrante externo en el nuevo PI de GIFIS.
- Docente/ investigador Alejandra Carrizo entregó el ICT-UNPA (requisito de la tutoría MIS-UNPA), el cual fue aprobado por la revisión de pares para su publicación (en prensa) [19], y continuará en el nuevo PI de GIFIS.
- Alumna de postgrado Silvia Villagra, retomó la MIS-UNPA, y continuará en el nuevo PI de GIFIS.
- Alumno de grado Hernán Sosa inicia la ejecución de su Proyecto Final de Carrera de la IenS-UNPA, y se incorporá al nuevo PI de GIFIS.
- Alumna de pregrado Rosalina Piñero, defendió su Proyecto TUDW-UNPA, y se incorporá al nuevo PI de GIFIS.

5. AGRADECIMIENTOS

A la UNPA por dar soporte para la concreción del PI Nº: 29/B222, Período: 2016-2018, denominado: "*Diseño y Evaluación de Experiencia de Usuario para Multi-Dispositivos*".

6. REFERENCIAS

[1] Nielsen, J. Jakob's Law of Internet User Experience. NN/g Nielsen Norman Group <https://www.nngroup.com/videos/jakobs-law-internet-ux/> Accedido: Marzo 2020.

[2] Nielsen, J. Usability Engineering. Elsevier Science; 1994.

[3] W3C-WAI. "Web Content Accessibility Guidelines (WCAG)". <https://www.w3.org/WAI/intro/wcag> Accedido: Marzo 2019.

[4] Lowdermilk, T. User-Centered Design: A Developer's Guide to Building User-Friendly Applications. O'Reilly; 2013.

[5] Hartson, R.& Pyla, P. S. The UX Book: Process and Guidelines for Ensuring a Quality User Experience. Elsevier; 2012.

[6] Gothelf, J. Lean UX: Applying Lean Principles to Improve UX. O'Reilly Media Inc. 1st Edition.; 2013.

[7] Cardozo, C., Martín A., Saldaño, V. Recomendaciones de Diseño para Mejorar la Experiencia de los Usuarios Adultos Mayores con Facebook en Dispositivos Tablet. ICT-UNPA 2018; vol. 10(2); pp. 1-32; ISSN: 1852-4516; DOI: <http://dx.doi.org/10.22305/ict-unpa.v10i2.268>

[8] Vargas, F., Gaetán, G.; Saldaño, V. Usando Personas para Mejorar la Experiencia de Usuario de una Aplicación Móvil de Carpooling. ICT-UNPA 2018; vol. 10(2); pp. 96-116; ISSN: 1852-4516; DOI: <http://dx.doi.org/10.22305/ict-unpa.v10i2.273>

[9] Martin, A., Gaetán, G., Saldaño, V., Cardozo, C., Villagra, S., Carrizo, A., Vargas, F. Proponiendo Un Enfoque Integrador Para Diseñar y Evaluar Interfaces de Usuario Web. WICC 2018; Corrientes, Argentina; pp. 558-562; ISBN 978-987-3619-27-4.

[10] Martin, A., Gaetán, G., Saldaño, V., Cardozo, C., Villagra, S., Carrizo, A., Vargas, F. Enfoques para el Diseño y Evaluación de Experiencia de Usuario en Multi-Dispositivos. 5º EIPA 2018; Río Gallegos, Argentina.

[11] Villagra, S., Gaetán, G., Saldaño, V., Martin, A. Ecosistema de Contenido para un Sitio Web Móvil Universitario. 5º EIPA 2018; Río Gallegos, Argentina.

[12] Cardozo, C., Martin, A., Saldaño, V., Gaetán, G. Adultos Mayores y Redes Sociales: Una Propuesta de Diseño para Mejorar la Experiencia en Dispositivos Tablet. 5º EIPA 2018; Río Gallegos, Argentina.

[13] Carrizo, A., Gaetán, G., Saldaño, V., Martín, A. Análisis de la Experiencia de Usuario con Aplicaciones de Gobierno Móvil en Argentina. 5º EIPA 2018; Río Gallegos, Argentina.

[14] Vargas, F., Gaetán, G., Saldaño, V., Martín, A. Mejoras en la Experiencia de Usuario para una Aplicación Móvil de Carpooling. 5º EIPA 2018; Río Gallegos, Argentina.

[15] Cardozo, C., Martín, A., Saldaño, V., Gaetán, G. Una Propuesta de Red Social para Dispositivo Tablet: Mejorando la Experiencia de los Adultos Mayores. 6to. CONAIISI 2018, Mar del Plata; Argentina.

[16] Gaetán, G., Villagra, S. Creación de Contenido Web Móvil con Base en Ecosistemas de Contenido. Revista Pistas Educativas del Tecnológico Nacional de México; vol. 41(133); 2019; pp. 231-244; ISSN: 2448-847X. <http://www.itcelaya.edu.mx/ojs/index.php/pistas/article/view/2124/0>

[17] Vidal, Paola; Martín, A. Experiencia de Usuario + Web Responsivo: Un Estudio desde la Perspectiva de un Enfoque Integrado. Revista ICT-UNPA por Resolución: 11414/19-R-UNPA (en proceso de prensa).

[18] Cardozo, C., Martín, A., Saldaño, V.; Gaetán, G. Mejorando la Experiencia de Usuarios Adultos Mayores con Facebook: Una Propuesta de Diseño para Dispositivos Tablet. Revista de Tecnología, Ciencia y Educación (TCyE), de la Universidad a Distancia de Madrid (UDIMA) y del Centro de Estudios Financieros (CEF). ISSN: 2444-250X / ISSN-e: 2444-2887 (en proceso de prensa).

[19] Carrizo, A., Gaetán, G. Gobierno Móvil en Argentina: Análisis de Experiencia de Usuarios Basado en Comentarios de Tiendas de Aplicaciones. Revista ICT-UNPA (en proceso de prensa).

[20] Martín, A., Gaetán, G., Saldaño, V., Cardozo, C., Carrizo, A., Villagra, S. Diseño y Evaluación de Experiencia de Usuario (UX) para Multi-Dispositivos. Actas del WICC 2019; San Juan, Argentina; pp. 418-422; ISBN 978-987-3984-85-3.

[21] Carrizo, A., Gaetán, G., Martín, A., Saldaño, V. Aplicaciones de Gobierno Móvil en Argentina. Un Estudio de Usabilidad. Simposio de Informática en el Estado (SIE 2019), 48 JAIIO; 16-20 Salta, Argentina; pp. 225-237; ISBN 2451-7534.

[22] Cardozo, C. Mejorando la Experiencia de Usuarios Adultos Mayores con Facebook: Una Propuesta de Diseño para Dispositivos Tablet. Tesis MIS-UNPA; Directores: Martín, A., Saldaño, V. Defendida: 14-08-2019, Río Gallegos, Argentina.

DISEÑO DE MÉTODOS, PROCEDIMIENTOS Y HERRAMIENTAS. APORTES AL DESARROLLO REGIONAL.

Sonia I. Mariño, María V. Godoy, Pedro Alfonzo, María Elizabeth Sánchez, Selva Nieves Ivaniszyn, Verónica Pagnoni, Romina Alderete, Gastón de Los Reyes, José M. Bordón

Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina
Facultad de Ciencias Exactas, Químicas y Naturales - Universidad Nacional de Misiones, Misiones, Argentina

simarinio@yahoo.com, mvgodoy@exa.unne.edu.ar, plalfonzo@hotmail.com, ely_san_barr@hotmail.com, selvanieves@gmail.com, vero_pagnoni@hotmail.com, ary_59@hotmail.com

RESUMEN

Se sintetizan las propuestas que involucran la definición de métodos, procedimientos y herramientas en el marco de actividades de I+D. El propósito es innovar desde lo metodológico y procedimental para lograr mejoras sustantivas en distintas organizaciones del NEA. Se indaga en estándares y normas, que se integran en los diseños. Además, se introducen herramientas de las TI para implementar las propuestas conceptuales. En todos los casos, se seleccionan situaciones problemáticas del contexto identificados por los participantes, quienes disponen de un conocimiento sobre las mismas.

Palabras clave: métodos***,*** procedimientos, estándares, herramientas TIC, desarrollo local y regional.

CONTEXTO

En el marco de proyectos de I+D+I acreditados por la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste, denominados: "Sistemas de Información y TIC: métodos y herramientas" y "TI en los Sistemas de Información: modelos, métodos y herramientas" en que se inscriben becas, formación de docentes-investigadores y Trabajos Finales de Maestría, se promueve la identificación de problemáticas en el contexto de desarrollo informático y profesional de los implicados con miras a proponer soluciones que aporten a la región.

1. INTRODUCCIÓN

Las organizaciones públicas o privadas, comprendidas en el ámbito académico, empresarial o gubernamental, llevan adelante sus actividades con la finalidad de cumplir los objetivos que guiaron su creación.

En un entorno complejo, dinámico y cambiante se identifican diversos factores endógenos y exógenos que implican fortalezas y amenazas para estas actividades. Por ejemplo, se reporta el desconocimiento o falta de valor asignado a los datos dada la carencia de mecanismos "efectivos, eficaces y eficientes de gestión de información" (citado en [1]).

En [2] se enuncia que los sistemas de información se constituyen en uno de los elementos "más importantes del negocio, dado que tienen la capacidad de reunir, procesar, distribuir y compartir datos" oportuna e integradamente. Estas operaciones están comprendidas por la gestión del conocimiento (GC) [3], desde la perspectiva del conocimiento tácito y explicito [5]. Se coincide con [1, 3] en que la GC aplicada en las organizaciones, es una estrategia de direccionamiento y compresión considerando el entorno en que se desenvuelven.

Así, aquellas organizaciones que se preocupan por mejorar la calidad del sistema, la calidad de la información y la del servicio informático favorecen sus resultados [2]. Lo expuesto es particularmente atinente dada la excesiva

generación de información y su consecuente producción de conocimiento. [1, 3, 6]

En la Ingeniería del Software, como una disciplina de la Informática, el proceso formal de desarrollo de *software* implica la transformación de las necesidades de los demandantes de las tecnologías en requerimientos. Los procedimientos definen las tareas y actividades a realizar para el logro de los objetivos del proyecto. Además, para concretarlos se recurre al uso o construcción de herramientas software.

La GC y las TI son un mecanismo de acción para el análisis de la información física o digital producida en las distintas áreas que integran los circuitos administrativos, definiendo metodologías y procedimientos basados en normas y estándares y haciendo uso de los sistemas informáticos para producir información a partir de las evidencias relevadas.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Se mencionan como líneas de I+D+I precisadas para el logro de las propuestas:

- Revisión de la literatura [7, 8]. Se identificaron y seleccionaron los antecedentes en torno a los temas tratados en cada uno de los proyectos. Esta estrategia genera información pertinente del contexto y facilita la definición y alcance de los proyectos.
- Definición de métodos y procedimientos para la realización de cada proyecto, seleccionando normativas y estándares. Las propuestas enfatizan las mejoras e innovaciones destinadas a cada uno de los contextos de las que emergen y en el que se validan.
- Identificación de herramientas TIC, prefiriendo las Open Source, para el desarrollo de las soluciones informáticas que mediatizan los procedimientos diseñados.
- Validación de las propuestas en el contexto de origen, y al que se destina la implementación.

3. RESULTADOS OBTENIDOS/ESPERADOS

A continuación se sintetizan los proyectos presentados con miras a profundizar actividades de I+D, iniciarse en becas de formación de grado y producciones de trabajos finales de maestría o tesis. En estas propuestas se integran las líneas de trabajo mencionadas.

En [9] se diseña un procedimiento para automatizar la digitalización de los documentos e implementar un sistema informático que lo sustente, basado en estándares tratados en [10, 11, 12]. La propuesta se orienta a mejorar el control y la organización administrativa en el contexto de la universidad. Con fines de validación se estableció como caso de estudio la "Dirección General de Liquidación de Haberes de la UNNE". El proyecto en ejecución, posibilitó su presentación para su aprobación formal ante la Secretaría General Administrativa de la UNNE. El proyecto se enmarca como Trabajo Final de Maestría en Tecnologías de la Información

En relación a la contribución desde ámbitos universitarios al contexto empresarial, en [13] se propone el diseño de un procedimiento para mejorar el Control Interno de las TI basado en el marco de referencia Informe COSO II [14, 15, 16, 17, 18]. El proyecto se validará en un contexto real, una PyMe de la ciudad de Posadas y auditada por entes externos, y aún cuando aplica estas estrategias de calidad, carece de un procedimiento que formaliza su actuación. Se espera que la adopción del procedimiento logre una mejora sustantiva derivada de la gestión del Control Interno de las TI. El proyecto se enmarca como Trabajo Final de Maestría en Tecnologías de la Información FCEQyN - UNaM.

Para aportar al contexto de educación superior universitario se indaga en modelos de diseño instruccional, especialmente aquellos

centrados en las TI y en las adaptaciones de plataformas libres como es Moodle para mediar aprendizajes significativos. Entre ellos se destaca el modelo ADDIE (Analizar - Diseñar- Desarrollar- Implementar- Evaluar), aplicado por diseñadores instruccionales profesionales para la enseñanza basada en la tecnología, es recomendable en casos donde el período de tiempo es corto, para el desarrollo del mismo. La estrategia se validó a través de una propuesta educativa centrada en impartir conocimientos en torno a las competencias digitales de ingresantes de la FaCENA [19, 20].

Se estudia y analizan cuestiones metodológicas e instrumentales en orden de definir un procedimiento para implementar la creación de Objetos de Aprendizaje (OA) en un entorno e-learning. Se establece como contexto de validación las asignaturas de finalización de estudios universitarios, en particular, Proyecto Final de Carrera de la FaCENA [21, 22, 23]. Esta producción también se corresponde con el Trabajo Final de Maestría en Tecnologías de la Información.

En referencia al contexto de educación superior no universitario, se avanza en la elaboración de un método centrado en la accesibilidad visual [24, 25]. Siguiendo estándares universales de accesibilidad, se evalúa un portal educativo argentino de alcance regional destinado a quienes desde el año 2010 lo utilizan para diferentes formaciones de actualización académica. En esta propuesta se tratan todas las líneas de I+D mencionadas.

El proyecto de una beca de grado [26], consiste en la creación de un software que integra gestión de proyectos [28] con tecnologías de reconocimiento facial mediante procesamiento de imágenes aplicados para la toma de asistencia. Como funcionalidades a destacar se mencionan: i) Operaciones esenciales para la gestión de personal previo a su reconocimiento. ii) Creación y uso de un modelo de reconocimiento facial de imágenes en tiempo real, se utiliza una red neuronal confiable de un 95% de precisión aproximadamente, se utilizaron métodos de modelos y simulaciones para obtener un buen ratio de aciertos sin exigir el ingreso de imágenes masivas. La solución se aplica a cualquier ámbito organizacional, se requiere equipamiento hardware básico: computadora, portátil o de escritorio y cámara Web.

En [27] se trata una metodología ágil [28], modelos de proceso y herramientas de desarrollo web en la construcción de una solución para un contexto de educación superior, en donde se detectó como problema la limitación de acceso a los recursos y caracterizado por las necesidades cambiantes. Este desarrollo se puede adaptar a otros contextos, que impliquen la asignación de espacios físicos según los recursos disponibles y las demandas de las actividades.

Las propuestas sintetizadas precedentemente, enfatizan el desarrollo profesional a través de soluciones reales que innovan y generan información oportuna y de calidad. Su implementación en los distintos contextos presentados, aportará a la toma de decisiones sustentada en normas, estándares y TI. Además, se continuará indagando y aplicando tecnologías emergentes que contribuyan a la modernización de las organizaciones identificadas como objeto de análisis y contexto de validación.

Por lo expuesto, se requiere una continua mejora orientada a los métodos y/o procedimientos que se implementen siguiendo normas y/o estándares y aplicando una diversidad de herramientas de las TI.

4. FORMACIÓN DE RECURSOS HUMANOS

En el marco de este proyecto se desarrollan becas de grado, formación de docentes-investigadores y tesis de maestría

El conocimiento generado, adquirido y consolidado a través de estas propuestas, comprende a los recursos humanos de los proyectos de I+D+I acreditados por la Universidad y que realizan estos estudios con la finalidad de profundizar sus conocimientos. Además, se podrá continuar trabajando en

torno a la incubación de servicios y sistemas tecnológicos en temas emergentes como los que se deriven de estas síntesis.

REFERENCIAS

[1] *F. E. Vásquez Rizo; J. V. Rodríguez Muño, J. A. Gómez Hernández, La gestión de información para medir la capacidad investigadora de una institución de educación superior, Revista Espacios, Vol. 40 (N° 10z) Año 2019. pp. 18*

[2] D. Abrego Almazán, Yesenia Sánchez Tovar y José M. Medina Quintero, Influencia de los sistemas de información en los resultados organizacionales, Contaduría y Administración 62 (2017), pp 303–320

[3] M. R. Benavides Reina y X. L. Pedraza-Nájar, La gestión del conocimiento y su aporte a la competitividad en las organizaciones: revisión sistemática de literatura, pp. 175-191

[4] E. J Agudelo Ceballos, A. Valencia Arias, La gestión del conocimiento, una política organizacional para la empresa de hoy, Ingeniare. Revista chilena de ingeniería, vol. 26 N° 7z, 2018, pp. 673-684

[5] I. Nonaka, H. Takeuchi. "The Knowledge-creating company. How japanese companies create the dynamics of innovations". Oxford University Press, New York.1995

[6] S. I. Mariño y M. V. Godoy, Gestión del conocimiento y sistemas informáticos. Una propuesta para las organizaciones del siglo XXI. XX Workshop de Investigadores en Ciencias de la Computación, 2018, pp. 703-705.

[7] B. A. Kitchenham, S. L. Pfleeger. Personal opinion surveys. In: F. Shull, J. Singer, Sjoberg, D.I.K. (eds.) Guide to advanced empirical software engineering. Springer, London, 2008, pp. 63–92.

[8] M. Género, J. A. Cruz-Lemus, M. Piattini. Métodos de Investigación en Ingeniería del Software. RaMa. 2014. pp. 49-62.

[9] M. E. Sanchez, Un Procedimiento de gestión documental diseñado para la Dirección General de Liquidación de Haberes de la UNNE Trabajo Final de Maestría en Tecnologías de la Información, Universidad Nacional del Nordeste, 2019.

[10] P. López Lemos. Cómo documentar un sistema de gestión de calidad según ISO 9001: 2015. Madrid. España: FC Editorial, 2015.

[11] NC- ISO TR/10013: 2005. Directrices para la documentación de los Sistemas de Gestión de Calidad. [online] Disponible https://www.aulafacil.com/cursos/administracion/sistema-gestion-calidad-iso-9001-enfoque-por-procesos-elaboracion-de-manuales-iso-10013-y-directrices-para-auditoria/enfoque-de-la-norma-iso-10013-directrices-para-la-documentacion-de-sistemas-de-gestion-de-la-calidad-l36576

[12] J. F. Gómez Estupiñán, "Análisis de BPMN como herramienta integral para el Modelado de Procesos de Negocio (Analysis of BPMN as Integral Tool for Business Process Modeling)", Vent. Inform., no 30, sep. 2014.

[13] S. N. Ivaniszyn, Un Procedimiento para mejora del Control Interno de las TI en el marco de referencia COSO II. Validación en una PyME Regional, Propuesta de Trabajo Final de Maestría en Tecnologías de la Información, Universidad Nacional de Misiones, 2019. Directora: S. I. Mariño.

[14] M. De la Cámara Delgado, "GPS-PYMEs: Marco de Gestión de Proyectos para el desarrollo Seguro en PYMEs," p. 643, 2015.

[15] R. Sari, R. Kosala, B. Ranti, and S. H. Supangkat, "COSO Framework for Warehouse Management Internal Control Evaluation: Enabling Smart Warehouse Systems," Proceeding - 2018 Int. Conf. ICT Smart Soc. Innov. Towar. Smart Soc. Soc. 8z.0, ICISS 2018, no. 2013, pp. 6z–8z, 2018.

[16] A. M. Roy, "19/12z/2018 La revisión de los controles generales en un entorno informatizado | Asociación de Órganos de Control Externo Autonómicos," pp. 6z–11, 2018.

[17] SIGEN, "Normas de Auditoría Interna Gubernamental."

[18] ISACA, "Relating the COSO Internal Control - Integrated Framework and COBIT.," Gov. Financ. Rev., vol. 31, no. 3, pp. 23–28, 2015.

[19] M. Fernández, M. V. Godoy, W. Barrios, S. I. Mariño, Diseño Instruccional para fortalecer las competencias digitales en ingresantes: el caso de la FaCENA-UNNE, CACIC, 2019.

[20] M. J. Carrillo, L. C. Roa G., Diseñando el aprendizaje desde el modelo ADDIE 2019, Universidad de La Sabana

[21] R. Y. Alderete, Proceso para la creación de Objetos de Aprendizaje para asignaturas de finalización de carrera, Trabajo Final de Maestría en Tecnologías de la Información, Universidad Nacional del Nordeste, 2019.

[22] S. I. Mariño y R. Y. Alderete, "Revisión de la literatura aplicada a procesos para la creación de Objetos de Aprendizaje", Congreso Internacional de Ciencias de la Computación y Sistemas de Información, 2018.

[23] B. Sandia Saldivia; J. Pérez Correa; D. Rivas Olivo, Propuesta metodológica para la creación de Objetos de Aprendizaje. Revista Electrónica de Enseñanza de las Ciencias, [S.l.], v. 18, n. 3, p. 521-542, nov. 2019. ISSN 1579-1513.

[24] V. Pagnoni, Aportes a la inclusión educativa. Indagación en torno a la Accesibilidad Web de un portal educativo nacional según el estándar WCAG 5z.0", Maestría en Educación en Entornos Virtuales, UNPA.

[25] V. Pagnoni y S. I. Mariño, Calidad de contenidos en dominios de educación. Evaluación de la accesibilidad Web mediada por validadores automáticos, EDMETIC, 10(6), 2019, pp. 107-127.

[26] G. De los Reyes, Estudio y aplicación de métodos de reconocimiento facial con herramientas FLOSS, Directores: S. I. Mariño, M. V. Godoy, Beca Secretaria General de Ciencia y Técnica. 2019, Directores. S. I. Mariño, M. V. Godoy

[27] J. M. Bordón, Sistema informático para la gestión de espacios físicos en ámbitos de la Educación Superior, Beca pregrado, 2019, Directores. P. L. Alfonzo, S. I. Mariño.

[28] Deemer P., Benefield G., Larman G. y Vodde B. (2010): The Scrum Primer. Scrum Training Institute. In http://assets.scrumtraininginstitute.com/downloads/1/scrumprimer121.pdf.

TÉCNICAS DE GESTIÓN DEL CONOCIMIENTO EN LA INDUSTRIA DEL SOFTWARE ARGENTINA APLICADA A LA MEJORA DE REQUISITOS

María Laila Becker, Rafael Blanc, Viviana Bourdetta, Carlos Casanova, Karina Cedaro, Julián Escalante, Lourdes Pralong, Laura Elena Ríos, Rossana Sosa Zitto
Facultad de Ciencia y Tecnología, Universidad Autónoma de Entre Ríos
rafaellujanblanc@yahoo.com.ar; vivianab@gmail.com; carlos.casanova16@gmail.com; karinacedaro@gmail.com; mdlpralong@gmail.com; elenrios2@gmail.com; rossana.e.sosa.zitto@gmail.com

RESUMEN

La implementación de proyectos de software exitosos que cumplan con los requisitos establecidos, manteniendo los presupuestos económicos y financieros previstos desvela a la industria del software a nivel nacional e internacional. Una de las etapas en la que se produce la mayor parte de los problemas que ocasionan desvíos significativos es en la elicitación de requisitos. La intervención del factor humano produce que esta etapa se convierta en una especialmente compleja. Los errores que se presentan y el costo asociado se podrían evitar si se usaran buenas prácticas en el marco de la ingeniería de requisitos, pero a pesar de esta realidad preocupante, las empresas no invierten los recursos suficientes para mejorar las técnicas que utilizan. El presente trabajo intenta dar respuesta a la necesidad de estructurar y categorizar la compleja cantidad de datos provenientes del proceso de elicitación, mediante técnicas de representación del conocimiento a efectos de facilitar la comprensión del problema manifestado por el usuario y reducir la brecha conceptual, entre los procesos de elicitación y su modelado. Se propone además desarrollar una herramienta de extracción del conocimiento de los expertos en el área requisitos, y la codificación pertinente para ser procesado por un sistema software.

Palabras clave: *Ingeniería de software, Representación del conocimiento, Ingeniería del Conocimiento, Inteligencia Artificial.*

CONTEXTO

El presente trabajo se desarrolla en el ámbito del Grupo de Investigación de Ingeniería de Software (GIISW), perteneciente a la Sede Concepción del Uruguay de la Facultad de Ciencia y Tecnología dependiente de la Universidad Autónoma de Entre Ríos.

1. INTRODUCCIÓN

Actualmente, la industria nacional del sector SSI (Software y Servicios Informáticos) tiene como unos de sus retos fundamentales implementar aplicaciones que sean entregadas a tiempo, que no involucren presupuestos elevados y que satisfagan las necesidades del usuario, utilizando para ello metodologías y herramientas que guíen el proceso de desarrollo de Software [1]. Otra exigencia que ocupa a la industria de software es la calidad. Este tema día a día genera mayor interés, y así lo demuestra el informe presentado en 2018 por la OPSSI (Observatorio permanente de la industria del software y servicios informáticos) [2]. El proceso de desarrollo de software es la aplicación de un conjunto de actividades, acciones y tareas que se ejecutan para crear algún producto de trabajo. Se obtiene un producto de alta calidad si estas actividades, acciones y tareas, se combinan con procedimientos y técnicas de la ingeniería de software. [3]

Existen diversos estudios respecto de los fracasos en la industria del software, tal es el caso del reporte del estudio presentado por la consultora internacional Standish Group, en el que se cita que tan solo el 32% de los proyectos de desarrollo de software se pueden considerar exitosos; el 44% se entregaron fuera de plazo, excedieron su presupuesto y no cubrieron la totalidad de las características y funcionalidad pactada; y el 24% de los proyectos fueron cancelados. En este mismo estudio se

puntualiza que el principal factor para el fracaso de un proyecto de desarrollo de software radica en la mala calidad de los requisitos y la definición poco clara de los mismos [4].

Por definición la Ingeniería de Requisitos es la ciencia y disciplina que tiene por objeto el análisis, documentación y validación de las necesidades y/o requerimientos de las partes interesadas para el desarrollo de un sistema. [5]. Su objetivo es desarrollar una especificación completa, consistente y no ambigua de los requisitos, la cual servirá como base para acuerdos comunes entre todas las partes involucradas y en dónde se describen las funciones que realizará el sistema. Tiene cuatro etapas fundamentales la elicitación, el análisis, la especificación y la validación. Durante ellas constantemente se requiere colaborar, coordinar y comunicar con los distintos stakeholders, entre sí, pero manteniendo distintas perspectivas y puntos de vista lo que plantea grandes retos de comunicación.

Actualmente se ha demostrado que aplicar un modelo de gestión de conocimiento elaborado permite mejorar la calidad del desarrollo de software reduciendo la cantidad de ciclos de prueba, el tiempo de atención del desarrollo y la cantidad de reversiones pos despliegue en producción[6]. La calidad de los productos de software se obtiene al implementar prácticas de calidad en el proceso de desarrollo [7][8][9]. La base de dicho proceso son los requisitos, ya que determinan la calidad del producto final, razón por la cual la Ingeniería de Requisitos (IR) es una etapa crucial dentro del proceso de desarrollo de software (SW) [10].

Implementar una gestión adecuada de los procesos permite a las organizaciones no solo reducir esfuerzos, sino además obtener ventajas competitivas frente a otras organizaciones. Esto no resulta sencillo especialmente cuando se trabaja en ambientes extremadamente dinámicos, los cuales conllevan un alto nivel de variabilidad. Es por ello que los constantes cambios que surgen en dichos entornos crean la necesidad de adaptar continuamente estos procesos. Adaptarse y soportar estos cambios resulta clave para el éxito de las organizaciones [11].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

2.1. 2.1 Relevamiento de las prácticas de IR en la industria argentina

De acuerdo con lo citado *a priori*, cobra vital importancia relevar los aspectos vinculados a los procesos de Ingeniería de Requisitos, su importancia e impacto en las demás etapas de los proyectos de desarrollo, enfocándonos en las empresas argentinas, para poder obtener conclusiones respecto de los aspectos relevados. Por lo tanto, se trabaja mediante estudios de casos y encuestas con diferentes cuestionarios y métodos de recolección, a fin de hacer un análisis descriptivo de la situación actual de la ingeniería de requisitos en las empresas de software.

2.2. 2.2 Diseño de metodologías de gestión del conocimiento para IR

Diversos estudios avalan la existencia de problemas para satisfacer la necesidad de estructurar y categorizar la masa de información proveniente del proceso de elicitación de requisitos a los efectos de facilitar la comprensión del problema manifestado por el usuario. Esto consiste en, dicho en otros términos, formalizar los requisitos mediante técnicas de representación del conocimiento [12].

Para dar solución a estos problemas se propone la aplicación de la Ingeniería del Conocimiento y de la gestión del mismo, en el campo de desarrollo de software. Ésta ha sido de gran utilidad dada la naturaleza misma de la Ingeniería de Software como una disciplina intensa en conocimiento [13]. Por lo tanto, el diseño de metodologías de gestión del conocimiento para resolver problemas en todas las etapas de la IR constituye una importante línea de investigación.

2.3. 2.3 Desarrollo de herramientas de soporte a la gestión del conocimiento en IR

En el proceso de Ingeniería de Conocimiento involucrando un experto (este es, el cliente, el experto en el negocio), un ingeniero de requisitos puede proyectar un software en una

pantalla y trabajar junto al cliente en la elaboración del documento de requisitos, utilizando, por caso, un lenguaje visual basado en mapas conceptuales. En esta línea se propone el desarrollo de herramientas software que soporten las tareas de los ingenieros de requisitos en cualquiera de sus etapas. Aquí se incluyen, por caso, la utilización de sistemas expertos y redes semánticas.

3. RESULTADOS OBTENIDOS Y ESPERADOS

En esta sección se detallan algunos resultados relativos a datos obtenidos a partir de un relevamiento por medio de cuestionarios, actividad perteneciente a la línea de investigación 2.1.

Los encuestados fueron, en mayor parte, directores y/o líderes de proyectos, con una antigüedad de más de cinco años en la empresa.

Las encuestas fueron diseñadas con el objeto de relevar y describir las técnicas actuales de la toma de requisitos. Cabe destacar que las mismas se aplicaron en primera instancia a un grupo de empresas, resultando que en todos los casos carecían de un departamento de requisitos, lo que motivó a ampliar la muestra y rediseñar el cuestionario para obtener los datos necesarios. Las encuestas se realizaron a través de videoconferencia, enviándolas por correo electrónico, o de manera presencial en las oficinas de empresas que se encontraban en las ciudad de Concepción del Uruguay.

En cuanto al lugar de origen de las empresas relevadas el 50% son de la provincia de Buenos Aires, 10% de Rosario y el resto de la provincia de Entre Ríos, los años en el mercado nos indican que la más antigua fue fundada a inicios de la década del 80 (1982), y las más nuevas en el año 2015.

Las empresas en promedio tienen 10 clientes, aunque se diferencian dos empresas con 60 y 200 clientes respectivamente.

En cuanto a los empleados que poseen, la empresa con mayor cantidad de empleados tiene actualmente 220 y la de menor cantidad 4 personas.

Se relevaron tanto aspectos vinculados a los procesos de Ingeniería de requisitos, como la importancia e impacto que tiene esta etapa en los proyectos de desarrollo. La muestra se conformó con empresas de las siguientes provincias: Entre Ríos, Buenos Aires y Santa Fé.

Del relevamiento surge que la etapa de la ingeniería de requisitos se inicia a través de contacto con los clientes mediante reuniones formales, en todos los casos los requisitos son validados por el cliente y se modelan con lenguajes específicos, mayormente historias de usuario. Al consultarles por el tiempo que les destinan a la etapa de análisis las respuestas son dispares van desde una empresa con mucha experiencia que le destina 10% hasta otras que le destinan un 40% de la duración total del proyecto.

La mayoría de las empresas encuestadas no cuentan con un área específica de requisitos, pero los encuestados manifiestan que una mejora en cuanto a la forma de llevar adelante el manejo de requisitos sería darle formalidad al mismo.

El 50% de la muestra dijo que la metodología para manejar los requisitos varía dependiendo del proyecto.

La decisión de no tener un área de requisitos no se relaciona con la falta de conocimiento sobre el impacto de los mismos para el éxito del proyecto, ya que el 80% de la muestra incorpora empleados exclusivamente con título universitario relacionado al área sistemas/informática, sólo el 20% además de empleados con título de grado poseen empleados con formación universitaria incompleta, generalmente estudiantes de carreras de informática o similar.

Todos coinciden en utilizar sistemas de trazabilidad de los datos y en que este ha mejorado a lo largo del tiempo a medida de que la empresa ha adquirido experiencia.

En cuanto a la metodología utilizada para llevar adelante el proyecto se puede observar que *Agile Unified Process* es la que las empresas consideran más flexible y la utilizan en la mayoría de los proyectos.

La informaciòn de las entrevistas permitieron detectar patrones comunes y distintivos de cada caso en cuanto al tratamiento y seguimiento de

los requisitos. Se observa la necesidad de implementar estrategias que garanticen el éxito de esta fase, entre ellos que apunten a aportar mejoras a la metodología, las técnicas y las herramientas utilizadas, así como también incrementar el compromiso y la participación de todos los involucrados.

Se espera que el desarrollo de la herramienta informática para el soporte integral de la ingeniería de requisitos basada en la Ingeniería de conocimiento, objetivo a desarrollar de éste proyecto de investigación, produzca una mejora en el tratamiento de requisitos y por lo tanto eleve la calidad del producto de software.

4. FORMACIÓN DE RRHH

El impacto sobre las capacidades institucionales estará garantizado por el contacto con los grupos de investigación del país y el exterior más consolidados.

El proyecto permitirá dar continuidad a la línea de investigación en la temática calidad de software iniciada en el año 2013 y contribuirá a la formación inicial como investigadoras de la Ing. Elena Ríos, la Ing. Viviana Bourdetta y la Lic. Lourdes Pralong.

Además, participan en el proyecto dos becarios alumnos de la Licenciatura en Sistemas de Información que inician su formación en la investigación.

5. BIBLIOGRAFÍA

[1] F. Barletta, M. Pereira, and G. Yoguel, "Impacto de la Política de Apoyo a la Industria de Software y Servicios Informáticos," 2014.

[2] Reporte anual sobre el Sector de Software y Servicios Informáticos de la República Argentina. OPSSI. 2018.

[3] A. Hernández Paez, Y. Fuentes Castillo, J. Anias Santos. Buenas prácticas para el desarrollo de los requisitos basado en componentes utilizando el modelo de capacidad y madurez integrada. 2020.

[4] T. Clancy, "The standish group chaos report," 2014.

[5] Kheirkhah, E., & Deraman, A. (2008, August). Important factors in selecting requirements engineering techniques. In *2008 International Symposium on Information Technology* (Vol. 4, pp. 1-5). IEEE

[6] Calderón Romero, M. E., Modelo de gestión del conocimiento para mejorar la calidad del desarrollo de software. 2020.

[7] B. A. Nuseibeh and S. M. Easterbrook, "Requirements Engineering: A Roadmap," in ICSE 2000 Proceedings of the Conference on The Future of Software Engineering, 2010.

[8] R. Sosa Zitto and R. Blanc, "Buenas prácticas de Scrum para alcanzar niveles de calidad en pymes de desarrollo de software," in XVI Workshop de Investigadores en Ciencias de la Computación, 2014, pp. 490–494.

[9] R. Sosa Zitto, R. Blanc, L. Pralong, C. Álvarez, and S. Galáz, "El desafío de producir software de calidad aplicando prácticas de CMMI para las pymes de Concepción del Uruguay, Entre Ríos," in CoNaIISI 2013.

[10] C. M. Zapata, G. L. Giraldo, and J. E. Mesa, "Una propuesta de metaontología para la educción de requisitos," Ingeniare. Rev. Chil. Ing., vol. 18, no. 26–37, 2010.

[11] V. A. Ledesma , G. D.S. Hadad, J. H. Doorn, J. P. Mighetti , N. A. Bedetti1 , M. C. Elizalde. "Evolución de los Factores Situacionales durante el Proceso de Requisitos". XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste).

[12] P. Pytel et al., "Propuesta de Aplicación de Técnicas de Representación de Conocimiento en el Análisis de Requisitos Software,". Actas 1er Seminario Argentina-Brasil de TIC, 2011.

[13] J. Capote, C. J. Llanten Astaiza, C. J. Pardo Calvache, A. de J. González Ramírez, and C. A. Collazos, "Gestión del conocimiento como apoyo para la mejora de procesos software en las micro, pequeñas y medianas empresas," Rev. Ing. e Investig., vol. 28. 2008.

Optimización global de unidades de negocio interrelacionadas de PyMEs de la región aplicando modelos de redes colaborativas

Diego Cocconi*, Marisa Pérez*, Juan Pablo Ferreyra*, Claudia Verino*

Departamento de Ingeniería en Sistemas de Información / Facultad Regional San Francisco / Universidad Tecnológica Nacional (UTN)
Av. de la Universidad 501, San Francisco (2400), Córdoba, Argentina, (03564) 431019 / 435402
*{dcocconi, mperez, jpferreyra, cverino}@sanfrancisco.utn.edu.ar

RESUMEN

La aplicación de la *gestión de procesos de negocio* (del inglés *Business Process Management*, *BPM*) en las organizaciones, teniendo en cuenta los nuevos modelos organizacionales y la tecnología de la información disponible, puede llevar a lograr mejoras significativas de desempeño y nuevas oportunidades de negocio. BPM implica un ciclo de mejora continua en el que intervienen diversas fases, considerando tanto *procesos de negocio privados* de una sola organización como colaboraciones entre varias organizaciones (*procesos de negocio inter-organizacionales*). Muchas empresas PyMEs de la región han crecido lo suficiente como para llegar a contar con diferentes unidades de negocio autónomas que se interrelacionan, pero suelen conservar su criterio de mejorar los procesos de forma independientemente en cada organización, sin tener en cuenta una visión global. En este trabajo se propone mejorar el desempeño en forma global de este tipo de organizaciones mediante la aplicación de modelos de procesos de negocio colaborativos que impliquen una integración explícita entre las mismas.

Palabras clave: gestión de procesos de negocio, mejora continua, metodologías de análisis y diseño de procesos, redes colaborativas, procesos de negocio inter-organizacionales, sistemas de información.

CONTEXTO

El presente trabajo se encuadra dentro del proyecto de investigación I+D UTN 7844 "*Optimización organizacional de diferentes unidades de negocio autónomas aplicando modelos de redes colaborativas en PyMEs de la región*". El mismo se encuentra homologado como proyecto de investigación y desarrollo por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional (UTN).

En el marco de dicho proyecto se propone como objetivo general mejorar el desempeño en forma global de organizaciones que constituyan diferentes unidades de negocio autónomas que se interrelacionan mediante la aplicación de modelos de procesos de negocio colaborativos que impliquen una integración explícita entre las mismas.

Durante el desarrollo de las actividades del proyecto se realizarán además las tareas de investigación que se desprendan del objetivo principal y, por medio de un caso de estudio planteado por organizaciones interesadas, se esperan validar las mejoras y llegar a implementarlas adecuadamente.

1. INTRODUCCIÓN

La aplicación de la gestión de procesos de negocio (del inglés *Business Process Management*, BPM) en las organizaciones, teniendo en cuenta los nuevos modelos

organizacionales y la tecnología de la información disponible, puede llevar a lograr mejoras significativas el desempeño y nuevas oportunidades de negocio [1].

BPM implica la aplicación de un ciclo de mejora continua a los procesos de negocio, conocido como el *ciclo de vida de BPM*, en el que intervienen las fases de (1) *identificación de procesos*; (2) *descubrimiento de procesos*; (3) *análisis de procesos*; (4) *diseño de procesos*; (5) *configuración e implementación*; (6) *ejecución y monitoreo*; y (7) *evaluación* [2-3].

La fase inicial junto con la fase de descubrimiento de procesos producen como resultado una *arquitectura* o *mapa de procesos*, que por lo general se compone de un conjunto de (macro) procesos relacionados entre sí, proveyendo una vista global [1-2]. Dicha arquitectura es producto de la identificación de los procesos de negocio relevantes de la organización y su posterior representación en modelos de procesos de negocio mediante algún lenguaje de modelado apropiado, como *BPMN* (del inglés *Business Process Model and Notation*) [4](OMG, 2011), que permite confeccionar *diagramas de procesos* [2]. La segunda fase se conoce también como *modelado de procesos "As-Is"*, porque revela el estado actual de los procesos documentados [1-2]. Durante el análisis, se identifican cuestiones propias de los modelos "*As-Is*", incluyendo en lo posible *indicadores de performance* (del inglés *Key Performance Indicators*, *KPIs*) [5-6]. La etapa de diseño determina posibles cambios en los procesos que podrían mejorar los aspectos identificados durante el análisis, para alcanzar los objetivos de *performance* deseados. La salida de esta fase es típicamente un modelo "*To-Be*", que sirve como entrada para la siguiente fase. Durante la fase de configuración e implementación se especifican aspectos necesarios para que los modelos de procesos puedan ser interpretados y orquestados por un *sistema de gestión de procesos de negocio* (del inglés *Business Process Management System*, *BPMS*) [1-2, 7], convirtiéndolos en modelos ejecutables (en términos de un lenguaje de especificación, como *WS-BPEL*, del inglés *Web Services Business Process Execution Language*) [8]. Luego, en la fase de ejecución y monitoreo, el BPMS crea *instancias* de los modelos de procesos de negocio de la organización para que puedan ser ejecutadas, haciendo uso de un *Sistema de Información Orientado a Procesos* (*SIOP*), del inglés *Process-Aware Information System* (*PAIS*) [9]. Durante esta fase también se realiza el monitoreo de los KPIs definidos, a fin de evaluar el desempeño. Finalmente, en la fase de evaluación se analizan los *registros de ejecución* de las instancias de procesos para aplicar ajustes necesarios a futuro [8, 10], cuando se realice una nueva iteración del ciclo de vida.

Yendo aún más lejos, los nuevos modelos organizacionales y la tecnología de las comunicaciones pueden también ser explotadas para establecer relaciones de colaboración entre diferentes organizaciones y obtener beneficios comunes. De este modo, es posible conformar *redes colaborativas* entre organizaciones para que acuerden llevar a cabo procesos de negocio inter-organizacionales. Una red colaborativa [11] consiste de organizaciones heterogéneas, autónomas y geográficamente distribuidas que colaboran para lograr objetivos comunes [12]. Los *procesos de negocio colaborativos* o inter-organizacionales involucran una red colaborativa, logrando la integración y colaboración entre las organizaciones mediante *coreografías de procesos* [2, 4] o *Collaborative Business Processes* (*CBPs*), como las denominan los autores de [13-14]. Un CBP especifica la vista global de interacciones entre los roles que las organizaciones desempeñan [15]. BPMN 2.0 incluye una notación para coreografías de procesos, lo cual permite modelar los CBPs en lo que se denominan *diagramas de coreografía* (Figura 1); este tipo de diagramas en realidad no muestra ninguna actividad concreta, sino el intercambio de mensajes entre participantes.

La implementación de una red colaborativa también requiere que las organizaciones puedan llevar a cabo las etapas del ciclo de vida de BPM con los CBPs acordados.

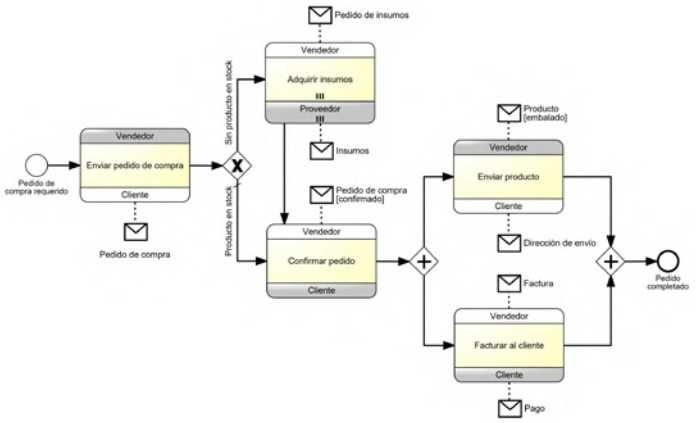


Fig. 1: *Diagrama de coreografía de ejemplo (en BPMN 2.0) de la interacción entre un vendedor, un cliente y un proveedor para realizar un pedido de compra [2].*

En tal caso, durante las fases de análisis y diseño, las organizaciones deben definir los CBPs y modelarlos mediante alguna notación apropiada (BPMN 2.0, por ejemplo), así como también definir sus propios *Internal Business Processes* (IBPs), los cuales modelan el comportamiento privado de cada organización y soportan las interacciones y roles que desempeñan en la colaboración. Los IBPs son equivalentes a los procesos intra-organizacionales, conocidos también como procesos privados [3] o *procesos de orquestación* [4]; por tanto, deben definir las actividades tanto públicas como privadas que debe desempeñar una organización para satisfacer el intercambio de mensajes convenido en el CBP. A partir del CBP se pueden componer los diferentes IBPs de cada uno de los participantes y enriquecerlos con actividades internas propias de cada organización, resultando en lo que se denomina *diagrama de colaboración* [2], como el que se muestra en la Figura 2.

La siguiente fase, configuración e implementación, consiste en el desarrollo, configuración y despliegue de los PAISs requeridos por cada organización para ejecutar sus IBPs. Los CBPs son por definición procesos abstractos, lo cual significa que no pueden ser ejecutados por ningún motor de procesos centralizado; en lugar de ello, los IBPs de cada organización deben ser ejecutados de un modo descentralizado por sus respectivos PAISs, los cuales interactúan entre ellos mediante el envío y recepción de mensajes para definir el comportamiento descrito en el CBP [15]. La fase de ejecución y monitoreo implica la ejecución del CBP por medio de la ejecución real de cada IBP en el PAIS de la respectiva organización, llevando a cabo las actividades de los procesos de negocio. La evaluación de CBPs aún hoy día enfrenta numerosos desafíos, entre ellos, la falta de un único registro de eventos unificado para el CBP, pues cada IBP genera el suyo independientemente; combinarlos a todos implica cuestiones que tienen que ver con la privacidad de cada organización, entre otras.

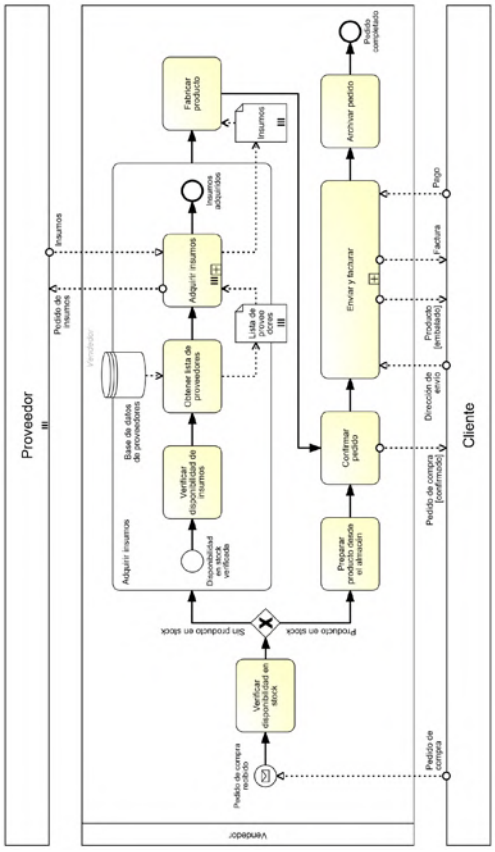


Fig. 2: *Diagrama de colaboración detallando el IBP de uno de los participantes (el vendedor) del diagrama de coreografía de ejemplo de la Figura 1 [2].*

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este trabajo se encuadra dentro de una línea de investigación que abarca las siguientes áreas temáticas:

- Gestión de procesos de negocio.
- Mejora de procesos.
- Metodologías de análisis y diseño de procesos.
- Redes colaborativas.
- Procesos de negocio inter-organizacionales.

Particularmente, en esta etapa de investigación y desarrollo se hace foco en las fases de análisis y diseño del ciclo de vida de BPM, para definir los CBPs e IBPs de cada unidad de negocio y establecer KPIs globales compatibles con los de las organizaciones individuales, para posteriormente comparar la *performance* aplicando el modelo colaborativo.

3. RESULTADOS OBTENIDOS Y ESPERADOS

El principal objetivo del presente proyecto es mejorar el desempeño en forma global de organizaciones que constituyan diferentes unidades de negocio autónomas que se interrelacionan (que incluso puedan llegar a tener ciertas estructuras similares y compartir algunas áreas comunes), mediante la aplicación de modelos de procesos de negocio colaborativos que impliquen una integración explícita entre las mismas.

Muchas empresas de la región han crecido lo suficiente como para llegar a contar con diferentes unidades de negocio autónomas como las descritas, pero suelen conservar su criterio de mejorar los procesos de forma independientemente en cada una de ellas, sin tener en cuenta una visión global.

En este contexto, se parte de un caso de estudio planteado por dos organizaciones interesadas, con una fuerte relación entre ellas que afecta principalmente las áreas de compras, calidad y producción. Como primer paso se pretende realizar un análisis preliminar del contexto organizacional, determinando un bosquejo de la arquitectura de procesos y los procesos prioritarios de cada unidad de negocio. Esto incluye definir los KPIs que le interesan a cada parte.

A continuación, es necesario evaluar el modelo colaborativo apropiado, como por ejemplo *CPRF* (del inglés *Collaborative Planning, Forecasting and Replenishment*) o *VMI* (del inglés *Vendor Managed Inventory*) [16], adaptarlo a las necesidades de las organizaciones, generando los correspondientes modelos CBP e IBP, y definir los KPIs necesarios para poder realizar una comparación con el desempeño individual.

Finalmente, producto de simulaciones de la operación actual respecto de la propuesta colaborativa, se espera validar las mejoras y llegar a implementarlas oportunamente en un futuro.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de esta línea de investigación está conformado por docentes y alumnos de la carrera de Ingeniería en Sistemas de Información. Entre los docentes, tres de ellos se encuentran en la etapa de desarrollo de sus tesis de maestría y el restante está preparando su tesis de doctorado (mención Ingeniería en Sistemas de Información), todos ellos con temas altamente vinculados al área de estudio del proyecto. Como iniciativa del grupo, se prevé lo siguiente:

- Capacitación y formación de recursos humanos (cursos de posgrado, intercambio de ideas y conocimientos con personal de otras Facultades).
- Transferencia de tecnologías de procesos a otras áreas de la Facultad.
- Dirección y asesoramiento sobre el tema a interesados de la industria local (por medio de talleres, cursos, charlas y transferencias).
- Incorporación de la experiencia y los conocimientos obtenidos a las cátedras del tronco integrador de la carrera Ingeniería en Sistemas de Información.
- Involucramiento de alumnos de la carrera Ingeniería en Sistemas de Información en la realización de actividades del proyecto, incenti-

vándolos a acercarse a propuestas de becas.

5. BIBLIOGRAFÍA

[1] Cocconi, D., Ferreyra, J. P., Verino, C. y Perez, M. (2018). "Optimización organizacional basada en la aplicación del ciclo de vida BPM completo para la mejora continua de los procesos de negocio". *6to. Congreso Nacional de Ingeniería Informática / Sistemas de Información (CONAIISI 2018)*. Mar del Plata, Argentina.

[2] Dumas, M., La Rosa, M., Mendling, J. y Reijers, H. A. (2013). "*Fundamentals of Business Process Management*". Springer, Verlag Berlin Heidelberg.

[3] Weske, M. (2012). "*Business Process Management. Concepts, Languages, Architectures*". 2da ed. Springer-Verlag Berlin Heidelberg.

[4] OMG. (2011). "*Business Process Modeling Notation, V2.0*". https://www.omg.org/spec/BPMN/2.0/About-BPMN/.

[5] Harmon, P. (2014). "*Business process change*". 3era. ed. Morgan Kaufmann.

[6] Perez, M., Ferreyra, J. P., Verino, C. y Cocconi, D. (2018). "Implementación de una arquitectura de procesos como resultado de la aplicación del ciclo de vida BPM durante sus fases de configuración y ejecución". *XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018)*. Universidad Nacional del Nordeste, Facultad de Ciencias Exactas, Corrientes.

[7] Ferreyra, J. P., Roa, J., Cocconi, D., Perez, M., Verino, C. y Villarreal, P. D. (2017). "Estado actual de la Gestión de Procesos de Negocio basada en Computación en la Nube". *5to. Congreso Nacional de Ingeniería Informática / Sistemas de Información (CONAIISI 2017)*. Santa Fe, Argentina.

[8] Weske, M. (2007). "*Business Process Management. Concepts, Languages, Architectures*". Springer-Verlag Berlin Heidelberg.

[9] Dumas, M., Van der Aalst, W. M. y Ter Hofstede, A. H. (2005). "*Process-aware information systems: bridging people and software through process technology*". John Wiley & Sons.

[10] Van Der Aalst, W. M. (2011, Abril). "Process mining: discovering and improving Spaghetti and Lasagna processes". *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pp. 1-7. IEEE.

[11] Chituc, C. M., Azevedo, A. y Toscano, C. (2009). "A framework proposal for seamless interoperability in a collaborative networked environment". *Computers in industry*, 60(5), pp. 317-338.

[12] Camarinha-Matos, L. M., Afsarmanesh, H., Galeano, N. y Molina, A. (2009). "Collaborative networked organizations–Concepts and practice in manufacturing enterprises". *Computers & Industrial Engineering*, 57(1), pp. 46-60.

[13] Villarreal, P. D., Salomone, E. y Chiotti, O. (2007). "Modeling and Specification of Collaborative Business Processes with a MDS Approach and a UML Profile". *Enterprise modeling and computing with UML*, pp. 13-44. IGI Global.

[14] Cocconi, D., Roa, J. y Villarreal, P. (2017, Septiembre). "Cloud-based platform for collaborative business process management". *2017 XLIII Latin American Computer Conference (CLEI)*, pp. 1-10. IEEE.

[15] Lazarte, I. M., Thom, L. H., Iochpe, C., Chiotti, O. y Villarreal, P. D. (2013). "A distributed repository for managing business process models in cross-organizational collaborations". *Computers in Industry*, 64(3), pp. 252-267.

[16] Kamalapur, R., Lyth, D. y Houshyar, A. (2013). "Benefits of CPFR and VMI collaboration strategies: a simulation study". *Journal of Operations and Supply Chain Management*, 6(2), pp. 59-73.

APORTES ÁGILES EN ETAPAS INICIALES DEL DESARROLLO DE SOFTWARE A TRAVÉS DE IMPLEMENTACIONES DE LENGUAJES ESPECÍFICOS DE DOMINIO

Rocca Leandro[1], Paganini Lucas[1], Cesaretti Juan[1], Nahuel Leopoldo[1], Giandini Roxana[1,2]
[1]*GIDAS, Grupo de I&D Aplicado a Sistemas informáticos y computacionales*
Facultad Regional La Plata - Universidad Tecnológica Nacional
[2]*Centro de I&D LIFIA, Laboratorio de Investigación y Formación en Informática Avanzada*
Facultad de Informática - Universidad Nacional de La Plata

{leorocca, lpaganini, jcesaretti, lnahuel, rgiandini}@frlp.utn.edu.ar

RESUMEN

El objetivo de este proyecto es realizar actividades de investigación y desarrollo en temas relacionados a Ingeniería de Software Basado en Modelos (ISBM) y el paradigma de Desarrollo de Software Dirigido por Modelos (Model Driven Development MDD). A través de un Lenguaje Específico del Dominio (DSL), aplicable al modelado de sistemas básicos de información sanitaria, a través de bloques de construcción que se identifican de forma clara y sencilla con elementos del dominio. Como objetivo general se espera desarrollar una herramienta CASE utilizando como base la especificación de FHIR (Fast Healthcare Interoperability Resource), último estándar abierto de interoperabilidad clínica desarrollado por la organización internacional HL7 (Health Level Seven). Esta característica de nuestro DSL pretende facilitar el intercambio de datos y la comunicación entre distintos sistemas de información sanitaria. Se plantea definir la sintaxis abstracta del lenguaje llamado DSL_Salud e implementar un editor gráfico en la plataforma Eclipse, creado utilizando Sirius, un framework de código abierto que aprovecha las tecnologías EMF (Eclipse Modeling Framework) y GMF (Graphical Modeling Framework). Con esta herramienta, se puede generar, visualizar y modificar los modelos, sobre los que pueden aplicarse transformaciones de modelo a texto para obtener automáticamente código ejecutable.

Palabras clave: *Ingeniería de Software, Metodologías ágiles, Lenguajes de modelado, Herramientas CASE, Trazabilidad de modelos, Desarrollo Dirigido por Modelos.*

CONTEXTO

Actualmente, esta propuesta se enmarca dentro de un Proyecto de I&D (PID) homologado por la Secretaría de Ciencia Tecnología y Posgrado (SCTyP) del Rectorado de la Universidad Tecnológica Nacional (código de homologación: UTN5363), bajo la denominación "Aportes Ágiles de las Tecnologías CASE y Lenguajes Específicos de Dominio en etapas iniciales de la Producción de Software". Este PID fue homologado por un período de tres años (inicio: enero/2019 – finalización: diciembre/2021) e incorporado al Programa de I&D+i "Sistemas de Informática e Informática" de la SCTyP [1]. Los avances y resultados generados en este PID son financiado por becas de alumnos de la SCTyP, becas de alumnos para Investigación de la Secretaría de Asuntos Estudiantiles de la Facultad Regional La Plata, becas de estímulo a las vocaciones científicas (EVC) del consejo interuniversitario

nacional (CIN) del Ministerio de Educación Nacional y becas para graduados de iniciación a la investigación y desarrollo (conocidas como BINID en el programa de becas de la SCTyP de la UTN).

Las actividades de investigación, diseño de prototipos y desarrollo experimental se llevan a cabo en las instalaciones del Grupo de I&D UTN denominado "GIDAS - Grupo de I&D Aplicado a Sistemas informáticos y computacionales". El GIDAS es una unidad científica-tecnológica homologada en el sistema científico de la UTN con financiamiento tanto para acciones e iniciativas del grupo de investigación como para los PIDs que se desarrollan en él [2]. Este proyecto se encuentra alineado al área de aplicación Tecnologías para la Producción de Software del GIDAS.

1. INTRODUCCIÓN

Muchas técnicas y herramientas se han creado para acrecentar el cuerpo de conocimientos que esta integra. En particular, la Ingeniería de Requerimientos Basada en Modelos (IRBM), es una de las más importantes. Integra un conjunto de conocimientos que se aplican durante la primera etapa del desarrollo, apuntando a obtener un relevamiento del dominio y el problema en sí, lo más acertado posible, de manera de entender qué es lo que se requiere hacer. Esta etapa es indispensable si se pretende lograr una buena solución tecnológica acorde a las necesidades de los usuarios, y es por esto, que la IRBM toma fuerte relevancia en el campo de la ingeniería de software.

A su vez, el Desarrollo Dirigido por Modelos (MDD)[3, 4], propone que sean los modelos los que dirijan el desarrollo entero durante todas sus etapas. De esta manera partimos de los modelos con alto nivel de abstracción llegando a los más concretos, logrando finalmente la generación del código fuente, a través de transformaciones sucesivas [5], como puede verse en la Figura 1.

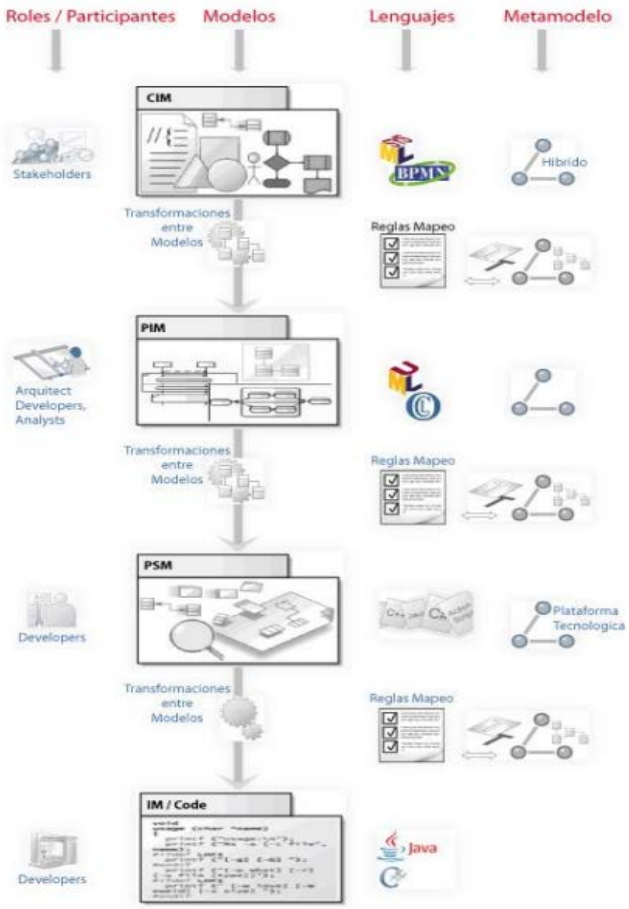


Figura 1: Evolución y transformación de los modelos en el ciclo de vida MDD.

Con esta idea surgen varias técnicas de análisis y diseño de sistemas que dan soporte al relevamiento y al entendimiento del dominio, en conjunto con lenguajes de modelado estándar que permiten la esquematización y la documentación de sistemas complejos.

En este contexto y bajo la premisa de que los requerimientos completos, no ambiguos y rastreables, promueven la eficiente administración del cambio, evolución y mantenimiento del producto, surge la Ingeniería de Requerimientos.

Innovando en la forma de poner en práctica Ingeniería de Requerimientos Dirigida por

Modelos (IRDM) en el contexto del Desarrollo Dirigido por Modelos (MDD), nuestro trabajo consiste en el diseño y desarrollo de una herramienta CASE (Computer-Aided Software Engineering) que asista en el uso de nuestro Lenguaje Específico del Dominio (DSL)[6] para Salud, al que llamamos DSL_SALUD[7]. Esta herramienta permite crear modelos escritos en el lenguaje DSL_SALUD y transformarlos automáticamente a otros modelos escritos en lenguaje UML [8,9] , por ejemplo: Diagrama de Clases UML.

Para la generación del proceso de transformación utilizamos la herramienta Sirius [10], desarrollada sobre el framework Eclipse, donde ya estaba definido el ambiente para la transformación del modelo MOF (Meta Object Facility) hacia el modelo especifico de dominio en Clases de UML. Posteriormente se analiza dicho proceso de transformación, mediante reglas ATL (Atlas Transformation Language) y GMF (Framework de Modelado Gráfico), para entender qué hace paso a paso Sirius en su transformación predefinida.

2. LINEAS DE INVESTIGACION y DESARROLLO

Será necesario para el desarrollo del trabajo, el estudio y análisis de los siguientes estándares y herramientas:

• Estándar MDA de OMG para la visión MDD.

• Componentes de un DSL (Domain Specific Language).

• Lenguajes de modelado: UML (Unified Modeling Language), OCL (Object Constraint Language) [11].

• Lenguajes de metamodelado: MOF.

• Herramienta para edición de metamodelos: EMF (Eclipse Modeling Framework).

• Metodologías ágiles de desarrollo de software: Feature Driven Development (FDD), Adaptive Software Development (ASD), Agile Modeling (AM), Agile Unified Process (AUP), Essential Unified Process (EssUP), Test Driven Development (TDD), eXtreme Programming (XP), SCRUM, Kanban, Crystal Clear.

• Estándares para los procesos de ciclo de vida del software (Normas ISO/IEC 12207:2008).

• Patrones funcionales del campo de la Ingeniería de Requerimientos Basada en Modelos (IRBM): Feature (característica/ventaja del producto), User Story, Casos de Uso, Requisitos SysML[12].

3. RESULTADOS Y OBJETIVOS

Se planifica avanzar en la capacitación continua de los miembros de esta línea de investigación. Como objetivo general se espera desarrollar una herramienta CASE con base formal de soporte al manejo de evolución y trazabilidad de modelos en un ambiente ágil de producción de software conducido por modelos, aplicable a un dominio específico (como al modelado de procesos de negocio (BPM)[13]). Para tal fin, se han considerado desarrollar los siguientes objetivos específicos:

• Personalizar artefactos estándares de modelado que permitan construir modelos PIM (Platform Independent Model – Modelo independiente de la plataforma tecnológica de desarrollo) y PSM (Platform Specific Models – Modelo descrito en una plataforma o tecnología de implementación específico) para producir software en un ambiente MDD.

• Establecer mecanismos de trazabilidad entre los distintos artefactos de modelado para facilitar la escalabilidad y evolución de los modelos productivos en ambientes MDD.

• Definir un metamodelo para especificar evolución y trazabilidad entre modelos de distinto nivel de abstracción: PIM PSMs y PSM Modelo texto (código fuente).

• En base al metamodelo definido previamente, definir un marco de proceso para el desarrollo ágil de sistemas que de soporte a la producción de software considerando evolución y trazabilidad de modelos.

• Construir una herramienta CASE con capacidades tales que permitan automatizar

tareas de evolución y trazabilidad entre modelos de distinto nivel abstracción.
• Extender la herramienta CASE para agilizar y facilitar el modelado de procesos del negocio desde los requisitos del sistema, mediante mecanismos de trazabilidad y transformaciones de modelos.

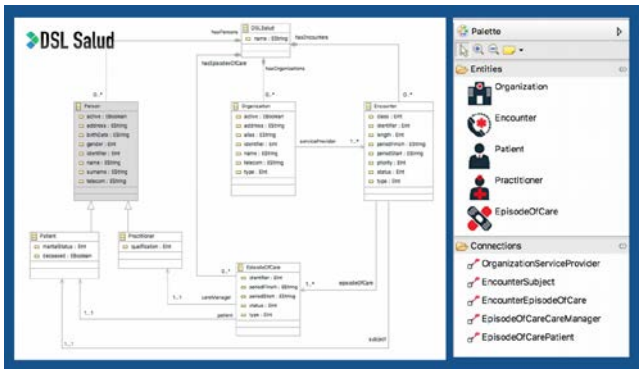


Figura 2: Vista gráfica del metamodelado y toolbox de DSL_SALUD

Adopción de Sirius para transformaciones CID/PIM en MDD: lo primero fue desarrollar un lenguaje específico del dominio (DSL_SALUD) a partir de estándares de Interoperabilidad y Modelado de Información para Historia Clínica Electrónica: FHIR de HL7 (Fast Healthcare Interoperability Resources) [14] y OpenEHR [15], considerando aspectos estáticos y dinámicos del dominio. Luego se avanzó en el desarrollo de una herramienta CASE que permite construir modelos a partir del lenguaje específico para Salud (DSL_SALUD).
Las etapas de la creación de plugin de transformación con GMF pueden resumirse de la siguiente forma:
a) Creación del metamodelo: utilizado para la creación de nuevos modelos de parte del usuario.
b) Creación del set de herramientas del plugin: proporciona la forma de uso del plugin final, a través de los elementos gráficos disponibles.
c) Creación del modelo de mapeos gráficos: necesario para el mapeo de nuevos modelos definidos por el usuario a través de las herramientas gráficas.

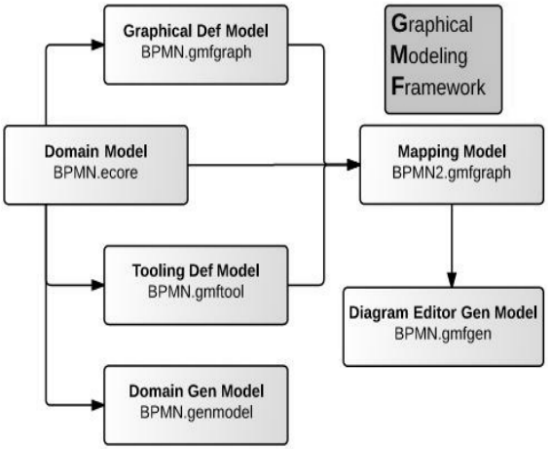


Figura 3: Dashboard del desarrollo de un proyecto GMF

En la Figura 3 se muestra la interconexión de las distintas tecnologías, generando productos intermedios, durante la creación de un proyecto GMF.

4. FORMACION DE RECURSOS HUMANOS

El Proyecto I&D en el que se enmarcan estas líneas de trabajo, es parte de un proceso de incentivación para el desarrollo de actividades de I&D, que se originan en el GIDAS (Grupo de I&D Aplicado a Sistemas informáticos y computacionales) del Departamento de Sistemas de Información de la UTN-FRLP. Consecuentemente, se está poniendo acento y esfuerzo en las siguientes actividades: - Desarrollo de seminarios abiertos de formación general en relación con temas, técnicas y tecnologías incluidos en esta línea de investigación, para alumnos avanzados en la carrera de Ingeniería en Sistemas y para becarios de este equipo de trabajo, realizados en UTN-Facultad Regional La Plata y UNLP - Facultad de Informática. A cargo de integrantes de este equipo de trabajo, se encuentran en curso 2 Tesis de Postgrado (Magíster en Ingeniería de Software, Facultad de Informáticas - UNLP) relacionadas con el campo de investigación de éste proyecto.
Los docentes integrantes de esta línea de investigación participan en el dictado de

asignaturas con contenidos relacionados a este campo I&D:
En grado: UTN-Facultad Regional La Plata, Ingeniería en Sistemas de Información y UNLP-Facultad de Informática.
En postgrado: UAI-Facultad de Tecnología Informática y UNLP-Facultad de Informática, Maestría en Ingeniería de Software.

5. BIBLIOGRAFÍA

[1] Portal Oficial de la SCTyP del Rectorado UTN con los PIDs homologados vigentes (código UTN5363): http://sicyt.scyt.rec.utn.edu.ar/scyt/proyectos/RESUMENF.ASP?VAR1=5363

[2] Portal Oficial de la SCTyP del Rectorado UTN con los Grupo de I&D homologados vigentes (Resolución CSU Nº 1770/18): https://utn.edu.ar/es/secretaria-sctyp/centros-y-grupos-utn/grupos-utn#LP

[3] Roxana Giandini, Matías Mangano, Lautaro Mendez, Leopoldo Nahuel. "La Producción de Software Dirigida por Modelos y la Filosofía Agil", PIPP, Año 2011.

[4] Pons Claudia, Giandini Roxana y Pérez Gabriela. "Desarrollo de Software Dirigido por Modelos: conceptos teóricos y su aplicación práctica". 1er. edición. EDULP & McGrawHill Education.

[5] Título del Proyecto: Lenguajes para transformaciones de Modelos en MDD. Acreditación: Proyectos de PID UTN 25-I025. Duración: 3 años. Fecha del Proyecto: 1/1/2006 hasta 31/12/2008. Disposición Resolutiva 119/08 para Programa de Incentivos. Grupo de trabajo integrado por 5 personas. Director: Roxana Giandini. Co-directores: C. Pons, C. López.

[6] Arie van Deursen, Paul Klint, Joost Visser, "Domain-Specific Language: An Annotated Bibliography", ACM SIGPLAN, Junio 2000.

[7] Ariste, C., Rocca, L., Caputti, M., Zugnoni, I. "Diseño de un Lenguaje Específico del Dominio para sistemas de Salud basado en estándares de interoperabilidad FHIR de HL7 y OpenEHR)". Congreso Argentino de Informática y Salud (2017).

[8] J. Rumbaugh, I. Jacobson, Grady Booch. "El Lenguaje Unificado de Modelado, Manual de Referencia". Addison Wesley, Primera Edición, 2000.

[9] UML: www.omg.org/spec/UML/

[10] Sirius – The easiest way to get your own Modeling Tool, https://www.eclipse.org/sirius/

[11] The object constraint language: getting your models ready for MDA. Jos B. Warmer, Anneke G. Kleppe. Editorial Addison-Wesley Professional, 2° Edition, 2003.

[12] Friedenthal S., Moore A., Steiner R., "A Practical Guide to SysML", The Systems Modeling Lang.

[13] Weske Mathias, "Business Process Management: Concepts, Languages, Architectures". Springer, Pag 3-67. ISBN 978-3-540-73521-2, 2008.

[14] Resourcelist – FHIR v4.0.0, https://hl7.org/fhir/resourcelist.html

[15] Open EHR https://www.openehr.org/

MÉTODOS EMERGENTES EN CALIDAD DEL SOFTWARE: PRUEBAS CONTINUAS Y MODELADO DE PROCESOS DE NEGOCIO

Irrazabal, Emanuel; Dapozo; Gladys N.; Greiner, Cristina;
Mascheroni, Agustín; Sambrana, Ivan; Company, Ana M.; Ferraro María; Medina, Yanina;
Badaracco Numa

Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste
{eirrazabal, gndapozo, cgreiner, eirrazabal, mascheroni}@exa.unne.edu.ar,
anamacom@gmail.com

RESUMEN

Este es el tercer año del proyecto F018-2017; una continuación de los proyectos F07-2009 y F10-2013, ambos enfocados en modelos, métodos y herramientas para la calidad del software. En este tercer período se trabaja temáticas de innovación y calidad de software, en particular, aspectos vinculados con la gestión de los procesos administrativos mediante herramientas software y en la entrega continua de productos software. El equipo de trabajo busca promover y generar métodos y herramientas que contribuyan a mejorar la calidad del proceso y del producto software.

En particular, se está trabajando en el desarrollo de un modelo de madurez para la gestión de las actividades de prueba continua dentro de la disciplina de entrega continua. Esto ha incluido el desarrollo de revisiones sistemáticas de la literatura y entrevistas a la industria. Finalmente también se atiende la problemática de la eficiencia de los sistemas en organismos públicos, mediante el enfoque de procesos de negocios y el desarrollo rápido de aplicaciones.

Palabras clave: calidad de software, gestión de procesos de negocio, entrega continua.

CONTEXTO

Las líneas de Investigación y Desarrollo presentadas en este trabajo corresponden al proyecto PI-17F018 “Metodologías y herramientas emergentes para contribuir con la calidad del software”, acreditado por la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE) para el periodo 2018-2021.

1. INTRODUCCIÓN

Las funciones de una aplicación o producto software se vuelven más complejas a lo largo del tiempo para satisfacer a los usuarios [1]. Esto genera avances constantes que se traducen en una gran cantidad de nuevas técnicas y tecnologías relacionadas con el desarrollo de software [2]. Pero el desafío sigue siendo llevar adelante un proceso de desarrollo que termine en un plazo acordado con unos atributos de calidad esperados, como, por ejemplo la mantenibilidad o la usabilidad.

En este sentido, la calidad del software está formada por diferentes factores y apoyados por herramientas que automatizan su medición. Por lo tanto, las nuevas tecnologías y técnicas emergentes muchas veces no tienen el conjunto de herramientas necesarias para la medición y mejora de la calidad del producto software resultante [3]. Por ello la necesidad de investigar en el uso y adecuación de las

tecnologías emergentes incorporando aspectos de calidad del software.

Autores como Pressman [4], McCall [5], y estándares tales como la ISO 25010 [6] han tratado de determinar y categorizar los factores que afectan a la calidad del software. Y en general se concluye que la calidad interna del producto software está directamente relacionada con diferentes atributos cuantificables a partir de su código fuente. Existen diferentes estudios que relacionan estos atributos con herramientas y métricas [7].

Asimismo el análisis del código fuente a partir de estas métricas puede ser tenido en cuenta para describir la complejidad del producto software y la probabilidad de encontrar errores [8]. Pero en las tecnologías emergentes esto no está del todo desarrollado. Así por ejemplo, no se encuentran herramientas o técnicas de pruebas software para productos software nativos móviles [3].

Siguiendo con este mismo enfoque y con el objetivo de lograr una mayor satisfacción del cliente, es crítico para las organizaciones entregar productos de calidad de manera aún más rápida. Esto dio lugar al surgimiento de un nuevo enfoque denominado "Entrega Continua de Software", más conocido en inglés como Continuous Delivery (CD). En este enfoque los equipos mantienen la producción de software en ciclos cortos de tiempo, asegurando que el producto pueda ser lanzado de manera fiable en cualquier momento [10] [11].

Nuevamente, uno de los principales problemas de estos enfoques estaría en la calidad del producto software. Ésta puede disminuir, dado que, al realizarse los despliegues del sistema con mayor frecuencia, aparecen más defectos en el producto [11]. Por tanto, es esencial desarrollar un enfoque de priorización de los diferentes aspectos en la calidad del producto software, teniendo en cuenta la forma de trabajo actual de las empresas de desarrollo software.

Para adoptar este enfoque, se utiliza el concepto de Tubería de Despliegue (DP - Deployment Pipeline), un estándar para automatizar el proceso de CD [12].

Desde el enfoque de los procesos de construcción también existen desafíos, no solamente desde el sector privado sin también en el sector público. La gestión de procesos de negocio se basa en la idea de que cada producto es el resultado de un conjunto de actividades que se realizan a fin de obtener dicho producto. Por este motivo, la correcta y eficiente gestión de los procesos de negocio es un aspecto importante para la productividad de toda organización, ya que permite identificar las tareas, el orden de ejecución de las mismas y las personas responsables de realizarlas [13].

Por lo tanto, las líneas de trabajo del proyecto tienen que ver con características de calidad de calidad del producto software y del proceso de desarrollo.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

A continuación, se describen las siguientes líneas de investigación y desarrollo:

Entrega continua

Se está desarrollando de un conjunto buenas prácticas nucleadas en el marco de una propuesta de modelo de madurez para pruebas continuas. Asimismo se está trabajando en el desarrollo de los procedimientos que cumplan con estas buenas prácticas teniendo en cuenta los principales desafíos de esta disciplina y los enfoques más actuales para agilizar el proceso de pruebas. En esta línea estamos especialmente orientados a las pruebas de tipo API Rest, las pruebas de visualización en navegadores y las pruebas de tipo flaky. Asimismo se está realizando un estudio de caso junto con el área de desarrollo software de la iniciativa SIU. Como primera medida se han realizado revisiones

sistemáticas de la literatura y como segundo paso el desarrollo de encuestas a grandes empresas mundiales.

El paso siguiente es la creación de un primer modelo formado por distintas etapas de pruebas basándose en la Tubería de Despliegue. Para validar el modelo, se procederá con el desarrollo de un marco de trabajo que implemente estas pruebas. Se hará una serie de experimentos con este marco para comparar la velocidad de ejecución de las pruebas contra las convencionales. Y, finalmente, se propone implementar el modelo en un ambiente real de desarrollo, para analizar los resultados obtenidos.

Gestión de procesos de negocios en el sector público

El desarrollo de sistemas en la Administración Pública es constante y cada vez se busca una mayor integración con la forma de trabajo del dominio de aplicación. Esto asegura el apoyo en la estructura del estado y su cultura organizacional [13].

En este sentido existen diferentes estrategias de modelización para los requisitos de software y aún antes, para la descripción de los pasos actuales del trabajo administrativo que se busca sistematizar. Una de las técnicas que han demostrado ser adecuadas es el Modelado de Procesos de Negocio [14].

En esta línea de investigación se busca realizar un marco de trabajo para la gestión de procesos de desarrollo software que favorezca la usabilidad en el ámbito interno del proceso administrativo de las instituciones públicas. Por un lado, el modelado del dominio de trabajo, su mejora, la propuesta de notación orientada al dominio de trabajo y por otro lado el desarrollo de procedimientos que permita la notación *Bussiness Process Model* (BPM).

3. RESULTADOS OBTENIDOS/ESPERADOS

En el marco de este proyecto y respecto de la línea de gestión de procesos se lograron los siguientes objetivos:

1. Relevamiento de información relevante sobre los conceptos involucrados, procesos de negocios y usabilidad
   En primer lugar, se realizó una revisión sistemática de la literatura (RSL) con el objetivo de sintetizar la literatura existente sobre procesos de desarrollo de software basados en procesos de negocio y su vinculación con el diseño y la usabilidad de las aplicaciones. Las fuentes seleccionadas: IEEE Xplore, Science Direct, en el área de Computer Science y ACM Digital Library.
   El conocimiento extraído de esta RSL será la base para la construcción de un marco de trabajo para el desarrollo de software basado en BPM para la mejora de la usabilidad de aplicaciones [15].

2. Diseño de un conjunto de procedimientos de gestión de requerimientos, diseño y desarrollo de software basado en la mejora de la usabilidad del sistema. Se seleccionó como caso de estudio el proceso de gestión de viáticos de empleados de un organismo estatal provincial.

En la línea de Ingeniería de Software para la Entrega Continua:

1. En [16] se realizó una encuesta a más de 200 empresas y equipos de trabajo de América y Eurupoa. Se resumieron e identificaron los principales problemas y desafíos que afectan a la calidad del software en las disciplinas de entrega continua y pruebas continuas teniendo en cuenta lo obtenido en revisiones sistemáticas de

la literatura publicadas en años anteriores.

2. En [17] se realizó una revisión sistemática de la literatura para obtener los mecanismos de personalización de técnicas ágiles que favorezcan la validación de los requisitos software.
3. En el marco de una tesis de maestría se realizó el estudio terciario de revisiones sistemáticas de la literatura para identificar los tópicos de investigación en torno a la integración continua, los métodos y herramientas más utilizados. Se encontraron 14 revisiones sistemáticas, las cuáles hicieron hincapié en la entrega y pruebas continuas, especialmente en la automatización de las mismas.
4. Se está planificando el desarrollo de un experimento controlado para conocer la percepción de la importancia que tienen las diferentes características de calidad del producto software a lo largo de la vida profesional. Se realizará el estudio mediante una encuesta a estudiantes de grado, de posgrado y profesionales del medio. Se utilizará como instrumento de recolección la comparación por pares de características; una técnica que ha demostrado tener una precisión adecuada [18] y que ha sido utilizada en experimentos similares [19].

En general, para esta línea de trabajo se espera desarrollar un modelo de buenas prácticas para la gestión de pruebas continuas en equipos de trabajo que tengan implantados la entrega continua de software así como un ecosistema de herramienta verificado en por lo menos un marco de trabajo real.

4. FORMACIÓN DE RECURSOS HUMANOS

En el Grupo de Investigación sobre Calidad de Software (GICS) están involucrados 6 docentes investigadores, 1 tesista de doctorado, 1 becario de investigación de pregrado y 2 tesistas de maestría que finalizaron el cursado en el 2017.

5. REFERENCIAS

[1] M. Lehman, "On Understanding Laws, Evolution and Conservation in the Large Program Life Cycle", J. of Sys. and Software, vol. 13, pp. 213-221, 1980.

[2] Randell, Brian. "Fifty Years of Software Engineering-or-The View from Garmisch." arXiv preprint arXiv:1805.02742 (2018).

[3] Kong, Pingfan, et al. "Automated testing of android apps: A systematic literature review." IEEE Transactions on Reliability 99 (2018): 1-22.

[4] R. Pressman , Ingeniería de Software. Un Enfoque Práctico, Sexta ed., MCGraw Hill, 2005.

[5] J. A. McCall, P. K. Richards y G. F. Walters, Factors in Software Quality, 1997.

[6] International Organisation for Standardisation, ISO/IEC 25010, Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models;Geneva, Switzerland, 2011.

[7] Arvanitou, Elvira Maria, et al. "A mapping study on design-time quality attributes and metrics." Journal of Systems and Software127 (2017): 52-77.

[8] R. Malhotra y A. Jain, «Software Fault prediction for Object Oriented Systems: A Literature Review,» ACM SIGSOFT Software Engineering Notes, vol. 36, nº 5, pp. 1-6 , 2011.

[9] Y. C. Cavalcanti, P. A. da Mota Silveira Neto y . do Carmo Machado, «Towards Understanding Software Change Request Assignment: a survey

with practitioners,» de Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering, Porto de Galinhas, Brazil, 2013

[10] L. Chen, "Continuous Delivery: Huge Benefits, but Challenges Too" in IEEE Software 03/2015. V. 32(2).

[11] J. Humble and D. Farley. “Continuous delivery: reliable software releases through build, test, and deployment automation”, 1st ed. Boston, US: Pearson Education, 2010.

[12] O. Prusak."Continuous Testing: The Missing Link in the Continuous Delivery Process". Blaze Meter. 2015. https://blazemeter.com/blog/continuous-testing-missing-link-continuous-delivery-process

[13] Corradini, Flavio, et al. "Innovation on public services using business process management." InternationalConferenc e on E-business, Management and Economics. Vol. 25. 2011.

[14] Bicevskis, Janis, and Zane Bicevska. "Business process models and information system usability." Procedia Computer Science77 (2015): 72-79.

[15] Moreno, Juan Carlos, Marcelo Martin Marciszack, and Juan Pablo Fernandez Taurant. "Usability through Business Model." 2016 11th Iberian Conference on Information Systems and Technologies (CISTI). IEEE, 2016.

[16] Mascheroni, Maximiliano, Irrazábal, Emanuel Carruthers, Juan, Pinto, Juan. Rapid Releases and Testing Problems at the industry: A survey. XXV Congreso Argentino de Ciencias de la Computación. CACIC 2019. Río Cuarto, octubre 2019.

[17] Pereyra Coimbra, Rodrigo, Carruthers, Juan, Pinto, Juan, Irrazábal, Emanuel. Personalización de técnicas ágiles en el desarrollo de software para la obtención de requisitos de calidad en Pymes: una revisión sistemática de la literatura. XX Simposio Argentino de Ingeniería de Software (ASSE 2019)- JAIIO 48. Salta, Septiembre de 2.019.

[18] Perini, A., Ricca, F., & Susi, A. (2009). Tool-supported requirements prioritization: Comparing the AHP and CBRank methods. Information and Software Technology, 51(6), 1021-1032.

[19] Febrero, F., Moraga, M. A., & Calero, C. (2017, July). Software Reliability as User Perception: Application of the Fuzzy Analytic Hierarchy Process to Software Reliability Analysis. In *2017 IEEE International Conference on Software Quality, Reliability and Security (QRS)* (pp. 224-231). IEEE.

Ingeniería de software para sistemas embebidos, sistemas críticos e Internet de las Cosas

Irrazábal, Emanuel; Bernal, Rubén; Ríos, José; Mascheroni, Maximiliano Agustín; Sambrana, Iván; Alonso, José Manuel ; Acevedo, Joaquín

Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura. Universidad Nacional del Nordeste

eirrazabal@exa.unne.edu.ar

Resumen

Esta línea de investigación aborda temas de ingeniería del software desde hace tres años, tanto en sistemas tradicionales y su traslado a sistemas embebidos y sistemas críticos. Se espera la aplicación de la tecnología de desarrollo software en sistemas embebidos para el ámbito regional. En este sentido se ha realizado la construcción de un modelo de procedimientos para la gestión de requerimientos en entidades agrícolas con una cultura organizacional jerárquica, específicamente las entidades yerbateras del nordeste argentino. Por otro lado en el estudio de arquitecturas software para el desarrollo de sistemas críticos con aplicación al ámbito ferroviario nacional. Asimismo se está trabajando en el desarrollo de sistemas embebidos con tecnología de comunicación LoRa y la construcción de prototipos para la prevención de incendios forestales.

Palabras clave: Ingeniería de software, sistemas embebidos, sistemas críticos

Contexto

La línea de Investigación y Desarrollo presentada en este trabajo corresponde al proyecto PI-F17-2017 "Análisis e implementación de tecnologías emergentes en sistemas computacionales de aplicación regional.", acreditado por la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE) para el periodo 2018-2021.

Introducción

Ingeniería de software para sistemas embebidos y sistemas críticos

Un sistema embebido es aquel sistema basado por lo general en un microprocesador, sensores y actuadores diseñado para realizar funciones dedicadas [1]. Y han cobrado gran importancia desde el punto de vista de los sistemas de información con el uso de plataformas tipo Arduino para el desarrollo rápido de prototipos [2].

En este sentido, el grupo de investigación desarrolla de soluciones para entidades regionales, en particular para la agricultura, las entidades públicas provinciales y los grupos de investigación de la universidad.

Uno de los objetivos para este año es la innovación en tecnología de incendios forestales. En este sentido, el nordeste tiene una alta incidencia de incendios; se estima que anualmente se queman aproximadamente 2.000.000 ha. Es común la quema en parches de grandes extensiones próximas a aguadas. Quemas mal manejadas que frecuentemente alcanzan a los bosques nativos y cultivados [3].

Las labores para el manejo y control del fuego forestal incluyen la detección como un primer paso. [4].

La importancia de lograr la detección temprana de incendios forestales (DTIF) radica en su fácil eliminación a un bajo costo, antes de que se quemen grandes extensiones de bosque. La DTIF está concebida para dar la alarma cuando se ha iniciado el fuego. Por esta razón es necesario utilizar sensores para las principales variables indicadoras de la existencia del fuego: la variable química (emisiones de humo, CO, CO2, CH4) y la variable física (radiación infrarroja y el calor) [5].

A su vez, las grandes extensiones de campo hacen imposible el uso de conectividad preexistente vía GSM o GPRS. Es decir, es necesario plantear una solución que tenga en cuenta la portabilidad de los sensores, su adecuación al ambiente forestal de la región y una solución de conectividad simple.

Asimismo, en el marco de los sistemas embebidos se encuentran los sistemas críticos cuyo mal funcionamiento o fallo pueden resultar en pérdida de vidas humanas, equipos, instalaciones o daños medio ambientales severos [6]. Por ello, uno de los ámbitos de mayor utilización de sistemas críticos es el ferroviario.

Una normativa importante para el desarrollo de software crítico ferroviario es la norma EN 50128 [7]. En la misma se trata la calidad del software en términos del ciclo de vida del desarrollo, especificando los procedimientos y requisitos técnicos. Abarca los aspectos de seguridad en cinco niveles desde el 0 al 4, de acuerdo a la gravedad de las consecuencias del fallo, siendo el nivel 4 el que tiene asociado una mayor cantidad de técnicas y características a cumplir. A estos niveles de seguridad se los conoce con la sigla en inglés SSIL derivada de la expresión *software safety integrity level*.

En particular, en el apartado 7.4 de la norma EN 50128 se especifican las buenas prácticas al construir la arquitectura del software en términos de actividades, documentación, especificación integral de cada módulo de la arquitectura y uso de buenas prácticas de programación.

En este sentido, los patrones de arquitectura que cumplen con las buenas prácticas antes mencionadas pueden ser complejos [8] y, a su vez, disminuir aspectos relacionados con la calidad del producto software, especialmente la mantenibilidad [9].

El propósito de esta línea de trabajo es, por tanto, desarrollar procedimientos aplicables a la construcción de sistemas críticos, así como su verificación y validación en ensayos junto con la Autoridad Ferroviaria Nacional, teniendo en cuenta el cumplimiento de la normativa EN 50128.

Ingeniería de Software para Entrega Continua

Tal y como lo describe la segunda ley de Lehman [10], el software necesita ser cada vez más complejo para satisfacer las necesidades de los usuarios. Esto, sumado a la importancia de detectar errores en la etapa de captura de requerimientos por sus costos [11], hace a la validación de los requerimientos una etapa crítica en el desarrollo de las aplicaciones [12].

Asimismo, las empresas regionales, como por ejemplo las empresas yerbateras o agrícolas tienen una estructura jerárquica tradicional, donde se identifican necesidades para gestión y seguimiento de tareas relacionadas a la Ingeniería de Software [13]. Estas empresas en su mayoría cuentan con una dirección que toma las decisiones, el cual conoce el negocio, pero no cuenta con el tiempo o la heurística suficientes para manifestar sus necesidades.

En este sentido, los desarrollos de metodologías ágiles representan un avance en la manera de construir sistemas; usando métodos, como, por ejemplo: el prototipado, las historias de usuario y los casos de uso. Esto hace posible la entrega temprana de valor, la respuesta rápida a los cambios y la colaboración constante del equipo de trabajo con los clientes [14][15].

Entonces, haciendo foco en la gestión de los requerimientos, es posible adaptar técnicas ágiles las cuales tendrán un fuerte impacto positivo [16][17], especialmente en PYMEs con problemas al momento de la gestión de sus requerimientos.

Las herramientas y prácticas genéricas deben ser seleccionadas y adaptadas para la empresa, tamaño y dominio específico. Y así, reducir los problemas causados por el no alineamiento tales como requisitos implementados incorrectamente, retrasos y esfuerzos desperdiciados [18].

Desde otro punto de vista, en el ámbito de fábricas de software se busca llevar adelante el desarrollo de las soluciones software con un enfoque evolutivo. Esto es cubierto por el enfoque de la ingeniería del software conocido como Entrega Continua de Software, en inglés Continuous Delivery (CD). Este es un enfoque en el cual los equipos mantienen la producción de software en ciclos cortos de tiempo, asegurando que el producto pueda ser lanzado de manera fiable en cualquier momento [19]. El objetivo es poder lanzar a producción un producto software libre de defectos "*con solo apretar un botón*" [20].

Uno de los principales desafíos de este enfoque es la calidad del producto software resultante. Ésta puede disminuir, dado que, al realizarse los despliegues del sistema con mayor frecuencia, aparecen más defectos en el producto. Por tanto, es esencial desarrollar un enfoque de priorización de los diferentes aspectos en la calidad del producto software, teniendo en cuenta la forma de trabajo actual de las empresas de desarrollo software.

Líneas de investigación y desarrollo

En la línea de Ingeniería de Software para Sistemas Embebidos y Sistemas Críticos:

- Diseñar y construir el prototipo funcional de sensores para la detección de focos de incendio y la construcción del software para la gestión de la red de sensores.
- Diseñar y construir un prototipo de antena Gateway LoRaWAN para la recepción de alarmas por parte de los sensores, que facilite la comunicación en entornos sin otro tipo de conectividad, así como la visualización de las alarmas.
- Realizar pruebas de concepto para la construcción de arquitecturas software y diseños específicos con patrones orientados a sistemas críticos de altas prestaciones.
- Planificar y ejecutar una Revisión Sistemática de la Literatura para identificar los estudios empíricos donde se verifican diferentes arquitecturas software de sistemas críticos y se comprueba el nivel de SSIL alcanzado.

En la línea de Ingeniería de Software y Entrega Continua:

- Utilizar el modelo de procedimientos para la gestión de requerimientos en PYMEs yerbateras a partir de técnicas ágiles.
- Desarrollar un modelo de madurez para el diseño y ejecución de pruebas continuas de software, especialmente orientado a flaky test y crossbrowsing testing.

Resultados obtenidos

El grupo de investigación es de reciente formación, por lo cual se enumeran los resultados de este último año. A continuación se indican:

En la línea de Ingeniería de Software para Sistemas Embebidos y Sistemas Críticos:

- En [21] se desarrolló un conjunto de procedimientos para la gestión de requisitos software en proyectos de sistemas críticos ferroviarios cumpliendo las buenas prácticas de la normativa EN 50128.
- En [22] se diseñó una prueba de concepto para arquitecturas software de sistemas críticos y se evaluaron los resultados en cuanto a complejidad y características del código fuente

resultante. De esta manera se pudo obtener un resultado comparable entre los diferentes patrones de arquitectura para software crítico, el nivel de seguridad asociado, la complejidad de su construcción y el grado de mantenibilidad resultante.

Asimismo, se está trabajando en el desarrollo de una red de sensores para la lectura automática de medidores eléctricos con comunicación LoRaWAN en el marco de un Acuerdo de Trabajo entre la Facultad de Ciencias Exactas y Naturales y Agrimensura (FACENA- UNNE) y la Dirección Provincial de Energía de la provincia de Corrientes [23].

En la línea de Ingeniería de Software para la Entrega Continua:

- En [24] se realizó una encuesta a más de 120 empresas y equipos de trabajo de todo el mundo para identificar el conjunto de problemas que afectan a la calidad del software en las disciplinas de entrega continua y pruebas continuas comparándolo con lo obtenido en Revisiones Sistemáticas de la Literatura anteriores
- En [25] se realizó una Revisión Sistemática de la Literatura para identificar las estrategias de personalización en las técnicas ágiles asociadas con la gestión de requisitos software.

Formación de recursos humanos

En el Grupo de Investigación en Innovación en Software y Sistemas Computacionales (GIISSC) están involucrados 5 docentes investigadores, 1 becario de investigación de pregrado, 1 tesista de doctorado y 3 tesistas de maestría. Tres alumnos de la carrera están realizando sus proyectos finales vinculado a estos temas.

La tesis de maestría seleccionada para ser presentada llevó el titulo: “Integración de múltiples herramientas de software open source para la gestión y análisis de rendimiento de clúster en ambiente HPC”, fue realizada por el Esp. Leopoldo José Ríos y dirigida por el Dr. Fernando Gustavo Tinetti.

El objetivo de la tesis fue el estudio, diseño e implementación de un sistema informático que integre múltiples herramientas de software Open Source para el sistema operativo Linux permitiendo la gestión global de recursos computacionales disponibles en un ambiente de Clúster HPC. El nivel de gestión que se desea obtener con esta aplicación, permitirá a los responsables operativos del Clúster, poder evaluar condiciones de usabilidad y planificación estratégica.

Los datos a obtener, provienen de los sistemas operativos que controlan el funcionamiento de la infraestructura general, determina la visualización del estado en tiempo real de Memoria RAM, Procesadores, estados de cachés, Consumo de la red de interconexión, entre otros. Serán utilizados también, programas como ‘perf’, que miden el rendimiento que tiene una aplicación mientras es ejecutada. Los resultados obtenidos con ‘perf’ serán almacenados e integrados a aquellos obtenidos desde el sistema operativo.

El almacenamiento de los datos, se hará sobre una base de datos de series de tiempo, con la herramienta RRDtool específicamente, a los efectos de poder elaborar y organizar información numérica a lo largo del tiempo. La misma herramienta nos servirá tanto para el proceso de registración, como para el proceso de generación de gráficos sobre consumos y usos que se consideren útiles para la gestión y análisis de rendimiento del clúster.

El tratamiento y gestión de los datos obtenidos, permitirá estudiar la demanda de equipamiento en un determinado período de tiempo, por ejemplo, un (1) año hacia atrás. Será posible la selección de criterios de búsqueda, ejemplo: uso de memoria RAM, uso de CPU, y los programas

utilizados en ese hecho, o usuarios del sistema involucrados.

Desarrollo y resultado.

La actividad de desarrollo fue solapada con actividades de gestión propias del clúster, del instituto tomado como caso de estudio de este trabajo. En el ámbito del instituto, la premisa siempre fue soportar el sistema de gestión de trabajos del clúster, para que los trabajos fluyan en el tiempo. La autenticación de usuarios fue desarrollada con técnicas RBAC (control de acceso basado en roles), con acceso a la ejecución de tareas por encima del acceso a objetos, es altamente aceptable y ha sido demostrada su efectividad. La plataforma generada con el framework "CakePHP", con un despliegue modelo vista controlador, constituye la base de funcionamiento del sistema, y es ampliable a tantas funcionalidades como requerimientos de los usuarios exista.

Las vistas definidas en la aplicación, es el resultado de procesos internos que se disparan en función a la lógica de negocio impuesto por los permisos, asociados a roles, y éstos a usuarios. La aplicación del concepto RBAC se ha dado de manera muy simple, la experiencia indica que es necesario concentrarse en la identificación de roles, en la declaración y definición de permisos, y en la asignación a las cuentas de usuario de los roles establecidos, sin importar la cantidad de usuarios que sean. La herramienta Torque PBS se integra a la aplicación propuesta, como el módulo de gestión de trabajos en modalidad batch del clúster. Se encarga de la organización, clasificación, ejecución y finalización de trabajos de usuario. Provee una serie de comandos que hacen de interface al ser ejecutados, informan en tiempo real el estado de los trabajos de usuario; lo cual registrado por la aplicación para reunir información, registrarla y difundirla al usuario bajo demanda.

En lo referente a la administración de datos, se ha logrado gestionar datos de usuario como ser scripts, historiales de trabajos ejecutados, con la herramienta de base de datos relacional MySQL. Los datos de monitoreo en línea se gestionan con la herramienta RRDtool, en la visualización de los distintos estados del sistema. Es posible demostrar que las herramientas se complementan perfectamente bien, sin observar interferencias.

Publicaciones relacionadas a la tesis.

Año 2017: V Jornadas de Cloud Computing & Big Data, desarrolladas en la Facultad de Informática de la UNLP, del 26 al 30 de junio de 2017. Trabajo "Acceso, Monitorización y Asistencia para el Desarrollo en Clusters para HPC: posible utilización en múltiples clusters y clouds".

Año 2016: "Round Robin Data Bases for Performance Evaluation of High Performance Applications and Clusters", Fernando G. Tinetti, Leopoldo J. Rios, The 2016 International Conference on Grid, Cloud, and Cluster Computing (GCC'16), July 25-28, 2016, Monte Carlo Resort, Las Vegas, USA., Eds. Hamid R. Arabnia, Ashu M. G. Solo, Fernando G. Tinetti, ISBN: 1-60132-436-7, CSREA Press, pp 45-48.

Año 2015: "Instalación, Configuración y Evaluación de un Cluster HPC en una Nube Pública", Fernando G. Tinetti, Leopoldo J. Rios, III Jornadas de Cloud Computing & Big Data, 29 de junio al 3 de julio de 2015, Fac. de Informática de la UNLP.

Conclusiones.

El enfoque de este trabajo de tesis fue elaborar una herramienta de gestión de clúster HPC basada en código abierto, para lo cual se ha incorporado otras herramientas de software que propician esta idea. El diseño previsto permite la interconexión entre las herramientas de manera muy simple, con la posibilidad de reemplazo de alguna de ellas y escasa re-configuración posterior. Es posible su migración a cualquier plataforma operativa: RedHat, Ubuntu, CentOS, y de gestión de datos como PostgreSQL. El resultado se pone a disposición de la comunidad para

permitir fácilmente la programación de tareas a ser desplegada por usuarios localizados en Internet. La programación de tareas requiere el conocimiento de programación PHP y MySQL para gestión de datos.

La aplicación permite gestionar equipos de computación configurados en clúster HPC, como un todo, o como partes de un todo. En el escenario del caso de estudio, el ámbito es el clúster completo: todos los nodos configuran colas de trabajo, que se gestionan desde la misma aplicación. Quizá en otros escenarios, determinen el uso de una parte del clúster, entonces la aplicación puede ajustarse específicamente a esos recursos, que luego los usuarios darán uso.

Referencias

[1] Heath, Steve. *Embedded systems design*. Elsevier, 2002.

[2] Jamieson, Peter. "Arduino for teaching embedded systems. are computer scientists and engineering educators missing the boat?." Proc. FECS 2010 (2010): 289-294.

[3] Minaverry, Clara María. La regulación de los incendios forestales en Argentina y Uruguay. Ars Boni et Aequi, 2016, vol. 12, no 1.

[4] Loureiro, Maria L.; Allo, Maria. Los Incendios Forestales Y Su Impacto Económico: Propuesta Para Una Agenda Investigadora. Revista Galega de Economía, 2018, vol. 27, no 3, p. 129-142.

[5] Oviedo, Byron, et al. Redes inalámbricas de sensores para detección temprana de incendios forestales. Revista Científica Ciencia y tecnología, 2019, vol. 18, no 20.

[6] EN 50126. Railway applications - The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS). 2005.

[7] EN 50128. Railway applications - Communication, signalling and processing systems - Software for railway control and protection systems. 2011.

[8] W. Wu y T. Kelly, «Safety tactics for software architecture design», en Proceedings of the 28th Annual International Computer Software and Applications Conference, 2004. COMPSAC 2004., 2004, pp. 368–375.

[9] L. Bass, P. Clements, y R. Kazman, Software architecture in practice. Addison-Wesley Professional, 2003.

[10] M. Lehman, "On Understanding Laws, Evolution and Conservation in the Large Program Life Cycle", J. of Sys. and Software, vol. 13, pp. 213-221, 1980.

[11] S. Maalem and N. Zarour, "Challenge of validation in requirements engineering", Journal of Innovation in Digital Ecosystems, vol. 3, no.1,pp. 15-21, 2016.

[12] G. Kotonya and I. Sommerville, Requirements engineering. Chichester: John Wiley & Sons, 1998.

[13] Plan Estratégico para el Sector Yerbatero – Yerba Mate Argentina", Yerbamateargentina.org.ar, 2016.

[14] A. Sillitti and G. Succi, Requirements Engineering for Agile Methods. In: Aurum A., Wohlin C. (eds) Engineering and Managing Software Requirements. Berlin, Heidelberg: Springer, 2005.

[15] I. Inayat, S. Salim, S. Marczak, M. Daneva and S. Shamshirband, "A systematic literature review on agile requirements engineering practices and challenges", Computers in Human Behavior, 2014.

[16] S. Dragicevic, S. Celar and L. Novak, "Use of Method for Elicitation, Documentation, and Validation of Software User Requirements (MEDoV) in Agile Software Development Projects", CICN 2014.

[17] One, V. “State of agile development survey results”, 2017.

[18] E. Bjarnason, P. Runeson, M. Borg, M. Unterkalmsteiner, E. Engström, B. Regnell, G. Sabaliauskaite, A. Loconsole, T. Gorschek and R. Feldt, "Challenges and practices in aligning requirements with verification and validation: a case study of six companies", Empirical Software Engineering, vol. 19, no. 6, pp. 1809-1855, 2013.

[19] L. Chen, "Continuous Delivery: Huge Benefits, but Challenges Too" in IEEE Software 03/2015. V. 32(2).

[20] J. Humble and D. Farley. “Continuous delivery: reliable software releases through build, test, and deployment

automation", 1st ed. Boston, US: Pearson Education, 2010.
[21] Pinto Luft, Cristian: Construcción de una metodología de gestión de requerimientos software y desarrollo de un ecosistema de herramientas de acuerdo con la norma EN 50128. Aplicación en el desarrollo de un prototipo para la Autoridad Ferroviaria Nacional. Maestría en Tecnologías de la Información FaCENA – UNNE. Tesis de Maestría
[22] Acevedo Duprat Joaquín, Irrazábal Emanuel, Evaluación de la calidad en patrones de arquitecturas software para sistemas críticos ferroviarios: un enfoque basado en la mantenibilidad, XXV Congreso Argentino de Ciencias de la Computación. Río Cuarto (Argentina).Octubre de 2019.
[23] Acuerdo de Cooperación entre la FACENA y la Dirección Provincial de Energía de Corrientes para el desarrollo de un prototipo de comunicación entre medidores domiciliarios con colector en subestación y centro para el registro remoto del consumo y otros parámetros en tiempo real. Res. 0766/19 CD Facena.
[24] Mascheroni, Maximiliano, Irrazábal, Emanuel Carruthers, Juan, Pinto, Juan. Rapid Releases and Testing Problems at the industry: A survey. XXV Congreso Argentino de Ciencias de la Computación. CACIC 2019. Río Cuarto, octubre 2019.
[25] Pereyra Coimbra, Rodrigo, Carruthers, Juan, Pinto, Juan, Irrazábal, Emanuel. Personalización de técnicas ágiles en el desarrollo de software para la obtención de requisitos de calidad en Pymes: una revisión sistemática de la literatura. XX Simposio Argentino de Ingeniería de Software (ASSE 2019)-JAIIO 48. Salta, Septiembre de 2.019

Mejora de procesos, Calidad de Software y Gobernanza Digital

Silvia Esponda, Ariel Pasini, Marcos Boracchia,
Julieta Calabrese[1], Rocío Muñoz[1], Juan Santiago Preisegger[1], Patricia Pesado

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
[1] Becario postgrado UNLP
50 y 120 La Plata Buenos Aires
Centro Asociado CIC
526 e/ 10 y 11 La Plata Buenos Aires
(sesponda, apasini, marcosb, jcalabrese, rmunoz, jspreisegger, ppesado) @lidi.info.unlp.edu.ar

RESUMEN

La constante demanda de la industria por lograr un software en plazos cada vez más acotados y con altos niveles de calidad, requiere que las organizaciones dirijan la atención hacia la mejora de procesos de software como una forma de aumentar la calidad, reducir sus costos o acelerar los procesos de desarrollo.
La aplicación de buenas prácticas, logran mejorar las organizaciones desde diferentes aristas, el proceso, su producto/servicio e incluso los datos que utiliza y genera a lo largo del proceso.

El III-LIDI (Instituto de Investigación en Informática LIDI) posee un grupo dedicado a la investigación y desarrollo de tópicos concernientes con la mejora de los procesos de gestión, la calidad en los procesos de desarrollo y productos de software, sus datos y gobierno digital.
En este sentido, se ha trabajado en la mejora de procesos de desarrollo del software y en el estudio de estándares internacionales de Calidad de Producto, haciendo hincapié en los datos utilizados y se han investigado y desarrollado distintos servicios de gobierno digitales (con foco en políticas de prestación de servicios y herramientas de uso que beneficien a la comunidad).

Palabras Claves

Ingeniería de Software – Calidad del proceso – Calidad del producto – Calidad de Datos - Normas de Calidad – ISO - Gobernanza Digital

CONTEXTO

Esta línea de investigación se enmarca en el proyecto "11/F023 Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso" y en el subproyecto "Gobernanza Digital. Mejora de Procesos" (2018-2021), acreditado por el Ministerio de educación de la Nación.
Además, la línea se encuentra abordada por el proyecto "Calidad de Datos" aprobado por la Facultad de Informática UNLP.
El Instituto posee diversos acuerdos de cooperación con varias Universidades de Argentina y del exterior y con empresas privadas del sector, interesadas en mejorar sus procesos de desarrollo aplicando mejoras y participa en proyectos internacionales, entre ellos ERASMUS "Fortalecimiento de la capacidad de gobernanza de ciudades sostenibles inteligentes".

1. INTRODUCCION

En la actualidad, la gestión de la calidad es una actividad fundamental dentro de cualquier organización. Dicha gestión

puede llegar a ser una ventaja competitiva que la fortalezca en el momento de prestar un mejor servicio o de tener un producto que cumpla las exigencias y expectativas del cliente. Puntualmente, la industria del software es una de las industrias de más rápido crecimiento en las últimas décadas, y aunque la producción de software tiene a su disposición un conjunto de normas específicas destinadas a evaluar diferentes aspectos del proceso y/o producto, muy a menudo es necesario establecer un Sistema de Gestión de la Calidad (SGC) de acuerdo con la norma IRAM-ISO 9001. Los problemas relacionados con alcanzar niveles óptimos de calidad en diferentes aspectos del software son de amplio conocimiento, y es mayor la criticidad, evidenciando la gran variedad de información que las diferentes organizaciones manejan, aspecto clave para la toma de decisiones.

Estas dependen directamente de los datos con los que operan. No disponer de recursos que evalúen la calidad de estos es uno de los principales problemas a los que se enfrentan actualmente, ya que influye de manera significativa en la efectividad y eficiencia de la organización y del negocio.

Existen diferentes estrategias de enfocar la calidad de estos servicios, y un conjunto de normas relacionadas a la calidad del software que se pueden clasificar en tres grupos explicados a continuación.

El primer grupo está comprendido por las asociadas a la calidad del producto de software: Dentro de este grupo se cuenta con una norma específica para evaluar la calidad de los datos utilizados, el segundo son normas relacionadas al proceso de desarrollo del software y el tercer grupo, más genérico que los anteriores, son las normas relacionadas con la gestión de la organización desarrolladora de software

Por otro lado, una ciudad digital es aquella en la que, utilizando los recursos propios de la infraestructura de telecomunicaciones y de la informática existentes, brinda a sus habitantes un conjunto de servicios digitales a fin de mejorar el nivel de desarrollo humano, económico y cultural de esa comunidad a nivel individual y colectivo. La infraestructura para brindar estos servicios, tales como redes telefónicas e Internet, se sustenta en los recursos propios de los usuarios e instituciones. Ligado al concepto de ciudad digital, se encuentran los términos gobierno y gobernanza digital.

La gobernanza se lleva a cabo a través de la prestación de servicios públicos. Se define como servicio público a la "actividad llevada a cabo por la administración o por una organización (con un cierto control) destinada a satisfacer necesidades de la comunidad." Si dichos servicios son brindados a través del uso de TICs, son considerados servicios públicos electrónicos. [1]

Dentro de la línea de la investigación del proyecto, se destacan los siguientes ejes principales:

1- Mejora de Procesos de gestión en el desarrollo de software

La gestión de la calidad es una actividad fundamental dentro de cualquier organización. Dicha gestión puede llegar a ser una ventaja competitiva que la fortalezca en el momento de prestar un mejor servicio o de tener un producto que cumpla las exigencias y expectativas del cliente. En el camino de obtener una buena gestión de la calidad de procesos en las PyMEs, es necesario establecer un Sistema de Gestión de la Calidad (SGC) de acuerdo con la norma IRAM-ISO 9001.

Bajo este contexto, la norma IRAM-ISO 9001:2015 "Sistemas de gestión de la calidad. Requisitos" [4] se funda en los retos que enfrentan las empresas de cualquier tamaño y sector hoy en día, centrándose en la eficacia del sistema de gestión para dar cumplimiento a los requisitos del cliente. La norma promueve la adopción de un enfoque basado en

procesos y exige cierta información documentada que la organización determine como necesaria para la eficacia del SGC.

2- Calidad en productos - Calidad de datos.

El número de empresas desarrolladoras de software ha experimentado un fuerte crecimiento, juntamente con el incremento de la demanda de productos del sector. Para este tipo de empresas, la calidad del software tiene un papel fundamental y las actividades relacionadas con la calidad de software junto con su evaluación están cobrando cada vez más importancia.

Esto nos lleva a centrar nuestra atención en los estándares definidos por ISO. La producción de software tiene a su disposición un conjunto de normas específicas destinadas a evaluar diferentes aspectos del software. La familia ISO/IEC 25000 (mayormente conocida como SQuaRE: Software Product Quality Requirements and Evaluation) nace en el año 2005 para dar respuesta a las necesidades actuales de las organizaciones. Su objetivo es la creación de un marco de trabajo común para evaluar la calidad de un producto software, sustituyendo a las anteriores ISO/IEC 9126 [1] e ISO/IEC 14598 [2]. Dentro de la familia ISO/IEC 25000, se presenta la norma ISO/IEC 25012 - "Data Quality Model" [3] y la ISO/IEC 25040 - "Evaluation process" [4]. La norma ISO/IEC 25012 – "Data Quality Model" especifica un modelo general de calidad para aquellos datos que se encuentran definidos en un formato estructurado dentro de un sistema informático. Puede utilizarse para detallar requisitos, establecer medidas, como así también para planificar y realizar evaluaciones de calidad de datos. La norma clasifica los atributos de calidad en quince características analizadas desde dos puntos de vista: inherentes y dependientes del sistema.

Dichas características tendrán una importancia y prioridad diferente para cada persona, siendo ésta la encargada de definir sus propias necesidades

Estableciendo que el activo más importante de las organizaciones es el gran volumen de datos que manejan, se trabajó en una guía de medición para aquellos interesados en reconocer el nivel de ciertas características de calidad asociadas a sus datos, empleando su uso en el contexto de una evaluación de producto de software. De esta forma, una organización podrá obtener una retroalimentación del estado de los datos con los que opera, pudiendo especificar previamente cuáles son los valores esperados.

3- Mejora de los servicios de gobierno digital en organismos públicos de gobierno

En el marco de este proyecto, el III-LIDI se enfocó en la investigación sobre las características de gobierno abierto y su aplicación en las distintas agencias gubernamentales, como así también en el diseño de soluciones y estrategias para integrar diferentes actividades mejorando la eficacia y eficiencia de las interacciones gobierno-ciudadanos.

Los servicios públicos se prestan a través de diferentes formas, entre ellos las aplicaciones de dispositivos móviles, se relevaron las aplicaciones móviles de diferentes municipios de nuestro país y se construyó una taxonomía aplicable a dichos servicios públicos [11].

Además, se analizó la aplicación del blockching en gobierno abierto [12].

Desde el año 2003, el III-LIDI trabaja en aplicaciones relacionadas con la gobernanza digital para la elección de autoridades mediante el voto electrónico. Se destacan la definición e implementación de tres modelos (presencial, semipresencial y remoto) en distintos tipos de votaciones (urnas

electrónicas, ambientes de votación, comunicaciones, entre otras).

4- Mejora en los procesos de gestión de la Facultad de Informática

El III-LIDI, en coordinación con el área de Gestión Electrónica Digital y Calidad de la Facultad de Informática, trabaja en este proyecto, con el objeto de analizar, definir y establecer un plan a ser aplicado a distintos procesos de la Gestión Universitaria.

La Facultad de Informática posee un Portal de Gestión Administrativa que provee a sus Alumnos, Docentes, No Docentes y Graduados una guía para la realización de trámites de los principales servicios que presta. De cada uno de estos trámites se brinda información estandarizada sobre: Descripción, Forma de Acceso, Requisitos y Contacto [3]

Desde el año 2011, la Facultad de Informática ha iniciado el camino hacia la certificación de distintos procesos.
El curso de Nivelación a Distancia logro en el año 2012 la certificación IRAM-ISO 9001:2008 del Sistema de Gestión de Calidad (SGC) del "Diseño y realización del curso de Nivelación a Distancia para el Pre-Ingreso a la Facultad de Informática". En 2018 se realizó la migración de la certificación a la versión IRAM-ISO 9001:2015 y se actualizo el alcance del SGC.
Asimismo, en 2016 se ha obtenido la certificación IRAM-ISO 9001:2008 del SGC del área de Concursos Docentes: "Llamado a Concursos Docentes Ordinarios", "Concurso Auxiliar Docente Ordinario" y "Concurso Profesor Ordinario" de acuerdo a la ordenanza 179 de la Universidad Nacional de La Plata, y las ordenanzas 303 y 308 de la Facultad de Informática. En 2018 se realizó la migración de la certificación a la versión IRAM-ISO 9001:2015.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

- Análisis y estudio de normas y modelos orientados a la Calidad del Producto, como por ejemplo la familia IRAM - ISO/IEC 25000
- Evaluación de procesos en organismos públicos y privados según los requisitos de IRAM - ISO 9001.
- Evaluación de procesos de software según las directrices de la ISO/IEC 90003
-Desarrollo de Voto electrónico presencial y Voto por Internet. Arquitecturas adaptadas a la legislación vigente.
- Análisis, discusión y estudio de normas de calidad relacionadas con certificación de servicios gubernamentales.
- Análisis, discusión y estudio de mejoras de proceso en el desarrollo de software combinando herramientas de gestión de proyectos.
- Evaluación de madurez de los servicios de gobierno digital de una unidad académica y de organismos públicos gubernamentales.

3. RESULTADOS OBTENIDOS/ ESPERADOS

-Se avanza con la construcción de patrones de errores en los datos utilizados para avanzar en la construcción de la herramienta de evaluación de calidad de datos.
- Se continúa con el mantenimiento de los procesos certificados bajo la norma IRAM-ISO 9001:2015: "Pre-Ingreso a Distancia de la Facultad de Informática" y "Concursos Docentes de la Facultad de Informática".
- Se analizan nuevos procesos de la Facultad, con posibilidad de generar guías/instructivos para facilitar la prestación de los servicios y analizar la posibilidad de su certificación.
- Se generó un Portal de Gestión Administrativa para la Facultad de Informática UNLP, donde se reúne

información y enlaces sobre los diferentes servicios que brinda la facultad.
- Se propuso un documento único que integra documentación utilizada en estándares conocidos y en la metodología ágil Scrum, con el fin de satisfacer la información documentada solicitada por la norma IRAM-ISO 9001:2015 para la eficacia del SGC.
- Capacitación y desarrollo de los documentos básicos de gestión de la calidad de productos.
- Se realizaron acciones de consultoría y asesoramiento en organismos públicos y privados.
- Se avanzó en la tesis de doctorado "Modelo de madurez de los servicios de gobierno electrónico en el ámbito universitario".
- Se aplicó el Voto electrónico presencial en elecciones universitarias y Voto por Internet en distintos organismos.

4. FORMACION DE RECURSOS HUMANOS

- El proyecto cuenta con becarios de Maestría de la UNLP en los temas del área.
- Se desarrollan tesis de doctorado y tesinas de grado en el área.
- Se dirigen 2 tesinas de grado en temas del área.
- Los integrantes de esta línea de Investigación participan en el dictado de asignaturas/cursos de grado/postgrado en la Facultad de Informática de la UNLP y en otras universidades del país. En particular, en la UNLP, se dicta la asignatura "Calidad de Sistemas de Software"

5. BIBLIOGRAFIA

[1] E. Estévez and T. Janowski, "Gobierno Digital, Ciudadanos y Ciudades Inteligentes.," 2016.

[2] A. Pasini and P. Pesado, "Quality Model for e-Government Processes at the University Level: a Literature Review," Proc. 9th Int. Conf. Theory Pract. Electron. Gov., pp. 436–439, 2016.

[3] A. Pasini, E. Estévez, P. Pesado, and M. Boracchia, "Una metodología para evaluar la madurez de servicios universitarios," Proc. Congr. XXII Congr. Argentino Ciencias la Comput. Congr. Argentino Ciencias la Comput., pp. 636–646, 2016.

[4] IRAM-ISO 9001:2015 "Sistemas de gestión de la calidad. Requisitos".

[5] ISO, "ISO/IEC 25000:2014 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Guide to SQuaREtle," 2014.

[6] ISO, "ISO/IEC 25040:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Evaluation process," 2011.

[7] ISO, "ISO/IEC 25010:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models," 2011.

[8] D. St-Louis and W. Suryn, "Enhancing ISO/IEC 25021 quality measure elements for wider application within ISO 25000 series," 2012, pp. 3120–3125.

[9] J. Calabrese, R. Muñoz, A. Pasini, S. Esponda, and M. Boracchia, "Asistente para la evaluación de características de calidad de producto de software propuestas por ISO / IEC 25010 basado en métricas definidas usando el enfoque GQM," pp. 660–671, 2017.

[10] Taxonomía para aplicaciones móviles de servicios de gobierno en municipios de Argentina R. Muñoz, J. S. Preisegger, A. C. Pasini, and P. M. Pesado, Actas del XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019), ISBN: 978-987-688-377-1, págs. 1374-1384, 2019.

[11] Blockchain y gobierno digital J. S. Preisegger, R. Muñoz, A. C. Pasini, and P. M. Pesado, Actas del XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019), ISBN: 978-987-688-377-1, págs. 1305-1315, 2019.

Reuso orientado a dominios: hacia un proceso de desarrollo basado en reusabilidad a nivel de subdominos

Agustina Buccella, Juan Luzuriaga, Alejandra Cechich,
Gabriel Cancellieri, Maximiliano Ceballos,
Matias Pol'la, Maximiliano Arias, Marcos Cruz
Rodolfo Martínez, Rafaela Mazalu
GIISCO Research Group
Departamento de Ciencias de la Computación
Universidad Nacional del Comahue
Neuquen, Argentina
agustina.buccella@fi.uncoma.edu.ar

1. Resumen

Las Líneas de Producto de Software (LPS) son un paradigma de desarrollo de software cuyo principal objetivo es la reusabilidad dentro de los dominios. Las metodologías existentes se basan en la construcción de artefactos o activos de software reusables de manera que puedan utilizarse cada vez que un nuevo producto de la línea es derivado. A su vez, es posible reutilizar fuera de un dominio, especialmente cuando los dominios están relacionados entre sí.

En nuestro proyecto actual hemos trabajado en la creación de un proceso integral para el desarrollo de LPSs proponiendo enfoques, métodos, técnicas y recursos que permitan mejorar la reusabilidad. En particular nos hemos centrado en el reuso a nivel del dominio geográfico, especializándonos a dominios más específicos como el de ecología marina o paleontología. Nuestros últimos avances en esta área han sido enfocados en la creación de un proceso de desarrollo basado íntegramente en la identificación y selección de artefactos de software previamente desarrollados. Estos nuevos avances permiten aprovechar nuestros artefactos para otros subdominios geográficos reusándolos en una nueva LPS, maximizando así el reuso efectivo.

Palabras Clave: Reusabilidad - Líneas de Producto de Software - Jerarquía de Dominios - Dominio Geográfico

2. Contexto

La línea presentada se inserta en el contexto del *Programa: Desarrollo de Software Basado en Reuso - Parte II (04/F009)*. Directora: Dra. Alejandra Cechich, y *SubProyecto: Reuso Orientado a Dominios - Parte II*. Incluído dentro del Programa. Directora: Dra. Agustina Buccella, Codirector: Mg. Juan Manuel Luzuriaga.

3. Introducción

Actualmente, el reuso es, en general, uno de los atributos del software más valorados y buscados dentro del área de ingeniería de software. Entre sus principales beneficios podemos citar mejoras en el tiempo de comercialización, la calidad del software y los costos. Existen varios paradigmas de desarrollo que están centrados en la reutilización de software, como el desarrollo de línea de productos de software [9], el desarrollo basado en componentes [1], el orientado a servicios [2], etc. Cada uno de ellos propone técnicas y/o métodos específicos

basados en maximizar la reutilización, interoperabilidad y estandarización.

En particular, los desarrollos en las líneas de productos de software (LPS) se centran en identificar similitudes y variabilidades dentro de dominios particulares para ser reutilizados cuando se desarrollan nuevos productos. A su vez, este reuso de dominios puede ser extendido a subdominos, en especial cuando entre ellos existen relaciones o aspectos similares. En este sentido, hemos trabajado en el dominio geográfico como un dominio genérico, especializándonos luego en subdominios específicos y relacionados. Sin embargo, lograr reusabilidad no es una tarea sencilla y no se obtiene solo por el hecho de seguir un desarrollo de LPS. Se debe trabajar en la creación de artefactos que puedan ser efectivamente reusables y luego en mecanismos que permitan identificarlos y aplicarlos, ya sea en su totalidad o parcialmente.

De esta forma, hemos trabajado en estos últimos años en mecanismos que asistan a la identificación de artefactos a ser reusados. Basados en nuestras experiencias de trabajos previos en cuanto a la creación de LPSs en los subdominios de ecología marina [5] y paleobiología [6], utilizamos los artefactos allí creados para definir un proceso de desarrollo basado en la reutilización de estos artefactos a nivel de subdominio. Luego, este proceso fue aplicado al subdominio de *patrimonio cultural paleontológico* con el cual se poseía muchos servicios y funcionalidades en común con los otros subdominos [8].

En la Figura 1 se muestra nuestro proceso de desarrollo basado en la reutilización de artefactos preexistentes para otros subdominios. En particular dos artefactos de software previamente construidos son considerados para analizar en la reutilización. Ellos son: la *taxonomía de servicios* y las *hojas de datos funcionales*. La primera, representa una taxonomía basada en una jerarquía de dominios [4, 5, 6]. Esta jerarquía comienza desde el dominio de geográfico y se especializa en un conjunto de subdominios que comparten algunos servicios. La taxonomía utiliza como estructura inicial las categorías definidas en el estándar ISO 19119 std[1] la cual se basa en servicios geográficos. Así se incluyen cinco categorías de servicios: *(HI) Interacción humana*, *(MMS) Modelo/gestión de información*, *(WTS) Flujo de trabajo/Tarea*, y *(PS) Servicios de procesamiento*, que a su vez se subdivide también en cinco subcategorías que incluyen servicios temáticos, espaciales y de comunicación. A partir de allí, estas categorias se han ampliado en un conjunto de servicios para subdominios particulares, específicamente en ecología marina, oceanografía [5] y paleobiología [6]. Para cada uno de estos subdominios, también utilizamos otros estándares disponibles como por ejemplo, la ISO 21127[2] y el Modelo conceptual de referencia (CRM) de CIDOC[3] que se han aplicado para el subdominio de paleobiología, por ejemplo.

El segundo artefacto, *la hoja de datos funcional*, se utiliza para diseñar las funcionalidades en cada subdominio. Cada hoja de datos contiene cinco elementos: (1) un *identificados* que es único, un *nombre* con una descripción de la función principal, un *dominio* en el que se incluye la funcionalidad para la que se creó en primer lugar, una lista de *servicios* (de la taxonomía del servicio) utilizada para representar la funcionalidad, y un conjunto de *modelos de variabilidad* que muestran las interacciones del servicio (como servicios comunes y variantes). Para el último elemento, la variabilidad se representa mediante la notación OVM [9].

En la Figura 1 podemos observar, a su vez, que el proceso contiene tres pasos en los cuales dos de ellos (pasos 2 y 3) se subdividen en dos actividades diferentes considerando la reutilización de artefactos o el desarrollo de otros nuevos. Los tres pasos son:

1. *Procesamiento de requerimientos*: las entradas de este paso son los *requerimientos del nuevo subdominio* proporcionados por usuarios expertos en el área; y *estándares, leyes y / o políticas*. Esta última in-

[1]Norma Internacional de Servicios 19119, ISO / IEC, 2005.

[2]Información y documentación - Una ontología de referencia para el intercambio de información sobre patrimonio cultural, 2014

[3]Modelo conceptual de referencia Versión 6.0 `http://www.cidoc-crm.org`

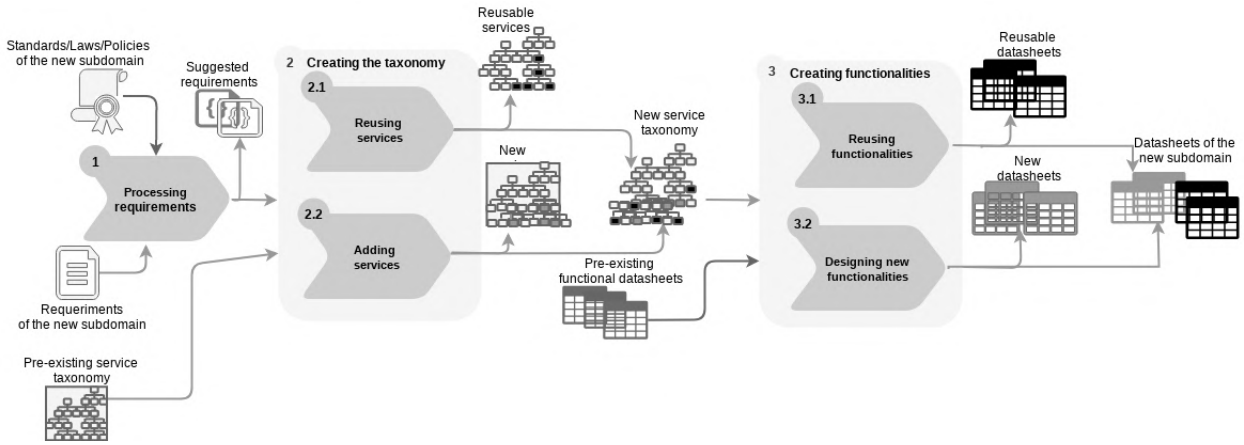


Figura 1. Proceso de desarrollo para nuevas funcionalidades basadas en el reuso a nivel de subdominio

formación es útil para determinar las regulaciones que deben cumplirse para todos los productos generados en el subdominio. El resultado de este paso es una lista de los *requerimientos sugeridos* que deben incluirse en la LPS.

2. *Creación de la taxonomía*: Las entradas de este paso son los *requerimientos sugeridos* obtenidos en el último paso y la *taxonomía preexistente* creada durante el desarrollo de otras LPSs en otros subdominios. El resultado de este paso es la nueva taxonomía que contiene la lista de servicios agregados/reutilizados del nuevo subdominio. Este paso se subdivide en dos subpasos:

    a) *Reutilización de servicios*: Aquí, es importante determinar cuáles de los requerimientos sugeridos se pueden traducir a un servicio ya definido por la taxonomía preexistente (para otros subdominios). De esta manera, el servicio puede ser reutilizado.

    b) *Adición de servicios*: cuando los requisitos sugeridos no coinciden con un servicio de la taxonomía preexistente, se deben agregar como nuevos. Esta adición no es trivial porque se pueden agregar como una categoría completamente nueva de la taxonomía o como una nueva especialización de otro servicio (o categoría) preexistente.

3. *Creación de funcionalidades*: en este paso, se deben diseñar las hojas de datos funcionales del nuevo subdominio basadas en el resultado del último paso (lista de servicios agregados/reutilizados en la nueva taxonomía). El resultado de este paso es el conjunto de hojas de datos agregadas/reutilizadas para el nuevo subdominio. Como en el paso 2, es importante analizar si la funcionalidad ya está definida para subdominios anteriores o si es nueva. Por lo tanto, este paso se subdivide en dos subpasos:

    a) *Reutilización de funcionalidades*: debemos identificar las funcionalidades que se pueden reutilizar. Para ello, se debe analizar el conjunto de hojas de datos funcionales preexistentes determinando si la nueva funcionalidad se puede reutilizar por completo, parcialmente o si no es posible.

    b) *Diseño de nuevas funcionalidades*: se diseña una nueva funcionalidad cuando no se puede reutilizar por completo. A veces, es posible reutilizar una funcionalidad sólo parcialmente, en cuyo caso es necesario crear también una nueva con el conjunto de servicios reutilizados.

4. Líneas de Investigación y Desarrollo

Durante los años de ejecución de nuestro proyecto de investigación, hemos realizado amplios avances en lo que respecta al área de LPSs definiendo y refinando una metodología de desarrollo a nivel de subdominos. Dentro de esta metodología, la cual posee un conjunto de actividades asociadas a la ingeniería de domino y de aplicación, nos hemos centrado en minimizar la complejidad del dominio geográfico mediante la división o jerarquía del mismo en subdominos más manejables [4, 5] que mejoraron su complejidad inherente. Al mismo tiempo, dentro de la metodología de desarrollo, nos hemos centrado en la construcción de artefactos de software específicos para cada una de las actividades. Se han presentado las bases de la metodología y los artefactos diseñados los cuales se utilizan en el análisis de dominios y en el análisis organizacional de una LPS [4] y tienen la particularidad de favorecer el reuso basado en una taxonomía de servicios. Es precisamente esta ventaja la que nos permitió luego realizar extensiones hacia otros subdominios. De esta forma hemos podido así avanzar en el desarrollo de múltiples LPSs basadas en la jerarquía de dominios definida. Sin embargo, para dicha extensión hemos tenido que formalizar varios aspectos respecto a los artefactos de software creados. En el caso de la taxonomía, se utilizaron los estándares reconocidos del subdomino, y se creó un proceso de desarrollo que se aplicó en la creación de las dos primeras LPSs [5, 6]. Luego, para el caso de funcionalidades, hemos definido a las hojas de datos funcionales como los artefactos más apropiados para su desarrollo.

Así, hemos trabajado en la definición y construcción de dichas hojas de datos de manera que permitan un diseño estandarizado y computable. En este sentido, nos hemos centrado en mejorar las técnicas para determinar la validez de estos modelos e identificar los problemas que contienen tanto a nivel de la definición de una LPS así como en la instanciación y generación de productos. En este contexto hemos realizado experimentaciones en cuanto a la forma y artefactos creados en el proceso de manera de poder luego realizar un análisis de las mejoras generadas a la hora de encontrar anomalías y defectos dentro de los modelos de variabilidad [3, 10]. Por último en [7] hemos propuesto un framework para combinar herramientas de análisis de variabilidad automáticas, que pueden adaptarse a los requisitos específicos de un caso de gestión de variabilidad.

Nuestros últimos avances se encuentran enfocados en el aprovechamiento de los artefactos creados de forma tal de facilitar la identificación de las partes reusables al realizar un nuevo desarrollo de LPS en otros subdominos.

5. Resultados Obtenidos/Esperados

El objetivo principal de la línea de investigación es *Desarrollar técnicas y herramientas que mejoren el desarrollo de software basado en reuso a nivel de dominios, favoreciendo el desarrollo de ambientes inteligentes que permitan crear LPSs según taxonomías de servicios*.

Dentro de este objetivo, se definen dos líneas de investigación específicas. En la primera, hemos obtenido resultados concretos y continuaremos trabajando dentro del área de Líneas de Productos de Software. Como se ha descripto previamente, hemos avanzado en el desarrollo de artefactos de software y herramientas que sirven de soporte para las actividades dentro de la metodología de LPS definida (meta-modelos, taxonomías de servicio, etc.), hemos definido y aplicado nuevos métodos y técnicas para la creación de LPSs con soportes inteligentes dentro del dominio geográfico que contemplan las particularidades de los subdominios incluidos, y hemos realizado formalizaciones de reglas y patrones para soportar el desarrollo asistido, de manera que sean lo suficientemente generales para ser aplicados en otros subdominios geográficos.

En la segunda línea derivada del objetivo principal, hemos avanzado en la creación de ambientes inteligentes que permitan garantizar un reuso efectivo a la hora de crear una nueva LPS. Hemos avanzado en un proceso para desarrollar LPSs basadas en la reutilización de

funcionalidades de una jerarquía de dominio, que incorpora elementos comunes y variables de varios subdominios. Así, construir una nueva LPS significa comprender y probablemente extender una taxonomía y funcionalidad de servicio en forma de activos reutilizables. El proceso propuesto ha sido aplicado al subdominio de *patrimonio cultural paleontológico*, mediante la identificación y selección de los posibles activos reutilizables. Nuestro trabajo a futuro se centra en continuar sobre este proceso de desarrollo basado en reuso para enfocarnos en casos de reuso más complejos, como el análisis de servicios en la taxonomía sin la funcionalidad correspondiente en las hojas de datos. Actualmente estamos trabajando para ampliar este enfoque y desarrollar una herramienta de apoyo que permita la automatización de la mayor parte del proceso.

6. Formación de Recursos Humanos

El proyecto reúne aproximadamente a 13 investigadores, entre los que se cuentan docentes y alumnos de UNComa, y colaboradores. A su vez, el proyecto cuenta actualmente con un doctor y un magister. Varios de los docentes-investigadores de GIISCo-UNComa han terminado o se encuentran próximos a terminar carreras de postgrado. Uno de ellos se encuentra finalizando su doctorado en el transcurso de este año. A su vez varios de los integrantes se encuentran finalizando sus tesis de grado. Por último, este año seguiremos con la supervisión del trabajo de 2 becarios EVC-CIN.

Referencias

[1] Mehmet Aksit. *Software Architectures and Component Technology*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[2] Michael Bell. *Service-Oriented Modeling: Service Analysis, Design, and Architecture*. Wiley Publishing, 2008.

[3] Germán Alejandro Braun, Matias Pol'la, Laura Andrea Cecchi, Agustina Buccella, Pablo R. Fillottrani, and Alejandra Cechich. A DL semantics for reasoning over ovm-based variability models. In *Proceedings of the 30th International Workshop on Description Logics, Montpellier, France, July 18-21, 2017.*, 2017.

[4] A. Buccella, A. Cechich, M. Arias, M. Pol'la, S. Doldan, and E. Morsan. Towards systematic software reuse of gis: Insights from a case study. *Computers & Geosciences*, 54(0):9 – 20, 2013.

[5] A. Buccella, A. Cechich, M. Pol'la, M. Arias, S. Doldan, and E. Morsan. Marine ecology service reuse through taxonomy-oriented SPL development. *Computers & Geosciences*, 73(0):108 – 121, 2014.

[6] Agustina Buccella, Alejandra Cechich, Juan Porfiri, and Domenica Diniz Dos Santos. Taxonomy-oriented domain analysis of gis: A case study for paleontological software systems. *ISPRS International Journal of Geo-Information*, 8(6), 2019.

[7] Agustina Buccella, Matias Pol'la, Esteban Ruiz de Galarreta, and Alejandra Cechich. Combining automatic variability analysis tools: An spl approach for building a framework for composition. In *Computational Science and Its Applications – ICCSA 2018*, pages 435–451, Cham, 2018. Springer International Publishing.

[8] Sofia Caballero, Agustina Buccella, and Alejandra Cechich. Reuso de funcionalidades en el subdominio de patrimonio cultural. In *Computer Science – CACIC 2019*. RedUnci, 2019.

[9] Klaus Pohl, Günter Böckle, and Frank J. van der Linden. *Software Product Line Engineering: Foundations, Principles and Techniques*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[10] Matias Pol'la, Agustina Buccella, and Alejandra Cechich. Automated analysis of variability models: The sevatax process. In *Computational Science and Its Applications - ICCSA 2018 - 18th International Conference, Melbourne, VIC, Australia, July 2-5, 2018, Proceedings, Part IV*, pages 365–381, 2018.

Innovación en Educación Informática

SCRUM como metodología de enseñanza y aprendizaje de la Programación

Nicolás Tymkiw, Juan Manuel Bournissen y Marisa Cecilia Tumino
Universidad Adventista del Plata, Libertador San Martín, Entre Ríos, Argentina
{nicolas.tymkiw, juan.bournissen, marisa.tumino}uap.edu.ar

Resumen

El objetivo del trabajo fue identificar el impacto de la utilización de la metodología de desarrollo de software SCRUM, como técnica de refuerzo de las estrategias de enseñanza y aprendizaje en la asignatura de Programación II, en el aprendizaje percibido de los estudiantes.
A partir de una intervención, consistente en la implementación de la metodología SCRUM como estrategia de enseñanza y aprendizaje de la programación y en la recolección de datos referidos al aprendizaje percibido de los estudiantes, los resultados permiten estimar que la aplicación de la metodología SCRUM tiene un impacto positivo en el aprendizaje percibido.

Contexto

El proyecto se inserta en la disciplina de estrategias de enseñanza y desarrollo de sistemas, contemplada en el área de Innovación en Educación en Informática de WICC. El trabajo fue coordinado por las cátedras de Tesina final y Programación II de la Licenciatura de Sistemas de Información de una universidad de Argentina. La institución que financió el proceso de investigación fue la Universidad Adventista del Plata.

1. Introducción

Para las carreras informáticas, la programación es una asignatura fundamental debido a que sustenta el progreso del alumno en la carrera y orienta su desempeño profesional. Una de las mayores preocupaciones del docente se centra en que el estudiante logre la comprensión completa de los contenidos de programación.

Investigadores como Costelloe [1] muestran que los estudiantes no logran conceptualizar todos los temas de programación enseñados, siendo esta una de las razones de abandono de la carrera.
Perazo [2] afirma que el porcentaje de deserción en las carreras informáticas se aproxima al 80%.
Barberis, y Del Moral Sachetti [3], señalan que desde que implementaron esta metodología en el año 2014, la tasa de deserción bajó un 23% con respecto a los años anteriores, mientras que el rendimiento académico aumentó un 27%.

1.1. Beneficios de SCRUM

Albaladejo [4] menciona que SCRUM cuenta con beneficios, fundamentos y requisitos que hacen al éxito de la aplicación de esta metodología:

- Entrega de resultados que debe ser dentro de un plazo corto, mensual o quincenal.
- Aporte a la gestión de las expectativas del cliente.
- Resultados anticipados, con los que el cliente puede utilizar los más importantes antes de que el proyecto termine.
- Flexibilidad y adaptación, donde el cliente va dirigiendo el proyecto en función de sus nuevas prioridades.
- Retorno de Inversión (ROI) mediante el que el cliente maximiza el beneficio del proyecto.
- Definición de los riesgos desde la primera iteración para poder mitigarlos.
- Mayor productividad y calidad puesto que el equipo mejora y simplifica su manera de trabajar en cada iteración.
- Trabajo conjunto entre cliente y equipo desde el planteamiento de los requisitos.

- Equipo autogestionado y motivado, donde las personas pueden usar su creatividad para resolver problemas.

1.2. Componentes de SCRUM

La metodología SCRUM se encuentra dividida en fases y roles. Las fases se pueden identificar como reuniones también conocidas como Sprint.
Reuniones. Gallego [5] divide las reuniones en tres fases: (a) en la primera se desarrolla una planificación del Backlog y se obtiene la lista de tareas a realizar; (b) para la segunda fase se genera un seguimiento del sprint. y (c) en la tercera se concluye el Sprint y se presentan los resultados finales.
Roles. Los roles se dividen en dos grupos: (a) los que están comprometidos con el proyecto y proceso de SCRUM y (b) los que no son parte del proceso pero se necesitan para la realimentación de la salida de los procesos y el planeamiento de los sprint:
Product Owner: es el encargado de tomar las decisiones del proyecto.
SCRUM Master: comprueba que la metodología funciona y se mantenga la fluidez del proceso y la interacción con el cliente y los gestores.
Equipo de Desarrollo: un grupo pequeño de personas que organizan y toman decisiones para conseguir su objetivo.
Dentro del segundo grupo se pueden encontrar los Usuarios que son los destinatarios finales del producto, el Stakeholders, que participa de las revisiones del sprint, y los Managers, encargados de la toma de decisiones finales.

1.3. SCRUM en el aula

Kuz, Falco y Giandini [6] explican que, dentro del ámbito educativo, los estudiantes requieren desarrollar capacidades y aptitudes que le servirán en su carrera profesional, por lo que se considera que SCRUM favorece la creación de un ambiente propicio para que los estudiantes experimenten un aprendizaje creativo.
Para Miller [7] un "aula ágil" precisa de la integración de cinco elementos fundamentales para alcanzar objetivos específicos: (a) clase visible, (b) ritmo de aprendizaje, (c) colaboración, (d) capacitación y (e) el camino para evolucionar un aula hacia la autoorganización.

1.4. SCRUM en programación

Varios autores como Chávez Andrade [8] muestran que la utilización de metodologías ágiles, como SCRUM, pueden ser altamente eficientes al desarrollar un software. Hervás Lucas [9] sostiene que el método prototipado ágil fomenta el aprendizaje y el desarrollo de competencias clave como el trabajo en equipo, la autodidactica y el análisis crítico.

1.5. Evaluación de Impacto

Para definir el concepto de impacto de un proceso educativo, se toma como modelo la definición de Aguilar [10].
El impacto de un proceso docente educativo se traduce en sus efectos sobre una amplia población: comunidad, claustro, entorno, estudiante y administración. Se identifican efectos científicos, tecnológicos, económicos, sociales, culturales e institucionales, centrados en el mejoramiento profesional y humano y en la superación social.
Dado que el impacto de la implementación de la metodología SCRUM debe evaluarse, se considera sustancial partir del concepto de evaluación de impacto, entendido como "el proceso evaluativo orientado a medir los resultados de las intervenciones, en cantidad, calidad y extensión según las reglas preestablecidas" Abdala [11].

2. Metodología

El presente estudio es de tipo explicativo y cuasiexperimental, puesto que busca identificar el nivel de impacto en el aprendizaje percibido en dos grupos definidos por la estrategia de enseñanza y aprendizaje implementada en una clase de programación. El diseño metodológico

estuvo basado en un estudio desarrollado por Barberis y otros [2].

La variable independiente en esta investigación fue la metodología utilizada y las variables dependientes: (a) el nivel de impacto en el aprendizaje percibido de los estudiantes y (b) las calificaciones promedio cuatrimestrales.

Durante los últimos años se ha identificado que en la asignatura de Programación II existía un alto porcentaje de estudiantes que perdían la cursada y otro porcentaje, menor pero significativo, de estudiantes que abandonaban.

Este plan de trabajo consistió en aplicar la metodología SCRUM, como estrategia de enseñanza y aprendizaje de la programación, en un curso de Programación II de la carrera de Ingeniería en Sistemas de Información de una universidad de Argentina. Se dividió al curso en equipos de tres estudiantes a quienes se les explico la metodología de trabajo. Los procedimientos tuvieron como objetivo potenciar el trabajo de cada uno de los estudiantes.

2.1. Instrumentos

El impacto fue medido mediante la encuesta de Nivel de Impacto de la Implementación de las estrategias de enseñanza y aprendizaje en el aprendizaje percibido de los estudiantes, construida, validada y adaptada por Tumino y Bournissen [12], compuesta por ítems vinculados a la percepción del aprendizaje (con 8 ítems) y por ítems relacionados con estrategias de aprendizaje (con 10 ítems). La escala utilizada en las dos dimensiones es de tipo Likert desde 1= muy en desacuerdo, 2= en desacuerdo, 3= neutro, 4= de acuerdo y 5= muy de acuerdo.

3. Resultados

El objetivo principal fue conocer si los estudiantes que trabajaron con la metodología SCRUM, muestran un mayor nivel de aprendizaje percibido, lo que se reflejaría tanto en las calificaciones promedio, como en el impacto de la implementación de la metodología en su aprendizaje percibido.

Mediante el análisis realizado con los datos de cursado de la asignatura obtenidos de las últimas diez cohortes, se pudo observar que en el año en que se realizó la experiencia, el 66,67% de los estudiantes promocionó la asignatura, mientras que el 33,3% restante la regularizó. De esta manera ningún estudiante perdió o abandonó la asignatura.

Los nueve años anteriores se utilizó una misma metodología y en el último año se implementó la metodología SCRUM. Se observa que tan solo en este último año ningún alumno perdió o abandonó la cursada, tal como se muestra en la Tabla 1.

Tabla 1: Datos de cursado correspondientes a los últimos diez años.

Año	Promoción Directa	Promoción Indirecta	Pierde la Materia	Abandona
2009	30,0%	30,0%	30,0%	10,0%
2010	11,8%	41,2%	35,3%	11,8%
2011	33,3%	40,0%	26,7%	0%
2012	20,8%	37,5%	33,3%	8,3%
2013	10,0%	50,0%	40,0%	0%
2014	29,4%	23,5%	41,2%	5,9%
2015	28,6%	28,6%	42,9%	0,0%
2016	9%	18,2%	54,6%	18,2%
2017	28,5%	52,4%	19,1%	0%
2018	66,7%	33,3%	0%	0%

3.1. Pruebas de hipótesis

A continuación, se presentan los resultados de las pruebas de hipótesis planteadas durante la investigación.

H_1: Existe diferencia estadística significativa de rangos promedios de rendimiento académico, en términos de calificación promedio obtenido en programación II, entre los estudiantes que utilizan la estrategia de enseñanza con metodología SCRUM y los que cursaron con las metodologías utilizadas previamente.

Se probó mediante la prueba no paramétrica U de Mann-Whitney para dos muestras independientes, tomando a la calificación final como variable de comparación y al tipo de metodología como variable de agrupación. Se observó que existe una diferencia estadísticamente significativa de rangos promedios de calificación promedio entre los dos grupos ($z > 1,96$ y $p < 0,05$). Se encontró un mayor rango promedio en los estudiantes que trabajaron con la metodología SCRUM (RP = 9.93), mientras que los estudiantes que trabajaron con otras metodologías obtuvieron un rango promedio de 5.07.

H_2: Existe diferencia estadística significativa en el nivel de impacto en el aprendizaje percibido en programación II entre los estudiantes que utilizan la metodología SCRUM y los que cursaron con las estrategias. Se probó también mediante la prueba no paramétrica U de Mann-Whitney, tomando el impacto percibido en sus dos dimensiones como variable de comparación. Se encontró que existe una diferencia estadísticamente significativa de rangos promedios en la dimensión aprendizaje percibido entre los dos grupos ($z > 1,96$ y $p < 0,05$) (ver Tabla 2). El mayor nivel de impacto estuvo asociado a los estudiantes del grupo experimental (RP = 10.07), mientras que los estudiantes que usaron otras metodologías obtuvieron un rango promedio de 4,93.

En cuanto a la dimensión estrategia de aprendizaje del Nivel del Impacto, no se obtuvo una diferencia estadísticamente significativa con un nivel de confianza del 95%, aunque sí con un 90% de confianza ($z > 1,64$ y $p < 0,1$). Con la metodología SCRUM el rango promedio fue mayor (RP = 9.5) que con las otras metodologías (RP = 5.5) como se observa en la Tabla 2.

Tabla 2: Rango promedio de aprendizaje y estrategia-aprendizaje.

Categoría	Metodología	Rango promedio
Aprendizaje	SCRUM	10,07
	Otras	4,93
Estrategia-Aprendizaje	SCRUM	9,5
	Otras	5,5

4. Discusión y Conclusión

Se pudo observar que existe una diferencia estadísticamente significativa de rango promedio en el aprendizaje entre los dos grupos estudiados. También existió una diferencia significativa de rango promedio de la calificación final de los alumnos que utilizaron la metodología SCRUM, respecto de quienes trabajaron con otras metodologías.

En el transcurso de los años 2009 al 2017 existía un predominio de la promoción indirecta, e incluso de la pérdida de la asignatura, mientras que, en el año 2018, con la implementación de la metodología SCRUM, se denotó el predominio de la promoción directa de los estudiantes.

Los estudiantes que tuvieron la posibilidad de trabajar con SCRUM percibieron un mayor impacto en el aprendizaje. Los docentes que implementaron la nueva metodología expresaron entusiasmo con los resultados obtenidos y decidieron mantenerla durante los próximos años.

La utilización de la metodología SCRUM en el ámbito académico, y específicamente dentro de las asignaturas de programación, resulta una práctica muy satisfactoria debido a su impacto positivo tanto en el aprendizaje percibido de los estudiantes como en las calificaciones promedio. Los estudiantes lograron asimilar los contenidos, trabajar en forma colaborativa y autoorganizada,

administrar los tiempos de entregas, asumir los diferentes roles y exponer satisfactoriamente el trabajo realizado.

La implementación de la metodología SCRUM facilitó a los docentes el mantenimiento de la dinámica de la clase, consiguiendo que todos los estudiantes pudiesen realizar un seguimiento constante de los contenidos presentados en la asignatura.

La implementación de la metodología SCRUM permite generar un ambiente de trabajo ideal dentro del aula donde los estudiantes pueden aprovechar los aportes proporcionados por los docentes y por sus pares.

5. Formación de recursos humanos

El equipo de investigación estuvo conformado por los docentes investigadores de las cátedras de Tesina final y Programación II, con la participación de un asistente de investigación, estudiante de la Licenciatura en Sistemas.

6. Bibliografía

1. Costelloe E. (2004). "Teaching Programming The State of the Art. CRITE. Technical Report". Department of Computing, Institute of Technology Tallaght, Dublin, Ireland.

2. Perazo C. (2013) Tecnología. Reporte del Ministerio de Educación. Diario La Nación. Recuperado de: http://www.lanacion.com.ar/1632045 -el-80-de-los-estudiantes-de-carreras-informatica

3. Barberis, Á. R., & Moral Sachetti, L. E. D. (2016, August). SCRUM como herramienta metodológica en el entrenamiento cooperativo de la programación: de la teoría a la práctica. In XI Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2016), recuperado de: http://sedici.unlp.edu.ar /bitstream/handle/ 10915/54603/Documento_ completo.pdf-PDFA.pdf?sequence=1 &isAllowed=y:

4. Albaladejo X. (2008) Fundamentos de SCRUM. Recuperado de https://proyectosagiles. org/fundamentos-de-SCRUM/

5. Gallego, M. T. (2012). Metodologia SCRUM. Universitat Oberta de Catalunya.

6. Kuz, A., Falco, M., & Giandini, R. S. (2018). Comprendiendo la aplicabilidad de SCRUM en el aula: herramientas y ejemplos. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología, (21), e07-e07.

7. Miller J. (2016), 5 Elements of Agile Classrooms. Recuperado de http://blog.agileclassrooms. com/2016/11/5-elements-of-agile-classrooms.html

8. Chávez Andrade, J. V. (2019). Estandarización de los procesos de desarrollo de software utilizando buenas prácticas de programación y SCRUM como marco de trabajo ágil en departamentos de TI (Master's thesis, Universidad Técnica de Ambato. Facultad de Ingeniería en Sistemas, Electrónica e Industrial. Maestría en Gerencia de Sistemas de Información).

9. Hervás Lucas, R. (2012). Método para el aprendizaje de entornos y lenguajes de programación basado en prototipado ágil. Jornadas de Enseñanza de la Informática (18es: 2012: Ciudad Real).

10. Aguilar M. A. (2009). *El impacto de la carrera de economía de la BUAP en el mercado laboral: la visión de los egresados de la generación 1995-2000*. México: Benemérita Universidad Autónoma de Puebla.

11. Abdala, E. (2004). *Manual para la evaluación de impacto en programas de formación para jóvenes*. Montevideo: CINTERFOR

12. Tumino, M. C. y Bournissen, J. M. (2020). Integración de las TIC en el aula e impacto en los estudiantes: elaboración y validación de escalas. IJERI: *International Journal of Educational Research and Innovation, 13*, 62-73.

Formación de docentes y alumnos de grado como Investigadores Científicos Iniciales en las áreas de Informática y Ciencias de la Computación

Bianchini Germán[1], Caymes-Scutari Paola[1,2], Ontiveros Patricia[3], Rotella Carina[3], Salinas Sergio[4], Tagarelli Sandra[4], Chirino Pamela[1], Galdamez Mariela[1]

[1]Laboratorio de Investigación en Cómputo Paralelo/Distribuido
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579
[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
[3]Laboratorio de Gobierno Electrónico, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza
[4]Laboratorio de Analítica de Datos, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza

gbianchini@frm.utn.edu.ar, pcaymesscutari@frm.utn.edu.ar, pontiveros@frm.utn.edu.ar, carinarotella@gmail.com, s4salinas@gmail.com, stagarelli@gmail.com, pamelaachirino@gmail.com, mariela.galdamez.16@gmail.com

RESUMEN

La capacidad investigativa resulta un elemento clave e imprescindible para hacer frente a los retos con los que los nuevos profesionales se encontrarán en el mundo laboral actual y futuro. Sin embargo, la educación argentina, aunque debiera estar orientada a generar pensamiento, se enfoca quizá más en profesionalizar a los alumnos. Esta es una situación que debe modificarse, y es responsabilidad de los propios docentes/investigadores abrir un nuevo panorama a los ojos de los estudiantes y docentes universitarios en general, de modo tal de ayudarlos a desarrollar sus capacidades de evaluación, crítica e inventiva.

El principal obstáculo que se encuentra es que las actividades de investigación se encuentran desvinculadas, en general, de las actividades académicas de grado, y por tanto esta disociación mantiene alejados a los alumnos y a muchos docentes de tareas de enfoque científico. Por tal motivo, en el presente proyecto se busca formalizar y poner en práctica el proceso de transferencia y formación de Investigadores Científicos Iniciales a través de su incorporación en actividades en el marco de tres grandes áreas: Cómputo Paralelo, Analítica de Datos y Gobierno Electrónico.

Palabras clave: Investigación, Formación, Alumnos, Cómputo Paralelo, Gobierno de Datos, IoT.

CONTEXTO

El presente proyecto se inició en enero de 2020, llevándose a cabo en el marco del LICPaD. El mismo se realiza en colaboración con docentes de los grupos ADA-Lab (Laboratorio de Analítica de Datos) y GE-Lab (Laboratorio de Gobierno Electrónico), ambos también pertenecientes al Departamento de Ingeniería en Sistemas de Información de la UTN-FRM. Cuenta con la acreditación, homologación y financiamiento de la Universidad Tecnológica Nacional a través del proyecto PID TEUTIME0007658TC. Mediante este trabajo conjunto, se busca combinar en el mismo proceso la formación teórica y la práctica de los investigadores iniciales para que éstos experimenten los pasos involucrados en un ámbito de investigación real, recorriendo todas las etapas necesarias que involucran el proceso científico. Una forma razonable es utilizar un trabajo de investigación específico como facilitador del proceso de aprendizaje, basado en la teoría del constructivismo y enfocado en las competencias específicas que deben adquirir los graduados de Ingeniería en Sistemas de Información.

Es importante señalar que, al fomentar las competencias investigativas, se contribuye también a incrementar el número y calidad de las investigaciones realizadas en la institución, además de mejorar el pensamiento crítico de docentes y alumnos frente a nuevos retos que no se presentan sólo durante su formación profesional sino en la vida laboral.

1. INTRODUCCIÓN

La tarea de investigar es un proceso sistemático, organizado y objetivo, cuyo propósito es responder a una pregunta para así incrementar el conocimiento y la información sobre algo desconocido. Es una actividad enfocada en obtener, mediante observación y experimentación, nueva información y conocimiento necesario para ampliar los diversos campos de la ciencia y la tecnología [1].

Dicho de otro modo, podemos definir la investigación como la acción y el efecto de realizar actividades intelectuales y experimentales de manera sistemática con el propósito de aumentar los conocimientos sobre una materia determinada, teniendo como fin ampliar el conocimiento científico, sin perseguir, en principio, ninguna aplicación práctica inmediata.

Abordar un proyecto de investigación científica es costoso, sobre todo considerando que los resultados suelen obtenerse a largo plazo. La investigación no suele ser una inversión con resultados o ganancia inmediata, como a menudo suele esperarse en otro tipo de ámbitos. Esto la convierte en una actividad en apariencia poco atractiva para quienes buscan impulsar solamente aquellos proyectos donde la premisa sea aportar a los intereses inmediatos. No obstante, es claro que, como inversión a largo plazo, resulta fundamental para el desarrollo de las instituciones y la sociedad.

De acuerdo a los principios mencionados, la investigación debería ser considerada pilar en todas las actividades académicas de nivel superior, y evidentemente esto debe incluir la participación activa tanto de docentes como de estudiantes. En diversas instituciones ha comenzado a entenderse esta idea, y paulatinamente se han impulsado planes que mejoran el nivel de la educación superior a través de modelos académicos que contemplan a la investigación científica como un eje fundamental para el desarrollo de una sociedad del conocimiento equitativa y democrática. Por citar un ejemplo, el

Centro Universitario de Los Lagos, de la Universidad de Guadalajara [2] (México), se ha propuesto como un centro regional que entiende a la investigación como una actividad sustantiva, necesaria para elevar el nivel y calidad de la educación. Una de sus políticas, en relación con la actividad docente de los investigadores, ha establecido que no debe centrarse en el posgrado, sino que antes debe asegurarse que cada investigador imparta por lo menos una asignatura en los programas de pregrado. De esta manera, se intenta propiciar un elemento más de conexión entre la docencia y la investigación, sobre todo en las licenciaturas, brindando así la posibilidad de acceder a conceptos o contenidos del ámbito de la investigación [3].

Otros ejemplos de cambio, son la Facultad de Medicina de la Universidad de Navarra (España) [4], donde se impulsa la actividad investigativa desde distintos departamentos y centros, con el objetivo de repercutir en la formación directa del alumno. La Universidad Maimónides [5] (Buenos Aires, Argentina) también busca acercar a los docentes y estudiantes a la investigación, al desarrollo y la transferencia tecnológica y la creación de empresas de base tecnológica o empresas intensivas en la aplicación de conocimiento. En el caso de la Universidad Nacional Autónoma de Honduras, se ha planteado un Plan Táctico de Reforma Universitaria, que establece una política institucional a nivel de investigación, desarrollo tecnológico e innovación, incorporando a estudiantes y catedráticos en los procesos de investigación a niveles de grado y posgrado [6], sólo por mencionar algunos ejemplos en distintos puntos geográficos.

Es de gran importancia que los estudiantes se incorporen a grupos y proyectos de investigación, pero es tan o más importante aún que aprendan de éstos los procesos de la ciencia, de modo que tengan la posibilidad de agregarlo a sus saberes y asimilarlo, y no que simplemente se los convoque para tareas mecánicas o generación de insumos.

En la UTN-FRM, por ejemplo, se efectúan anualmente ferias de ciencia y encuentros en los que se exponen los avances en materia de investigación que se llevan a cabo en los distintos centros y grupos de los diversos departamentos. La mayoría de los estudiantes (salvo aquellos pocos que participan de algún grupo consolidado) están normalmente ajenos a estas actividades, las cuales no son del todo bien difundidas, o son enfocadas a otro público, y se realizan en lugares o fechas que no facilitan al alumno de grado el acercarse para informarse sobre qué otras actividades se desarrollan en la institución en la que estudia y forma su perfil profesional.

En vista de esta realidad, es fundamental que los docentes proporcionen a los estudiantes las herramientas necesarias para conocer, interpretar y comprender la realidad de la región, del país y el mundo, que se encuentran en constante transformación. Sólo de esta forma éstos podrán intervenir de forma comprometida y crítica en su entorno para mejorar la calidad de vida de las personas a través de nuevos aportes a la sociedad.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Dentro del área del Cómputo Paralelo, las líneas de investigación planteadas en el presente proyecto por los alumnos-becarios son diversas: Paralelismo en Algoritmos de Aprendizaje para Redes neuronales [7]; Transformación automática de imágenes satelitales para su aplicación en sistemas de predicción de

incendios forestales [8] y Paralelización del Método de Tránsitos para la detección de exoplanetas [9].

Por su parte, en la rama de Analítica de Datos, las posibles temáticas son Gobierno Abierto y Datos abiertos [10], Gestión por procesos, Mejora de procesos continua, Indicadores de procesos y medición, Decisiones, Datawarehouse, Datamining, Dashboard, Sistemas basados en reglas y contenidos vinculados con la Maestría en Inteligencia de Negocios y Analítica de negocios (por ej. Predictive Analytics, Big data y Social Networks, Bi- Strategy, Information Visualization, Applied Statistics, etc.).

En el área de Internet de las Cosas, es posible indagar en temas relacionados con Sistemas de Comunicación Descentralizados, Aspectos de diseño de protocolos de comunicación, programación de Raspberry's y Arduinos. Además, es posible analizar el uso de modelos de comunicación basados en Internet de las Cosas en escenarios de catástrofes.

Finalmente, otra temática novedosa y de interés para docentes y alumnos es la aplicación de blockchain sobre la gestión de documentos [11].

3. RESULTADOS ESPERADOS

La contribución principal que se busca es la transferencia de formación por parte de docentes-investigadores hacia docentes y alumnos de grado y al medio educativo (y de manera directa a las cátedras de la carrera Ingeniería en Sistemas de Información vinculadas como Sistemas de Gestión, Análisis de Sistemas, Modelado de procesos de negocio, Gobierno Electrónico y Computación Paralela). Además, se busca que los alumnos aprendan y experimenten el trabajo en colaboración y por metas, ya que, en general, el trabajo de investigación no es algo que pueda realizarse de forma individual: normalmente requiere de espacios inter y transdisciplinarios. No obstante, como ya se indicó, el objetivo y ganancia a largo plazo al involucrar a los estudiantes en grupos y otras comunidades y redes de investigación, es el desarrollo de proyectos que puedan alcanzar un alto impacto y el tratamiento de temas de interés dentro de las distintas áreas del conocimiento. Por lo tanto, este es una segunda contribución, no menor, que estará asociada a los temas específicos que se desarrollen.

4. FORMACIÓN DE RECURSOS HUMANOS

Este es un proyecto que busca contribuir al crecimiento y desarrollo de investigadores nóveles, para lo cual se contará con los directores e investigadores formados como guía. Es importante para los estudiantes contar con la posibilidad de acceder a becas de investigación, y por esta razón, en principio, el proyecto está pensado con alumnos-becarios, lo cual les permitirá dedicar cierta cantidad de horas semanales a la tarea de investigar de modo tal de llegar a cumplir los objetivos formales establecidos. Además, dicha actividad les permitirá la realización de proyectos iniciados por ellos mismos y contar con fondos para lograr su concreción (compra de equipamiento específico y bibliografía, inscripción en congresos, gastos de traslados y estadías, etc.), lo que es extremadamente valioso para aquellos que quieran iniciarse en el camino de la investigación, estén en la etapa de desarrollo de su proyecto de fin

de carrera o manifiesten una vocación científica.

En la era del conocimiento y de los avances tecnológicos, las habilidades del pensamiento crítico y lectura analítica proporcionarán a los jóvenes las herramientas necesarias para enfrentar los desafíos de la actualidad, así como las del campo laboral, dado que el mundo necesita de jóvenes profesionales que puedan resolver tales problemas y que, de la misma forma, presenten nuevos paradigmas.

5. BIBLIOGRAFÍA

[1] Murguía, Eugenia Amador; López Reyes, Luis Javier (2011). Incorporación temprana a la investigación. Cuadernos de Educación y Desarrollo. Vol.3, N°28.

[2] Centro Universitario de Los Lagos. Universidad de Guadalajara. http://www.lagos.udg.mx/ (fecha de consulta: abril de 2019)

[3] Castelán Rueda, Roberto (2009). Rector de Centro. Informe Anual de Actividades 2009. Centro Universitario de los Lagos.

[4] Facultad de Medicina, Universidad de Navarra. https://www.unav.edu/web/facultad-de-medicina/la-investigacion-en-la-facultad (fecha de consulta: abril de 2019)

[5] Universidad Maimónides. https://www.maimonides.edu/investigacion/ (fecha de consulta: abril de 2019)

[6] Escalante Héctor, Corrales Rafael, Barahona Miguel (2011). La investigación científica en los estudiantes de grado y posgrado. Revista Portal de la Ciencia, UNAH. Nro.11. pp.19-25.

[7] Pamela Chirino, Mariela Galdamez, Jesús Hidalgo, Germán Bianchini, Paola Caymes Scutari (2019). Propuesta de paralelización en redes neuronales. CONAIISI 2019, La Matanza, Buenos Aires, Argentina. En prensa.

[8] Zúñiga Agustín, Aranciba José, Méndez-Garabetti Miguel, Bianchini Germán, Caymes-Scutari Paola, Tardivo María Laura (2018). Propuesta de Metodología de Generación Automática de Perímetros de Área Quemada usando Imágenes Satelitales MODIS para su Aplicación en un Sistema de Predicción de Incendios. CONAIISI 2018. Mar del Plata. Publicación on line. ISSN 2347-0372.

[9] Jesús Hidalgo, Mariela Galdamez, Pamela Chirino, Paola Caymes Scutari, Germán Bianchini (2019). Detección de exoplanetas y cómputo paralelo. CONAIISI 2019, La Matanza, Buenos Aires, Argentina. En prensa.

[10] Cohen, R 2006. BI Strategy: What's in a Name? Data Governance Roles, Responsibilities and Results Factors. DM Review, viewed 12 Jan 2007 http://www.dmreview.com/article_sub.cfm?article Id=1057220

[11] Lemieux, Victoria (2017). Blockchain and distributed ledgers as trusted recordkeeping systems: An archival theoretic evaluation framework. Conference Paper.

Enseñanza y Aprendizaje del Pensamiento Computacional y la Programación en los distintos Niveles Educativos

Ana Casali[1,2], Claudia Deco[1,3], Pamela Viale[1,3], Cristina Bender[1,3], Dante Zanarini[1,2], Natalia Monjelat[4]

[1] Facultad de Ciencias Exactas, Ingeniería y Agrimensura,
Universidad Nacional de Rosario, Rosario, Argentina.
{acasali, dante, deco, bender}@fceia.unr.edu.ar
[2] Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas
(CIFASIS: CONICET-UNR)
[3] Facultad de Química e Ingeniería del Rosario,
Universidad Católica Argentina, Campus Rosario, Argentina.
{cdeco, cbender}@uca.edu.ar
[4] Instituto Rosario de Investigaciones en Ciencias de la Educación (IRICE: CONICET-UNR)
monjelat@irice-conicet.gov.ar

Resumen

En este trabajo se presenta una línea de investigación orientada al desarrollo de estrategias y dispositivos para la enseñanza y aprendizaje del Pensamiento Computacional y la Programación en los distintos niveles educativos. Estas estrategias coinciden en combinar un enfoque desconectado (unplugged) con actividades que implican el uso de computadoras y otros dispositivos. Por un lado, se presenta el diseño y la implementación de una Especialización Docente de Nivel Superior en Didáctica de las Ciencias de la Computación, destinada a docentes de nivel primario de Santa Fe (Argentina). Por otro lado, se describen experiencias exitosas llevadas a cabo en Medellín (Colombia) a nivel primario y secundario. Finalmente, y en relación con el nivel universitario, se presentan experiencias desarrolladas en los cursos iniciales de carreras de ingeniería en la Facultad de Química e Ingeniería del Rosario, UCA (Argentina).

Palabras clave: Pensamiento Computacional, Programación, Carreras de Ingenierías, Formación Docente, Educación Primaria.

Contexto

Esta línea de I+D se está llevando a cabo en un grupo interdisciplinario a través de varios proyectos que comparten algunos de sus miembros docentes e investigadores, conformando una red de colaboración:

- PID-UNR 80020180300062UR "Aprendizaje y Enseñanza de las Ciencias de la Computación en el Nivel Primario" Directora A. Casali (2019-2022).
- PID-UCA "Pensamiento Computacional Aplicado a Educación" Directora C. Deco (2019-2022).
- Especialización Docente de Nivel Superior en Didáctica de las Ciencias de la Computación: Aprendizaje y Enseñanza del Pensamiento Computacional y la Programación en el Nivel Primario. Coordinadora A. Casali, Convenio Fundación Sadosky, UNR y Ministerio de Educación Pcia. Santa Fe (2016-2019).
- Proyecto de Investigación Científica y Tecnológico (PICT) Nº 1530: "Aprender a programar en primaria: Hacia la construcción de Tecnologías para la inclusión social", N. Monjelat (2017-2019).

Introducción

En las últimas décadas ha adquirido gran importancia en el campo educativo internacional, la relevancia de introducir la enseñanza y aprendizaje de las Ciencias de la Computación (CC) en todos los niveles del currículum escolar. Una visión de la preparación para el siglo XXI, supone que los alumnos superen el enfoque instrumental de las TIC y comprendan cómo funcionan las computadoras y cómo procesan la información. Asimismo se enfatiza el aprendizaje para la formulación de problemas y expresar su solución de forma que un computador o un humano pueda ejecutarla. Este modo de pensar propio de las CC fue nombrado como Pensamiento Computacional (PC). Wing [1] plantea que el aprendizaje del PC beneficia no sólo a los futuros informáticos, sino a toda la sociedad y por lo tanto, su enseñanza debería estar presente en todos los niveles educativos. Esto contempla el desarrollo de las habilidades vinculadas al PC, la apropiación de conceptos de programación y de la tecnología involucrada.

Por otra parte, se presenta el problema de tener pocos estudiantes en las carreras vinculadas a las áreas de ciencia, tecnología, ingeniería y matemáticas, usualmente denominadas STEM, y el problema de la alta deserción en los estudiantes universitarios de estas carreras que ocurre en muchos países latinoamericanos

Por esto, las universidades deben buscar diferentes estrategias para que el alumnado pueda motivarse y visualizar un futuro mejor a través de lo que les puede brindar ésta y el programa de formación al cual pertenece; por ejemplo, estrategias para desarrollar las destrezas del PC en los estudiantes. Una de estas alternativas consiste en clubes académicos y en semilleros de investigación. Pero en la mayoría de los casos estas iniciativas llegan demasiado tarde a los estudiantes, por lo que muchas universidades y secretarías de educación han planteado estas estrategias no sólo para los más grandes, sino también para los niños que hasta ahora están en el proceso de desarrollo de su parte lógica.

En el contexto europeo es posible relevar diferentes estudios donde se analizan procesos y situaciones vinculadas a la enseñanza de las CC, la programación y el PC [2,3], mientras que en el contexto latinoamericano se observan iniciativas aisladas, siendo aún necesario implementar políticas públicas al respecto [1].

En nuestro país, el Consejo Federal de Educación señaló al aprendizaje de la programación de importancia estratégica para el Sistema Educativo Nacional durante la escolaridad obligatoria (Res. CFE Nº 263/15, 2015) y en setiembre de 2018 se aprobaron los "Núcleos de Aprendizajes Prioritarios de Educación Digital, Programación y Robótica" (NAP EDPR) para los diferentes niveles de la educación obligatoria (Res. CFE Nº 343/18, 2018). En el Anexo II de la citada resolución se establece un plazo de dos años para la adecuación curricular y la formación docente en esta dirección. En este sentido, es posible señalar que desde el marco institucional existen dos aspectos clave e interrelacionados que condicionan la introducción de nociones de las CC en el sistema educativo argentino. Por un lado, la necesidad de modificar los planes de estudio para incluir estos contenidos y por otro, la implementación de propuestas de formación docente que atiendan al aprendizaje y enseñanza del PC y la programación. En particular, en el contexto de la provincia de Santa Fe, es posible observar que la formación y capacitación docente se ha enfocado en un primer nivel de apropiación instrumental de TICs en la práctica educativa [8].

Teniendo en cuenta esta realidad, se consideró importante llevar adelante una

primera Especialización Docente de Nivel Superior en Didáctica de las Ciencias de la Computación, atendiendo a las particularidades del contexto

Por otro lado, en la actualidad existen muchas instituciones, públicas y privadas, preocupadas por las múltiples dificultades que tienen los estudiantes universitarios al enfrentarse con materias como física, matemática, programación, entre otras. Para esto, en Colombia se han propuesto diferentes estrategias a nivel universitario como ser semilleros y clubes académicos, y también a nivel de educación básica secundaria con propuestas de la Secretaría de Educación y de empresas privadas. Algunas de las propuestas son Innobótica, Ingeniería[n] e Interchange. Estas propuestas promueven el aprendizaje significativo en varias áreas STEM del conocimiento. En [4] se presenta una propuesta para la Institución Universitaria Pascual Bravo, que propone crear un club enfocado en los más niños, hacia el aprendizaje de software específico para el logro del PC.

Por otra parte, para introducir conocimientos sobre el PC en los primeros años de las carreras de ciencia, tecnología, ingeniería y matemáticas, se propuso el proyecto actual en UCA. Enseñar las bases de las CC a los alumnos del primer año de las carreras de ingeniería industrial, ingeniería ambiental, licenciatura en química y licenciatura en tecnología de los alimentos de la Facultad de Química e Ingeniería del Rosario (UCA) resulta de fundamental interés, ya que, si bien no son conocimientos específicos de estas especialidades, el proceso de pensamiento involucrado es similar al que necesitan estos profesionales para resolver los problemas inherentes a sus profesiones.

Líneas de Investigación y Desarrollo

Formación Docente de Nivel Primario en Didáctica de las Ciencias de la Computación

A partir de un Convenio suscripto entre la Fundación Sadosky, la Universidad Nacional de Rosario y el Ministerio de Educación de la provincia de Santa Fe se diseñó la "Especialización Docente de Nivel Superior en Didáctica de las Ciencias de la Computación: Aprendizaje y Enseñanza del Pensamiento Computacional y la Programación en el Nivel Primario" (Res. ME N° 1565/17). La primera cohorte se dictó en el Instituto Superior de Formación Docente N° 36 "Mariano Moreno" de la ciudad de Rosario entre agosto de 2017 y setiembre de 2019, con una carga horaria total de 400 hs, con el 80% presencial. Esta formación tiene por objetivo formar docentes capaces de experimentar y reflexionar críticamente acerca de los procesos de desarrollo del PC y la programación, a los fines de construir las competencias adecuadas al nivel primario que posibiliten una práctica educativa innovadora con énfasis en la resolución de problemas mediante la producción colaborativa e interdisciplinaria de Tecnologías para la Inclusión Social [3, 5]. Dado que se ha probado que el desarrollo de procesos de enseñanza y aprendizaje a través de actividades "unplugged" o desconectadas, resulta efectivo para desarrollar habilidades cognitivas del PC [2, 6], este postítulo combina los enfoques unplugged y plugged en sus diferentes módulos. La Especialización se organizó en cuatro semestres dictando tres módulos en formato de taller en cada uno, en los que se articulan los contenidos de manera espiralada. Los ejes de los mismos son: el PC, la programación y conceptos tecnológicos. Se incluyeron también proyectos integradores, con el fin de que faciliten a los maestros llevar al aula estos aprendizajes [7, 8].

Enseñanza del PC en el Nivel Secundario

Respecto al nivel primario y secundario, se están analizando algunos juegos para el logro de PC y sus principales características. Estos fueron separados en dos grandes grupos: juegos básicos de lógica y juegos enfocados al aprendizaje de la programación de sistemas informáticos. Dentro de los juegos de lógica, se pueden mencionar dos tipos de juegos: aquellos que pueden ser abordados dentro de un aula convencional (deducir valores, mover fósforos, operaciones matemáticas regresivas, sudoku, etc.) y aquellos donde es necesario el uso de un computador (Candy Crush Saga, Tetris, etc.). Entre los juegos enfocados al aprendizaje de la programación algunos de los programas en análisis son Scratch, RoboMind y Js Robot, entre otros.

Experiencias del desarrollo del PC en Carreras de Ingeniería

Este proyecto, que se desarrolla en el ámbito universitario, comenzó concentrándose en las materias Informática e Informática General que se dictan en el primer año de las carreras de Ingeniería Industrial, Ingeniería Ambiental, Licenciatura en Química y Licenciatura en Tecnología de los Alimentos. Se desarrolló material específico para estas materias haciendo hincapié en aquellos conocimientos relacionados al PC, como ser: abstracción, descomposición de problemas en subproblemas, reconocimiento de patrones y algoritmos. Se continuará con la elaboración y mejora de este material durante los próximos meses.

Antes de comenzar el último cuatrimestre, en agosto de 2019, se evaluó a los alumnos sobre PC utilizando un test propuesto inicialmente por el equipo de Marcos Román-González, con el objetivo de recabar información acerca de los conocimientos previos sobre las CC con los cuales los alumnos llegan a la Universidad. Al finalizar dicho cuatrimestre, en diciembre de 2019, los alumnos volvieron a responder el test.

Resultados

Se ha diseñado e implementado una formación superior para docentes del nivel primario en CC, la primera en la provincia de Santa Fe y una de las tres primeras en el país, ya que se encuentran en curso formaciones similares en Río Cuarto, Córdoba y en Tandil, Buenos Aires. Los resultados son sumamente alentadores, tanto desde la apropiación de los maestros de los conocimientos de CC, como desde su percepción de estar capacitados para implementar estos contenidos en el aula de forma inclusiva. Resultados preliminares de las opiniones docentes en cuanto a la didáctica y contenidos de distintos módulos son altamente positivos y pueden verse en [8]. A partir de la información recogida en un focus group, realizado al final del cursado los siguientes conceptos surgen como relevantes: *"las herramientas que brinda el postítulo para pensar en las prácticas docentes desde otro ángulo", "las actividades desenchufadas, que brindan la posibilidad de trabajar sin la computadora", "la formación en pensamiento computacional" y "el software pilas bloques, se destaca como una herramienta adecuada para los alumnos".* Las observaciones realizadas en instancias áulicas del Proyecto Final, también han sido evaluadas positivamente. Destacamos además, que en este proceso se ha conformado un equipo docente interdisciplinario y competente en la formación docente en CC para el nivel primario. Lo cual hará posible replicar y expandir esta experiencia en la provincia. Se espera que la presente especialización se constituya como una propuesta de referencia para el

campo de la enseñanza de CC en Argentina, cuyo carácter innovador posibilite un trayecto transformador de las currículas y prácticas docentes actuales.

Con respecto al PC en el nivel inicial y secundario, en Colombia se lleva adelante una propuesta para el desarrollo de PC para niños de 5 años y más. En particular, la propuesta Programando Robots que se desarrolló en 2017 en el Instituto Técnico Industrial Pascual Bravo donde se observó que el uso de robótica educativa en la adquisición de competencias y desarrollo de PC, permitió observar en los alumnos grandes conocimientos en la parte eléctrica y electrónica, así como en la parte mecánica y de programación, lo cual se evidenció en el lenguaje utilizado y el desarrollo propio del proyecto [4, 9].

Con respecto a la incorporación del PC en el primer año de carreras de ingeniería, se están analizando las diferencias entre las respuestas al test de Marcos Román-González, pre y post cursos de Informática e Informática General. Esta comparación nos brindará valiosa información acerca de la aprehensión de los conocimientos sobre el PC por parte de los alumnos y dará una pauta sobre la calidad del material desarrollado. Se trabajará sobre los contenidos de los apuntes de estas cátedras, considerando estos resultados.

Formación de Recursos Humanos

El equipo de trabajo está integrado por las doctoras Ana Casali, Claudia Deco, Natalia Monjelat, Pamela Viale, la magister Cristina Bender y los licenciados Dante Zanarini y Martín Ceresa. Dentro del marco de esta línea de I+D se desarrollan actualmente un doctorado en Informática (Lic. Ceresa), un doctorado en Ciencias de la Educación (Ing. Maritza García Angarita) y una maestría en Educación (Lic. Sebastián Velazquez).

Cabe destacar que la formación que se lleva a cabo de docentes del Depto. de Ciencias de la Computación en la enseñanza de la disciplina a docentes de nivel primario y secundario, es un trabajo interdisciplinario con especialistas de las ciencias de la educación.

Referencias

[1] Wing, J. Computational Thinking Benefits Society. Social Issues in Computing. 2014.

[2] Brackmann C., Román-González M., Moreno-León J., Robles G., Casali A. and Barone D. Computational Thinking Unplugged: Teaching and Student Evaluation in Primary Schools. In WIPSCE Nijmegen, The Netherlands, ACM. November, 2017.

[3] Monjelat N. & San Martín P. Programar con Scratch en contextos educativos: ¿Asimilar directrices o co-construir Tecnologías para la Inclusión Social? Praxis Educativa, 20(1), 61–71. 2016.

[4] García M., Deco C., Bender C., Collazos C. Invited paper: Robotics Based Strategies to Support Computational Thinking: The Case of the Pascual Bravo Industrial Technical Institute. Journal JCS&T, 17(1), 59-67. 2017.

[5] Monjelat, N. "Programming Technologies for Social Inclusion," En LACLO, 2017. La Plata, Argentina. 2017.

[6] Bell, Witten & Fellows. CS Unplugged: An enrichment and extension programme for primary-aged students. 2015.

[7] Casali A., Zanarini D., Monjelat N. & San Martín P. Teaching and Learning Computer Science for Primary School Teachers: an Argentine Experience, LACLO 2018, San Pablo, Brasil, IEEE Xplore. 2018.

[8] Casali A., Zanarini D., Monjelat N. & San Martín P. Formación docente en Ciencias de la Computación: experiencias de la primera Especialización para el nivel Primario de la Provincia de Santa Fe, VIII WIE-CACIC 2019, Río Cuarto, Argentina, RedUnci. 2019.

[9] García M., Deco C., Bender C., Collazos C. Herramientas de Diseño para el Desarrollo de Competencias en Educación Básica, Media y Tecnológica: Experiencia en el Instituto Técnico Industrial Pascual Bravo de Colombia. Revista TE&ET, Argentina. 2017.

“Tecnicatura Superior en Programación (TSP) Modalidad Semipresencial”

Mg. Ing. Carrizo, Blanca Rosa
GICCAP. Dpto. Ing. Industrial. Facultad Regional Córdoba (F.R.C.)
Universidad Tecnológica Nacional (U.T.N.).
bcarrizo@frc.utn.edu.ar / brcarrizo@yahoo.com.ar

RESUMEN

La presente propuesta se fundamenta en un análisis de requerimientos de la carrera TSP, que se dicta en modalidad presencial en la F.R.C. de la U.T.N. desde 1999, y que hoy exige un cambio de paradigma: migrar de la modalidad presencial a la semipresencial.

Fundamenta este proyecto el ingreso masivo de estudiantes a distintas ofertas académicas, la saturación y desborde de espacios físicos ; provocando falta de capacidad áulica y de laboratorios.

Para ello, es necesario desarrollar una metodología de trabajo que facilite, a modo de guía de buenas prácticas, el análisis, diseño e implementación de Programas Educativos Virtuales (Fernández A., 2014) para esta carrera, validándola con una prueba piloto.

Un Programa Educativo Virtual planteado como modelo pedagógico, implica desarrollar competencias profesionales integrales que permitan a estudiantes y egresados desempeñarse con eficiencia, calidad, liderazgo y compromiso social en el entorno donde ejercen su profesión.

El presente trabajo se aprobó con el desarrollo de una prueba piloto en una instancia inicial denominada “Nivelación de Contenidos” para el ciclo lectivo 2019.

El proyecto inicial se extendió e implementó en toda la carrera; y hoy dada la contingencia sanitaria a nivel mundial, se está desarrollando en modalidad virtual.

Palabras Claves: Tecnologías de Información y Comunicación (TIC), Educación a distancia (EaD), Sistema Institucional de EaD (SIED).

CONTEXTO

La U.T.N. es una Universidad pública que presenta dos características que la distinguen del resto del sistema universitario argentino.

En primer lugar, es la única del país cuya estructura académica tiene a las Ingenierías como objetivo prioritario.

La segunda característica está dada por su carácter federal dado que está presente en todas las regiones del país a través de sus treinta (30) Facultades Regionales que se ubican a lo largo y ancho del mapa.

Actualmente, TSP se dicta en dieciséis (16) Regionales.

Esto se traduce en una permanente e íntima vinculación con los sistemas productivos regionales y un fecundo intercambio académico a nivel nacional.

El perfil mediterráneo de Córdoba está marcado por la industria metalmecánica, informática, alimenticia, cultivos, ganadería y turismo, y en este contexto el perfil programador es muy requerido.

Las empresas tecnológicas tienen problemas para conseguir estos perfiles por falta de formación técnica y experiencia.

En respuesta a ello, la U.T.N. ha asumido un rol activo y determinante en relación con las prioridades que debe afrontar la educación en el siglo XXI, aprobando dentro de su marco institucional, la Ord. N° 1627 del 26/06/18 que crea el SIED y la Ord. 1688 del 25/10/18, que establece que los títulos de pregrado se denominarán “Técnicos Universitarios”.

1. INTRODUCCIÓN

Las Tecnicaturas surgen en el seno de la U.T.N. (Res. N° 278/94 y 131/95) como respuesta a demandas actuales del medio, cubren una real necesidad y son ofertas que permiten a los egresados una rápida inserción en el mercado del trabajo.

El Técnico en Programación estará capacitado para realizar programas o componentes de sistemas de computación, interpretar especificaciones de diseño, documentar los productos realizados, verificar los componentes programados, buscar causas de malfuncionamiento y corregir los programas o adaptarlos a cambios a nivel especificaciones técnicas, integrando equipos de trabajo con Analistas y Testers.

El título del proyecto a desarrollar es "TSP Modalidad Semi Presencial" y la línea de aplicación del conocimiento es la línea 1 denominada Diseño, Gestión y Desarrollo Curricular a través de las TIC. / Proyecto 1. Proyecto para la generación o actualización de programas educativos.

El alcance es una propuesta de mejora, con diseño e instrumentación del proyecto de forma parcial.

TSP lidera, desde su nacimiento (ciclo lectivo 1999), el segmento Tecnicaturas, en la Facultad Regional Córdoba.

En estos últimos seis (6) años, la cantidad de ingresantes de cada ciclo lectivo se vio incrementada por la demanda laboral para este perfil en nuestra provincia (en este ciclo lectivo hay 400 ingresantes, lo cual representa el 40% de la población de pregrado).

Simultáneamente a la evolución de los números de la carrera, se diseñó una Encuesta a medida para ser integrada por aspirantes a la carrera, de la cual se obtuvieron las siguientes conclusiones:

- El perfil Programador es muy requerido en el mercado laboral regional, dado que Empresas de Informáticas Internacional radicadas en Córdoba y las Pymes locales así lo exigen (desarrollador web, programador ASP.NET, entre otros similares).
- TSP en una carrera corta de dos años de duración con un Trabajo Final orientada a "hacer", que tiene título oficial y es de validez nacional, los cual la posiciona como una alternativa válida; frente a una carrera de grado que es un proyecto a largo plazo y cuya formación es más genérica.
- Los potenciales ingresantes constituyen una población muy heterogénea y está constituida por adultos mayores (idóneos, gente que abandonó carreras similares y necesita el título), adolescentes que egresan del ciclo secundario (de colegios de formación similar), alumnos de otras carreras de la F.R.C. y otras Universidades, pases Inter-facultades y de otras Instituciones privadas que requieren la modalidad semipresencial y/o virtual dado que en un 90% están trabajando actualmente.
- Los adultos mayores solicitan articulación de la TSP con la carrera de grado de la F.R.C. y/o la implementación de un 3er. Nivel que amplíe el campo laboral (hay un 3er nivel aprobado para formar un Técnico en Análisis. De esta forma, es factible articular con tres (3) años de Tecnicatura una Licenciatura en un área afín tanto en la U.T.N. como en otras Instituciones del medio.

Dado que en junio del 2018 la U.T.N. aprobó los lineamientos del SIED a través de la Ord. 1627, este proyecto educativo se inscribe en el marco de una política que le da sentido y significación, basado en la mejora continua de los procesos formativos.

Se sustenta en un enfoque pedagógico socio constructivista y sistémico mediado por las tecnologías apropiadas y se concibe como un sistema flexible y dinámico que pueda absorber tanto los rápidos cambios tecnológicos como los avances científicos y en el conocimiento.

El SIED contiene a las carreras de pregrado, grado y posgrado, sustentado en un enfoque pedagógico socioconstructivista y sistémico mediado por las tecnologías apropiadas, que atendiendo las líneas estratégicas, garantizará que las propuestas académicas que adoptan la modalidad no presencial o mixta cumplan los criterios de excelencia y calidad propios de la U.T.N., en un continuo proceso de actualización, reflexión, investigación e innovación.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

- Plataforma de Educación Virtual
- Bases de Datos y Almacén de Datos.
- Leguajes de Programación basados en la filosofía Software libre y Propietario.
- Minería de Datos.
- Diseño de Encuestas a medida
- Relevamiento y análisis de datos

Se utilizarán los canales oficiales de la U.T.N.:

- Autogestión: canal de comunicación, validado a través del Sistema Académico, donde se pueden dejar mensajes oficiales que los estudiantes están acostumbrados a revisar.
- Aula virtual (Moodle, UV): allí se puede establecer foros para responder preguntas, dejar material organizado por clases, y plantear mecanismos de seguimiento y evaluación, entre otras actividades.
- Videoconferencia: la U.T.N. ha optado por el sistema de Zoom (https://utn.edu.ar/es/secretariatic/servicios/tic-servicios/zoom), allí los docentes pueden crear una cuenta a partir del mail institucional (xx@frc.utn.edu.ar). Zoom no tiene límite de participantes, puede pasar simultáneamente una presentación en Powerpoint y explicarla, con un límite de 40 minutos, vencido el mismo, se reinicia la sala con las mismas características. Se puede informar la Id de sala por Autogestión.

3. RESUTADOS OBTENIDOS / ESPERADOS

El Objetivo Principal de esta tesis era: "Migrar el programa educativo TSP hacia la modalidad semipresencial para atender la demanda de formación en la U.T.N. – F.R.C., a través de la implementación de una prueba piloto."
Este objetivo general se cristalizó en los siguientes objetivos específicos:

- Definir una estructura organizacional para la gestión de esta nueva modalidad que se ajuste al perfil de la U.T.N., donde se definan los roles necesarios para cada función: Coordinador, Diseñador Instruccional (DI), Diseño Web (DW), apoyo administrativo y Tutores, entre otros.
- Diseñar capacitaciones a medida con el objetivo de comprometer al docente en esta nueva modalidad y desarrollar las competencias necesarias para el diseño e implementación modular de cursos en ambientes virtuales.
- Definir y seleccionar el perfil DI que guie la producción de contenidos en función de sus antecedentes y experiencias laborales comprobables en empresas del medio.
- Diseñar materiales a medida organizados por cuatrimestres, integrados por módulos o asignaturas, para ser publicados en el Aula Virtual conforme imagen institucional
- Definir, junto al Centro de Cómputos, el acceso a una nueva área de producción desde un sitio institucional exclusivo para carreras académicas, segmentadas en: Pregrado, Grado y Posgrado; con la finalidad de migrar el aula virtual del área de prueba a la plataforma institucional (Moodle).
- Implementar una prueba piloto denominada NC para los ingresantes a la carrera, con la finalidad de homogeneizar los conocimientos de cada ingresante.

Tanto el objetivo principal como los específicos se cumplieron en un 100% nivel prueba piloto; y se amplió a toda

la carrera segmentada en cuatrimestres (4) como prevee el diseño curricular de la misma.
Actualmente, el Centro de Cómputos definió un sitio institucional homogéneo para los tres segmentos académicos: pregrado, grado y posgrado; que responda a la imagen institucional de la U.T.N., en nuestro caso Tecnicaturas en general y TUP en particular (Link: https://uvs.frc.utn.edu.ar/login/index.php)
En este sitio institucional, se ha cargado el material actualizado y mejorado de NC y de todos cuatrimestres:

- 1er. Cuatrimestre (Programación I / Laboratorio de Computación I / Inglés I / Sistemas de Procesamiento de Datos / Matemáticas).
- 2do. Cuatrimestre (Programación II / Laboratorio de Computación II / Inglés II / Arq, y Sistemas Operativos / Estadística / Met. Investigación).
- 3er. Cuatrimestre (Programación III / Laboratorio de Computación IV / Org. Contable / El. Inv. Operativa / Org. Empresarial.
- 4to. Cuatrimestre: en proceso de elaboración de materiales.

Simultáneamente se están desarrollando talleres docentes con un coach organizacional, donde se profundizan los siguientes tópicos.

- Diseño Estratégico TSP
- Competencias del Tutor
- Normas APA aplicadas al diseño de materiales
- Seguridad en ZOOM
- Evaluaciones en Moodle
- Software interactivo para optimizar el dictado de teóricos como Miro.
- Edición de Videos.

A medida que se avanza en la implementación, se proponen nuevos tópicos que se tratan en el seno de la carrera como capacitaciones a medida.
Como parte del proceso de retroalimentación, para tender a la mejora continua, es fundamental la construcción de un Sistema de Gestión de Conocimiento (SGC), que reciba, almacene, procese y distribuya conocimiento y que incluya la gestión de recursos humanos, tecnológicos e informativos, interrelacionados dinámicamente.
Para ello, la F.R.C. debe elaborar un SGC propio que represente una combinación creativa de elementos culturales, tecnológicos y de procesos, con miras a favorecer la estrategia organizacional.
Queda pendiente, a nivel F.R.C., institucionalizar un Área de TIC´s responsable de planificar, organizar, dirigir y coordinar las actividades derivadas de la Ord. 1627; así como mejorar los mecanismos internos para desarrollar procedimientos estándares tendientes a agilizar futuras propuestas académicas a diseñar.
En este contexto, hoy confluyen en este momento académico varias situaciones (oportunidades) que enriquecen la propuesta.
Algunas de ellas son:

- La demanda de programadores viene creciendo exponencialmente en Argentina, donde se requieren más del doble de los técnicos que ofrece el mercado laboral actual. En Córdoba, la industria del software es la 2da. Fuente de empleo en la provincia.
- Las Tecnicaturas han sido reconocidas como Tecnicaturas Universitarias (TUP) conforme Ord. 1688 del 25/10/18, que establece que los títulos de pregrado pasarán a denominarse Técnicos Universitarios; dado que responden a demandas sociales, que integran el espacio existente entre el nivel medio, que no prepara laboralmente, y un título de grado, que implica condicionamientos y tareas diferenciadas.
- La implementación de TSP está potenciada por la Ord. 1627 de junio de 2018 mediante la cual, se crea del SIED de la U.T.N., que contiene a las carreras de pregrado, grado y posgrado, sustentado en un enfoque pedagógico socioconstructivista y

sistémico mediado por las tecnologías apropiadas; permitiendo la implementación de propuestas académicas en modalidad semipresencial o a distancia.

Firma de un convenio con la Secretaría de Políticas Universitarias (SPU) para implementar la carrera en modalidad semipresencial en Jesús María en un Centro Regional de Educación Superior (CRES).

4. FORMACIÓN DE RECURSOS HUMANOS

Tesista: Ing. Blanca Carrizo ("C" y "III").

Directores de Tesis:

- Mtro. Jorge Stefan Cruz León
- Dra. Erendira Yaretni Mendoza Meza
- Mtra. Laura Olivia Cerón González.
- Mtro. Sergio Olguín Aguirre.
- Mtra. Linda Gladiola Flores Flores

Tutores: docente de la carrera TSP:

- Esp. Ing. Kunda Beatriz
- Esp. Ing. Perez Rita
- Mg. Caminos Constanza
- Ing. Serrano Diego
- Mg. Ing. Arietti Adrián
- Esp. Ing. Di Gianantonio Alejandra
- Lic. Freudenreich Graciela
- AS. Botta Oscar
- AS. Santoro Ezequiel
- Esp. Ing. Stefanich Clarisa
- Ing. Olmedo Sergio

5. BIBLIOGRAFÍA

Belloch, Consuelo. (2011). Universidad de Valencia. Unidad de Tecnología Educativa. Diseño Instruccional. Recuperado: https://www.uv.es/bellochc/pedagogia/EVA4.pdf.

Bellucci, Marcelo. (2017). Faltan programadores, cada año quedan 5 mil puestos vacantes y buscan hasta chicos. Recuperado: https://www.clarin.com/sociedad/falta-programadores-quedan-mil-puestos-vacantes-pais_0_B1FHsMlB-.html.

Tropical Server. (Junio 2018). ¿Qué es Moodle? La guía definitiva – 2018. Recuperado: https://www.tropicalserver.com/ayuda/que-es-moodle/

Universidad Tecnológica Nacional Resolución. N° 278/1994 del Consejo Superior. Aprueba la existencia de carreras cortas en la U.T.N. Recuperado de http://csu.rec.utn.edu.ar/docs/php/salida.php3?tipo=RES&numero=278&anio=1994 &facultad=CSU&pagina=1

Universidad Tecnológica Nacional Resolución. N° 278/1994 del Consejo Superior. Pautas Curriculares para el Desarrollo de Carreras Cortas en la U.T.N. Recuperado de: http://csu.rec.utn.edu.ar/docs/php/salida.php3?tipo=RES&numero=131&anio=1995 &facultad=CSU&pagina=1

Universidad Tecnológica Nacional Resolución Junio 2018 del Consejo Superior. Lineamientos del Sistema Institucional de Educación a Distancia de la Universidad Tecnológica Nacional. Recuperado de: http://csu.rec.utn.edu.ar/docs/php/buscador.php3

Zubieta García y Rama Vitale. (2015). La Educación a Distancia en México.

Plataforma para la Programación Tangible

Celeste Ramos Jorge Rodríguez Laura Cecchi

{celeste.ramos, j.rodrig, lcecchi }@fi.uncoma.edu.ar

Grupo de Investigación en Lenguajes e Inteligencia Artificial
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

Resumen

En las últimas décadas la enseñanza de las Ciencias de la Computación en el ámbito de la educación ha ido evolucionando. En este contexto, se fue incorporado a la currícula la utilización de distintas herramientas que permitan a los estudiantes comenzar a programar.

Para desarrollar un programa, en la mayoría de los lenguajes existentes, se requiere de ciertos conocimientos y habilidades previas, como por ejemplo saber contar, leer y escribir.

La Programación Tangible permite la utilización de piezas físicas como componentes de programación. Cada pieza tiene características y funcionalidades específicas que al secuenciarlas es posible generar un programa.

En este trabajo se presenta una Línea de Investigación que propone definir y diseñar un lenguaje de Programación Tangible. Asimismo, se diseñará e implementará un entorno Web que permita, a los estudiantes no alfabetizados, programar en dicho lenguaje.

Palabras Clave: PROGRAMACIÓN TANGIBLE - ENSEÑANZA DE LA PROGRAMACIÓN - EDUCACIÓN EN CIENCIAS DE LA COMPUTACIÓN - PLATAFORMA DE PROGRAMACIÓN TANGIBLE - ENTORNO WEB PARA LA PROGRAMACIÓN TANGIBLE.

Contexto

Esta línea de investigación se desarrolla, por un lado, en el contexto de los temas de interés que promueve el Grupo de Investigación en Lenguajes e Inteligencia Artificial(GILIA), de la Facultad de Informática. En particular, se enmarca en el ámbito de dos proyectos de investigación miembros del GILIA, ambos financiados por la Universidad Nacional del Comahue y con una duración de cuatro años a partir de enero del 2017: *Agentes Inteligentes. Modelos Formales y Aplicaciones para la Educación (04/F015)* y *Agentes Inteligentes y Web Semántica (04/F014)*.

Por otro lado, el trabajo está en el contexto del Convenio Marco de Colaboración firmado durante 2016 entre la Facultad de Informática y el Ministerio de Educación de la Provincia del Neuquén. Particularmente, se trabaja con el Consejo Provincial de Educación de la Provincia de Neuquén.

1. Introducción

Hoy en día es muy común ver a niños desde los 3 años utilizar celulares, tablets o computadoras para acceder a distintas aplicaciones con fines lúdicos o formativos.

En la última década en muchas escuelas la educación en informática tradicional se ha ido modificando. Progresivamente, la formación en ofimática está perdiendo relevancia dando lugar a la incorporación de tópicos que ayuden a desarrollar el Pensamiento Computacional [5, 6, 10].

El Pensamiento Computacional es una forma de trasponer prácticas y conceptos fundamentales de las Ciencias de la Computación para la resolución de problemas no necesariamente computacionales [12, 11]. En este contexto se están utilizando distintas herramientas que permiten a los estudiantes comenzar a programar alrededor de los 8 años.

Una de las formas más amigables para desarrollar el Pensamiento Computacional, en niños desde los 3 años de edad, es la Programación por medio de Interfaces de Usuario Tangibles(TUI) o simplemente Programación Tangible. Se fundamenta en que los niños aprenden de una manera activa: aprenden a través de experiencias con personas, objetos y demás cosas que puedan percibir de su mundo.

Las TUI [4] son una forma de interacción hombre máquina, novedosa y poco explorada, donde la interfaz se amplia al integrar la computación con objetos físicos y entornos que están vinculados a representaciones digitales. En definitiva, permiten que el usuario pueda manipular de forma física la información digital [8].

En la Programación Tangible los estudiantes construyen código ensamblando, encadenando y/o conectando objetos físicos, que representan instrucciones precisas, que luego se traducen a un lenguaje de programación generando un programa [9].

Actualmente, existen múltiples herramientas de Programación Tangible [7] que permiten a los estudiantes aprender a programar desde los primeros años de edad.

Si bien, estas herramientas están adaptadas a las distintas edades, muchas no son fáciles de combinar debido a la exclusividad de su diseño, ya que sus licencias de software y/o hardware suelen ser propietarias. Por otro lado, su alto costo las hace poco accesibles para la gran mayoría de los niños o para las escuelas a las que éstos concurren.

En esta Línea de Investigación se propone definir y diseñar un Lenguaje de Programación Tangible, que sea comprensible para estudiantes no alfabetizados. A diferencia de las herramientas que se estuvieron analizando, para su definición se priorizará la utilización de elementos de bajo costo, accesibles y de uso cotidiano.

Además, se desarrollará una herramienta Web con licencia Open Source, que permita a los estudiantes programar en el lenguaje definido. Dicha herramienta será accesible desde cualquier navegador, por lo que sólo se necesitará tener un dispositivo con browser y conexión a internet, lo que permitirá su fácil acceso sin necesidad de instalarla. Por otra parte, para su utilización no se requerirá de dispositivos de hardware específicos o exclusivos.

El trabajo presentado está estructurado como sigue. En la siguiente sección se introducen los proyectos de investigación que dan contexto al trabajo y la línea en desarrollo. En la sección 3 se detallan los avances en este trabajo y los trabajos futuros. Finalmente, se comentan aspectos en relación a la formación de recursos humanos.

2. Línea de investigación y desarrollo

Esta línea de investigación, desarrollo y transferencia propone definir y diseñar un lenguaje de Programación Tangible. Asimismo, tiene como objetivo diseñar e implementar un Entorno Web de Programación Tangible que permita a estudiantes no alfabetizados programar en dicho lenguaje.

Este trabajo se desarrolla en el marco de las actividades de los proyectos de investigación *Agentes Inteligentes. Modelos Formales y Aplicaciones para la Educación* y *Agentes Inteligentes y Web Semántica*. Por un lado el proyecto *Agentes Inteligentes. Modelos Formales y Aplicaciones para la Educación*, busca desarrollar modelos teóricos que contribuyan a la producción de un marco conceptual de referencia, a fin de asistir la inclusión de la computación en la educación. En este sentido se plantea como objetivo el diseño e implementación de herramientas que los soporten. Por otra parte, el diseño de lenguajes caracterizados por gramáticas formales y el desarrollo de entornos que cumplan los principios de la Web Semántica son objetivos específicos del proyecto de investigación *Agentes Inteligentes y Web Semántica*.

Así, la definición y caracterización formal de un lenguaje de Programación Tangible, como el desarrollo de un entorno Web de programación, que contribuye a hacer posible la construcción de saberes del área de conocimiento algoritmos y programación, a sujetos no alfabetizados, es de interés para ambos proyectos, colaborando en su diseño e implementación. Los resultados de este trabajo permitirán ampliar la población estudiantil facilitando el desarrollo de los proceso de enseñanza y de aprendizaje de la programación.

En este sentido se plantean los siguientes ejes de trabajo para el desarrollo de la plataforma:

Lenguaje de Programación Tangible

Los Lenguajes de Programación Tangible (LPT) asumen qué piezas físicas pueden constituirse en elementos sintácticos de un lenguaje (variables, instrucciones, operadores, entre otros). La colección de piezas físicas proporciona al programador, posiblemente un sujeto no alfabetizado, la posibilidad de secuenciar u ordenar elementos físicos para construir un algoritmo [7, 9].

Como punto de partida de este trabajo se define un LPT simple, que no involucra el uso de variables, ni de las estructuras de control alternativa y repetitiva, y que se compone de un conjunto reducido de primitivas. Se espera que a través de este lenguaje se pueda programar un comportamiento sencillo de un agente o bot, en un mundo representado en forma de cuadrícula en un entorno Web. Las instrucciones para modelar el comportamiento del agente incluirán, entre otras, avanzar o girar a derecha.

En esta plataforma, las piezas físicas con las que se trabajará serán trozos de papel, cartulina u otros recursos de uso habitual en la escuela en los que se pueda distinguir símbolos impresos con un fondo blanco.

Los estudiantes construyen programas en el LPT colocando una secuencia de tarjetas que representan las instrucciones del lenguaje.

Entorno Web

En el momento en que el programa físico está elaborado, se digitaliza usando una cámara fotográfica convencional. La imagen capturada se procesa para producir un programa digital que luego se ejecuta. Finalmente, el resultado de la ejecución se mapea directamente sobre el mundo representado en el entorno web.

El desarrollo del entorno Web está divido en tres etapas:

1. Digitalización del Programa Tangible (PT): es el proceso que permite la lectura del PT por medio de una cámara web o de un dispositivo móvil.

2. Procesamiento del Programa: es el proceso por el cual la versión digital del PT, esto es, una imagen, es representado en un lenguaje susceptible a ser ejecutado sobre el mundo.

3. Ejecución y Visualización de los Resultados del PT: realiza la ejecución del programa mostrando el comportamiento del agente dentro del mundo, representado en forma de grilla.

Mediante este desarrollo, se espera que la herramienta propuesta contribuya a simplificar y enriquecer los procesos de enseñanza y de aprendizaje de la programación, alcanzando una población no alfabetizada. Se busca evitar que los estudiantes interactúen con la sintaxis de un lenguaje de programación textual. Por otra parte, al ser una herramienta Web se eliminan los aspectos relacionados a la instalación y configuración del sistema, sin demandar recursos costosos ni específicos.

3. Resultados obtenidos y esperados

En una primer etapa se definió y diseñó un LPT simple, prestando especial atención a la utilización de recursos de bajo costo y de una sintaxis que resulte accesible y familiar a niños no alfabetizados.

Dicha sintaxis está compuesta de símbolos que son intuitivos para los alumnos y que no requieren saber leer ni escribir para poder ser

interpretados. Cada símbolo es un gráfico de color negro que se ubica en el centro de un cuadro con fondo blanco, cuyo límite se establece con un borde de color negro, como muestra la Figura 1. La simplicidad de la representación de los símbolos permite que puedan ser impresos, dibujados o construidos sobre piezas físicas (papel, cartulina, cartón, etc).

Inicialmente, se definieron como símbolos flechas, con cuatro direcciones distintas, donde su semántica indica las acciones posibles a ejecutar por el bot, sobre un mundo representado en forma de cuadrícula. En este sentido las cuatro acciones iniciales definidas para el bot son: avanzar, retroceder, girar a la izquierda y girar a la derecha.

Un programa tangible (PT) desarrollado en este lenguaje se interpreta de izquierda a derecha con sus instrucciones, piezas físicas, ubicadas en forma secuencial. En la Figura 1 se puede observar un PT representado por diez piezas (o instrucciones). Cuando ese PT sea ejecutado en el entorno Web, los movimientos del bot serán mapeados en el mundo como avanzar tres casilleros, girar a la derecha, avanzar dos casilleros, luego girar a la izquierda, retroceder un casillero, girar a izquierda y finalmente, avanzar otro casillero.

Una vez definido el LPT, se diseñó la arquitectura del entorno Web basada en el Modelo Vista Controlador (MVC), que permitió clasificar sus componentes y definir las relaciones entre ellos. Los componentes de esta arquitectura son:

- *Modelo:* contiene una representación de los lenguajes utilizados:
  - Lenguaje de Programación Tangible (LPT): está compuesto de símbolos impresos en piezas físicas, cuya semántica se corresponden con las acciones para el bot a manipular.
  - Lenguaje Intermedio (LI): es un lenguaje que, como su nombre lo indica hace de intermediario entre el LPT y el mundo. El objetivo de utilizar este tipo de lenguaje es desacoplar el LPT de la forma en que se representa el mundo (interfaz de usuario).

    Para este desarrollo se optó por representar el LI con el formato de texto JSON[3] (JavaScript Object Notation), ya que es una notación simple, muy utilizada para el intercambio de datos.
- *Vista (o interfaz de usuario):* tiene dos funciones. La primera consiste en obtener la imagen del programa definido por el usuario y enviarla al controlador. La segunda, consiste en mostrar la ejecución de las acciones del bot en el mundo.
- *Controlador:* contiene tres funcionalidades
  - Recibe la imagen de la vista para analizar su contenido y, en base a éste, genera el PT.
  - Una vez generado el PT, lo traduce al LI.
  - Finalmente, interpreta el LI para poder ejecutar sus instrucciones en la interfaz de usuario.

Considerando las etapas desarrollo descriptas en la sección 2, se comenzó con la etapa de digitalización del PT, que tiene como objetivo la implementación de una aplicación que permita leer las tarjetas (o instrucciones) que lo componen. Esto se logró utilizando una cámara convencional y herramientas de Realidad Aumentada implementadas con A-Frame y AR.js [1, 2]. A-Frame es un framework Web para el desarrollo de ambientes de Realidad Virtual y AR.js es una biblioteca JavaScript para Realidad Aumentada, que posibilita el seguimiento de marcadores.

Actualmente, se encuentra en desarrollo la segunda etapa, Procesamiento del Programa, donde se interpreta el contenido del PT para luego traducirlo a un programa en LI. Asimismo, se están evaluando las tecnologías que se utilizarán en la interfaz de usuario, para mostrar el mundo y los movimientos que el agente realice en base a la acciones obtenidas del LI.



Figura 1: Ejemplo de un Programa en LPT

Por último, se desea que el trabajo realizado pueda ser tomado como base para futuros trabajos de investigación, es decir que sea susceptible a ser extendido o modificado para implementar nuevas funcionalidades, sin la necesidad de realizar cambios complejos sobre la arquitectura general del sistema.

4. Formación de Recursos Humanos

Uno de los autores de este trabajo se encuentra desarrollando su tesis de grado de la Licenciatura en Ciencias de la Computación, en la temática de esta línea de investigación.

Por otra parte, otro de los autores de este trabajo cursa la Maestría en Enseñanza en Escenarios Digitales que desarrollan de manera conjunta las Universidades Nacionales de Cuyo, Comahue, Patagonia Austral, Patagonia San Juan Bosco, San Luis, Chilecito y La Pampa.

Referencias

[1] A-FRAME Homepage `https://https://aframe.io/`. Accedido por última vez Marzo 2020.

[2] AR-JS Homepage `https://github.com/AR-js-org/AR.js`. Accedido por última vez Marzo 2020.

[3] JSON Homepage `https://www.json.org/json-es.html`. Accedido por última vez Marzo 2020.

[4] A. N. Antle and A. Wise. Getting down to details: Using learning theory to inform tangibles research and design for children. *Interacting with Computers*, 25(1):1–20, 2013.

[5] C. P. de Educación de la Provincia de Neuquén. Diseño Curricular Jurisdiccional de los tres primeros años de la Escuela Secundaria Neuquina. Resolución N°1463/18, 2018.

[6] S. Furber. *Shut down or restart? The way forward for computing in UK schools*. The Royal Society Education Section, 2012.

[7] S. Papavlasopoulou, M. N. Giannakos, and L. Jaccheri. Reviewing the affordances of tangible programming languages: Implications for design and practice. In *2017 IEEE Global Engineering Education Conference (EDUCON)*, pages 1811–1816. IEEE, 2017.

[8] J. M. Ruzafa. *Una arquitectura para aplicaciones educativas basadas en mundos virtuales e interfaces tangibles*. PhD thesis, Universidad Autónoma de Madrid, 2018.

[9] T. Sapounidis, I. Stamelos, and S. Demetriadis. Tangible user interfaces for programming and education: A new field for innovation and entrepreneurship', innovation and entrepreneurship in education (advances in digital education and lifelong learning, volume 2), 2016.

[10] R. Society. After the reboot: Computing education in UK schools. *Policy Report*, 2017.

[11] J. Wing. Research notebook: Computational thinking—what and why. *The Link Magazine*, 6, 2011.

[12] J. M. Wing. Computational thinking. *Communications of the ACM*, 49(3):33–35, 2006.

Recursos Educativos Desenchufados para la Enseñanza de las Ciencias de la Computación en la Escuela Secundaria

Daniel Dolz[1] Rodolfo Martínez[2] Gerardo Parra[1] Jorge Rodríguez[1]

{ddolz, rodolfo.martinez,gparra,j.rodrig}@fi.uncoma.edu.ar

[1]*Grupo de Investigación en Lenguajes e Inteligencia Artificial*
Departamento de Teoría de la Computación - Facultad de Informática
[2]*Grupo de Investigación en Ingeniería de Software*
Departamento de Ingeniería de Sistemas- Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

Resumen

Las actuales tendencias curriculares para la incorporación de contenidos de Ciencias de la Computación en la educación obligatoria, proponen un recorrido amplio por las distintas áreas de conocimiento.

Los recursos educativos desenchufados (RED) demuestran ser una opción adecuada para establecer un primer contacto entre los estudiantes y las Ciencias de la Computación.

Si bien existe una amplia variedad de Recursos Desenchufados elaborados en el marco de CSUnplugged, estos no cubren completamente la selección de contenidos propuestos para la educación obligatoria.

La insuficiente evidencia sobre su efectividad en la escuela secundaria argentina y la ausencia de un modelo consolidado que permite diseñar este tipo de RED, plantea la necesidad de abordar estos temas.

En este trabajo se propone desarrollar una línea de investigación y desarrollo, en el marco de investigación acción participativa, para abordar el problema.

Palabras Clave: EDUCACIÓN EN CIENCIAS DE LA COMPUTACIÓN, ESCUELA SECUNDARIA, RECURSOS EDUCATIVOS DESENCHUFADOS, PARTICIPATORY DESIGN.

Contexto

Esta línea de investigación se desarrolla, por un lado, en el contexto de los temas de interés que promueven el Grupo de Investigación en Lenguajes e Inteligencia Artificial (GILIA) y el Grupo de Investigación en Ingeniería de Software (GIISCo), de la Facultad de Informática de la Universidad Nacional del Comahue (UNCo). En particular, se enmarca en el ámbito de dos proyectos de investigación de la Facultad de Informática, ambos financiados por la UNCo y con una duración de cuatro años a partir de enero de 2017: *Agentes Inteligentes. Modelos Formales y Aplicaciones para la Educación (04/F015)* y *Desarrollo de Software basado en Reuso (04/F009)*.

Por otro lado, el trabajo se desarrolla teniendo en cuenta el Convenio Marco de Colaboración firmado durante 2016 entre la Facultad de Informática y el Ministerio de Educación de la Provincia del Neuquén. Particularmente, se trabaja con el Consejo Provincial de Educación de la Provincia de Neuquén.

1. Introducción

Existe un amplio consenso, aún con diferente grado de avance en los distintos países, para in-

corporar contenidos de Ciencias de la Computación (CC) en las propuestas curriculares para la educación obligatoria. En este sentido, la enseñanza de la informática en la escuela secundaria está transitando una profunda transformación [6, 11, 12]. Las actuales tendencias curriculares para la incorporación de las CC en la educación obligatoria proponen un recorrido amplio por las áreas de conocimiento de las CC.

Los recursos educativos desenchufados (RED) demuestran ser una opción adecuada para establecer un primer contacto entre estudiantes y las CC. Esto se debe, fundamentalmente, a que no se requiere aprender programación ni hacer uso de un dispositivo digital y que, por lo general, el ambiente en el que se desarrollan tiene un enfoque de juego que plantea desafíos para el estudiante [1, 13].

Más allá de su uso cada vez más creciente, existe relativamente poca evaluación sistemática de la efectividad de estos recursos educativos. No existe este tipo de RED para cubrir todas las áreas de conocimiento de las CC que forman parte de las propuestas curriculares para la educación secundaria [3, 6]. Tampoco existe suficiente evidencia sobre su efectividad en la escuela secundaria argentina. Además, es necesario establecer un modelo consolidado que permite diseñar este tipo de RED.

La propuesta de este trabajo es la de iniciar una línea de investigación y desarrollo, en el marco de investigación acción participativa, para abordar el problema. Se intentará definir un modelo consolidado para RED y se tratará de evaluar de manera sistemática su efectividad en la escuela secundaria.

El resto del artículo está organizado de la siguiente manera. A continuación, se presenta la línea de investigación y desarrollo. En la sección 3, se comentan los resultados iniciales obtenidos y los que se espera obtener producto del desarrollo de esta investigación. Finalmente, en la sección 4, se comentan aspectos relacionados a la formación de recursos humanos en el marco de esta línea, así como del proyecto de investigación en su conjunto.

2. Línea de investigación y desarrollo

Los estudios a desarrollar en esta Línea de Investigación consisten en definir un enfoque didáctico disciplinar que permita orientar los procesos de diseño y evaluación de REDs, definir los mecanismos fundamentales para el diseño y elaboración de nuevos REDs en el marco de Participatory Design Framing [4, 5] y desarrollar estudios tendientes a evaluar la efectividad de este tipo de recursos en el ámbito de la enseñanza de las Ciencias de la Computación en la Escuela Secundaria.

Para el desarrollo de esta línea se propone trabajar en forma articulada sobre los siguientes aspectos:

- *Definición de un enfoque para RED*, la teoría del aprendizaje experiencial enfatiza sobre el rol que juegan las experiencias inmediatas y concretas en los procesos de aprendizaje. En este contexto (ver Figura 1), el aprendizaje es concebido como un ciclo de cuatro fases, la Experiencia Inmediata y Concreta es la base para la Observación y Reflexión. El estudiante utiliza estas reflexiones para la Formalización de Conceptos Abstractos basados en sus observaciones. Finalmente, el estudiante prueba las Implicaciones de los conceptos elaborados en nuevas situaciones [7, 8].

  CSUnplugged, es un enfoque para la enseñanza de las Ciencias de la Computación diseñado para presentar, a estudiantes sin formación previa, conceptos a través de experiencias prácticas que no requieren el uso de computadoras. En general las actividades, no necesitan de conocimientos previos sobre programación u otra área de las Ciencias de la Computación, son de carácter altamente cinestésico e involucran el sentido de juego o historia [1, 10, 13].

  En base al análisis de estas perspectivas metodológicas, en esta Línea de Investigación se propone definir un enfoque que resulte útil como guía para desarrollar nuevos REDs y para revisar los existentes. En



Figura 1: Teoría del aprendizaje experiencial [8]

este contexto se busca articular las perspectivas expuestas en la Teoría del Aprendizaje Experiencial, las definiciones elaboradas en el enfoque CSUnplugged y los aprendizajes elaborados en el marco de ésta Línea.

- *Diseño participativo de RED*, se estructura el proceso de diseño usando Participatory Design Framing. Un marco de trabajo innovador para Educación en Ciencias de la Computación, donde los docentes de escuelas secundarias se involucran activamente en el proceso de elaboración de recursos educativos [4, 5].

  En el proceso de diseño participativo, definido para esta Línea de Investigación y Desarrollo, el equipo de investigación diseña una primer versión para un RED a partir de explorar el área de conocimiento, determinar enlaces curriculares para la actividad y del conocimiento del que dispone acerca de la enseñanza de la computación en el ámbito de la escuela secundaria.

  En segunda instancia, se convoca a un grupo de profesores de informática de escuelas secundarias a una sesión piloto que busca recuperar sus percepciones acerca del recurso. Los docentes evalúan el recurso y luego informan sobre las características que valoran positivamente y acerca de las que requieren ajustes.

  Producidos los ajustes, los docentes disponen de REDs para enseñar Ciencias de la Computación en sus aulas. Un nuevo ciclo de adecuación del recurso se produce al revisar los resultados obtenidos en el trabajo de campo.

- *Estudiar la efectividad de los RED,* tanto la teoría de aprendizaje experiencial, como el enfoque CSUnplugged se utilizan frecuentemente como forma de exponer nuevas prácticas y conceptos. Sin embargo, existe poca investigación sistemática acerca de su efectividad, sobre todo en el ámbito de la educación secundaria [1].

  Desde la perspectiva metodológica, esta línea se ubica en el ámbito de la Investigación Acción Participativa [9, 2]. La participación comunitaria se expresa en este caso en la acción desplegada por los docentes de escuelas secundarias.

  Como primer aproximación, el equipo de investigación elabora una colección de conjeturas de carácter teórico práctico en relación a la posibilidades didácticas de los REDs.

  Estas elaboraciones son ajustadas a partir una serie de sesiones, de las que participa un grupo de docentes, que buscan situarlas a las particularidades de la escuela secundaria. En forma conjunta, investigadores y docentes, avanzan en un proceso de diseño experimental que tiene como objetivo estudiar la efectividad de este tipo de recurso educativo.

  Finalmente, se revisa el resultado del trabajo de campo desarrollado como forma de avanzar en la producción de conocimiento.

3. Resultados obtenidos y esperados

Esta línea de investigación dio el marco necesario para efectuar pruebas de campo que

permitieron evaluar, con docentes de educación secundaria, diversos recursos didácticos en un entorno desenchufado basados en la teoría del aprendizaje experiencial [7]. También se pudo evaluar si estos mecanismos eran accesibles al docente y con qué facilidad ellos consideraban que eran trasladables al aula.

Se toma como ejemplo una de las pruebas llevadas a cabo para abordar la enseñanza del tema redes de computadoras, y específicamente, se evaluó como introducir el tópico Medios Físicos de Transmisión de Datos.

Del experimento participaron 46 docentes de informática de educación secundaria, a los que se le dio una introducción al tema de transmisión de datos, y luego se les presentaron diversos materiales (entre ellos: hilo, linterna y silbato) que debían utilizar para representar los diversos medios físicos de transmisión de datos en un entorno de red. A partir de estos elementos, debían transmitir un mensaje por ”la red”, teniendo en cuenta que cada punto o nodo de la red (representado por un docente) debería transmitir el mensaje utilizando un medio físico de transporte de información distinto hacia otro nodo. La transmisión de datos se realizaba en forma codificada, en donde cada letra era representada en forma distinta dependiendo del instrumento que utilicen para enviar el dato. Por ejemplo, si utilizaban un hilo, éste estaba conectado entre dos nodos y la cantidad de veces que tiraban del hilo representaba el dato que estaban transmitiendo, o si utilizaban una linterna, la cantidad de veces que se encendía indicaba el dato que se estaba enviando. Esta actividad constituye la fase de Experiencia Concreta.

A continuación se concretó un momento de carácter reflexivo que buscó revisar de manera retrospectiva la experiencia vivida. El proceso buscó analizar, por ejemplo, cómo un mensaje se representaba para poder ser transmitido o cómo participaban diferentes medios físicos, hilo, luz y sonido del proceso que conduce un mensaje. Este momento constituye la fase de Observación y Reflexión.

En una próxima fase, los conceptos son presentados formalmente conectando los aprendizajes elaborados en instancias previas con definiciones formales y rigurosas acerca del tópico. Esta fase se ubica como momento de Formalización de Conceptos de acuerdo a la Teoría del Aprendizaje Experiencial.

Finalmente, se discute sobre las formas en que estos conceptos participan de actividades de la vida cotidiana como forma de aplicar el conocimiento recientemente construido a la comprensión de situaciones de la vida cotidiana. Esta etapa refiere a la transferencia de los aprendizajes elaborados a nuevos contextos, es decir conforma la etapa de Prueba de la Implicaciones de los conceptos elaborados. Posiblemente se inicie un nuevo ciclo experiencial.

Se comprobó que este recurso didáctico no sólo es transferible al aula en forma directa, sino que abre la puerta a que los docentes propongan otros instrumentos en este recurso didáctico, que posibilite a los alumnos expandir el alcance de los medios físicos de transmisión de datos, incorporando nuevos elementos para transmitir el mensaje. Así, se logra generalizar el concepto abstracto y permite crear nuevas hipótesis con la información recibida de la experiencia.

Además, se logra preparar al alumno para introducir, con los mismos materiales didácticos, los conceptos de Representación de la Información, Sincronización, Ruteos de Información y otros.

El diseño de actividades utilizando recursos desenchufados permite efectuar una nueva experiencia del aprendizaje. Los docentes evaluaron que este tipo de actividades es factible de ser utilizado en el aula en un contexto educativo.

Al finalizar la sesión con los docentes, se realizó una indagación con intención de recuperar las percepciones de los docentes acerca de este recurso educativo. Se les consultó sobre la utilidad, el impacto y la organización de la actividad, obteniendo apreciaciones positivas para las tres categorías.

Consideramos que la utilización de estos recursos en la enseñanza de las Ciencias de la Computación y su posterior evaluación, valorizará la labor del docente, ya que los resultados obtenidos en las pruebas de campo ha demostrado que el docente ha logrado extrapolar la experiencia vivida, hacia nuevas vivencias que le permitirán resolver futuras situaciones de en-

señanza.

4. Formación de Recursos Humanos

Se espera que el desarrollo de esta línea de Investigación contribuya a la formación de recursos humanos en el campo de las Ciencias de la Computación en la Educación. En este sentido, tres de los autores de este artículo cursan maestrías orientadas a conocer, comprender y analizar procesos relacionados con las tecnologías en la educación.

Además, otro de los autores de este trabajo está inscripto en la Maestría en Enseñanza en Escenarios Digitales que desarrollan de manera conjunta las Universidades Nacionales de Cuyo, Comahue, Patagonia Austral, Patagonia San Juan Bosco, San Luis, Chilecito y La Pampa.

Un integrante docente del proyecto de investigación realizará su tesis de Licenciatura en la temática del grupo. Además, los integrantes alumnos del proyecto también desarrollan sus tesis en temas relacionados. Por otra parte, actualmente se están desarrollando al menos cinco tesis de Licenciatura en Ciencias de la Computación en temas de interés del grupo de investigación.

Referencias

[1] T. Bell and J. Vahrenhold. Cs unplugged—how is it used, and does it work? In *Adventures Between Lower Bounds and Higher Altitudes*, pages 497–521. Springer, 2018.

[2] A. M. Colmenares E. Investigación-acción participativa: una metodología integradora del conocimiento y la acción. *Voces y Silencios. Revista Latinoamericana de Educación*, 3(1):102–115, 2012.

[3] C. P. de Educación de la Provincia de Neuquén. Diseño Curricular Jurisdiccional de los tres primeros años de la Escuela Secundaria Neuquina. Resolución N°1463/18, 2018.

[4] B. DiSalvo, J. Yip, E. Bonsignore, and D. Carl. Participatory design for learning. In *Participatory design for learning*, pages 3–6. Routledge, 2017.

[5] M. Guzdial and B. Naimipour. Task-specific programming languages for promoting computing integration: A precalculus example. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research*, pages 1–5, 2019.

[6] K-12 Computer Science Framework Steering Committee. *The K–12 Computer Science Framework*. ACM, 2016.

[7] A. Y. Kolb and D. A. Kolb. Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of management learning & education*, 4(2):193–212, 2005.

[8] D. A. Kolb. Learning styles and disciplinary differences. *The modern American college*, 1:232–255, 1981.

[9] J. Martí. La investigación-acción participativa: estructura y fases. 2017.

[10] T. Nishida, S. Kanemune, Y. Idosaka, M. Namiki, T. Bell, and Y. Kuno. A cs unplugged design pattern. *ACM SIGCSE Bulletin*, 41(1):231–235, 2009.

[11] F. Sadosky. *CC – 2016 Una propuesta para refundar la enseñanza de la computación en las escuelas Argentinas* . Fundación Sadosky, Argentina, 2013.

[12] R. Society. After the reboot: Computing education in UK schools. *Policy Report*, 2017.

[13] R. Taub, M. Armoni, and M. Ben-Ari. Cs unplugged and middle-school students' views, attitudes, and intentions regarding cs. *ACM Transactions on Computing Education (TOCE)*, 12(2):1–29, 2012.

Estrategias de enseñanza basadas en Aprendizaje Activo para favorecer la accesibilidad académica en la enseñanza de la programación

M. V. Rosas, H. Viano, M. E. Zuñiga
Facultad de Ciencias Físico, Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 (San Luis)
{mvrosas, hviano, mezuniga}@unsl.edu.ar

Resumen

En lo últimos años nos enfrentamos al desafío de repensar el sistema educativo a fin de que responda a las expectativas sociales actuales. Desde las Universidades se plantea como fundamental un desplazamiento desde la enseñanza tradicional al aprendizaje activo del estudiante. Este aprendizaje de habilidades y aptitudes específicas como aprender a aprender, resolver problemas, trabajar en equipos, pensar en forma crítica y comunicar claramente las ideas, supone una formación acorde de los docentes en las distintas disciplinas. En este contexto, los docentes de programación en materias de los primeros años de carreras vinculadas a la informática están condicionados a la búsqueda, desarrollo e implementación de nuevas estrategias de enseñanza pertinentes al nuevo modelo de aprendizaje.
Es así que se considera relevante generar espacios interdisciplinarios de investigación educativa en el ámbito académico donde se contemple la formación integral del docente, en la didáctica disciplinar y en capacidades interpersonales e intrapersonales para favorecer además la accesibilidad académica.
Finalmente, se espera que estas propuestas redunden en estudiantes más motivados para el estudio, con un mejor rendimiento académico y con un mayor dominio de aquellas habilidades que son esenciales para una carrera exitosa.

***Palabras clave*:** Enseñanza de la programación, Aprendizaje Activo, Formación docente, Accesibilidad Académica.

Contexto

La línea presentada surge como resultado de las acciones de investigación y de producción desarrolladas por sus integrantes en los proyectos "Realidades Alternativas como Lenguaje Generativo aplicado a la solución de Problemas Reales" (PROICO 03-0818) e "Innovación educativa y práctica reflexiva mediante Recursos Educativos Abiertos y herramientas informáticas libres" (PROICO 03-1616) pertenecientes a la Facultad de Ciencias Físicas y Matemáticas (FCFMyN) de la UNSL. Habiendo encontrado un vínculo en las temáticas que se abordan en otros proyectos de la misma facultad, respecto especialmente al proceso de enseñanza y aprendizaje en el ingreso universitario de carreras en ciencia e ingeniería se plantea la conformación de un nuevo proyecto I/D interdisciplinario de la FCFMyN denominado "Estrategias de enseñanza basadas en Aprendizaje Activo para Física, Matemática y Ciencias de la Computación". Dicho proyecto está integrado por tres líneas diferentes representando a cada una de las áreas temáticas nombradas, siendo la línea computacional la presentada en este trabajo.

1. Introducción

Actualmente, la educación está siendo atravesada por diferentes cambios y

reestructuraciones. Una ruptura destacada se observa sobre las estrategias convencionales de enseñanza a partir de las nuevas concepciones del aprendizaje, que consideran al sujeto como eje central en los procesos de enseñanza y aprendizaje. Esto provoca un cambio metodológico y establece un desafío para la enseñanza. Los docentes deben aprender a comunicarse con los nuevos lenguajes y acorde a los nuevos estilos de aprendizaje, lo que podría resultar en un desplazamiento del paradigma de enseñanza tradicional [1, 2]. En este punto, cabe destacar la importancia de determinar qué se entiende por enseñanza tradicional y por enseñanza para el aprendizaje activo planteado en esta nueva línea de investigación. La enseñanza tradicional en las ciencias experimentales, supone esencialmente que el alumno por repetición aprenderá cada uno de los conceptos de la disciplina y formará con ellos la estructura conceptual de la ciencia. La instrucción es generalmente deductiva, con el docente impartiendo conocimientos, mientras que el alumno debe recibirlos y asimilarlos, en una actitud esencialmente pasiva. Recuperando los aportes del campo de la pedagogía se puede definir esta forma de entender la enseñanza dentro de la "educación bancaria". En esa perspectiva el docente es el centro y sujeto de la educación, los educandos son meros receptores que reciben los conocimientos sin importan sus necesidades, saberes, experiencias o contextos. Estos por medio de la repetición y la memorización se apropian de conocimientos que les son ajenos [3]. Muchas veces se realizan prácticas de enseñanza con alto potencial de significación pero el aprendizaje no se produce. Es por esto que en este nuevo modelo se busca que el estudiante sea protagonista de su aprendizaje, en un proceso dinámico y activo. Él posee conocimientos en base a los cuales elaborar nuevos. Debe participar activamente en la construcción de su propio aprendizaje, partiendo de un estado inicial de conocimientos proveniente de sus experiencias de vida, el cual debe ser modificado por la instrucción formal para llegar al estado final de comprensión deseado para cada disciplina. A partir de lo dicho surge el término *Aprendizaje activo,* usado en la literatura en una variedad de contextos y usualmente no definido. La línea presentada y el proyecto del cual es parte se basan en la definición de Freeman *"Aprendizaje Activo es aquel que compromete al estudiante en el proceso de aprendizaje a través de actividades y/o discusiones en clase, como opuesto a la escucha pasiva a un experto. Se enfatiza el pensamiento de orden superior y usualmente involucra el trabajo en grupo"* [4].

En este marco de referencia nuestra posición es que, si se pretende que los estudiantes practiquen estrategias de aprendizaje activo que mejoren drásticamente su aprendizaje conceptual, un primer paso de este proceso es que la formación de sus docentes sea congruente con el modelo educativo propuesto. Para ello se necesita una visión superadora, que integre los conocimientos de contenido, tanto disciplinares como los pedagógicos, relacionados directamente con el nuevo conocimiento acerca de cómo se aprenden los distintos conceptos de la disciplina, lo que en la actualidad se denomina *Conocimiento Pedagógico del Contenido* [5].

Por otro lado, la sociedad también presenta un desafío a las Universidades en relación a la educación inclusiva, entendiéndose por la educación que hace posible el acceso a una educación de calidad para todos los estudiantes, asegurando la eliminación de las barreras (académicas, comunicacionales y físicas) y motivando a la participación para mejores aprendizajes. Nuestro rol docente nos interpela a construir un espacio donde las políticas educativas se concretan en prácticas pedagógicas diversificadas, teniendo en cuenta la Accesibilidad Universal como el grado en el que todas las personas pueden utilizar un objeto, visitar un lugar o acceder a un servicio, independientemente de sus capacidades cognitivas o físicas [7, 8, 9].

El trabajo interdisciplinario pretende favorecer el diseño de acciones de formación docente que contemplen la *didáctica*

disciplinar y el desarrollo de capacidades asociadas a lo *interpersonal* e *intrapersonal*. En relación a lo disciplinar, la programación es una disciplina de las Ciencias de la Computación, cuya dificultad principal reside en la resolución de problemas. Al poseer una naturaleza ligada a la resolución de diferentes problemas del mundo real, se requiere de una capacidad de abstracción que permita operar sin que los conocimientos técnicos limiten al programador a resolver adecuadamente dichos problemas [10]. Gary Stager sostiene que la resolución de problemas mediante la programación demanda de los estudiantes encontrar diversas maneras de abordarlos y de plantear soluciones. Además, el desarrollo de habilidades para visualizar rutas de razonamiento divergentes, anticipar errores y evaluar rápidamente los diferentes escenarios mentales [11]. En este sentido la línea propuesta va a ser ampliamente enriquecida con las experiencias en el área de Física y de Matemática que cuentan con una importante producción científica sobre la *Resolución de problemas* y además sigue siendo un tema central en la investigación sobre su enseñanza en esas ciencias [12, 13, 14].

Adicionalmente a las competencias relativas al conocimiento disciplinar, un docente universitario debería desarrollar capacidades asociadas a lo *interpersonal* e *intrapersonal*, como el manejo de grupos, buena organización y comunicación asertiva, empatía, disponibilidad, integridad y otras competencias sociales [15].

En síntesis, desde el proyecto de investigación se propone llevar adelante una visión integradora del aprendizaje estudiantil, trabajando desde tres de áreas de conocimiento [16].

2. Líneas de Investigación y Desarrollo

Los ejes que delimitan esta línea de investigación están enfocados en:

Eje 1: El Aprendizaje activo de la programación en el primer año de carreras universitarias vinculadas a las ciencias de la computación.

Eje 2: La Resolución de problemas, como tema transversal que atañe a las ciencias experimentales que corresponden al proyecto.

Eje 3: La Formación integral de docentes en servicio, en la didáctica disciplinar y en capacidades interpersonales e intrapersonales para favorecer la accesibilidad académica.

Eje 4: El uso de las tecnologías emergentes para generar pedagogías en el aprendizaje de la programación.

3. Resultados obtenidos/esperados

En concordancia con los ejes planteados se pretende:

- Estudiar y desarrollar estrategias didácticas de enseñanza activa, tendientes a mejorar el aprendizaje conceptual y significativo de la programación en cursos introductorios y avanzados universitarios.
- Estudiar y desarrollar estrategias de enseñanza que incorporen en los estudiantes habilidades profesionales básicas: de resolución de problemas, interpersonales y de comunicación oral y escrita.
- Estudiar y desarrollar estrategias didácticas para mejorar la formación de recursos humanos idóneos en las nuevas estrategias educativas de la programación, aplicadas a docentes en servicio.
- Fomentar las prácticas educativas mediadas por la tecnología para la resolución de problemas generales favoreciendo el aprendizaje significativo.
- Estudiar y desarrollar estrategias didácticas para mejorar la formación de recursos humanos en accesibilidad académica.
- Contribuir a la generación de conocimiento que aporten a la formación académica en relación a las temáticas planteadas.

Los resultados esperados deberían identificar las condiciones para obtener una mejor

docencia de grado y pregrado, que resulte en un mejor desempeño estudiantil.

4. Formación de Recursos Humanos

En la línea presentada participan docentes con formación de grado y posgrado en carreras relacionadas a la Informática y a la Educación Superior. Como la línea forma parte de un proyecto interdisciplinario se contará con la participación y colaboración de investigadores en el área de Matemática y Física.

Los trabajos a realizar harán posible la definición de una tesis de Especialización y una de Maestría en Educación Superior, como así también el desarrollo de un plan de trabajo final de Doctorado en Educación. Además, se prevé la formación de tutores y becarios en investigación educativa en ciencias de la computación con orientación a la accesibilidad académica.

5. Bibliografía

[1] Gisbert, M., "Digital Leaners: la competencia digital de los estudiantes universitarios." La Cuestión Universitaria, 48-59. 2011.

[2] Adell, J., "Los estudiantes universitarios en la era digital: la visión del profesor". La Cuestión Universitaria, pp 97-100, 2011

[3] Leilani A. Arthurs & Bailey Zo Kreager "An integrative review of in-class activities that enable active learning in college science classroom settings", International Journal of Science Education, 39:15, 2073-2091. 2017.

[4] Freeman, S., Eddy, S. L., Mcdonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P., "Active learning increases student performance in science, engineering, and mathematic s." Proceedings of the National Academy of Sciences, 111(23), 8410–8415. 2014.

[5] Etkina E., "Pedagogical content knowledge and preparation of high school physics teachers, Physical Review Special Topics" - Physics Education Research 6, 020110. 2010

[6] Gallardo, G., & Morales, Y., "Una universidad para el aprendizaje de todos. Orientaciones para el desarrollo de una docencia inclusiva en primeros años." Santiago, Chile: Pontificia Universidad Católica de Chile, Dirección de Asuntos Estudiantiles (DAE), Observatorio de Juventud Universitaria. Recuperado 03 agosto 2011 de: http://vidauniversitaria.uc.cl/images//libr o_obse rvatorio_2011.pdf.

[7] C.A.S.T. "Guía para el diseño universal del aprendizaje (DUA)" Versión 1.0. . Universal design for learning guidelines versión 1.0. Wakefield, MA: autor. 2008.

[8] Grzona, M. A. "La responsabilidad de la universidad para proporcionar estrategias generadoras de espacios inclusivos". Publicación digital de las VI Jornadas Discapacidad y Universidad: Los sujetos, los procesos y los contextos. Universidad Nacional de Cuyo, 2010. Programa de Inclusión de Personas con Discapacidad en la UNCUYO y Comisión. ISBN 978-987-575-099-9. 2010.

[9] Zuñiga M., Rosas M.V., Viano, H., "Disminución Visual y Educación Inclusiva a través de Prácticas Educativas Abiertas". Docentes Conectados, vol. 2, N°4. ISSN 2618-2912. pp 34-49. 2019.

[10] Gries D., "The Science of Programming". Monographs in computer Science Series. Springer Verlag, 1981.

[11] Stager, G., "En pro de los computadores", 13 enero 2004. [En línea]. Available: http://www.eduteka.org/ProComputadore s. php.

[12] Artigue, M.; Douady, R.; Moreno, L. Y Gómez, P. (Ed.). Ingeniería Didáctica en Educación Matemática. Bogotá/México: una empresa docente/Grupo Editorial Iberoamérica. 1995.

[13] Sirur Flores J. and Benegas J, "Deep and superficial conceptual learning: the effect of the teaching approach".

International Conference on Physics Education (ICPE). August 18-22, 2014, Córdoba, Argentina, Aceptado para exposición oral y publicación Conference on Physics Education ICPE 2014. www.icpe2014.org. 2014.

[14] Helminen, J., Ihantola, P., Karavirta, V., & Malmi, L. "How Do Students Solve Parsons Programming Problems?". An Analysis of Interaction Traces. Proceedings of the Eighth Annual International Computing, pp. 119-126. 2012.

[15] Kereki I. & Adorjan A. "Ayudantes de Cátedra en los primeros cursos de Programación: selección, formación, evaluación y reflexiones". 47JAIIO - Simposio Argentino de Enseñanza Superior en Informática (SAESI).ISSN: 2618-3269 - 22-33. 2018.

[16] Scott, C.L., "El futuro del aprendizaje ¿Qué tipo de aprendizaje se necesita en el siglo XXI?" Investigación y Prospectiva en Educación UNESCO, París. [Documentos de Trabajo ERF, No. 14] 2015.

TECNOLOGÍAS DEL APRENDIZAJE Y EL CONOCIMIENTO APLICADO A LA ENSEÑANZA DE INFRAESTRUCTURAS IT

Baldino Guillermo, Ferrara Damián, Añasco Iván, Heredia Luciano,
Baez Nahuel, Nahuel Leopoldo, Marchesini Javier

[1]*GIDAS, Grupo de I&D Aplicado a Sistemas informáticos y computacionales*
Facultad Regional La Plata - Universidad Tecnológica Nacional

{gbaldino,dferrara,ianasco,lheredia,nbaez,lnahuel, jmarchesini}@frlp.utn.edu.ar

- RESUMEN

En actuales ámbitos universitarios, observamos que la mayoría de los alumnos son nativos digitales, por lo que es difícil pensar su vida académica sin relacionarla con la diversidad de herramientas cloud masivas tanto para la comunicación como el trabajo colaborativo a través de entornos digitales. En este contexto, los docentes universitarios atraviesan y actúan en diversos escenarios de comunicación áulicos, como también en apertura de trabajo colaborativo digital remoto. Esto constituye un modo de Transformación en el Proceso de Enseñanza-Aprendizaje, apoyándose en las nuevas Tecnologías de la Información y la Comunicación (TICs) basadas en la nube, propiciando interacción y transmisión de conocimientos a partir de estrategias pedagógicas apoyadas por tecnologías informáticas. Cubriendo estos aspectos particulares surge el concepto Tecnologías del Aprendizaje y del Conocimiento (TACs) y Tecnologías para el Empoderamiento y la Participación (TEPs). Por lo expuesto, es prioritario reconocer el impacto y la necesidad del manejo e inclusión de las Cloud TACs (este concepto incluye a las TICs) en la gestión docente, permitiendo configurar los espacios digitales de trabajo tanto para el equipo docente, los alumnos, y sus interacciones.

Palabas Claves: Cloud Computing, TACs, Infraestructura IT, OpenStack, Nubes privadas.

- CONTEXTO

Actualmente, esta línea de investigación se encuadra dentro de un Proyecto de I&D (PID) con Incentivos del Ministerio de Educación Nacional y homologado por la Secretaría de Ciencia Tecnología y Posgrado (SCTyP) del Rectorado de la Universidad Tecnológica Nacional (código de homologación: UTI4447), bajo la denominación "Innovación Informática en Tecnologías del Aprendizaje y Conocimiento aplicado al Mejoramiento de Procesos Educativos". Este PID fue homologado por un período de cuatro años (inicio: enero/2017 – finalización: diciembre/2020) e incorporado al Programa de I&D+i "Electrónica, Informática y Comunicaciones" de la SCTyP [1]. Los avances y resultados generados en este PID son financiado por becas de alumnos de la SCTyP, becas de alumnos para Investigación de la Secretaría de Asuntos Estudiantiles de la Facultad Regional La Plata, becas de estímulo a las vocaciones científicas (EVC) del consejo interuniversitario nacional (CIN) del Ministerio de Educación Nacional y becas para graduados de iniciación a la investigación y desarrollo (conocidas como BINID en el programa de becas de la SCTyP de la UTN).

Las actividades de investigación, diseño de prototipos y desarrollo experimental se llevan a cabo en las instalaciones del Grupo de I&D UTN denominado "GIDAS - Grupo de I&D Aplicado a Sistemas informáticos y computacionales". El GIDAS es una unidad científica-tecnológica homologada en el sistema científico de la UTN con financiamiento tanto para acciones e iniciativas del grupo de investigación como para los PIDs que se desarrollan en él [2]. Este proyecto se encuentra alineado al área de aplicación Tecnologías para Educación y Aula Siglo XXI del GIDAS.

1. INTRODUCCIÓN

El uso de sistemas y herramientas tecnológicas en estos días, apunta hacia el mundo del Cloud Computing, un paradigma que está en constante crecimiento. Durante estos últimos años, cada vez más compañías y grupos de investigación trabajan en conjunto con el fin de explotar las oportunidades ofrecidas por el mismo [3], en el cual los recursos (procesamiento, memoria, almacenamiento, DB's, etc.) se consumen como servicios: SaaS, PaaS, o IaaS (Infrastructure as a Service).

Este trabajo propone implementar una plataforma IaaS basada en OpenStack, que nos permita ofrecer recursos de Cloud Computing a docentes y alumnos de la comunidad académica UTN-FRLP. Esto permite promover el uso de software libre y de código abierto, dado que servicios como Amazon Web Service, Microsoft Azure y Google Cloud, son de difícil acceso por los costos elevados de sus licencias.

El alumno que necesite recursos de infraestructura informática en la nube para sus actividades curriculares y académicas, los podrá obtener mediante la plataforma que comenzamos a montar. De esta forma, acompañamos y apoyamos los procesos de enseñanza y aprendizaje sobre la gestión de Infraestructuras IT [4], ya que los docentes deberán orientar a los alumnos a gestionar y configurar las solicitudes de IaaS.

Los alumnos habitualmente necesitan recursos de Infraestructura IT para realizar las prácticas de desarrollo de software de las cátedras o talleres académicos. Para dar soporte a esta necesidad se ofrece máquinas virtuales con diferentes sistemas operativos, herramientas específicas, disponer de una red (VPN), entre otras.

El beneficio que obtiene la comunidad académica, es poder agilizar y facilitar el proceso de enseñanza-aprendizaje ya que los alumnos son partícipes activos de la nube privada. Además, mostrar concretamente la teoría de manera práctica, de modo que el alumno pueda acercarse a la experiencia a través de un contacto real y práctico.

Estos recursos podrían obtenerse sencilla, rápida y fácilmente mediante Cloud Computing a través de servicios de IaaS. Las máquinas virtuales, cómputo, almacenamiento, entre otros, podrán ser pedidos y gestionados por el mismo alumno bajo demanda desde la propia nube privada. La gestión y manejo de estas tecnologías permiten al alumno empoderarse de conocimientos técnicos-tecnológicos tan necesarios como el análisis, diseño y desarrollo de sistemas software [5].

Buscamos desplegar y prestar servicios a través de una nube privada en el marco de iniciativas promovidas por el área de I&D Tecnologías para la Educación y Aula Siglo XXI del Grupo de I&D GIDAS. Este trabajo permite acercar el uso de la herramienta OpenStack, no sólo brindando la posibilidad de utilizar instancias virtuales al resto de la comunidad académica, sino también promoviendo su uso y buscando generar en alumnos el conocimiento sobre las ventajas del uso de este tipo de tecnologías. Por otro lado, consideramos que acercamos una herramienta para centralizar el despliegue de proyectos de software y trabajos en distintas cátedras, con necesidad de correrlos sobre distintas

plataformas. Simplificamos esta tarea dando a todos los alumnos la posibilidad de desplegar sus propios servidores virtuales como parte de una nube privada dentro de la Facultad.

Uso de OpenStack en la implementación Cloud TACs

OpenStack es una plataforma de software libre, escalable y diseñada para ofrecer nubes públicas o privadas, permitiendo IaaS. Posee diferentes componentes con funciones específicas y se pueden instalar de manera separada o conjunta según la distribución. Esta integración es a través de APIs, que cada servicio ofrece y consume. Gracias a estas APIs, los servicios pueden comunicarse entre ellos y posibilita que un servicio sea reemplazado por otro de similares características siempre que se respete la forma de comunicación. Es decir, OpenStack es extensible y se ajusta a las necesidades de quien desee implementarlo [6]. Los principales componentes son [7]:

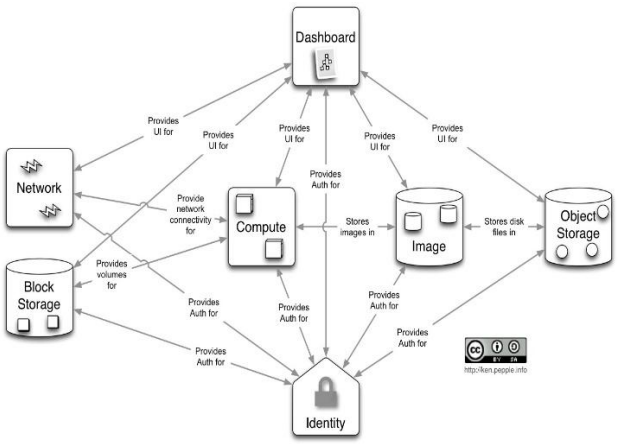


Fig. 1. Imagen representativa de la intercomunicación de módulos en OpenStack [7].

Compute (Nova): núcleo de OpenStack, diseñado para gestionar y automatizar los pools de los recursos del equipo, pudiendo trabajar con tecnologías de virtualización.

Object Storage (Swift): módulo encargado del almacenamiento redundante, escalables y tolerante a fallos de objetos y archivos.

Networking (Neutron): encargado de la gestión de redes.

Block Storage (Cinder): proporciona dispositivos de almacenamiento a nivel de bloque persistentes, permitiendo la búsqueda y recuperación de máquinas virtuales.

Identity Service (Keystone): servicio que ofrece autenticación de usuarios y políticas de seguridad.

Image Service (Glance): provee servicio de creación, búsqueda y recuperación de máquinas virtuales. Gestiona todas las imágenes de los sistemas operativos.

Dashboard (Horizon): proporciona a los administradores y usuarios una interfaz gráfica para el acceso, la provisión y automatización de los recursos en la nube

2. LINEAS DE INVESTIGACION y DESARROLLO

En la tarea de investigación que estamos llevando adelante, será necesario, un avance en el estudio y análisis de estándares y herramientas que nos permitan la integración de las TACs con herramientas de Cloud Computing para la mejora del proceso de enseñanza - aprendizaje de Infraestructuras IT. Bajo esta línea de trabajo, serán analizados los siguientes estándares y herramientas:

- Topología de las diversas Infraestructuras IT: estudio y comparación de casos.

- Herramientas que provee una solución de IaaS para la elaboración de nubes

públicas o privadas: evaluación y análisis comparativos.

- Caracterización de nube privadas para ámbitos de aplicación académicos.

- Evaluación de prestaciones de servicios como Amazon Web Service, Microsoft Azure, Google Cloud y su comparación y complementación con OpenStack.

- Enfoques metodológicos aplicables a Tecnologías del Aprendizaje y el Conocimiento (TAC).

- Estimación de recursos de hardware para el despliegue y funcionamiento sostenido de una nube privada en ámbitos académicos.

3. RESULTADOS OBTENIDOS / ESPERADOS

A partir del diseño de estas tecnologías Cloud en al ámbito de la Facultad esperamos lograr una efectiva utilización del uso de las TICs, poniendo no solo el conocimiento sino también, las herramientas desde la perspectiva de la usabilidad, despliegue y mantenimiento. Durante el cursado de la carrera Ingeniería en Sistemas de Información los alumnos aprenden entre otras cosas, a diseñar y desarrollar soluciones informáticas, sin profundizar en conocimientos teóricos y prácticos de la infraestructura tecnológica requerida. Es por ello que este trabajo captura esa gran oportunidad de enseñar haciendo uso de tecnologías y sus configuraciones, completando el ciclo de vida del software, incluyendo actividades de despliegue y adecuaciones para el futuro mantenimiento.

4. FORMACION DE RECURSOS HUMANOS

Este Proyecto de I&D se origina a partir de la necesidad de formar investigadores y áreas temáticas vinculadas la gestión, despliegue y manipulación de tecnologías de hardware, en el ámbito de su aprovechamiento en las diferentes cátedras de la Carrera. Como parte de los objetivos del proyecto buscamos formar docentes y alumnos en temáticas actuales de gestión de hardware, ya sea *on premise* o como tecnologías Cloud. En este aspecto hemos elegido como punto de inicio el uso de Openstack dado qué consideramos qué su versatilidad y uso nos permiten desarrollar las activadas de forma tal de lograr una adecuada transferencia de conocimientos tanto dentro del grupo de I+D como luego a las distintas Cátedras.

En las líneas de trabajo futuras el grupo de trabajo de este proyecto ha definido como una sus metas lograr el desarrollo de una Nube Privada utilizando recursos tecnológicos propios. Siguiendo con esta meta, el grupo de trabajo fomenta la formación de recursos humanos capaces de administrar la seguridad y el despliegue de todos los recursos de la Nube Privada.

5. BIBLIOGRAFÍA

[1] Portal Oficial de la SCTyP del Rectorado UTN con los PIDs homologados vigentes (código UTN5363): http://sicyt.scyt.rec.utn.edu.ar/scyt/proyectos/RESUMENF.ASP?VAR1=4447

[2] Portal Oficial de la SCTyP del Rectorado UTN con los Grupo de I&D homologados vigentes (Resolución CSU Nº 1770/18): https://utn.edu.ar/es/secretaria-sctyp/centros-y-grupos-utn/grupos-utn#LP

[3] Murazzo, M.: Análisis de una infraestructura Cloud Computing open source. In: http://sedici.unlp.edu.ar/handle/10915/53514 (2015).
[4] Bustichi, G., Mosconi, E.: Metodologías aplicadas para el desarrollo del aprendizaje autónomo. ISBN: 978-987-34-1796-6. IIº Jornadas sobre las Prácticas Docentes en la Universidad Pública (La Plata, 2018).
[5] López, M.: De las TICs a las TACs: la importancia de crear contenidos educativos digitales. Revista Didáctica, Innovación y Multimedia (DIM). In: http://www.pangea.org/dim/revista.htm
[6] Galarza, B., Zaccardi, G., Belizán, M., Duarte, D., Morales, M., Encinas, D.:Performance de cloud computing para HPC: despliegue y seguridad. ISBN: 978-987-42-5143-5. In: http://sedici.unlp.edu.ar/handle/10915/62576 (2017)
[7] Servicios y Componentes de Openstack. Último acceso 2020/11/06. htps://www.openstack.org/software/project-navigator/openstack-components.

Tecnologías Emergentes para la Educación

Claudia Russo[1], Mónica Sarobe[1], Paula Lencina[1], Tamara Ahmad[1], Mariana Ado[1], Piergallini Rosana[1], Pompei Sabrina[1]

Instituto de Investigación y Transferencia de Tecnología (ITT) - UNNOBA- CIC

Junín - Buenos Aires - Argentina

Sarmiento Nro. 1119 3er Piso, Junín (B) – TE: (0236) 4477050 INT 11610

{claudia.russo, monica.sarobe, paula.lencina, tamara.ahmad, mariana.ado, rosana.piergallini, sabrina.pompei}@itt.unnoba.edu.ar

Resumen

El desarrollo del hombre está impregnado por la presencia de innovaciones que impactan directa e indirectamente en su forma de vida, provocando la adaptación del mismo a los cambios generados por la tecnología, siempre en pos de su bienestar.

Las Tecnologías Emergentes han crecido y continúan creciendo de manera acelerada permitiendo desarrollo y avances en campos como la comunicación, la medicina, agricultura, el comercio, la educación, entre tantos otros.

En el caso particular de la educación las tecnologías emergentes han generado un nuevo escenario para el desarrollo del proceso de enseñanza-aprendizaje proporcionando diversas posibilidades y desafíos a todos los niveles educativos. El espacio de aprendizaje trasciende el aula, ahora está también en la web siendo accesible mediante múltiples herramientas disponibles tanto en dispositivos móviles como computadores, provocando la necesidad de nuevas formas de promover la educación. También las tecnologías emergentes permiten generar indicadores dinámicos proporcionando nuevas métricas tanto para el análisis del proceso de aprendizaje como para la gestión educativa.

El presente artículo brinda una síntesis del trabajo que se está realizando en el ITT en relación al desarrollo e impacto de las Tecnologías Emergentes al servicio de la Educación; específicamente robótica educativa, analítica de aprendizaje y asistentes virtuales.

Palabras Claves: Tecnologías Emergentes, Educación

Contexto

Las líneas de investigación a describir se enmarcan en el proyecto de investigación "Informática y Tecnologías Emergentes", con lugar de trabajo en el Instituto de Tecnología y Transferencia (ITT), presentado en la convocatoria de Subsidios a la Investigación Bianuales (SIB) 2019, aprobado y financiado por la Secretaría de Investigación de la UNNOBA.

[1] Investigador ITT-UNNOBA-CIC

El objetivo del mismo es estudiar cómo la informática impacta en el desarrollo de tecnologías emergentes de manera de analizar, definir y desarrollar herramientas y estrategias innovadoras que incidan de manera responsable en el desarrollo de la sociedad.

1. Introducción

La estrecha relación que naturalmente existe entre la educación y la creación y comunicación de conocimientos, explica las grandes oportunidades que las TIC ofrecen al sector educativo. En los distintos niveles de la educación formal vemos como poco a poco, y a diversos ritmos, se van incorporando el uso de las TIC tanto en modalidad presencial como a distancia. [1]

A fin de definir aquellas tecnologías todavía poco difundidas y utilizadas, cuyo impacto en distintos ámbitos es incipiente, pero que generan grandes expectativas, surge el término de "tecnologías emergentes". Gregory Day y Paul Schoemaker en el libro Gerencia de tecnologías emergentes, las define como "innovaciones científicas que pueden crear una nueva industria o transformar una existente. Incluyen tecnologías discontinuas derivadas de innovaciones radicales, así como tecnologías más evolucionadas formadas a raíz de la convergencia de ramas de investigación antes separadas".

Para el contexto de la educación, George Veletsianos propone la siguiente definición: "Las tecnologías emergentes son herramientas, conceptos, innovaciones y avances utilizados en diversos contextos educativos al servicio de diversos propósitos relacionados con la educación." Además, plantea que las tecnologías emergentes son organismos en evolución que experimentan ciclos de sobreexpectación y, al tiempo que son potencialmente disruptivas, todavía no han sido completamente comprendidas ni tampoco suficientemente investigadas. Por tanto, las tecnologías emergentes en educación pueden ser nuevos desarrollos de tecnologías ya conocidas o aplicaciones a la educación de tecnologías bien asentadas en otros campos de la actividad humana. [4][5]

Existen numerosas tecnologías emergentes aplicadas a la educación, a continuación describiremos brevemente aquellas que abordamos en el marco de nuestro proyecto de investigación como lo son la robótica, realidad aumentada, asistentes virtuales, analítica aplicada a la educación.

La robótica es una ciencia que genera curiosidad tanto a grandes como a chicos. Esta es una disciplina compleja dado que incluye para su desarrollo a la electrónica, la informática, la inteligencia artificial, la mecánica, el control, la locomoción, la visión por computador y por supuesto la calidad. Actualmente los robots son utilizados en los distintos niveles de la educación, desde el nivel inicial hasta la educación superior. La robótica educativa es un método interdisciplinario, su carácter transversal posibilita el desarrollo de pensamiento lógico, la creatividad, la lingüística, el trabajo en equipo, entre otras.

De acuerdo al Horizon Report: 2019 de educación superior, los asistentes virtuales se posicionan entre los principales recursos con mayor potencial de impacto a largo plazo en el sector de educación superior. Un asistente virtual es básicamente un programa informático capaz de reconocer el lenguaje del ser humano e interactuar con este a través de audio o texto. El informe resalta la capacidad de estos para interactuar con los estudiantes en conversaciones cada vez más naturales.

Además, pueden ser usados para investigar, tutorizar, escribir y editar trabajos; así como para brindar a la comunidad universitaria información sobre servicios de apoyo las 24 horas, tarea ejecutada, por ejemplo, por chatbots. Se prevee que pronto tutores y facilitadores virtuales puedan generar experiencias formativas personalizadas en diversas plataformas de aprendizaje adaptativo. [7]

Por último, los procesos de minería de datos permiten a las organizaciones descubrir conocimiento para la toma de sus decisiones. Esta tecnología es aplicada en el ámbito educativo con diversos propósitos. De acuerdo a un estudio de la Universidad de Morón, en primer lugar se resuelven problemas de rendimiento académico con el objetivo de estudiar el desempeño de un estudiante universitario, ya sea durante en el transcurso o final de una carrera. En un segundo lugar, se aplica con el propósito de clasificar a los estudiantes y poder realizar un análisis sobre los distintos grupos por sus similitudes y diferencias entre ellos. En un tercer lugar, se utiliza para analizar políticas institucionales universitarias, a partir del conocimiento obtenido de la institución como herramienta para la toma de decisiones estratégicas. [8]

La Minería de Datos Educativos da lugar a la Analítica de Aprendizaje que supone un paso más allá de la recolección, ya que sirve para analizar e interpretar esos datos, mediante técnicas pedagógicas y algoritmos propios de la Minería de Datos, con la finalidad de obtener información para mejorar la práctica educativa, optimizando el rendimiento de alumnos y profesores, así como el modelo educativo. La finalidad última de la Analítica de Aprendizaje es personalizar el aprendizaje de los estudiantes. Esta permite, entre otras cosas: guiar a los estudiantes a través de itinerarios personalizados; predecir el abandono y desempeño de los participantes en un curso o asignatura; adaptar las clases a los conocimientos reales de los alumnos; identificar necesidades de aprendizaje; conocer el comportamiento de los estudiantes en relación a la plataforma de trabajo; evaluar cursos o asignaturas en línea para adaptarlos según las posibilidades de mejora detectadas.

2. Líneas de Investigación y Desarrollo

Este trabajo se enmarca en un proyecto mayor que tiene como finalidad investigar cómo la informática impacta en el desarrollo de tecnologías emergentes de manera de analizar, definir y desarrollar herramientas y estrategias innovadoras que impacten de manera responsable en el desarrollo de la sociedad.

El presente equipo se centra en las tecnologías emergentes aplicadas a la educación, entre estas estudia:

- Robótica Educativa; diseño de una interfaz para un Robot Educativo Programable (REP).
- Minería de Datos; aplicada a estrategias para minimizar la deserción universitaria y detección de causas de rezago.
- Asistentes virtuales; uso de avatares como tutores virtuales.

3. Resultados Obtenidos/Esperados

Como resultado del trabajo realizado surgieron diversos proyectos de transferencia y tesis de posgrado.

Fue aprobado el FoNIETP “Proyectos Institucionales en Red” (PIR 2019), mediante

el cual se espera potenciar, el vínculo entre la actualización en Saberes Digitales y la retención estudiantil mediante una capacitación académica dirigida a docentes de los Ciclos Básico y Superior en la Formación Profesional modalidad Informática de la EEST N° 2. La finalidad es que los estudiantes del nivel medio lleguen a los últimos años de la modalidad en Informática Personal y Profesional, promoviendo hacia el nivel superior. Para ello se trabajac on dispositivos, aspectos facilitadores y obstaculizadores vinculados a la alfabetización digital y a un hacer digital crítico vinculado al pensamiento computacional, la robótica pedagógica y el diseño y fabricación digital.

Por otra parte fue defendida y aprobada la Tesis Doctoral "Minería de datos aplicada a estrategias para minimizar la deserción universitaria en carreras de Informática de la UNNOBA". En la misma se propone, a partir de la utilización de la minería de datos, un modelo de intervención áulica virtual que permite diseñar, e implementar, estrategias tecnológicas tendientes a realizar un seguimiento de las trayectorias educativas de aquellos estudiantes con rezago académico o riesgos de deserción.. Se espera que este modelo contribuya a minimizar la deserción en las carreras de Informática de la Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA). Como líneas futuras se proponen estudios comparativos entre diversas cohortes, desde la etapa fundacional de la carrera hasta la actualidad.

También se espera durante el corriente año la defensa de la tesis de maestría "Aprendiendo a programar con el Robot Educativo Programable (REP). Estudio de caso: Introducción a la Programación Imperativa, asignatura perteneciente a las carreras de informática de la UNNOBA". Ante la dificultad vista en la introducción de los estudiantes a la programación imperativa se propuso el diseño de una interfaz para el Robot Educativo Programable (REP) a fin de integrarlo y utilizarlo como herramienta educativa. El objetivo del trabajo se centró en establecer estrategias de enseñanza aprendizaje con respecto al paradigma imperativo de la programación de computadoras utilizando como herramienta al REP, en el primer año de estudios de las carreras del área de informática de la UNNOBA, durante el periodo 2018-2019. Se pudo observar que el uso de robots es de gran potencial para facilitar la aplicación más tangible y concreta de las habilidades de programación. A futuro se espera avanzar en el diseño de la interfaz incorporando nuevas funcionalidades entre las cuales se aspira poder desarrollar una versión para teléfonos smartphones; así como la posibilidad de conectar la interfaz con más de un REP, continuar la investigación de los sensores de REP, a fin de por ejemplo trabajar con sonidos para personas ciegas y luces para personas sordas.

Por último se espera presentar el trabajo de tesis de maestría "Avatares como tutores virtuales. Estudio de caso: curso Análisis y Diseño de Sistemas, asignatura de las carreras de Informática de la UNNOBA". El objetivo del trabajo fue definir estrategias para la utilización de avatares en el desarrollo de actividades colaborativas dentro de un entorno virtual 3D. En lo que respecta al aporte, se espera poder definir estrategias que permitan incorporar avatares pedagógicos en el desarrollo de actividades colaborativas para la asignatura ADS de las carreras informáticas de UNNOBA. Como futuras líneas se espera poder realizar experiencias con la utilización de los entornos, previos ajustes técnicos, en otras asignaturas y/o cursos de la universidad,

con el objetivo de analizar otras estrategias desde el punto de vista académico y/o pedagógico.

4. Formación de Recursos Humanos

El equipo de trabajo está compuesto por investigadores formados y en proceso de formación, becarios del sistema de Ciencia y Tecnología, Becarios alumnos, Graduados e investigadores externos.

En el transcurso del primer año se defendió una tesis doctoral, y se presentaron dos tesis de maestría las cuales están a la espera de fecha para su defensa.

Se está trabajando en una tercertesis de maestría cuya propuesta ya fue aprobada, y en el desarrollo de dos tesinas de grado correspondiente a la Licenciatura en Sistemas, tres Prácticas Profesionales de la Ingeniería Informática y un Trabajo Final Integrador de Especialización

5. Bibliografía y Referencias

[1] Díaz Patricia. "El cambio posible en la educación" en Aprender y educar con las Tecnologías del siglo XXI. Colombia: Colombia Digital. 2012. Disponible: http://www.aprendevirtual.org/centro-documentacion-pdf/Aprender_y_educar_con_las_tecnologias_del_Siglo_XXI.pdf

[2]Medina Velandia Lucy. "Tecnologías Emergentes al Servicio de la Educación" en Aprender y educar con las Tecnologías del siglo XXI. Colombia; Colombia Digital. 2012. Disponible: http://www.aprendevirtual.org/centro-documentacion-pdf/Aprender_y_educar_con_las_tecnologias_del_Siglo_XXI.pdf

[3] G. Day y P. Schoemaker. ·Gerencia de Tecnologías Emergentes. 1ra edición. Buenos Aires: Javier Vergara, 2001.

[4] G. Valetsianos. Emergence and innovation in digital learning: foundations and applications. Canada: AU Press, Athabasca University, 2016.

[5] J. Adel y L. Castañeda. Tecnologías emergentes, ¿pedagogías emergentes?. 1ra edición. Barcelona: Asociación Espiral, Educación y Tecnología, 2012.

[6] J. Cabero, J. Barroso y C. Llorente. "La realidad aumentada en la enseñanza universitaria". Revista de Docencia Universitaria. vol. 17(1), pp 105-118, junio 2019.

[7] B. Alexander, K. Ashford-Rowe, N. Barajas-Murphy, G. Dobbin, y otros. EDUCAUSE Horizon Report: 2019 Higher Education Edition. Disponible: https://library.educause.edu/-/media/files/library/2019/4/2019horizonreport.pdf?la=en&hash=C8E8D444AF372E705FA1BF9D4FF0DD4CC6F0FDD1

[8] M. Panizzi. "Establecimiento del estado del arte sobre la Minería de Datos Educacional en el Nivel Superior: Un Estudio de Mapeo Sistemático". RICUM. Vol. 4 (2), pp 51-60, julio 2019.

Innovación en Sistemas de Software

Impacto de las TICs en la sustentabilidad del uso del agua potable en la República Argentina

Leandro Crespi, Cristian Schiffino, Iris Sattolo, Marisa Panizzi
Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales. Universidad de Morón.
Cabildo 134. Morón, Argentina.
leandro.crespi@e-shell.com.ar, cristian.schiffino@gmail.com, iris.sattolo@gmail.com, marisapanizzi@outlook.com

Resumen

La línea de investigación se enmarca en las líneas propuestas por la UM, Informática Aplicada y Nuevas Tecnologías, la cual atiende una temática prioritaria a nivel nacional. Esta consiste en la aplicación de Tecnologías de la Comunicación e Información (TICs) a la sustentabilidad del agua potable. En base a la problemática de la escasez del recurso hídrico para las generaciones actuales y futuras; tanto a nivel mundial como nacional se propone la captación de agua de lluvia como fuente alterna de abastecimiento para satisfacer actividades socioeconómicas. Tiene como objetivo contribuir a dicha problemática mediante el diseño de un prototipo que permita la captación de agua de lluvia para uso de descarga de sanitarios, así como también, la evaluación de su viabilidad mediante estudios de casos.

Palabras clave: TICs, sustentabilidad, uso del agua potable.

Contexto

La línea de investigación que se reporta en este artículo es financiada por el Proyecto de Investigación Orientada (PIO) titulado: *"Impacto de las TICS en la sustentabilidad del uso del agua potable en la República Argentina"* de la Secretaria de Ciencia y Tecnología de la Universidad Morón.

Este PIO tiene por objetivo contribuir a la sustentabilidad del uso del agua potable en nuestro país mediante la aplicación de las TICs que permitan la captación del agua de lluvia y su uso para la descarga de sanitarios.

El grupo de investigación se propone como desafío esta línea de investigación prioritaria la cual atiende a una problemática tanto a nivel nacional como a nivel mundial.

Introducción

"El acceso a agua apta para el consumo es una necesidad humana fundamental y, en consecuencia, un derecho humano básico". Kofi Annan[1]. Se estima que actualmente se consume al año el 54 % del agua dulce disponible y, según la UNESCO[2], a mediados del siglo XXI la población mundial alcanzará los 12.000 millones de habitantes previstos, la demanda se habrá duplicado y las reservas hídricas de nuestro planeta llegarán a su tope (Ambientum, 2019).

El agua es esencial para la vida. Todas las personas, animales y plantas necesitan agua para vivir y crecer. En el caso particular del hombre, el agua es primordial para el desarrollo de muchas actividades productivas. Sin embargo, en numerosos lugares del mundo, la población no cuenta con el agua necesaria para mantener un nivel de vida aceptable. Es común encontrar que sectores importantes de la población mundial, deben recorrer grandes distancias para recolectar el agua disponible, la cual no siempre es potable. Cuando la población no cuenta con el agua necesaria para la vida diaria, se enfrenta a muchas dificultades y está en peligro de contraer enfermedades graves. Cuando una comunidad tiene acceso al agua potable de forma fácil y segura, la salud de las personas es notablemente mejor (Herrara Monroy L., 2010).

El agua no se distribuye equitativamente, es decir, en algunos sitios abunda y en otros es escasa. Cerca de mil millones de personas en el mundo carecen de este recurso y deben recorrer grandes distancias diariamente para

[1] Economista ghanés, séptimo secretario general de las Naciones Unidas, cargo que ocupó entre 1997 y 2006, galardonado, junto a la ONU, con el Premio Nobel de la Paz de 2001.
[2] Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

obtenerla. En América Latina, a pesar de que en las últimas décadas ha habido una mejora considerable, aún el 10 % de la población (aproximadamente 60 millones de personas) no tiene acceso al agua en su vivienda (Adler, I., 2008).

Para vivir adecuadamente, una persona necesita contar con al menos 20 litros diarios de agua, lo que implica un volumen de 7.300 litros por persona al año. Desafortunadamente, muchas personas no tienen acceso a esta cantidad, no sólo por las condiciones ambientales sino por la falta de infraestructura o la calidad del agua, que no la hace apta para el consumo humano La relación entre los componentes de los ecosistemas y el ciclo del agua, las altas tasas de deforestación, así como la contaminación de mares, lagos y ríos por la falta de saneamiento de aguas negras e industriales, hacen que ciertas regiones sean todavía más vulnerables a la escasez y baja disponibilidad de agua de calidad (Organización Mundial de la Salud, 2018).

Los Objetivos de Desarrollo Sostenible también conocidos como ODS, están diseñados para mejorar la calidad de vida a nivel mundial; se crearon en la Conferencia de las Naciones Unidas. Los ODS son un conjunto de 17 objetivos que permiten superar desafíos globales como la pobreza, la desigualdad, el cambio climático entre otros, además buscan armonizar tres componentes centrales para el futuro: crecimiento económico, inclusión social y protección ambiental, bajo esta perspectiva la captación del agua de lluvia para diferentes usos se alinea con los ODS (Organización de las Naciones Unidas, 2018).

La seguridad hídrica es esencial para alcanzar muchos de los Objetivos de Desarrollo Sostenible de la ONU, no solo el ODS número 6 sobre el acceso al agua y saneamiento, pero la amenaza del cambio climático surge como un obstáculo (Novo C., 2019).

Dado que es una nueva línea de investigación propuesta en la UM, presenta un gran desafío para el grupo de investigación. Se realizó una investigación exploratoria documental que ha permitido evidenciar algunos hallazgos de la literatura existente respecto a los sistemas de captación de agua de lluvia y sus potenciales sus usos.

La captación de agua de lluvia es la recolección, transporte y almacenamiento del agua de lluvia que cae sobre una superficie de manera natural o hecha por el hombre. Las superficies que captan el agua en las ciudades pueden ser techos de casas y edificios, techumbres de almacenes y de tiendas, explanadas, etc. El agua almacenada puede ser usada para cualquier fin, siempre y cuando se realicen los tratamientos adecuados para cada uso. Para usos básicos como limpieza de ropa, de pisos, sanitarios y riego puede usarse un filtro muy sencillo; para aseo personal y para agua que se pretenda beber, se deberá tener un sistema de filtros diferente, adecuados para estos fines (Adler I. et al., 2008).

La captación de agua de lluvia fue muy utilizada por las sociedades antiguas en todo el planeta y en muchas ocasiones, se utilizó como único procedimiento para el abastecimiento de agua en algunas regiones. Muchos edificios antiguos estaban construidos de tal forma que el agua que caía en los tejados se canalizaba a un gran depósito subterráneo o semi subterráneo (Carvajal Gómez D., 2016). Dentro de los potenciales usos del agua de lluvia, se pueden enumerar:

- Incendio
- Ducha
- Consumo humano
- Descarga de sanitarios
- Riego como actividad secundaria
- Aseo de superficies
- Lavado de Autos
- Lavado de Ropa

El agua de lluvia para la descarga de sanitarios ha sido utilizada en la solución de Estupiñán Perdomo J. L. y Zapata García H. O. (Estupiñán Perdomo J. L. y Zapata García H. O., 2010), En la propuesta de Olarte Mira D. (Olarte Mira D., 2017) se emplea el agua de lluvia para la descarga de sanitarios. Arengas Reines M. F. y Bernal Meneses J. A. (Arengas Reines M. F. Bernal Meneses J. A.,

2016) proponen el uso de agua de lluvia para descarga de sanitarios en Bogotá.

En Latinoamérica, países como Colombia son propensos a tener zonas con falta de agua, y en este caso, muchos estudiantes alumnos de diferentes niveles universitarios han propuesto en sus trabajos de tesis, proyectos de reutilización de agua de lluvia en descargas sanitarias en los campus o instalaciones universitarias (Carvajal Gómez D., 2016).

Existe cierto grado de automatización en la recolección y uso del agua de lluvia, se puede mencionar el trabajo de Duque Bonilla J. y Salazar Gil J. (Duque Bonilla J. y Salazar Gil J., 2013) en el cual se utilizan sensores y electrónicas de realizar aperturas de válvulas. En el trabajo de Riatiga Fadiño H. (Riatiga Fadiño H., 2016) se presenta una automatización mediante el uso de sensores y bombas para llenar tanques automáticamente.

El grupo de investigación, se plantea la siguiente pregunta de investigación (PI): ¿Es posible contribuir a la sustentabilidad del agua potable mediante un sistema de captación de agua de lluvia y darle uso para la descarga de sanitarios?

El desafío de esta línea de investigación consiste en dar respuesta a la pregunta presentada.

Líneas de Investigación, Desarrollo e Innovación

La línea de investigación aborda un problema que enfrenta la población mundial, no solamente por tener escasez de agua, ya sea potable o no. Se toma como punto de partida los antecedentes sobre la captación de agua de lluvia, sus posibles usos y en especial su destino a la descarga de sanitarios.

En este contexto, se pretende lograr una descentralización parcial del suministro del agua en las zonas habitadas y el lógico hábito de emplear el agua que cae del cielo, en vez de desperdiciarla y utilizar agua potable de red en aplicaciones que no deberían ser necesarias.

Por las razones mencionadas, se pretende que las TICs contribuyen a la sustentabilidad del agua potable mediante un prototipo que permita la captación de agua de lluvia para uso de recarga de sanitarios.

Para dar respuesta a nuestra pregunta de investigación planteada en la sección Introducción, se proponen los siguientes objetivos:

Objetivo general: obtención de un prototipo de un sistema de gestión de agua de lluvia para descarga de sanitarios.

Objetivos específicos:

1. Elaborar el estado del arte sobre la gestión de agua de lluvia y sus usos.
2. Identificar y especificar los requisitos funcionales y no funcionales para el prototipo de gestión de agua de lluvia.
3. Conceptualizar el sistema de gestión de agua de lluvia mediante la construcción de modelos.
4. Implementar el prototipo.
5. Probar la solución.
6. Evaluar la viabilidad del prototipo en un caso real.
7. Concientizar a la comunidad UM y extramuros sobre la sustentabilidad del agua potable.

Para el desarrollo del trabajo, se seguirá un enfoque de investigación clásico (Riveros H. et al., 1985) con énfasis en la producción de tecnologías (Sábato J. et al., 1982); identificando los métodos y materiales necesarios.

- Métodos: estudio de mapeo sistemático de la literatura (en inglés, Systematic Literature Mapping o SMS) método DESMET (Kitchenham B. et al., 1996), estudios de casos (Runeson P. et al., 2012), encuestas (Genero Bocco M. et al., 2014), prototipado evolutivo experimental (Basili V., 1993). Para el desarrollo de los formalismos y procesos de Ingeniería de software, se utilizarán los modelos conceptuales propuestos en el Lenguaje de Modelado Unificado (UML) (Booch G. et al., 2006) y los procesos técnicos, de soporte y de gestión definidos en el estándar ISO/IEC/IEEE 12207 (ISO/IEEE 12207, 2017).
- Materiales. Para llevar a cabo la búsqueda de la literatura se emplearán las librerías digitales que se encuentran disponibles en

la Biblioteca de la Universidad de Morón: IEEE Xplore, Scopus y ACM. Para la documentación de los modelos conceptuales del prototipo se utilizará una herramienta CASE (Software Engineering Asisted by Computer), la cual será evaluada en ese hito del proyecto.

A continuación, se presenta el abordaje metodológico para el desarrollo del trabajo:

- Para el Objetivo Específico 1, se realizará un estudio de mapeo sistemático (SMS).
- Para los Objetivos Específicos 2, 3, 4 y 5, se seguirán los procesos técnicos, de soporte y de gestión propuestos en la ISO/IEC/ IEEE 12207, se emplearán los formalismos de UML se utilizará una herramienta CASE y para la comparativa de tecnologías se empleará el método DESMET.
- Para el Objetivo Específico 6, se utilizarán estudios de casos.
- Para el Objetivo Específico 7, se utilizarán encuestas.

Resultados y Objetivos

Dado que es una línea de investigación que se inicia en la UM, no se presentan ***resultados concretos*** a describir.

Los ***resultados estimados*** a lograr son:

1. Académicos: lograr un trabajo de fin de carrera de grado.
2. Producción Científica: presentar los avances de la investigación en eventos científicos de alcance nacional (CACIC[3] 2020, CACIC 2021), y en el ámbito internacional, CIACA[4] 2020, y ICAETT[5] 2021, etc.
3. Proyectos de Investigación: esta línea se enmarca en el PIO titulado: "Impacto de las TICS en la sustentabilidad del uso del agua potable en la República Argentina" (período 2020-2021).
4. Formación en I+D+I: es un desafío para el grupo de investigación afrontar la resolución de un problema de la sociedad mediante la aplicación de las TICS. Se adquirirán competencias en el uso de métodos de investigación experimentales como, revisiones sistemáticas, estudios de casos y encuestas. El carácter de la línea de investigación requiere que el grupo trabaje con otras disciplinas como, por ejemplo, la electrónica, la arquitectura y la ingeniería ambiental.
5. Transferencia al medio socioeconómico: un prototipo que permita la gestión de la captación del agua para la descarga de sanitarios. Otro de los ejes a ejes a cumplir, consiste en la concientización a la comunidad UM y extramuros sobre la sustentabilidad del agua potable.

Formación de Recursos Humanos

El equipo de investigación se encuentra conformado por un Director, un Co-Director, dos investigadores en formación, dos becarios de grado.

Se estima la formación de dos Licenciados en Sistemas de la UM.

Referencias

Adler, I., Carmona, G. & Bojalil, J. A. (2008). Manual de Captación de Aguas de Lluvia - Uso y mantenimiento para su sistema residencial. Centros Urbanos. Instituto Internacional de Recursos Renovables AC. (IRRI). Disponible en: http://www.irrimexico.org/biblioteca. Último acceso: 7 de diciembre 2019. (2008).

Ambientum. Portal Profesional del medioambiente. Disponible en: https://www.ambientum.com/category/ambientum. Último acceso: 13 de diciembre 2019. Madrid, España (2019).

Arengas Reines M. F. Bernal Meneses J. A. Estudio de Prefactibilidad para el montaje de una empresa dedicada a la venta de sistemas de aprovechamiento de aguas de lluvia en el sector químico, agroindustrial y residencial de Bogotá D.C. Tesis de Especialización en Gestión de Proyectos en Ingeniería. Universidad Distrital Francisco José de Caldas. Bogotá (2016).

Basili V. (1993). The Experimental Paradigm in Software Engineering. En Experimental

3 Congreso Argentino de Ciencias de la Computación.

4 Conferencia Iberoamericana de Computación Aplicada.

5 Congreso Internacional sobre Avances en Nuevas Tendencias y Tecnologías.

Software Engineering Issues: Critical Assessment and Future Directions. En Computer Science, Vol. 706.

Booch G., Rumbaugh J., Jacobson I . El Lenguaje Unificado de Modelado, UML 2. 0: Guía de Usuario: Aprenda UML Directamente de sus creadores. · Addison-Wesley (2006).

Carvajal Gomez Danny Alejandro. Prototipo sistema automatizado de recolección de agua lluvia doméstico. Tesis de Grado de Tecnólogo en Automatización Industrial. Corporación Universitaria Minuto De Dios - Tecnología en Automatización Industrial. Bogotá (2016).

Duque Bonilla Juan Andrés y Salazar Gil Julián. Factibilidad económica y financiera de un sistema de captación, recolección y reutilización de agua pluvial aplicable en las instituciones educativas básicas primaria y secundaria en la ciudad de Pereira. Tesis profesional de Ingeniero Comercial. Universidad Libre Seccional Pereira (2013).

Estupiñán Perdomo J. L., Zapata García H. O. Requerimientos de Infraestructura para el Aprovechamiento Sostenible del Agua Lluvia en el Campus de la Pontificia Universidad Javeriana, sede Bogotá. Tesis de Maestría en Ingeniería Civil. Pontificia Universidad Javeriana Pontificia Universidad Javeriana. (2010).

Genero, M., Piattini, M., & Cruz Lemus, J. A. (2014). Métodos de investigación en Ingeniería del Software. Madrid: Ra-Ma S.A. Editorial y Publicaciones (2014).

Herrera Monroy Luis Alberto. Estudio de alternativas para el uso sustentable del agua de lluvia. Tesis de Maestría de Ingeniería Civil. Instituto Politécnico Nacional. México (2010).

IEEE ISO/IEC/IEEE 12207:2017(E). Systems and software engineering — Software life cycle processes (2017).

Kitchenham, B. (1996). "DESMET: A Method for Evaluating Software Engineering Methods and Tools " Department of Computer Science, University of Keele, Technical Report TR96-09, August 1996. ISSN: 1353-7776

Kitchenham, B. y Chartes, S. (2007). Guidelines for performing systematic literature reviews in Software engineering, Keele University, EBSE-2007-01.

Novo C. Experts reveal build up to global water crisis as water scarcity increases. Smart Water Magazine. Disponible en: https://smartwatermagazine.com/news/smart-water-magazine/experts-reveal-build-global-water-crisis-water-scarcity-increases. Último acceso: 21 de noviembre 2019 (2019).

Olarte Mira Diomar Alesis. Implementación de un sistema automático para el aprovechamiento del agua de lluvia y el agua residual de la lavadora. Tesis de Grado de Ingeniería Electromecánica. Instituto Tecnológico Metropolitano de Medellín. Disponible en: https://repositorio.itm.edu.co/jspui/handle/ITM/71. Último acceso: 10 de noviembre 2019 (2017).

ONU - Organización de las Naciones Unidas. Objetivos de Desarrollo Sostenible. Disponible en: https://www.un.org/sustainabledevelopment/es/. Último acceso: 2 de diciembre 2019 (2018).

Organización Mundial de la Salud. Guías para la calidad del agua de consumo humano. Cuarta edición que incorpora la primera adenda, 4a ed + 1a adenda. Organización Mundial de la Salud. Disponible en: https://apps.who.int/iris/handle/10665/272403. Último acceso: 16 de diciembre 2019 (2018).

Riatiga Fadiño Hugo. Evaluación del uso racional del agua y energía en una institución educativa. Tesis de la Maestría en Sistemas Energéticos Avanzados. Universidad de Santander. Bucaramanga, Colombia (2016).

Riveros H., Rosas L. (1985). El Método Científico Aplicado a las Ciencias Experimentales. Editorial Trillas. México.

Runeson P, Höst M, Rainer A, Regnell B. (2012). Case study research in software engineering: guidelines and examples. Wiley Publishing, Hoboken.

Sábato J., Mackenzie M. (1982). La Producción de Tecnología. Editorial Nueva Imagen. México.

TECNOLOGÍAS VESTIBLES APLICADAS AL CUIDADO DE LA SALUD – CONSTRUCCION DE UN PROTOTIPO DE MONITOREO

Pablo M. Vera, Rocío A. Rodríguez, Mariano Kaimakamian Carrau

CAETI - Centro de Altos Estudios en Tecnología Informática
Universidad Abierta Interamericana (UAI)
Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina

{pablomartin.vera, rocioandrea.rodriguez} @uai.edu.ar
mariano.kaimakamiancarrau@alumnos.uai.edu.ar

RESUMEN

Existen muchas personas que requieren de un cuidado y control permanente ya que de otra forma podría correr riesgo su salud. En muchos casos los adultos mayores por cuestiones inherentes a la edad sufren deterioro físico y/o mental que puede traer ciertos inconvenientes. Por ejemplo pueden sufrir desorientamiento momentáneo, caídas, dolores que no le permitan moverse, etc. El cuidado de estas personas requiere en muchos casos de un acompañante terapéutico permanente pero en muchos otros no es necesario pero se los debe acompañar para prevenir inconvenientes. En estos casos la persona se siente disminuida, incómoda por no tener privacidad o porque siente que no confían en sus capacidades. Además los costos que requiere acompañar constantemente a una persona son muy elevados ya sea en tiempo o dinero. En estos casos un sistema automatizado que realice un monitoreo no invasivo donde los familiares y amigos sean solo contactados ante una emergencia es una posible solución dando más libertad a la persona sintiéndose autónoma pero a la vez protegida porque sabe que ante cualquier problema tendrá la ayuda necesaria

Palabras clave: Interfaz, Tangible**,** Interacción Física, Dispositivos Móviles, Smartphone

CONTEXTO

Esta línea de I+D forma parte de los proyectos radicados en el Laboratorio de Algoritmos y Software del Centro de Altos Estudios en Tecnología Informática (CAETI) de la Universidad Abierta Interamericana (UAI). El proyecto cuenta con financiamiento asignado y una duración de 2 años.

1. INTRODUCCIÓN

En este proyecto se trabaja con tecnología vestible más conocida por su término en inglés wearable. “La palabra wearable posee una raíz inglesa cuya traducción significa “llevable” o “vestible“, en el argot tecnológico hace referencia a pequeñas computadoras que van siempre con el usuario. Bajo esta concepción, el PC deja de ser un dispositivo extraño para el usuario que solo lo usaba en un espacio definido pasando a ser un factor que se incorpora e interactúa de forma continua con él, además de acompañarlo a todas y cada una de las partes” [1] .

Basado en estadísticas mundiales, es posible afirmar que continuará creciendo en el 2020 las ventas de dispositivos wearables, “Las previsiones colocan en la primera posición de la lista a los auriculares bluetooth, con una cifra de ventas prevista para 2018 de 168 millones de unidades. La segunda y tercera posición son para las pulseras y los relojes inteligentes, ambos con una cifra de ventas cercana a los 49 millones de unidades” [2].
Así como un reloj inteligente (smartwach), tiene una serie de sensores que le permiten

conocer a un deportista la ruta realizada, la cantidad de kilómetros recorridos, también puede permitirle conocer a una persona con problemas de taquicardia cuál es su ritmo cardiaco; es decir que esta tecnología podrá tener usos muy variados. Esto permite plantear un universo de posibilidades de uso y un gran abanico de aplicaciones que pueden ser desarrolladas contando con los sensores que acompañan a estas tecnologías vestibles. "Los dispositivos wearables ahora son la nueva apuesta del futuro tecnológico" [3].

Actualmente la tecnología puede ser muy útil para poder sensar datos que puedan contribuir con cuestiones médicas, evitar accidentes que puedan ocurrirles a personas mayores ó personas que sufran determinadas patologías como por ejemplo pérdida temporal de la memoria. La tecnología actualmente está muy aferrada a la sociedad, siendo los más jóvenes los que no pueden prescindir de ella. "Actualmente la tecnología es una manera de vivir para muchos jóvenes [4].". Por lo cual las soluciones tecnológicas resultan ampliamente aceptadas por los jóvenes. Pero en el caso de los adultos mayores, adoptar una solución tecnológica puede resultar complejo. En algunos casos porque la tecnología les es ajena, les resulta difícil de manejar quedando excluidos tecnológicamente. Si bien hay otros motivos de exclusión basados en cuestiones socio-económicas. "la exclusión digital afecta principalmente a las personas adultas mayores" [5].

Este proyecto asume el desafío de incorporar tecnología en forma ubicua para la asistencia a adultos mayores. Para ello se desarrolla una solución que permite acompañar las actividades diarias de un adulto mayor sin resultar invasivo, por medio de un dispositivo vestible como lo es el smartwach, pudiendo ser habilitadas diversas funciones de monitoreo según las necesidades propias de cada usuario. La solución dispara alertas las cuales puedan ser recibidas por el usuario u otros contactos (que pueden ser configurados) en sus dispositivos móviles y de esta manera poder brindar ayuda al adulto mayor. Siendo ideal para esto las tecnologías vestibles como es el caso del smartwatch ya que puede ser utilizado como un reloj tradicional y una persona de confianza más familiarizada con la tecnología podría colaborar con el adulto mayor en el monitoreo de los datos vinculados con las funcionalidades que se consideren relevantes. En la sección 3 se presentan brevemente las funcionalidades actualmente desarrolladas en el prototipo de aplicación que hemos denominado Vigia.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Los ejes principales del trabajo son:

- Análisis de los sensores disponibles en los equipos móviles comprendiendo su funcionamiento. De esta forma se podrá determinar las posibilidades y limitaciones al momento de acceder a la información de los sensores mediante una aplicación
- Diseño de interfaces innovadoras que mediante el uso de los sensores permitan manejar distintas aplicaciones.
- Aplicación de sensores a tecnologías vestibles
- Análisis del impacto de los datos sensados en pos de alertar posibles problemas de salud.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se realizó un primer prototipo en Android el cual está funcionando para teléfonos celulares y será migrado a WearOS que por estar basado en Android esta migración será de simple realización.

A continuación en la figura 1 se muestra la pantalla de configuración con las opciones disponibles.

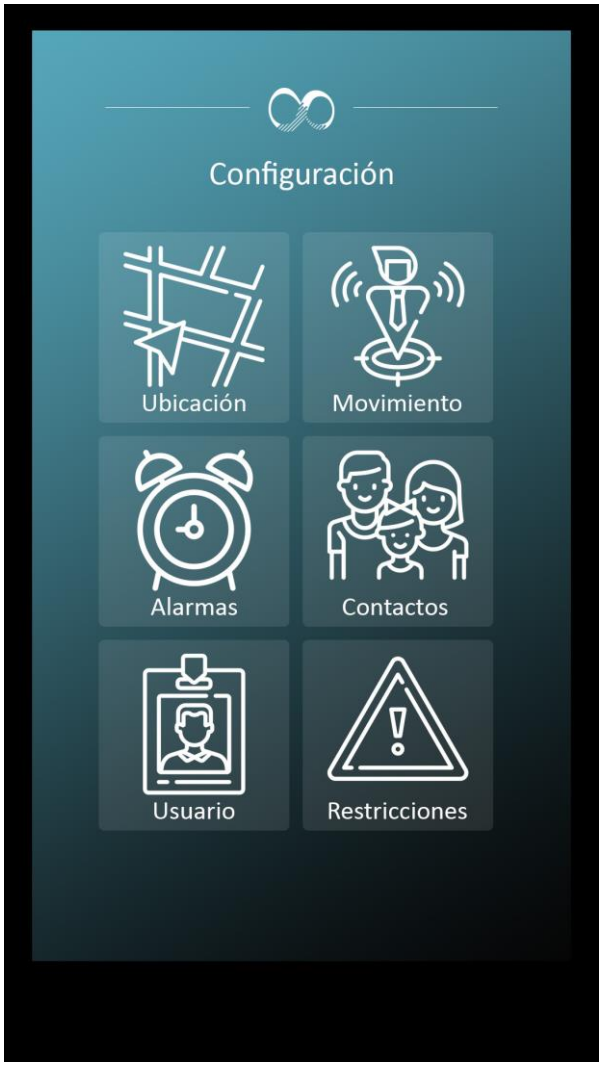


Figura 1. Pantalla de Configuración con Opciones disponibles

- Ubicación: Se define la zona de seguridad, una área en el que el usuario puede permanecer ilimitada cantidad de tiempo sin disparar ninguna alarma (generalmente el hogar o residencia en la que se encuentre), un área de control que es un ámbito en el que es posible que el portador de Vigía pueda realizar alguna actividad durante un período acotado de tiempo (por ejemplo, una tienda de conveniencia), y el tiempo de permanencia máximo antes de dar aviso a los contactos de confianza sobre el exceso de permanencia en la zona de control. Todas las transiciones entre las zonas disparan mensajes informativos.

  Es posible desactivar esta funcionalidad si no resulta de interés.

- Movimiento: Independientemente del geoposicionamiento, es posible mantener el monitoreo de las actividades del usuario a través del acelerómetro, algo crucial cuando por algún motivo no es posible hacer uso del posicionamiento satelital; en adición, es válido recordar que el geoposicionamiento no brinda información si no se produce una traslación. El funcionamiento se basa en la definición de una cota de tiempo a partir del cual el sistema, si no registra movimiento, dispara una alerta a los contactos de confianza para que tomen las medidas que consideren adecuadas. Para evitar falsos positivos, el apartado incluye la opción de definir rangos horarios en los que es plausible que el usuario se encuentre en reposo. Como ocurre con la funcionalidad de ubicación, también es posible desactivar esta característica. En futuros lanzamientos se incluirán mecanismos que detecten caídas, de modo tal de poder ofrecer un escenario más acabado de la situación del portador de Vigía.

- Alarmas: Este módulo va más allá de las funcionalidades básicas ofrecidas por el sistema operativo al incorporar el uso de tags NFC (near field communication) como mecanismo para validar que se haya realizado aquella actividad para la cual la alarma fue configurada. Imaginemos el escenario en la que una persona requiera medicarse y que, por algún tipo de distracción eventual, se olvidara de tomar el medicamento una vez desactivado el recordatorio; dependiendo del tratamiento, esto podría ser extremadamente perjudicial. Poder configurar un tag NFC [6] adherible al envase del medicamento con la orden de detener la alarma cuando el teléfono se encuentre lo suficientemente cerca, para lo cual se requiere un acto consciente, podría minimizar el olvido de la ingesta. Este es uno de varios usos que el sistema de

alarma con corte NFC -configurable- puede recibir.

- Contactos: La solución requiere que se configure al menos un contacto de confianza encargado de recibir las notificaciones sobre las distintas actividades, a través de distintos medios configurables. Actualmente es posible definir como canal de comunicación el correo electrónico o sms, aunque en lanzamientos posteriores será posible emplear otro tipo de medios como redes sociales. La nómina de posibles contactos es recuperada de la agenda del dispositivo móvil, razón por la que no hay necesidad de imputar manualmente ningún dato. Al momento de definir los contactos de confianza es posible indicarle al sistema cuáles de ellos formarán parte de la nómina de contactos de acceso rápidos, los que aparecerán en la pantalla principal para que el usuario de Vigía pueda llamar en caso de necesidad.

- Usuario Este apartado está reservado para incluir información acerca del usuario del dispositivo, importante para los casos en que la persona no se encuentre en condiciones de proveer información acerca de su identidad o afecciones.

- Restricciones: Hay ciertos lugares de la casa que pueden ser peligrosos para adultos mayores o personas con movilidad reducida, por ejemplo, un segundo piso, una terraza o un sótano. Sin embargo, muchas veces las personas acceden igualmente a esas zonas con el riesgo que ello acarrea. La aplicación mediante el uso del barómetro puede monitorear cuando la persona está cambiando de planta y dar una alerta avisando que corre algún riesgo y además puede notificar a los contactos si se accede a una zona inapropiada.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo está formado por 5 personas, docentes de grado, postgrado y alumnos.

En el área de dispositivos móviles se encuentran en realización 2 tesis de maestría en la UAI (Universidad Abierta Interamericana).

5. BIBLIOGRAFIA

[1] ¿Que es Wearable? – Los dispositivos vestibles . 2014. http://www.dispositivoswearables.net/

[2] Statista. Previsión de las ventas de dispositivos wearables a nivel mundial de 2017 a 2021, por categoría de producto (en millones de unidades), 2017 https://es.statista.com/estadisticas/637982/ventas-de-dispositivos-wearable-en-el-mundo-por-categoria/

[3] Monge Rodríguez, Janneth Mariela. Estudio de las características y el impacto de los wearables centrado en el paradigma del internet de las cosas. BS thesis. PUCE, 2017.

[4] Iliana Gpe. ¿Cómo afecta la tecnología a los jóvenes en la vida cotidiana?, 2017 https://medium.com/@142Gomeziliana/c%C3%B3mo-afecta-la-tecnolog%C3%ADa-a-los-j%C3%B3venes-en-la-vida-cotidiana-b58e97b52679

[5] Reverol, Carmen Laura Paz, et al. "El uso de las TIC en adultos mayores en Maracaibo (Venezuela)." Opción: Revista de Ciencias Humanas y Sociales 12 (2016): 169-188

[6] NFC, Near Field Communication. http://nearfieldcommunication.org/

Aplicaciones Web Progresivas Enfocadas en el Uso y Optimización de Cache

Rocío A. Rodríguez, Pablo M. Vera, M. Roxana Martínez, Claudia G. Alderete, Mariano G. Dogliotti

CAETI - Centro de Altos Estudios en Tecnología Informática
Universidad Abierta Interamericana (UAI)
Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina

{rocioandrea.rodriguez, pablomartin.vera, roxana.martinez, claudia.alderete, mariano.dogliotti}@uai.edu.ar

RESUMEN

La gran inserción de dispositivos móviles pone en foco el análisis de las soluciones diseñadas para los mismos. Los usuarios móviles descargan gran cantidad de aplicaciones y de hecho la tendencia indica que prefieren las aplicaciones móviles por sobre los sitios web móviles, este efecto se irá reduciendo con el desarrollo de PWA (Aplicaciones Web Progresivas). Las PWA permiten tener un ícono de acceso, poder trabajar offline utilizando información almacenada internamente en el dispositivo, pueden ofrecer notificaciones; una PWA bien desarrollada dará el mismo aspecto que una aplicación móvil instalada en el dispositivo. La ventaja de las PWAs es que si se optimizan bien los esquemas de cache, permitirá disminuir el consumo de datos. Sobre este tema de optimización y esquemas de uso de cache se centra el presente proyecto.

Palabras clave: Web Móvil, Dispositivos Móviles, PWA, Cache

CONTEXTO

Esta línea de I+D forma parte de los proyectos radicados en el Centro de Altos Estudios en Tecnología Informática (CAETI) de la Universidad Abierta Interamericana (UAI). En este proyecto participan docentes y alumnos tanto de sede Centro como de la Castelar (ambas en la provincia de Buenos Aires). El proyecto cuenta con financiamiento asignado y una duración de 2 años.

1. INTRODUCCIÓN

La tecnología móvil ocupa un papel primordial en la vida de las personas hasta el punto por el cual en la actualidad el teléfono móvil combinado con internet es el principal medio por el cual se accede a la información. Una gran parte de los usuarios están acostumbrados a instalar y utilizar distintas aplicaciones en sus dispositivos aun cuando la versión web esté adaptada para ser utilizada correctamente en la pantalla reducida. Según el reporte global que analiza el estado de las tecnologías móviles creado por la consultora Comscore, "más del 80% de los minutos de uso en los dispositivos móviles, tomando en cuenta todos los mercados, son utilizados en apps". Particularmente en Argentina según este mismo informe el porcentaje de tiempo de uso de apps asciende al 94% [1].

Es esta preferencia de los usuarios que obliga a los desarrolladores a construir aplicaciones nativas, teniendo en su mayoría como foco los dos sistemas operativos más populares actualmente que son Android y IOS. Tener dos versiones de una misma aplicación hace que se multiplique el esfuerzo de desarrollo y más aún el de mantenimiento y distribución de nuevas versiones y mejoras. Una solución son los entornos de desarrollo híbrido que permiten programar en un único lenguaje y generar apps para ambos sistemas operativos. Pero este enfoque tiene el inconveniente de que la parte web queda por otro lado por lo que aún hay un doble trabajo. Para solucionar

este problema nacen las aplicaciones web progresivas.

Para acercar más a las aplicaciones web de las nativas "un nuevo concepto ha surgido en los últimos años denominado Aplicaciones Web Progresivas (PWA por sus siglas en inglés). Una PWA es una aplicación web que utiliza las últimas tecnologías disponibles en los navegadores para ofrecer en dispositivos móviles una experiencia lo más parecida posible a la de una aplicación nativa. Los objetivos que persiguen las PWA son: lograr el mayor rendimiento posible en dispositivos móviles, que la aplicación cargue de manera casi instantánea, que la interfaz de usuario se parezca lo máximo posible a una nativa, que se pueda trabajar sin conexión (offline first) y que se puedan enviar notificaciones a los usuarios, como en una aplicación nativa" [2].

"Las aplicaciones web progresivas son una evolución natural de las aplicaciones web que difuminan la barrera entre la web y las aplicaciones, pudiendo realizar tareas que generalmente solo las aplicaciones nativas podían llevar a cabo. Algunos ejemplos son las notificaciones, el funcionamiento sin conexión a Internet o la posibilidad de probar una versión más ligera antes de bajarte una aplicación nativa de verdad" [3].

Resulta importante tener un enfoque integral que conjugue aspectos de hardware y software. "Una de las características importantes de la gran mayoría de los desarrollos móviles es su corta duración. Esto se debe a factores como la gran competencia en el sector, los cambios en el mismo con la aparición, casi constante, de novedades tanto software como hardware" [4].

El poder almacenar en cache el contenido web reduce considerablemente el consumo de datos, haciendo la aplicación más rápida y siempre disponible para el usuario final. Nicolas Gallagher un desarrollador de la red social Twitter en su artículo «How we built Twitter Lite» [5], indica que gracias a la versión PWA consiguieron reducir el consumo de datos hasta fracciones irrisorias en comparación con sus aplicaciones nativas.

Las PWA permiten realizar una aplicación 100% desconectada, pero si los datos cambian dicho cache debe ser actualizado. Por lo que es necesario estudiar los distintos enfoques para el manejo de cache estableciendo la mejor estrategia para cada tipo de aplicación. Los enfoques posibles son:

- Solo cache: Utilizado para páginas estáticas que no cambian y siempre serán levantadas de la cache
- Solo red: Por el contrario del enfoque anterior siempre se irá a buscar el contenido a la red, muy útil para contenidos dinámicos que siempre cambian.
- Cache y si falla red: Esquema combinado en el que se prioriza la cache y en caso de no encontrar allí el contenido se recurre a la red. Al recuperarlo adicionalmente puede ser cargado en cache para futuros accesos.
- Red y si falla cache: Esquema combinado en el que se prioriza la red y si el usuario no tuviese conectividad el contenido a mostrar es el de cache.
- Carrera entre cache y red: Con carrera se refiere a que se consulta en cache y red en simultáneo y lo que primero se consiga es lo que se muestra. Por ejemplo, si se cuenta con equipos de acceso lento al almacenamiento, pero con una excelente conexión a la red para recursos pequeños puede ser más rápida la red que la cache.
- Cache y después red: Se muestra al usuario rápidamente lo de cache y mientras se busca en la red si hay actualización del contenido.
- Contenido de reserva: Si no está disponible en cache y no hay acceso a la red, tener contenido de reserva para

mostrarle al usuario (página de sin conexión ó indicación por defecto).

Los componentes principales de una PWA serán un archivo de manifiesto, service worker, almacenamiento local y notificaciones. De hecho es el service worker el que se ocupará de trabajar con el manejo de esquemas de cache o del acceso a la red. A modo de ejemplo se presenta la figura 1 (tomada de [6]), puede observarse que se recurre al service worker el cual accede a cache y si el contenido no está disponible accede a la red para buscarlo y poder servirlo. Este esquema prioriza la cache, con la posibilidad de actualizar la cache en los casos que se requiera acceder a la red, lo cual está identificado con la línea punteada (que se le ha agregado a la figura del autor).

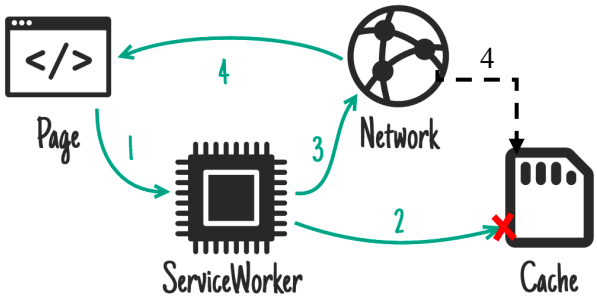


Figura 1. Esquema de acceso con prioridad a cache y luego red.

PWA hace que el usuario no distinga de si se trata de una aplicación nativa o una web, incluso pudiendo tener acceso al hardware disponible como GPS, vibración entre otros, lo que brinda una experiencia de uso más completa. “La brecha entre las capacidades de las aplicaciones nativas sobre las aplicaciones web se está reduciendo, especialmente impulsada por los estándares web de acceso al hardware” [7].

Actualmente existen diversos trabajos en el área de PWA entre ellos [8], [9] y [10] se enfocan en el manejo y optimización de cache. Otros se basan en distintas tecnologías y frameworks como [11], [12] y [13]. Por otra parte [14] y [15] realizan una comparativa entre las PWA y las aplicaciones web tradicionales.

Estos trabajos son solo algunos ejemplos y muestra de que el estudio de las PWA está en pleno auge, por lo tanto resulta de suma importancia su estudio y comparativa con otras tecnologías existentes.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Los ejes principales del trabajo son:

- Analizar las ventajas y desventajas de las PWA en comparación con las APP hibridas
- Analizar la evolución de los estándares para el manejo de hardware e información almacenada en el dispositivo mediante una PWA
- Profundizar en esquemas de Cache para manejar el versionado
- Comparar los distintos de entornos de desarrollo en contrapartida al desarrollo de PWA determinando la mejor opción según el tipo de APP a desarrollar.

3. RESULTADOS OBTENIDOS/ESPERADOS

Dado que este proyecto ha comenzado en el presente año aún no hay resultados obtenidos. Pero los resultados a obtenerse serán producto de analizar y probar los distintos esquemas de cache disponibles proponiendo la mejor forma de administrar la misma según el tipo de aplicación que se quiera realizar. No se aplicará el mismo esquema para un sitio estático que para un uno dinámico, pero a su vez, si el sitio tiene páginas estáticas se busca analizar qué pasa cuando se quiere incorporar una nueva versión por alguna modificación. El service worker al cambiar la versión hace

que las páginas estáticas se vuelvan a descargar lo que puede incurrir en gasto de tiempos y datos por un cambio menor. Uno de los objetivos es generar un esquema de manejo de cache que permita actualizar solo lo necesario para que el usuario no descargue datos que ya tiene. Esta optimización se realizará directamente desde el service worker trabajando con los eventos y métodos disponibles. Es importante que el usuario tenga disponible siempre la aplicación y que el desarrollador pueda actualizarla con un mínimo costo para el usuario cuando sea necesario.

En cuanto a los entornos de desarrollo híbridos se planea analizar los más difundidos actualmente y categorizarlos según sus características por ejemplo en los que usan web View, los que usan los controles nativos y los que renderizan sus propios controles.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo está formado por 5 docentes: 2 de ellos doctores en Ciencias Informática graduados en la Universidad Nacional de La Plata (UNLP), 1 con maestría finalizada (UAI) con la cursada del doctorado finalizada y que se encuentra actualmente haciendo su tesis doctoral (UNLP) y 1 realizando actualmente su maestría en UAI. Esto implica que 4 de los 5 docentes que componen el grupo tienen estudios de posgrados finalizados o en progreso.

Este proyecto también cuenta con la participación de alumnos de grado y posgrado de la UAI (actualmente en el proyecto se encuentran vinculados 4 alumnos).

En el área de dispositivos móviles se encuentran en realización 4 tesis de maestría (2 en la UAI y 2 en UNLaM – Universidad Nacional de La Matanza), siendo directores de dichas tesis miembros del equipo de investigación.

5. BIBLIOGRAFIA

[1] Comscore, 2019 Report Global State of Mobile https://www.comscore.com/Insights/Presentations-and-Whitepapers/2019/Global-State-of-Mobile

[2] Thomas Pablo, Delia Lisandro, Corbalán Leonardo, et al. "Tendencias en el desarrollo de Aplicaciones para Dispositivos Móviles". Instituto de Investigación en Informática LIDI (III-LIDI) Facultad de Informática – Universidad Nacional de La Plata. Centro Asociado a la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC). Abril 2018. https://digital.cic.gba.gob.ar/bitstream/handle/11746/8316/11746_8316.pdf?sequence=1&isAllowed=y

[3] Ramírez Ivan (2018), ¿Qué es una aplicación web progresiva o PWA?. https://www.xataka.com/basics/que-es-una-aplicacion-web-progresiva-o-pwa

[4] Vique, R. R. (2012). Métodos para el desarrollo de aplicaciones móviles.

[5] N. Gallagher, "How We Built Twitter Lite," Apr. 2017. https://blog.twitter.com/engineering/en_us/topics/open-source/2017/how-we-built-twitter-lite.html

[6] Archibald Jake (2019), La guía de soluciones sin conexión. https://developers.google.com/web/fundamentals/instant-and-offline/offline-cookbook/?hl=es

[7] Rodríguez Rocío A, Vera Pablo M, Martínez M. Roxana, Pons Claudia, Valles Federico E, de La Cruz Luis Verbel. "Reducing the Gap between Native and Web Applications". SETESEC – Italia2014

[8] Rojas, C. (2020). Caching Strategies. In Building Progressive Web Applications with Vue. js (pp. 67-81). Apress, Berkeley, CA.

[9] Hajian, M. (2019). Advanced Angular Service Worker and Runtime Caching.

In Progressive Web Apps with Angular (pp. 107-139). Apress, Berkeley, CA.

[10] Kim, H. G., Park, J. T., Choi, M. H., & Moon, I. Y. (2019). Design of Web Content Update Algorithm to Reduce Communication Data Consumption using Service Worker and Hash. The Journal of Advanced Navigation Technology, 23(2), 158-165.

[11] Satyal, A. (2020). Designing and Developing a Website with ReactJS: Progressive Web Application.

[12] Hajian, M. (2019). App Shell and Angular Performance. In Progressive Web Apps with Angular (pp. 169-199). Apress, Berkeley, CA.

[13] Khan, A. I., Al-Badi, A., & Al-Kindi, M. (2019). Progressive Web Application Assessment Using AHP. Procedia Computer Science, 155, 289-294.

[14] Said Tahirshah, F. (2019). Comparison between Progressive Web App and Regular Web App.

[15] Kvist, J., & Mathiasson, P. (2019). Progressive Web Apps and other mobile developing techniques: a comparison.

DISEÑO Y DESARROLLO DE APLICACIONES MÓVILES BASADAS EN GEOFENCING

Pablo M. Vera, Rocío A. Rodríguez, Hernán A. Viavattene, M. Roxana Martínez

CAETI - Centro de Altos Estudios en Tecnología Informática
Universidad Abierta Interamericana (UAI)
Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina

{pablomartin.vera, rocioandrea.rodriguez, roxana.martinez} @uai.edu.ar
hernanalberto.viavattene@alumnos.uai.edu.ar

RESUMEN

Hoy en día los dispositivos móviles incorporan en su gran mayoría la posibilidad de ubicarse espacialmente mediante el GPS y además incluyen otros sensores que pueden ser aprovechados en el desarrollo de aplicaciones. Aprovechando dichos componentes este proyecto busca diseñar y desarrollar aplicaciones móviles basadas en geofencing complementadas con el uso de sensores adicionales disponibles en los dispositivos móviles. El foco principal estará dado por detectar cuando un dispositivo entra o sale de una zona delimitada pudiendo desencadenar distintos eventos y alertas cuando esto ocurra. Adicionalmente se trabajará el uso de otros sensores que permitan complementar la experiencia del usuario por ejemplo al desencadenar acciones con gestos o mediante el análisis de los datos de los sensores de movimiento detectar la situación actual de una persona.

Palabras clave: Aplicaciones Móviles, APPs, Geofencing, Sensores

CONTEXTO

Esta línea de I+D forma parte de los proyectos radicados en el Centro de Altos Estudios en Tecnología Informática (CAETI) de la Universidad Abierta Interamericana (UAI). El proyecto cuenta con financiamiento asignado y una duración de 2 años.

1. INTRODUCCIÓN

Los dispositivos móviles, en especial los Smartphone son cada vez más potentes, no solo en cuanto a velocidad de procesamiento y memoria disponible, sino también a la variedad de sensores y formas de comunicación que incorporan. Llevar al alcance de la mano equipos tan completos abre un abanico de posibilidades para los desarrollados permitiéndoles construir aplicaciones más completas y con formas de interacción novedosas.

"Los dispositivos móviles forman parte de la vida cotidiana y son cada vez más sofisticados, su poder de cómputo genera posibilidades hasta hace años no pensadas"[1].

La selección del componente de hardware a utilizar dependerá del tipo de aplicación que se quiera desarrollar. Por ejemplo, utilizando el GPS no solo es posible establecer la posición actual y generar mapas de recorrido, sino que también es posible delimitar cercos virtuales que permitan disparar acciones cuando un dispositivo ingresa o sale de un área determinada, concepto denominado "geofencing" [2]. "Geofencing es una pequeña área geográfica que se describe para producir un evento de ubicación tan pronto como un usuario ingrese o salga de esta geovalla" [3]. Se trata entonces de poder definir un perímetro virtual, que establecerá un área de interés en las que se monitorea al

usuario, ofreciéndole distintos servicios ó incluso simplemente notificaciones. Este perímetro virtual puede tener diversas formas, un círculo, línea ó polígono. En la figura 1 (tomada de [4]) se muestra a modo de ejemplo algunas áreas definidas.

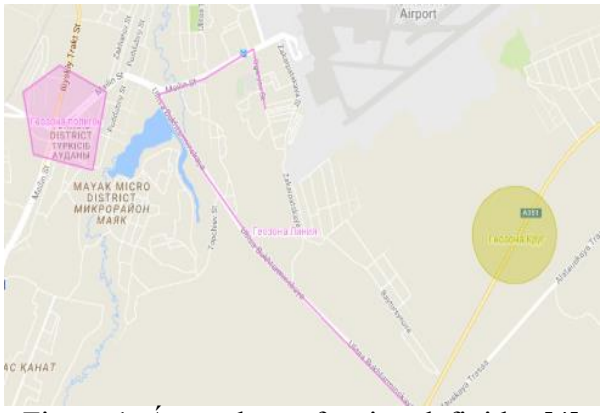


Figura 1. Áreas de geofencing definidas [4]

Los entornos de desarrollo para apps nativas incorporan APIs para facilitar a los desarrolladores incluir geofencing en sus proyectos. Por ejemplo, Android dispone de una guía completa para el desarrollador [5] que indica como crear cercas, como monitorear usuarios, cuestiones de ahorro de energía, etc. De la misma forma IOs también brinda estas características a sus desarrolladores [6].

Como alternativa a definir y controlar las cercas en el propio dispositivo, ya sea por cuestiones de procesamiento o de ahorro de energía, se puede recurrir a distintos servicios disponibles en la web. Estos servicios permiten definir las cercas virtuales y enviar la información de posición actual del usuario para determinar cuando ingresa o sale del área. Algunos ejemplos son: TomTom Geofencing API [7], Proximi.io [8], Radar.io [9]. Estas plataformas permiten gestionar múltiples áreas en forma simultánea, reduciendo además el esfuerzo de desarrollo, pero a expensas de un costo económico por el servicio.

Si bien ya existen servicios y APIs dedicadas a geofencing es necesario un análisis de las mismas para determinar cuál es la mejor opción según la aplicación y también desarrollar algoritmos propios de detección en el caso de que sea necesario.

Actualmente existen diversos trabajos referidos a geofencing aplicados a distintas áreas por ejemplo en [10] se utiliza geofencing para el control de vehículos autónomos. Con una temática relacionada en [11] se utiliza el geofencing para indicar a los conductores cuando están ingresando a un área con algún accidente vial, con el objetivo de evitar inconvenientes y embotellamientos.

Incluso existen patentes derivadas de la aplicación de geofencing como por ejemplo en [12] se establece una patente para poder realizar geofencing con consumo de energía reducido, mientras que en [13] se propone el uso de esta técnica en dispositivos médicos ante emergencias y en [14] se propone el uso de geofencing para compartir contenido en base a eventos.

Los planteados anteriormente son solo algunos ejemplos del poder de geofencing, este proyecto planea utilizar esta técnica para desarrollar aplicaciones como el seguimiento y control de personas, disparar alertas ante problemas médicos y de seguridad entre otras.

El uso de otros sensores adicionales permitirá complementar las aplicaciones de geofencing desarrolladas para brindar funcionalidades adicionales o permitir formas de interacción diferentes por ejemplo basadas en gestos utilizando los sensores de movimiento entre otras posibilidades.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Los ejes principales del trabajo son:

- Estudiar las APIs, servicios y técnicas de geofencing actuales para establecer la mejor opción a la hora del desarrollo de las APPs.

- Diseñar y desarrollar aplicaciones que hagan uso del GPS mediante Geofencing aplicado a distintos entornos
- Determinar que sensores pueden ser utilizados para complementar las aplicaciones desarrolladas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Este proyecto está en sus primeras etapas por lo que aún no se cuentan con resultados. No obstante, como antecedente el equipo de investigación ha desarrollado una aplicación VIGIA [15] en donde se muestra el radio en donde debería encontrarse una persona adulta mayor con problemas de orientación, por fuera de ese radio dispara una alerta a un familiar (esta aplicación ha sido pensada para ser migrada a un smartwatch). La zona definida era por medio de una circunferencia se analizará la posibilidad de definir zonas más complejas utilizando polígonos o calles específicas por medio de líneas. Esto aportará mayor conocimiento en el área de Geofencing. Este conocimiento permitirá realizar nuevos desarrollos y que a su vez sean complementados con el uso de sensores adicionales como los sensores de luz y movimiento dando mayores funcionalidades a la aplicación o haciendo interfaces más intuitivas y fáciles de utilizar.

Algunas de las posibles aplicaciones a desarrollar incluyen el aviso de salida de una persona (por ejemplo) un niño de una zona delimitada, aplicaciones de seguridad para buscar ayuda cercana, recibir ofertas de negocios al estar en las proximidades entre otras.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo está formado por 5 personas, docentes de grado, postgrado y alumnos.

En el área de dispositivos móviles se encuentran en realización 2 tesis de maestría en la UAI (Universidad Abierta Interamericana).

5. BIBLIOGRAFIA

[1] Thomas, P., Galdamez, N., Delía, L. N., Cristina, F., Dapoto, S., & Pesado, P. M. (2014). Dispositivos móviles: desarrollo de aplicaciones y conectividad. In Workshop de Investigadores en Ciencias de la Computación (WICC) (Vol. 16).

[2] Statler, S. (2016). Geofencing: Everything you need to know. In Beacon Technologies (pp. 307-316). Apress, Berkeley, CA.

[3] Zuva, K., & Zuva, T. Tracking of Customers using Geofencing Technology. 2019

[4] GTS4B. Sistema de monitoreo GPS GTS4B https://gts4b.com/es/docs/start/ui/geofences

[5] Android Developers, Cómo crear y supervisar el geovallado, Diciembre del 2019 https://developer.android.com/training/location/geofencing

[6] Apple Developer, Monitoring the User's Proximity to Geographic Regions, https://developer.apple.com/documentation/corelocation/monitoring_the_user_s_proximity_to_geographic_regions

[7] TomTom for developers, Geofencing API Documentation ,Noviembre 2019 https://developer.tomtom.com/geofencing-api/geofencing-api-documentation

[8] Proximi.io, API-first Geofencing for Mobile Apps https://proximi.io/geofencing-geofences/

[9] Radar Labs, Geofences, https://radar.io/documentation#context-geofences

[10] M. Maiouak and T. Taleb, "Dynamic Maps for Automated Driving and UAV Geofencing," in IEEE Wireless Communications, vol. 26, no. 4, pp. 54-59, August 2019.

[11] Nayak, B., Mugali, P. S., Rao, B. R., Sindhava, S., Disha, D. N., & Swarnalatha, K. S. (2019). GeoFencing-Based Accident Avoidance Notification for Road Safety. In Emerging Research in Computing, Information, Communication and Applications (pp. 379-386). Springer, Singapore.

[12] Kolavennu, S., & Weirich, K. L. (2019). U.S. Patent No. 10,271,284. Washington, DC: U.S. Patent and Trademark Office.

[13] Istoc, E. (2020). U.S. Patent No. 10,532,156. Washington, DC: U.S. Patent and Trademark Office.

[14] Matiash, B., Butko, N., Cohen, D. M., & Chang, E. (2019). U.S. Patent No. 10,257,291. Washington, DC: U.S. Patent and Trademark Office.

[15] P. M. Vera, M. K. Carrau and R. A. Rodríguez, "Mobile follow-up system for elderly and disabled people," 2018 Congreso Argentino de Ciencias de la Informática y Desarrollos de Investigación (CACIDI), Buenos Aires, 2018, pp. 1-5.

Enfoque para el diseño de aplicaciones móviles usando técnicas de Design Thinking dirigido a usuarios con dificultades/desafíos en habilidades sociales (Fase I).

Corso Cynthia, Hug Mercedes, Cecilia Challiol, Evin Diego, Carreras Xilenia.

Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información
Departamento Ingeniería en Sistemas de Información
Centro de Investigación y Transferencia Acústica
Facultad Regional Córdoba. Universidad Tecnológica Nacional
Maestro M. López esq. Cruz Roja-Ciudad Universitaria-Córdoba
cynthia@bbs.frc.utn.edu.ar/mercehug@gmail.com/ceciliac@lifia.info.unlp.edu.ar/
diegoevin@gmail.com/ carrerasxilenia@gmail.com

RESUMEN

El *Design Thinking* es un enfoque que fomenta la creación de soluciones innovadoras, además considera las Tecnologías de Información y Comunicación (TIC) como recurso para la resolución de necesidades y demandas reales de los usuarios. Este proyecto de investigación considera la aplicación del *Design Thinking* y el uso de la tecnología, mas precisamente el uso de aplicaciones móviles, para satisfacer necesidades de usuarios con dificultades/desafíos en habilidades sociales. En el que el objetivo principal del proyecto es proponer un enfoque para abordar el diseño de aplicaciones móviles dirigidos a niños y jóvenes que presentan dificultades/desafíos en la interacción social; el cual será planteado teniendo en cuenta técnicas utilizados en el *Design Thinking*.

Palabras claves: *Design Thinking, Enfoque de desarrollo de software, Dificultad en habilidades sociales, aplicaciones móviles.*

CONTEXTO

Este trabajo hace referencia al proyecto de investigación "Enfoque para el diseño de aplicaciones móviles usando técnicas Design Thinking dirigido a usuarios con dificultades/desafíos en habilidades sociales (Fase I)" PID-SIUTNCO0006612, que ha sido homologado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional. El contexto de desarrollo de la presente investigación es el Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información (GIDTSI) en forma conjunta con el Centro de Investigación y Transferencia Acústica (CINTRA), ambos radicados en la U.T.N Facultad Regional Córdoba.

1. INTRODUCCIÓN

Design Thinking se ha convertido en los últimos años en una disciplina muy valiosa que se está implementando gradualmente en las organizaciones, como una forma de crear productos, servicios, proyectos y/o estrategias que tienden a satisfacer de mejor manera las necesidades de los usuarios; haciéndolos parte activa del proceso de creación. Esta disciplina utiliza la sensibilidad y los

métodos de los diseñadores para hacer coincidir las necesidades de las personas con lo que es tecnológicamente factible y con una estrategia viable de negocios para convertirla en un valor agregado para el cliente, así como una gran oportunidad de mercado [1].

Design Thinking utiliza instrumentos como la empatía, la observación y la experimentación para lograr soluciones innovadoras basándose en necesidad reales de los usuarios, considerando la utilización de prototipos con la finalidad de conseguir ideas más rápidas y eficientes. Si bien el ámbito inicial de aplicación de esta disciplina fue el contexto empresarial, gracias a su efectividad comenzó a ser tenido en cuenta en otras situaciones, como las de carácter social. Independientemente del ámbito, este enfoque promueve que el equipo de personas responsable de generar soluciones esté formado por un grupo multidisciplinar, el cual involucre en alguna de las etapas a los destinatarios finales. Un mismo problema abordado desde la perspectiva de diferentes perfiles y experiencias profesionales, puede ser resuelto de distintas maneras; lo cual aumenta el campo de oportunidades y la generación de soluciones.

Un ámbito de aplicación novedoso del Design Thinking, que aborda este proyecto de investigación, está dirigido a personas que por diferentes causas presentan dificultades/desafíos en las habilidades sociales; y en particular diseñar aplicaciones móviles que les sean útiles.
El estudio, entrenamiento y evaluación de habilidades sociales ha sido foco de interés en múltiples áreas (educativa, terapéutica y de investigación) [2], modificándose con el tiempo el modo en que han sido definidas y sus dimensiones principales. Algunas de las definiciones más clásicas las describen como conjuntos de competencias y comportamientos empleados en la vida diaria para relacionarse satisfactoriamente con otras personas [2][3][4]. Por otro lado, los modos de evaluarla se basan principalmente en escalas, auto informes y observaciones de docentes, terapeutas y cuidadores. La importancia de intervenir en esta área se debe a que un déficit en habilidades sociales tiene un profundo impacto en el desarrollo psicosocial del individuo [5].

En la actualidad gracias a los avances tecnológicos, se han generado y desarrollado diversas opciones de Tecnologías de la Información y Comunicación (TIC), las cuales han influido de manera favorable en la vida de las personas que presentan esta condición en las habilidades sociales [6][7]. Particularmente, los niños y jóvenes con TEA, poseen déficit en habilidades para la interacción social pero manifiestan un alto interés en la tecnología, por lo que intervenciones que partan de éste interés pueden resultar más motivadoras y atractivas para ellos.

Existe una tendencia emergente de integrar la tecnología en las intervenciones para mejorar las habilidades sociales. Los materiales asistidos por computadora usados en este contexto, incluyen videos y multimedias interactivos, recientemente las redes sociales y herramientas de aprendizaje virtual [8][9][10]. Sin embargo, en general el proceso de diseño de las mismas ha sido dirigido por metodologías formales que no están necesariamente centradas en el destinatario, es decir, en sus demandas y necesidades específicas. Si bien existen antecedentes de estudios e investigaciones [11][12][13] que han

abordado el diseño e implementación de aplicaciones móviles con el fin de facilitar el desarrollo de habilidades necesarias para la interacción social; el diseño y validación de estos sistemas desde la perspectiva de destinatarios con diagnóstico de dificultades en el neurodesarrollo es un campo poco explorado.

El proceso de diseño de aplicaciones móviles en estos casos implicaría la convergencia de acciones conjuntas por parte de un grupo de profesionales como psicólogos, fonoaudiólogos, docentes de educación especial, y padres. En este sentido los recursos (o técnicas) que ofrecen Design Thinking son capaces de encontrar soluciones creativas y beneficiosas fundamentalmente centradas en el destinatario [14].

A partir de este proyecto de investigación las personas que presentan dificultades/desafíos en las habilidades sociales se verán beneficiadas, ya que por medio del enfoque de diseño propuesto, será posible aplicarlo a diferentes grupos con la finalidad de atender y satisfacer necesidades particulares de este tipo de destinatarios. Así, se pretende aportar en el área temática de las TICS destinadas a niños y jóvenes con dificultades/desafíos en la interacción social, generando aportes metodológicos aplicables a novedosas tecnologías de diseño, enriqueciendo además a la transferencia de herramientas destinadas a niños y jóvenes con trastornos en el neurodesarrollo.

2. LINEAS DE INVESTIGACIÓN y DESARROLLO

Por las particularidades temáticas del proyecto de investigación hay dos líneas de investigación y desarrollo principales que se abordarán, las cuales se mencionan a continuación, detallando temas específicos de cada una de ellas.

- Ingeniería de Software
  - Metodologías de Ingeniería de Software aplicadas a comunidades con desafíos en las habilidades sociales. Esto permitirá poder considerar otros abordajes para poder tomar de ahí las mejores técnicas.
  - En relación a la temática de *Design Thinking* se investigará:
    - Pensamiento de diseño y Creatividad.
    - Desarrollo de la confianza creativa.
    - Modelo general de Design Thinking
    - Desarrollo de la Empatía y la definición de necesidades.
    - Procesos de Ideación.
    - Orientación a la misión.
    - Desarrollo de prototipado.
    - Procesos de evaluación.
    - Estrategias para implementar Design Thinking y esquemas de innovación.
- Habilidades sociales y TEA
  - Concepto y tipos de habilidades sociales.
  - Desarrollo de niños y jóvenes con TEA.
  - Las habilidades sociales en niños y jóvenes con TEA.
  - Enfoques terapéuticos y TIC orientados a habilidades sociales y TEA.

3. RESULTADOS OBTENIDOS/ESPERADOS

Este proyecto se ha planificado en su primera fase en dos años de ejecución

(2020-2021) para el desarrollo de los siguientes objetivos:

Objetivo General

Proponer un enfoque para abordar el diseño de aplicaciones móviles dirigidas a niños y jóvenes que presenten desafíos en la interacción social; el cual será planteado usando recursos utilizados en el *Design Thinking*.

Objetivos Particulares

- Analizar cómo abordar las técnicas de Design Thinking en la etapa de relevamiento de requerimientos de aplicaciones móviles dirigidas a destinatarios que presentan desafíos en habilidades sociales.
- Definir el enfoque de diseño propuesto mediante el uso de técnicas de *Design Thinking* mas apropiadas para co-diseñar el contenido de una aplicación móvil en forma conjunta con los destinatarios.
- Realizar un prototipo de maquetado para validar el mismo con los destinatarios que presenten desafíos en la interacción social, pero no posean diagnóstico de TEA (Trastorno de Espectro Autista); en base a esta validación enriquecer el diseño inicial considerando los requerimientos detectados. Esta decisión obedece a la necesidad de resguardar a los niños con trastornos en el neurodesarrollo en la etapa inicial del proyecto.

El enfoque de diseño propuesto en el presente proyecto de investigación será validado mediante el desarrollo de una aplicación móvil para ser usado en tablets. Esta elección se fundamenta en la necesidad de maximizar las posibilidades que tienen los niños y jóvenes con esta problemática de interactuar con un dispositivo de mayor dimensión y así garantizar una usabilidad óptima.

A estos tres primeros objetivos específicos se continuarán otros dos más que estarán vinculados con el desarrollo y validación de la aplicación móvil considerando el enfoque de diseño propuesto en esta primera fase. Para ello, se prevé presentar una nueva solicitud de financiamiento correspondiente a una Fase II.

Resultados esperados

Como este proyecto está al inicio de su primer año de ejecución los resultados que se han logrado hasta el momento están vinculados con la documentación del estado del arte relacionado con de los diferentes frameworks y métodos de *Design Thinking*. Como así también tomar conocimiento respecto a las características de los niños con TEA en diferentes áreas del desarrollo como la interacción social, la comunicación verbal y no verbal, la cognición, la capacidad imaginativa, las conductas repetitivas , los movimientos estereotipados y los intereses restringidos, las alteraciones motoras y las sensoriales [15].

A continuación se detallan algunas de las actividades planificadas para el desarrollo del proyecto de investigación:

- Preparación y selección de métodos que serán utilizados en las entrevistas.
- Entrevistas con los destinatarios y personas involucradas.
- Tabulación de los datos relevados en las entrevistas.
- Diseño colaborativo del enfoque de diseño de software.
- Validación del enfoque de diseñado con los destinatarios y personas involucrados.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto prevé la formación de los recursos humanos que forman parte de los mismo, y que al participar en el proyecto contribuyan a su formación. Dentro del contexto de este proyecto de investigación se esta desarrollando el trabajo de tesis de Doctorado en Ciencias Informática de la directora del proyecto.

Se incorporarán al equipo de trabajo docentes e investigadores, cuyo ambito de trabajo es la UTN- Facultad Regional Córdoba, en el rol de investigadores de apoyo con la finalidad de fortalecer su formación en investigación científica y tecnológica; como así también becarios alumnos de investigación, quienes colaborarán en el desarrollo de este marco metodológico.

Los estudiantes tendrán la posibilidad de hacer la Práctica Supervisada, de quinto año, en el marco del proyecto. Los avances y propuestas, estarán disponibles para su transferencia y aplicación en el Centro de Investigaciones, Desarrollo y Transferencia de Sistemas de Información - CIDS. Como este proyecto es de carácter interdisciplinario formarán parte del equipo de trabajo integrantes con formación en investigación en el campo de la Psicología que pertenecen al Centro de Investigación y Transferencia en Acústica CINTRA. En este marco, se solicitará una beca Posdoctoral CONICET destinada a una de las investigadoras formadas que integran el equipo.

5. REFERENCIAS

[1] Brown, T., & Wyatt, J. (2010). Design thinking for social innovation. Development Outreach, 12(1), 29-43.

[2] Caballo, V.& Salazar, I. (2017). Desarrollo y validación de un nuevo instrumento para la evaluación de las habilidades sociales: el cuestionario de habilidades sociales (CHASO) Behavioral Psychology / Psicología Conductual, Vol. 25, Nº 1, 2017, pp.

[3] Del Prette, Z. A. P., & Del Prette, A. (2002). Psicología de las habilidades sociales: terapia y educación. Revista Evaluar, 3(1).

[4] Reyna, C., & Brussino, S. (2011). Evaluación de las habilidades sociales infantiles en Latinoamérica. Psicologia em Estudo, 16(3), 359-367.

[5] Alfageme, B. & Sánchez P. (2002). Aprendiendo habilidades con juegos. Comunicar, 19, 2002. Revista científica de Comunicación y Educación pp. 114-119. ISSN 1134-3478.

[6] Moreira Rodríguez, L. Y., & Pesántez Flores, K. D. (2018). Desarrollo de una aplicación móvil como herramienta de apoyo en el proceso de enseñanza aprendizaje para la educación básica en niños con problemas de autismo (Doctoral dissertation, Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas. Carrera de Ingeniería en Sistemas Computacionales).

[7] Rodríguez Montalvo, M. D. C. (2018). EmoRutinas: Aplicación móvil para el fomento de las rutinas y el reconocimiento de emociones en personas con TEA.

[8] Rodrigo Marcos José Manuel, Romero Corral David. (2019).AraBoard Player: cargando y utilizando nuestro tablero de comunicación.

[9] Gómez Llorente, Adrián; Ezpeleta Mateo, Joaquín. (2014).AraSuite: Integración de las aplicaciones de Tico y Araword. Recuperado de https://sourceforge.net/projects/arasuite/

[10] Marcos José Manuel. (2015).Pictotraductor, un fantástico procesador de textos con pictogramas on

line con múltiples opciones. Informática para educación especial.
[11] Moya, D. P. R., Lima, J. S. G., & Pérez, Y. G. (2018). Sistema De Comunicación Con Intercambio De Imágenes Pecs, A Través De Una Interfaz Nui Como Parte De La Terapia De Lenguaje En El Trastorno De Espectro Autista. Open Journal Systems en Revista: REVISTA DE ENTRENAMIENTO, 3(1), 31-48.
[12] Romero Martínez, S. J., González, I., García, A., & Lozano, A. (2018). Herramientas tecnológicas para la educación inclusiva.
[13] Cortés Monroy, P. A., Quintero, M., & Sharay, S. (2018). Construcción de un sistema alternativo de comunicación de alta tecnología para un usuario con EMOC: un estudio de caso.
[14] Castillo, M., Álvarez, A., y Cabana, R. (2014). Design thinking: cómo guiar a estudiantes, emprendedores y empresarios en su aplicación. Ingeniería Industrial, 35(3), 301-311.
[15] Ramos, M. C. (2018). Trastornos del Espectro Autista: Grado de información sobre Indicadores y Estrategias de intervención en docentes de Nivel Inicial.

Interfaces no convencionales aplicadas a la captura de datos en procesos productivos industriales aplicados a las PyMEs

Panessi Walter, Ortiz Claudia, Juan Ignacio Tonelli, Pablo Chale, Eugenia Céspedes, Santiago Ricci
Departamento de Ciencias Básicas, Universidad Nacional de Luján (UNLu)
wpanessi@unlu.edu.ar; cortiz@unlu.edu.ar; juanitonelli@gmail.com; pcchale@gmail.com; eugeniaccespedes@gmail.com; sricci@unlu.edu.ar

Resumen

El proyecto que se lleva a cabo en la Universidad Nacional de Luján, se propone estudiar y desarrollar interfaces de usuario no convencionales aplicables en la captura de datos en procesos productivos industriales en PyMES, con el objetivo de contribuir al mejoramiento de la calidad y eficiencia del sector.

Palabras clave:

Interacción Humano-Computadora; Interfaces No Convencionales; Producción; PyMEs.

Contexto

Este trabajo de investigación se encuentra radicado en la Universidad Nacional de Luján, en su Departamento de Ciencias Básicas. El inicio fue en enero del 2020 y tiene como duración 5 años (2020-2024).

Introducción

Este proyecto de investigación se encuadra dentro de los proyectos de investigación aplicada. En particular, se pretende aplicar investigación y desarrollo de un área del conocimiento sobre otra área. Por el lado de la aplicación, posee una fuerte relación con los Procesos Industriales, principalmente con el estudio de la Organización y del Trabajo. La Organización del Trabajo es la ciencia que se ocupa del estudio de los procesos de producción (tanto de productos como de servicios), centrado en cómo participa el hombre en el mismo (Rojas Díaz, 2014). Su objetivo es maximizar la efectividad del trabajo del hombre en su interrelación con otras personas y la tecnología. Comprende el análisis de qué se hace, dónde, cómo y con qué, con el fin de diseñar e implantar métodos que perfeccionen la participación del hombre en el proceso de producción. Para ello, es necesario la obtención de datos sistemáticos que propicien un estudio que permita conocer con exactitud cómo se realizan las tareas y cuáles son las limitaciones que impiden lograr una mayor efectividad en el trabajo, con vistas a buscar y proponer las soluciones más adecuadas que permitan el incremento de la productividad.

Uno de los mecanismos que viabiliza la calidad de los procesos es la trazabilidad (Briceño, 2013). La trazabilidad es definida por la Organización Internacional de Estandarización ISO, en su estándar ISO 8402:1994 como *"... la capacidad de recuperar los antecedentes y uso o localización de un artículo o una actividad a través de una identificación registrada"*. Ese estándar fue luego reemplazado por el ISO 9000:2000, que amplía la definición en relación con los sistemas de administración de calidad de la siguiente manera: *"7.5.2. Identificación y Trazabilidad: La organización debería dar pasos para identificar el estado del producto/ servicio en lo que concierne a las actividades de verificación y mediciones requeridas y debería, cuando sea necesario, identificar el producto / servicio utilizando los medios adecuados*

a lo largo del proceso. Esto debería aplicarse a todas las partes involucradas en el producto y/o servicio cuando la interacción de ellas tiene peso en relación con el cumplimiento de los requerimientos. Cuando la trazabilidad es un requisito, la organización debería controlar y registrar la identificación única e inequívoca del producto y/ o servicio".

De manera que los estudios realizados por los métodos de la Organización del Trabajo, en los cuales se recolectan datos respecto de qué se hace, dónde, cómo y con qué, permite aportar trazabilidad a los productos. El problema entonces es determinar cómo se capturan esos datos. Los métodos manuales requieren de un proceso de digitalización posterior para que sean realmente útiles. Los computarizados frecuentemente funcionan con captura de datos a partir de teclados o pantallas táctiles integradas a terminales, generando una tarea "extra" al trabajador. La intención de este trabajo es ir más allá de las interfaces tradicionales.

Desde la perspectiva de las Ciencias de la Computación, este trabajo se vincula con Ingeniería de Software y en ésta, específicamente con la interacción Humano-Computadora.

Martig enuncia que *"La incorporación de nuevas tecnologías en los sistemas de control y los avances en las tecnologías de recolección y comunicación de datos han impactado en la forma en que los operadores interactúan con los sistemas."* … y luego indican que *"...En la actualidad hay una gran cantidad de software disponible que provee una manera cómoda y flexible de visualizar procesos industriales. La mayoría de estas herramientas se basan en una representación mímica del proceso sin aprovechar los avances realizados en las áreas de Visualización de Información y en Interacción Humano-Computadora…"* (Martig et al, 2004)

Un desprendimiento de la Interacción Humano-Computadora (HCI), el Diseño Centrado en el Usuario (DCU) plantea técnicas y herramientas que permiten integrar al usuario como parte del equipo de diseño de las interfaces para que éstas sean más usables. Marmolejo dice que *"Cuando se trata de la persona como centro de un proceso, es inevitable pensar en el análisis de las dimensiones que dan lugar a las características del individuo, como su conducta, modelos mentales o la cultura propia del contexto en que se ha formado"* (Marmolejo, 2019) para hacer que el diseño de las interfaces sea más usable por éstos (Nielsen, 2012). Uno de los aspectos a tener en cuenta es el Factor Humano de la Usabilidad, que se enfoca en entender el proceso que ocurre en las personas desde el momento en que un artefacto impacta en sus sentidos, hasta que el cerebro comprende lo que está ocurriendo y decide que hacer al respecto (Montero, 2009). El entender este hecho hizo que la interacción del usuario con los sistemas sea cada vez más fácil (o se perciba como tal).

En el último tiempo, el avance de la ciencia y la tecnología permitió ir más allá. Martig expresa que *"La comunidad en Interacción Humano Computadora (HCI) ha mostrado que las Interfaces de Usuario están trascendiendo los escritorios. Estas interfaces emergentes emplean técnicas novedosas para el ingreso de los datos, ejemplo de esto lo constituyen las interfaces hápticas, las tangibles, las basadas en cámaras, entre otras. Por otro lado la presentación de información se realiza a través de una variedad de dispositivos cada vez más amplia y tendientes a embeberse en el entorno del usuario."* (Martig et al, 2012). Turk desarrolla el tema de las interfaces perceptuales como otro mecanismo alternativo de interacción con el usuario (Turk, 2001). A estas nuevas formas de interactuar es la que denominamos *"Interfaces de Usuario No*

Convencionales". Desde este punto de vista es como se pretende intervenir en el proceso de captura de datos para trazabilidad y productividad. Se piensa que generando interfaces más transparentes al usuario, los datos pueden ser tomados con mayor exactitud y menor impacto en los procesos productivos.
Respecto del estado de la práctica, y sin haber profundizado en el tema porque es parte del objetivo de este proyecto, parece que en Argentina, puntualmente en la región noroeste de la provincia de Buenos Aires, podrían darse casos similares a los descriptos por (Sarache, 2005) en Colombia donde no se detectaron estrategias de mejora de la producción. No obstante, es necesario estudiarlo.

Líneas de Investigación, Desarrollo e Innovación

La realización del trabajo estará dividida en dos etapas, en las cuales se implementarán distintas estrategias de abordaje. Para la primera parte, donde se realizará un análisis del estado del arte en términos de sistemas de producción y calidad y en términos de interfaces de usuario de software para éstos sistemas (Kim, 2015), se seguirá un proceso sistemático y lineal. En cambio, para la segunda etapa, que implica la producción de nuevos diseños, sus implementaciones y validaciones, se conducirá un enfoque iterativo e incremental.
Como se mencionó anteriormente, en la primera etapa se realizará un análisis bibliográfico para establecer el estado del arte en materia de sistemas productivos, de calidad y las distintas formas de establecer trazas entre los productos. Esto permitirá tener una visión más amplia de los posibles requisitos que debería tener un sistema informático de control de gestión en estos aspectos. Se consultarán también expertos de campo para entender la viabilidad de las prácticas desarrolladas. Promediando la mitad del primer año ya estaremos en condiciones de comenzar el estudio de mediciones de las metodologías tradicionales. Estos datos serán resguardados para permitir las comparaciones futuras.
Se replicarán algunos de los experimentos planteados por otros investigadores para poder percibir su utilidad en aplicaciones de producción al tratar de identificar patrones de comportamiento de los usuarios para soportar luego herramientas automáticas de detección de estos patrones (Rauterberg, 1996). Con estos incentivos, se formularán pequeños objetivos de aplicación, se diseñarán e implementarán y evaluarán, con estos datos, se conducirán nuevos experimentos con nuevos objetivos, diseños, implementaciones y evaluaciones.
Estamos conscientes que no toda interfaz es apta para todos los procesos productivos. En algunos casos podría implementarse reconocimiento de imágenes con procesos de aprendizaje automático, en otros interfaces basadas en movimiento, en otros se podría utilizar redes neuronales profundas, y así, para cada caso particular se pueden establecer estrategias distintas. Este proyecto pretende obtener como resultado no solo la creación de prototipos funcionales de estas interfaces, sino obtener un conjunto de buenas prácticas de implementación en procesos productivos reales.

Hipótesis de Trabajo

Se partirá de la hipótesis de que a través de interfaces no convencionales, se pueden capturar datos de los procesos productivos para trazabilidad y control directo o por resultados, minimizando la posibilidad de cometer errores en la captura y, al mismo tiempo, reducir el impacto improductivo que conlleva el proceso.
Para comprobar la hipótesis, se establecerán o seleccionarán indicadores para medir el grado de incidencia de la metodología de captura de datos. Se

seleccionará una muestra de personas que utilicen sistemas tradicionales de captura de datos. Se medirán sobre esta muestra varias veces los tiempos utilizados para la captura. Luego se implementarán sistemas utilizando interfaces no tradicionales y se volverá a medir el tiempo que se utiliza en el mismo proceso. También será útil observar la calidad de los datos generados por los nuevos procesos de captura a la luz de cómo se utilizan estos datos en el sistema de calidad o de producción de la empresa. Para eso, se establecieron contactos con una empresa que prestó conformidad para realizar el estudio y solventar los costos que conlleve la compra de hardware específico para la construcción de las interfaces. Esto último es importante de mencionar porque sin contar con una empresa real, que permita el estudio de las operaciones, la apertura del sistema de productividad y de trazabilidad, y la instalación de las nuevas interfaces para probar efectividad sería imposible comprobar la hipótesis.

Resultados y Objetivos

Dada la situación planteada y observando el avance de las nuevas tecnologías, con la consecuente aparición de métodos inteligentes y automáticos de reconocimiento de personas, objetos y animales, parece razonable pensar que estas tecnologías puedan ser utilizadas para automatizar parte de las tareas de trazabilidad de la producción. En las últimas décadas, pero sobre todo en la última, se ha estado investigando mucho sobre distintos tipos de interfaces de interacción Humano-Computadora, en cómo debe ser esa interacción para que sea efectiva y también, en interfaces de software no tradicionales.

En este proyecto, se pretende utilizar el conocimiento que se ha desarrollado en los últimos años sobre reconocimiento de imágenes e Inteligencia Artificial e Interfaces de Software convencionales y no convencionales, para producir interfaces de usuario que viabilicen el disponer de sistemas de captura de datos que permita la trazabilidad, posibilitando el establecer sistemas de calidad efectivos en PyMEs. Además se pretende que al minimizar el tiempo improductivo del registro de las tareas, se maximice el tiempo productivo, haciendo más rentable la producción. Por otro lado, al disminuir el grado de interacción entre el empleado y el sistema, se minimice el grado de capacitación y/o habilidades digitales necesarias en el empleado y, al mismo tiempo aumente la calidad y exactitud del registro haciendo la captura de datos menos susceptible de errores, producto de errores humanos.

La realización de este proyecto de investigación permitirá evaluar formas alternativas de captura de datos más automáticas y con menos intervención del usuario, desarrollar prototipos de interfaces de usuario no convencionales utilizando nuevas tecnologías y evaluarlas en sistemas productivos reales, midiendo el impacto en términos de tiempo empleado por el usuario para la captura de datos. Permitirá también seleccionar la metodología de captura de datos más adecuada para cada parte del proceso, fundamentando esa selección con análisis basados en datos. Finalmente permitirá generar módulos de conexión para integrar estas interfaces con software de gestión de producción existente o a desarrollar.

Objetivos Específicos

- Relevar y analizar sistemas productivos en PyMEs.
- Seleccionar una métrica que permita evaluar objetivamente el grado de intromisión de la captura de datos en los procesos industriales de los sistemas informatizados.
- Identificar interfaces existentes de captura de datos productivos, evaluar su

desempeño y el grado de intromisión en la producción.
- Desarrollar interfaces de software que permitan efectuar la toma de datos y, de esta manera, viabilicen la trazabilidad y el cálculo de la productividad de los empleados, con la menor incidencia posible en el tiempo de producción y,
- Evaluar la mejora en la disminución de la incidencia de las interfaces diseñadas respecto de las identificadas y la efectividad en la captura de datos.

Formación de Recursos Humanos

El proyecto está integrado por docentes estudiantes. Se planifica la realización de dos tesinas de grado de estudiantes de la carrera Licenciatura en Sistemas de Información y la incorporación de docentes, auxiliares y estudiantes en actividades de investigación. Se espera también generar temas para dos tesis de maestrías para integrantes del grupo.

Referencias

Briceño, Blas (2013) Trazabilidad: el "pequeño inconveniente", En: http://www.alimentacion.enfasis.com/articulos/67479-trazabilidad-el-pequeno-inconveniente recuperado el 30-05-2019

Kim, G. (2015) Human-Computer Interaction: Fundamentals and Practice. Boca Raton: CRC Press.

Marmolejo, M. C. (2019) "Modelo de Integración Genérica para Diseño Centrado en el Usuario MIG-DCU" recuperado el 05-06-2019 en http://www.nosolousabilidad.com/articulos/modelo_integracion_generica.htm

Martig, S.; Castro, S and Larrea, M (2004), "Diseño de Interfaces Industriales", Proceedings of the IV Congreso Argentino de Enseñanza de la Ingeniería: 331-339, Instituto Tecnológico de Buenos Aires.

Martig, S.; Castro, S.; Larrea, M.; Escarza, S.; Escudero, M.; Ganuza, M.L.; Urribarri, D. Schneider , J.; Flores Choque, d. y Di Luca, S. (2012) "Interfaces No Convencionales. Su impacto en las interacciones", XIV Workshop de Investigadores en Ciencias de la Computación, Red de Universidades con Carreras en Informática (RedUNCI), p. 331-339

Montero Y.H. y Santamaría S.O. (2009) "Informe APEI sobre usabilidad" Gijón: Asociación Profesional de Especialistas en Información. Informe APEI 3

Nielsen J. (2012) Usability 101: Introduction to Usability. Recuperado el 5 de Mayo de 2019 de https://www.nngroup.com/articles/usability-101-introduction-to-usability/

Rojas Díaz (2014), "Estudio de Organización del Trabajo en la UEB de Estructuras Metálicas de METUNAS". Universidad "Vladimir Ilich Lenin" Las Tunas, Facultad de Ciencias Técnicas, Departamento de Ingeniería Industrial. Cuba. Enero de 2014

Rauterberg, M. & Steiger, P. (1996) Pattern Recognition as a Key Technology for the Next Generation of User Interfaces. 1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929, vol.4. pags 2805-2810), Beijing, China

Sarache, W., Cárdenas, D. M., (2005) Giraldo, J., "Procedimiento para la definición y jerarquización de prioridades competitivas de fabricación: aplicaciones en la industria metalmecánica", Revista Ingeniería y Competitividad., Vol.7, No.2, 2005, pp. 8491.

Turk M. (2001) Perceptual User Interfaces. En: Earnshaw R.A., Guedj R.A., Dam A.., Vince J.A. (eds) Frontiers of Human-Centered Computing, Online Communities and Virtual Environments. Springer, London.

Desarrollo, Evaluación y Puesta en Marcha de Herramientas para la Toma de Decisiones del Sistema Productivo Ganadero de la Pcia. de Santa Cruz

[1]Sandra Casas, [1,3]Daniela Ferrante, [2]Antonio Collazo, [1,3]Paola Vargas, [3,1]Ricardo Bonil, [3]Alan Schorr, [1]Marycarmen Díaz Labrador, y [3]Miguel Andrade

sicasas@conicet.gov.ar, ferrante.daniela@inta.gob.ar, acollazo@concet.gov.ar, vargas.paola@inta.gob.ar, bonil.ricardo@inta.gob.ar, mlabrador@unpa.edu.ar, schorr.alan@inta.gob.ar, andrade.miguel@inta.gob.ar

[1]*Universidad Nacional de la Patagonia Austral. Unidad Académica Río Gallegos. Piloto Lero Rivero S/N. CP 9400. Río Gallegos. Santa Cruz. Argentina Tel/Fax: +54-2966-448176.*

[2]*Centro de Investigación y Transferencias Sata Cruz (CONICET-UNPA-UTN FRSC). Lisandro de la Torre N° 860. CP 9400. Río Gallegos. Santa Cruz. Argentina Tel/Fax: +54-2966-442313/17.*

[3]*Instituto Nacional de Tecnología Agropecuaria. Estación Experimental Agropecuaria Santa Cruz. Mahatma Gandhi N° 1322. CP 9400. Río Gallegos. Santa Cruz. Argentina Tel/Fax: +54-2966-442305/06.*

RESUMEN

Este línea de I+D+i se origina a partir de la carencia de métodos para identificar los principales problemas de producción y cómo afecta los resultados económicos-financieros de los establecimientos ganaderos de Santa Cruz. Resulta muy difícil evaluar escenarios para mejorar el desempeño de la actividad. Resulta complejo caracterizar el sistema productivo actual con una cobertura territorial completa, lo que apareja un déficit en la definición e implementación de políticas públicas especificas. A partir de diversos trabajos técnicos, distintos actores de sector agropecuario, tanto del ámbito público como privado, de Santa Cruz, colaborativamente diseñaron un modelo matemático que define un conjunto de variables significativas cuya aplicación permitirá dar respuesta a las necesidades de toma de decisión a los distintos actores del sector. Este proyecto propone poner en marcha herramientas para la toma de decisiones del sistema productivo ganadero de Santa Cruz, que dará solución específica a los problemas de los productores y gestores públicos locales. El proyecto tiene por objeto el desarrollo de este sistema que completará la implementación e implantación del modelo, incluyendo la etapa de adopción de la misma por los actores del sector.

Palabras claves: Toma de Decisión, Simulador, Aplicación web.

CONTEXTO

El modelo matemático-económico que constituye la base de esta propuesta fue diseñado muy recientemente como una versión superadora y ampliada al utilizado en el informe técnico "Análisis Económico del Sistema Ovino Extensivo del Departamento Magallanes – Provincia de Santa Cruz - Región Agroecológica Meseta Central Santacruceña (Agosto 2015) [3]. En el mismo sentido, la puesta en marcha del sistema de soporte de decisión se ha presentado como la siguiente etapa del informe técnico "Análisis Económico de Establecimientos Modales Producción Ovina Extensiva Santa Cruz 2017 (Julio 2018) [4], y en colaboración y en articulación con los distintos técnicos y profesionales de las distintas Agencias de Extensión Rural de la Estación Experimental Santa Cruz del INTA, con la misma simbiosis, se trabajó con otras instituciones referentes del área, como el Consejo Agrario Pcial. (CAP), la Federación de Instituciones Agropecuarias de Santa Cruz (FIASC), entre otros, para poder realizar una mayor cobertura territorial y permitir una caracterización del sistema productivo actual.

Desde el punto de vista del software, existen algunos programas o simuladores económicos, que calculan y muestran resultados económicos primarios pero no representan la naturaleza del sistema ganadero ovino de Santa Cruz, se pueden mencionar "SSD Neuquén" y "Software para Evaluación Económica de Sistemas

El modelo matemático-económico para las distintas áreas productivas de Santa Cruz, que se utilizarán para dar sustento al sistema de decisión que este proyecto propone implementar, fue desarrollado a pedido de la FIASC. Este pedido surge como una necesidad de la federación de evaluar el nivel de adopción de tecnología por parte de las distintas unidades productivas, estadísticas de producción a nivel zonal, identificar principales problemas de producción: producción de lana, nivel de señalada, pérdida de animales, etc., y cómo esto afecta los resultados económico-financiero de los

establecimientos ganaderos, y con el objetivo de definir estrategias para la mejora del sistema productivo.
La cantidad de establecimientos que tienen producción ovina extensiva en Santa Cruz son 709, según el registro de Senasa del 2016 (lo más actualizado y oficial), en dicho número están incluidos los establecimientos que tiene producción ovina y producción mixta, es decir que incluye la explotación de ovinos.
El objetivo general planteado es brindar herramientas para la toma de decisiones del sistema productivo ganadero de la Pcia. de Santa Cruz. Para ello se consideran objetivos específicos: (i) Diseñar e implementar artefactos software que implanten el modelo matemático que representa el sistema productivo ganadero de Santa Cruz; (ii) -Desarrollar un simulador de escenarios productivos-económicos para la evaluación de decisiones (económicas-financieras-productivas); (iii) Evaluar la funcionalidad y usabilidad del sistema de decisión; y (iv) Realizar la transferencia de las herramientas al sector productivo.
La propuesta se ha presentado y aprobado como un proyecto de desarrollo tecnológico-social (PDTS) cuya ejecución inició el 01/12/2019 y es acreditado por UNPA y cofinanciado por UNPA e INTA. El adoptante es FIASC.

INTRODUCCION

La definición de un PDTS debe responder las consignas: problema, pertinencia y novedad. A continuación se detallan.

Problema: El tipo de producción ganadera, de la pcia de Santa Cruz, se diferencia del resto de Patagonia, principalmente por el tamaño de los campos que suelen tener más de 10.000has, aunque se manejen majadas reducidas, lo que plantea desafíos para afrontar los costos y lograr explotaciones rentables. A la vez, se carece de métodos para identificar los principales problemas de producción como, producción de lana, nivel de señalada, perdida de animales y cómo esto afecta los resultados económicos-financieros de los establecimientos ganaderos. Además resulta muy difícil evaluar escenarios al variar algunos índices por ejemplo para mejorar la finura de la lana, mejorar los niveles de señalada, cambio de categoría de animales vendidos, etc. que ayuden a mejorar el desempeño de la actividad. En consecuencia resulta complejo caracterizar el sistema productivo actual con una cobertura territorial completa, lo que conlleva un déficit en la definición e implementación de políticas públicas especificas para el sector como ser los instrumentos de emergencia agropecuaria en periodos extremadamente secos, o por incidencia de disturbios tales como erupciones volcánicas o incendios, que por lo general no puede evaluar el impacto a nivel productivo. Para revertir esta situación, distintos actores del sector (INTA, CAPS, FIAS, etc.) de Santa Cruz, colaborativamente diseñaron un modelo matemático que a partir de un conjunto de variables significativas permitirá dar respuesta a las necesidades de toma de decisión a los distintos actores del sector.
Relevancia: La implantación del sistema de decisión, representa contar con un modelo interactivo para la representación e interpretación de las principales variables productivas y económicas de la producción ovina extensiva en la provincia de Santa Cruz, así permitirá a los diferentes actores (productores y sectores de gobierno) del sector disponer de información de los niveles de producción y de ingresos. La generación y disposición de los datos de los distintos resultados productivos de los establecimientos o unidades de explotación ganadera, permitirá trazar lineamientos, políticas, estrategias de intervención y/o acciones oportunas y más acertadas frente a los distintos fenómenos climáticos, macroeconómicos y de la región que condicionan al sistema ganadero, en especial al ovino extensivo. Donde los distintos actores podrán identificar, describir y hacer una mejor interpretación de la situación en la que se encuentra la actividad en la provincia de Santa Cruz.
Cómo ya se indicó, actualmente no se cuenta en la pcia. con una herramienta que cumpla estas funciones, sino que además no se aplica ningún método que brinde y/o genere información sistematizada de los resultados productivos de los establecimientos ganaderos ovinos de la provincia de Santa Cruz. Este proyecto colabora directamente con el Plan Ganadero 2020-2030 [5], elaborado por Instituto para la Promoción de la Ganadería creado por Ley N° 3144 integrado por CPA, INTA, Federación Lanera, Cámara de Frigoríficos, Senasa. Este plan busca elevar en un 30% la producción de cantidad y calidad ovina y bovina de Sta Cruz

Pertinencia: La propuesta del proyecto se basa en el desarrollo, evaluación y puesta en marcha de un sistema de soporte de decisión para el sistema productivo ganadero de la Pcia de Santa Cruz, como método para la mejora productiva y por ende económica de este sector. La capacidad para resolver los problemas identificados está basado en el modelo que implementará, como ya se ha indicado, ha sido diseñado en forma colaborativa por los distintos actores del sector, y ha sido identificado como una solución en informes técnicos elaborados recientemente. Este modelo esta compuesto por variables de decisión, parámetros de configuración, procesos de cálculo y variables de resultado, que reflejan lo más cercanamente posible la realidad del ciclo productivo-económico de los establecimientos, de modo que las variables de resultado tengan una interpretación que agregue valor a la toma de decisiones. Por otro lado ofrecerá varias

herramientas, como el simulador, la generación de información cuantitativa-cualitativa, etc. para su aplicación en distintos ámbitos

Novedad: Como se ha señalado, el modelo matemático – económico de reciente definición (2018) representa e interpreta las principales variables productivas y económicas de la producción ovina extensiva en la provincia de Santa Cruz, en consecuencia actualmente no existe ningún tipo de implementación del mismo. Así, el sistema que se propone permitirá a los diferentes actores analizar los niveles de producción y de ingresos, entre otros factores. La implantación de este modelo como sistema de soporte de decisión además constituye un método específico para los actores locales (productores, sectores de gobierno, etc.). Es importante destacar que luego de la implantación del modelo, se comenzará a generar y obtener información que actualmente no se dispone, que dará lugar no solo a la propuesta de políticas públicas para este sector, sino también información para la propuesta de nuevos proyectos de I+D, productivos, planes de desarrollo/trabajo/negocio, etc.

2. LINEAS DE INVESTIGACION

Esta propuesta, por tratarse de un PDTS, claramente refiere a una investigación aplicada de I+D+i. Avances en el campo de sistemas soporte de decisiones, sistemas de recomendación, simuladores, y desarrollo de aplicaciones web serán aplicados al igual que diversos métodos de evaluación de calidad, específicamente usabilidad.

3. RESULTADOS ESPERADOS

El proyecto propone el diseño, implementación y puesta en marcha de un sistema de soporte de decisión [1-2] destinado a los distintos actores del sistema productivo ganadero de Santa Cruz. La herramienta constará de diversos componentes:
(i) implantación de un modelo matemático que representa las variables que afectan críticamente al sistema productivo ganadero de Santa Cruz;
(ii) un simulador de escenarios para la evaluación del resultado económico de los establecimientos ganaderos en función de las principales variables que caracterizan el sistema productivo;
(iii) identificación y determinación de los valores óptimos de ciertas variables/indicadores;
(iv) generación de una base de datos para el análisis de información cuantitativa y cualitativa territorial y definición de políticas públicas para el sector.
A través de esta herramienta, será posible simular escenarios y determinar el conjunto de decisiones necesarias para llevar a los establecimientos productivos de una situación de negocio a otra. Estas capacidades se convierten en un soporte importante para los propietarios y/o administradores del Sistema productivo ganadero de Santa Cruz

Resultados Preliminares

- Documento de especificación de parámetros de entrada/salida
- Definición de stackholders
- Configuración de servidor, definición de herramientas de desarrollo.
- Prototipación de la Interfaz (Fig. 1 a 4)



Figura 1. Acceso al sistema



Figura 2. Carga y selección de productores



Figura 3. Carga de parámetros de entrada



Figura 4. Pantalla de visualización de resultados

4. FORMACION DE RECURSOS HUMANOS

Participan de este proyecto un tesista doctoral, becario doctoral CONICET y una ayudante de docencia universitaria.

REFERENCIAS

[1] Bhargavaa H., Power D., Sunc D., "Progress in Web-based decision support technologies", Decision Support Systems, (2007), 43(4), 1083-1095-https://doi.org/10.1016/j.dss.2005.07.002

[2] Power D., Kaparthi S., "Building Web-based decision support systems", Studies in Informatics and Control, (2002), 11(4), 291-302

[3] Análisis Económico del Sistema Ovino Extensivo del Departamento Magallanes–Provincia de Santa Cruz-Región Agroecológica Meseta Central Santacruceña-(Agosto-2015)

[4] Análisis Económico de Establecimientos modales Producción Ovina Extensiva Santa Cruz 2017(Julio-2018)

[5] Plan Ganadero Ovino de la Provincia de Santa Cruz, CAP- FIAS – INTA – SENASA. Comisión Redactora: Alan Schorr, Jorge Santana y Miguel Andrade.(2016)

Computación Afectiva aplicada a la valoración emocional en contextos gastronómicos.

Jorge Ierache, Iris Sattolo, Gabriela Chapperón, Rocío Ierache, Facundo Nervo, Fernando Elkfury, Ramiro Nicolosi, Gonzalo Ponce.

Instituto de Sistemas Inteligentes y Enseñanza Experimental de la Robótica (ISIER)
Secretaria de Ciencia y Tecnología -Universidad de Morón

Cabildo 134, Buenos Aires, Argentina

jierache@unimoron.edu.ar

Resumen

Este proyecto de investigación aplicada se orienta a la captura de estados emocionales de personas interactuando en contextos gastronómicos, a través de un Framework Multimodal, con la capacidad de registrar datos tanto fisiológicos como biométricos, contribuye a reflejar el estado emocional con una mayor fidelidad que la obtenida cuando se utilizan pocos medios de recolección de datos o encuestas. Se pretende obtener información en un contexto multimodal, a fin de valorar el grado de placer-rechazo, e intensidad que se producen en una persona frente a diferentes comidas o alimentos.

Palabras clave: Computación Afectiva Alimentación, Datos Biométricos, Datos Fisiológicos, Framework Multimodal, Neuromarketing.

Contexto

Este proyecto de investigación aplicada se presentó en la convocatoria PIO-UM, se radica en el Instituto de Sistemas Inteligentes y Enseñanza Experimental de la Robótica ISIER UM y capitaliza las bases iniciales del proyecto PING/17-03-JI-002 UM. Existen conjuntos de imágenes de alimentos estandarizados con sus evaluaciones subjetivas. La base de imágenes de OLAF (Open Library of Affective Food) tiene el propósito específico de estudiar las emociones hacia la comida. Se ha argumentado que las evaluaciones de alimentos no son válidas entre individuos y grupos, a menos que los sentimientos hacia las señales alimentarias se comparen con los sentimientos hacia experiencias intensas no relacionadas con la comida, que sirven como puntos de referencia. Este proyecto continua las líneas I+D del ISIER-UM; particularmente en el contexto de la computación cognitiva y afectiva, en este orden se desarrollaron investigaciones con el empleo de interface cerebro-máquina (BCI) aplicada a la domótica y a la robótica[1][2][3], influencia en el estado biométrico emocional de personas, explotación de datos electroencefalograma (EEG) y parámetros fisiológicos de usuarios interactuando en contextos virtuales[4] [5], y valoración del grado de atención de alumnos en contextos áulicos[6].

Introducción

Rosalind Picard define a la computación afectiva como “la informática que se relaciona con las emociones, no sólo con

las consideradas más importantes, como la alegría o la tristeza, sino también con el interés, el aburrimiento o la frustración, que son las que se dan en relación con los ordenadores" [7]. Sandra Baldasarri plantea que "los sistemas "afectivos", deben ser capaces de: a) capturar y reconocer los estados emocionales del usuario a través de mediciones sobre señales generadas en la cara, la voz, el cuerpo, o cualquier otro reflejo del proceso emocional que se esté llevando a cabo; b) procesar esa información clasificando, gestionando, y aprendiendo por medio de algoritmos que se encargan de recoger y comparar gran cantidad de casos, y que tienen en cuenta los estados emocionales del usuario y, en su caso, del ordenador; y, por último, c) generar las respuestas y las emociones correspondientes, que pueden expresarse a través de diferentes canales: colores, sonidos, robots, o personajes virtuales dotados de expresiones faciales, gestos, voz, etc." [8] De acuerdo con los tipos de datos que se utilizan en los sistemas propuestos, pueden clasificarse en sistemas unimodales -los que exploran una sola fuente de datos-, y multimodal -los que combinan dos o más fuentes de datos-. Para deducir el estado emocional de un individuo en un contexto multimodal se tiene que registrar, simultáneamente, diversa información biométrica.

Como trabajos unimodales se pueden citar como ejemplo a: captura de rostro con cámaras o a través de videos [9],[10], [11]; captura de electrocardiograma (ECG) [12][13][14][15]; captura de cadencia de tecleo [16]; captura de rostro con posición de la cabeza [17]; captura de movimientos de mouse, tecleo y micrófono [18][19].

Como trabajos multimodales que integran información fisiológica de diferentes fuentes se pueden citar a [20],[21].

En el uso de estados afectivos utilizando enfoques multimodales, existen distintos desafíos, como lo son: (a) decidir qué modalidades combinar; (b) recopilar datos de entrenamiento; (c) manejo de datos faltantes, manejo de diferentes tasas de muestreo e interdependencia de modalidades al construir modelos; (d) decidir cómo fusionar datos de diferentes modalidades; y (e) decidir cómo evaluar los estados emocionales. Particularmente en las líneas I+D de computación afectiva trabajamos con: EEG-BCI pasivos -los que recolectan información del estado del usuario (a partir de biomarcadores) con el objeto de evaluar/mejorar la interacción entre el hombre y el ambiente virtual, real o mixto-, sensores fisiológicos (ECG-Ritmo Cardíaco, Conductancia), Imágenes reales y de base de datos – IAPS [22] [23], OLAF [24], [25], encuestas SAM (*por sus siglas en inglés Self-Assessment Manikin*) [26], y el "circunflex model" de Rusell [27].

Se reconoce ampliamente que la expresión afectiva humana consiste en una coordinación compleja de señales. Se espera que analizar múltiples señales y su interdependencia mutua produzca modelos que reflejen con mayor precisión la naturaleza subyacente de la expresión afectiva humana. Los enfoques multimodales contribuyen a mejorar la capacidad de detección de emociones, al basar sus decisiones en los distintos

canales disponibles.

Líneas de Investigación, Desarrollo e Innovación

Las líneas de I+D se centran en un framework multimodal para la captura de estados emocionales, con capacidad de registrar datos biométricos y fisiológicos de personas interactuando en contextos gastronómicos, a fin de facilitar la explotación de información, para la integración y desarrollo de aplicaciones en el contexto de la computación afectiva.

Resultados y Objetivos

Se espera alcanzar una valoración del grado de excitación-valencia (EEG), inferencia de estado emocional en función de la captura de rostro, y valoración SAM en contextos virtuales empleando imágenes del banco OLAF, imágenes propietarias, imágenes IAPS, en contexto reales de degustación empleando platos preparados en conjunto con imágenes (IAPS) previas y posteriores a la degustación. Sintéticamente esperamos ver el grado de placer-rechazo que produce una comida, en particular, en una persona.

En trabajos contribuyentes se desarrollaron frameworks con aplicación a contextos virtuales [5] y de educación [6], estas experiencias permiten plantear para este dominio gastronómico los siguientes objetivos específicos:

- Integrar interfaces BCI
- Integrar tracking de ojos
- Integrar sensores fisiológicos Variación de Ritmo Cardíaco (HRV) y Respuesta Galvánica de la Piel (GSR).
- Integrar imágenes del contexto gastronómico, imagen del rostro del sujeto de estudio, y datos de inferencia de estado emocional.
- Integrar test de Rusell y encuesta SAM
- Capturar ambientes (imágenes OLAF, imágenes propietarias, o situaciones reales de degustación)
- Integrar herramientas de explotación de datos y descubrimiento de datos.

1. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto cuenta con un investigador formado, un investigador en formación vincula una doctoranda en Ciencias Informáticas, una maestrada en Lenguajes y Sistemas Informáticos y una maestranda en marketing estratégico digital, y la participación de un Chef profesional Técnica superior en gastronomía. Como así también suma tres estudiantes de grado en el marco de las líneas de investigación del área de computación afectiva

Referencias

[1] Ierache, J., Pereira, G., Iribarren, J., & Sattolo , I. (2012). Robot Control on the Basis of Bioelectrical Signals. En Robot Intelligence Technology and Applications 2012 (pp. 337-346). Korea: Springer

[2] Ierache, J., Pereira, G., Sattolo, I., Guerrero, A., DÁlotto, J., & Iribarren, J. (2011). Control vía Internet de un Robot ubicado en un sitio remoto aplicando una Interfase Cerebro-Máquina. XVII CACIC 2011 (págs. 1373-1382). La Plata: REDUnci

[3] Ierache, J., Nervo, F., Pereira, G., & Iribarren, J. (2014). Estado Emocional Centrado en Estímulos,Aplicando Interfase Cerebro-Maquina. XX CACIC, 2014. Bs.As.: REDUnci

[4] Ierache, J., Cervino, C., Eszter , E., Fortin , D., & Castro Menna, A. (2018). Influencias del estado biométrico- emocional de personas interactuando en contextos de entornos virtuales. WICC, pp 785-789). Corrientes

[5] Ierache, J., Nicolosi, R., Ponce, G., Cervino, C., & Eszter, E. (2018). Registro emocional de personas interactuando en contextos de entornos virtuales. XXIV CACIC 2018, (págs. 877-886). i

[6] Ierache, J., Ponce, G., Nicolosi, R., Sattolo, I., & Chapperón, G. (2019). Valoración del grado de atención en contextos áulicos con el empleo de interfase cerebro-computadora. CACIC 2019, en prensa

[7] Picard, R. (2000). Affective Computing. En T. M. press, Affective Computing (págs. 4-8). Cambridge Massachusetts: The Mit Press

[8] Baldasarri, S. S. (15 de 9 de 2016). Computación afectiva: tecnología y emociones para mejorar la experiencia del usuario. Obtenido de http://hdl.handle.net/10915/53441

[9] van der Haar, D. T. (2019). Student Emotion Recognition. International Conference on Human-Computer interaction (págs. 301-311). Springer Chan

[10] Zatarain Cabada, R., Barron Estrada, M. L., Halor-Hernandez, G., & Reyes-García, C. A. (2014). Emotion Recognition in Intelligent Tutoring Systems. Mexican International Conference on Artifitial Intelligence (págs. 494-504).
Mexico: Springer

[11] Wei-Long, Z., & Bao-Liang, L. (2015). Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. En I. X. Library, IEEE Transactions on Autonomous Mental Development, vol. 7, no. 3, 9/ 2015 (págs. 162-175). IEEE

[12] Agriafioti, F., Hatzinakos, D., & Anderson,, K. (2012). «ECG Pattern Analysis for emotion detection. Transactions of affective computing (págs. 102-115). IEEE

[13] Bong, S., Murugappan, M., & Yaacob, S. (2012). Analysis of Electrocardiogram (ECG) Signals for Human Emotional Stress Classification. International Conference on Intelligent Robotics, Automation, and Manufacturing (págs. 198-205).
Springer

[14] Rodriguez Aguiñaga, A., López Ramirez, M., & Baltazar Flores, M. (2015). Classification model of arousal and valence mental states by EEG signals analysis and Brodmann correlations. International Journal of Advanced Computer Science and Applications, (págs. 1-9 volumen 6 N| 6). SAI

[15] Ali, M., MachoT, F., Mosa, A., & Kyamakya, K. (2016). A novel EEG-based emotion recognition approach for e- healthcare applications. 31st Annual ACM Symposium on Applied Computing (págs. 162-164). Pisa Italia: ACM.

[16] CALOT, E., IERACHE, J., HASPERUÉ, W.. 2019. Robustness of keystroke dynamics identification algorithms against brain-wave variations associated with emotional variations. En Advances in Intelligent Systems and Computing. Springer, c. En prensa

[17] Xu, R., Chen, J., Han, J., Tan, L., & Xu, L. (2019). Towards emotion-sensitive learning cognitive state analysis of big data in education deep learning-based. En Computing (págs. 1-16). Austria: Springer Viena.

[18] Salmeron-Majadas, S., Santos, O., & Boticario, J. (2014). An evaluation of mouse and keyboard interaction indicators towards non-intrusive and lowcost affective modeling in an educational context. 18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

[19] Calot, E., Ierache, J., & Hasperué, W. (2019). Robustness of keystroke dynamics identification algorithms against brain-wave variations associated with emotional variations. En Intelligent Systems and applications (págs. 194-211). Londres: Springer Cham.

[20] Wiew, M., & Lachiri, Z. (2017). Emotion Classificationin Arousal Valence Model using MAHNOB-HCI Database. nternational Journal of Advanced Computer Science and Applications, 1-6.

[21] Calot, E., & Ierache, J. (2017). Multimodal biometric recording architecture for the exploitation of applications in the context of affective computing. XXIII CACIC 1030-1039 La Plata REDUnci.

[22] Lang, P.J., Bradley, M.M., & Cuthbert, B.N. International Affective Picture System (IAPS): Technical Manual and Affective Ratings. NIMH ,1997

[23] Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). Affective ratings of pictures an instruction manual. International affective picture system (IAPS) Technical Report A-8. Gainesville: University of Florida.

[24] OLAF, the Open Library of Affective Foods in ADULTS: http://digibug.ugr.es/handle/10481/41499 Vigente: 29/03/2020

[25] OLAF, Open Library of Affective Foods: A tool to investigate the emotional impact of food in adolescents https://zenodo.org/record/10202#.XoC024hKiUl Vigente: 29/03/2020

[26] Lang, P. J. The cognitive psychophysiology of emotion: Fear and anxiety. In A. H. Tuma & J. D. Maser (Eds.), Anxiety and the anxiety disorders (pp.131-170). Hillsdale, NJ,US. (1985).

[27] Posner, J., Russell , J., & Peterson, B. (2005). “The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology.”, Development and psychopathology, vol. 17, n° 3, págs. 715-34, 2005, ISSN: 0954-5794. Development and psychopathology, 715-34 Vol 17 n° 3.

Integración de una BCI para rehabilitación de problemas cognitivos en pacientes neurológicos infantiles implementando VRPN como protocolo de comunicaciones

Javier J. Rosenstein, Rodrigo Gonzalez, Nicolás Ignacio Zárate Alvarez, Matías Campos

Instituto de Investigaciones, Facultad de Informática y Diseño, Universidad Champagnat, Belgrano 721, 5501 Godoy Cruz, Mendoza, Mendoza, Argentina.

rosensteinjavier@uch.edu.ar, gonzalezrodrigo@uch.edu.ar, nicolaszarate23@gmail.com, mgcampos9785@gmail.com

RESUMEN

En el desarrollo de sistemas de realidad virtual uno de los inconvenientes que se encuentran es la comunicación entre las aplicaciones y los dispositivos de adquisición de ondas cerebrales. Ya sea por no disponer de un método de acceso en forma directa o por necesitar independencia entre ambos, es decir que las aplicaciones corran en una plataforma y los dispositivos en otras. Para lograr esta independencia y a su vez permitir la integración de todo el sistema de realidad virtual, es necesario la implementación de algún protocolo de comunicaciones que permita esta vinculación heterogénea en tiempo real. Los dispositivos generalmente están asociados a funciones o características de los individuos que los utilizan y se necesita integrar los movimientos que estos representan a la aplicación de realidad virtual correspondiente. El presente trabajo trata del análisis e implementación del protocolo de comunicaciones VRPN (Virtual Reality Protocol Network) entre las partes de un entorno multimedia donde interactúan la adquisición de movimientos del usuario y la representación visual en un escenario virtual que permita la retroalimentación al usuario en tiempo real logrando una experiencia interactiva e inmersiva. Esto tiene aplicación directa en los tratamientos de rehabilitación en pacientes de patologías neurológicas y cognitivas, principalmente en niños, este enfoque es importante ya que está demostrado que es la mayor causa de discapacidad en niños, según estadísticas como las publicadas por el Indec de su último informe del 2018 (ver figura1), en argentina dentro de las principales discapacidades, la mental-cognitiva es la mayor de las discapacidades entre las edades de 6 a 14 años.

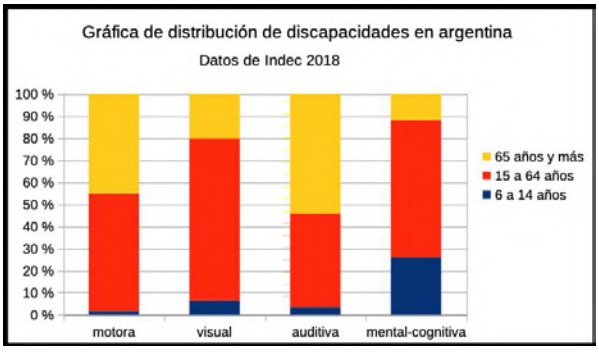


Figura 1: Discapacidades según Indec en argentina

Palabras clave: VRPN, EOG, EEG, BCI, Serious Games, Mirror Neurons, Neurofeedback

CONTEXTO

El presente proyecto forma parte del trabajo de tesis de Maestría en Teleinformática de la Universidad de Mendoza (Ciudad, Mendoza), correspondiente al tesista Javier Rosenstein, el mismo se desarrolla en el Instituto de Investigaciones de la Facultad de Informática y Diseño de la Universidad Champagnat (Godoy Cruz, Mendoza), en el marco de la Licenciatura en Sistemas de Información; en cooperación con el Laboratorio de I+D+i en Neurotecnologías de la empresa Neuromed Argentina S.A. (Godoy Cruz, Mendoza).
Este trabajo es parte del proyecto de investigación que dio inicio en Diciembre de 2017 denominado "Diseño y desarrollo de un prototipo de Serious Game destinado a la rehabilitación de problemas neurológicos implementando VRPN para la comunicación de la BCI".

1. INTRODUCCIÓN

El objetivo principal de la presente línea de investigación consiste en el diseño y desarrollo de una BCI (Brain Computer Interface) [1] [2], que permita interactuar entre las señales generadas por un paciente neurológico [3] y una interfaz de realidad virtual. Este sistema debe permitir lograr el principio de neurofeedback [4], o retroalimentación hacia el paciente. De este modo se podrán mejorar sus capacidades cognitivas correspondientes. Incluso, estudios indican la posibilidad de tratar patologías psiquiátricas como la depresión [5] [6].

La implementación del trabajo se organiza de acuerdo a las siguientes etapas:

a) Adquisición de señales mediante técnicas de Electroencefalografía (EEG) [7] y adquisición de movimientos oculares mediante las técnicas de electrooculografía (EOG) [8] [9] [10] [11] [12].
b) Análisis de estas señales en tiempo real para poder identificar la voluntad de movimiento del individuo, así como la dirección del movimiento.
c) Transformar la voluntad de movimiento en comandos del protocolo VRPN (Virtual Reality Protocol Network) que permitan transmitir la información al componente software / hardware que la requiera.
d) Como continuación al resultado obtenido en el punto anterior, poder comandar una interfaz gráfica de aprendizaje o interfaz Cerebro / Computadora (BCI – Brain Computer Interface).

Finalmente se presenta la BCI como un sistema de adquisición de datos, procesamiento del protocolo serie a VRPN y luego la representación en nuestro modelo de prototipo de Serious Game [14] [15]. Una vez adquiridas las señales EEG/EOG, estas se analizan y codifican con las bibliotecas desarrolladas como parte de este proyecto. Esto permite su comunicación mediante VRPN con las interfaces virtuales que interpretan este protocolo.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Este trabajo está compuesto por tres etapas o fases de trabajo. La primera de ellas corresponde a la implementación de VRPN entre un sistema simulado de captura de datos provenientes de un paciente, y una interfaz virtual básica. La segunda etapa del proyecto pretende avanzar sobre la captura de datos reales para que, luego de procesados, se transfieran a través de VRPN hacia la interfaz virtual de neurofeedback. Una breve descripción de cada una de las etapas se describe a continuación:

1. La primer parte consiste en la implementación de un simulador de señales EEG y EOG necesarias para el

análisis e interpretación de la voluntad del usuario de la BCI. Estas señales una vez procesadas se deben codificar en comandos de VRPN para poder ser transmitidas hacia una interfaz virtual. Esta debe poder interpretar las señales transmitidas y representar la voluntad inicial del usuario correspondiente. De este modo se cumplen los objetivos de captura, análisis, procesamiento, transmisión, recepción y representación, lo cual produce el efecto de neurofeedback deseado sobre una interfaz virtual de capacitación a nivel prototipo.

2. La segunda etapa consiste en el desarrollo y mejora del proyecto mediante la adquisición real de señales EEG y EOG por electrodos ubicados superficialmente sobre la cabeza del paciente. Esta modificación al sistema requiere de un diseño e implementación electrónica así como del desarrollo del firmware que permita adquirir, analizar y procesar estas señales, que luego serán transmitidas a través del protocolo VRPN hacia una interfaz virtual. Así, el sistema diseñado se incorporará en forma transparente al proyecto implementado en la primer etapa.
3. La etapa final de este proyecto consiste en el diseño y desarrollo del escenario virtual de rehabilitación cognitiva para pacientes neurológicos según las indicaciones concretas por parte del especialista en neurología. Se partirá de un relevamiento de las técnicas de aprendizaje que se requieren implementar y los resultados que se pretenden obtener, indicados por el neurólogo o experto afín.

3. RESULTADOS ESPERADOS

Los resultados esperados se pueden dividir en dos grandes grupos:

1. BCI e interfaz virtual de neurofeedback: se espera obtener como producto final un sistema de retroalimentación a un supuesto usuario simulado por la generación de señales que comandan el escenario de realidad virtual via comandos codificados en el protocolo VRPN que se transmiten a través de una red Ethernet.

2. Adquisición de datos reales de EEG y EOG: análisis y preprocesamiento; se identifican los comandos necesarios para su transmisión hacia el equipo generador de tramas VRPN para que finalmente se tenga el proyecto totalmente terminado y funcional que implemente el efecto de neurofeedback propuesto.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de I+D presentada está vinculada con el desarrollo de una tesis de postgrado, por parte del estudiante de maestría en teleinformática de la Universidad de Mendoza, Javier J. Rosenstein. Dicha tesis se centra en la implementación del protocolo VRPN demostrando su uso en un sistema BCI.

Además cuenta con la dirección del Lic. Javier J. Rosenstein (UCH) y como investigador principal el Dr. Rodrigo Gonzalez (UCH). En lo que respecta a la formación de estudiantes de la licenciatura, esta línea de investigación cuenta con dos tesis de grado en curso, pertenecientes a los estudiantes Nicolás Ignacio Zárate Alvarez y Matías Campos, cuyos planes de tesis se encuentran específicamente dentro del marco de este proyecto. Ambos cursan la licenciatura en Sistemas de Información en la Universidad Champagnat.

Una vez que el proyecto se encuentre implementado desde su primer etapa será utilizado como recurso para el dictado de talleres de comunicaciones, redes, programación de microcontroladores, Programación en C/C++ y Diseño y programación de interfaces virtuales de capacitación/rehabilitación en general, tanto para estudiantes de la universidad, como así

también para aquellos interesados externos a la UCH.

5. BIBLIOGRAFÍA

1. Y. Wang, X. Gao, B. Hong, and S. Gao, "Practical designs of brain-computer interfaces based on the modulation of EEG rhythms," in Brain-Computer Interfaces. Springer, 2009, pp. 137–154.
2. J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," Clinical neurophysiology, vol. 113, no. 6, pp. 767–791, 2002.
3. J. A. Pineda, "The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing"," Brain Research Reviews, vol. 50, no. 1, pp. 57–68, 2005.
4. S. Enriquez-Geppert, R. J. Huster, and C. S. Herrmann, "Boosting brain functions: Improving executive functions with behavioral training, neurostimulation, and neurofeedback," International Journal of Psychophysiology, vol. 88, no. 1, pp. 1–16, 2013.
5. R. Ramirez, M. Palencia-Lefler, S. Giraldo, and Z. Vamvakousis, "Musical neurofeedback for treating depression in elderly people." Frontiers in neuroscience, vol. 9, pp. 354–354, 2014.
6. W. Rief, "Getting started with neurofeedback," 2006.
7. J. D. Kropotov, Quantitative EEG, event-related potentials and neurotherapy. Academic Press, 2010.
8. A. Bulling, J. A. Ward, H. Gellersen, and G. Troster, "Eye movement analysis for activity recognition using electrooculography," IEEE Transactions on pattern analysis and machine intelligence, vol. 33, no. 4, pp. 741–753, 2011.
9. H. Singh and J. Singh, "A review on electrooculography," International Journal of Advanced Engineering Technology, vol. 3, no. 4, pp. 115–122, 2012.
10. D. P. Bautista, I. A. Badillo, D. De la Rosa Mejía, and A. H. H. Jiménez, "Interfaz humano-computadora basada en señales de electrooculografía para personas con discapacidad motriz," ReCIBE, vol. 3, no. 2, 2016.
11. S. Yathunanthan, L. Chandrasena, A. Umakanthan, V. Vasuki, and S. Munasinghe, "Controlling a wheelchair by use of EOG signal," in 2008 4th International Conference on Information and Automation for Sustainability, IEEE, 2008, pp. 283–288.
12. V. C. C. Roza, "Interface para tecnologia assistiva baseada em eletrooculografia," 2014.
13. A. C. Gaviria, I. C. Miller, S. O. Medina, and D. R. Gonzales, "Implementación de una interfaz hombre-computador basada en registros EOG mediante circuitos de señal mixta PSoC," in V Latin American Congress on Biomedical Engineering CLAIB 2011 May 16-21, 2011, Habana, Cuba. Springer, 2013, pp. 1194–1197.
14. P. Rego, P. M. Moreira, and L. P. Reis, "Serious games for rehabilitation: A survey and a classification towards a taxonomy," in 5th Iberian Conference on Information Systems and Technologies. IEEE, 2010, pp. 1–6.
15. J. S. Breuer and G. Bente, "Why so serious? On the relation of serious games and learning," Eludamos. Journal for Computer Game Culture, vol. 4, no. 1, pp. 7–24, 2010.

Un marco de trabajo para el desarrollo de software de dominio específico en el contexto de Gobierno Digital

Mauro Cambarieri[1], Luis Vivas[1], Nicolás Garcia Martinez[1], Gustavo Guimerans[2], Alejandra Viadana[1]

[1] Universidad Nacional de Rio Negro. Sede Atlántica
Laboratorio de Informática Aplicada
{mcambarieri,lvivas,ngarciam}@unrn.edu.ar, alejandraviadana@gmail.com

[2] Centro de Ensayos de Software, Montevideo, Uruguay.
gusguime@ces.com.uy

RESUMEN

Los gobiernos desarrollan iniciativas de Gobierno Digital con el objetivo de buscar técnicas, herramientas, enfoques y/o estrategias digitales que ayuden a resolver los problemas que se le presentan. Desde la perspectiva de una institución, en el contexto de Gobierno Digital, es deseable lograr la reutilización de software de manera sistemática con el fin de lograr beneficios asociados a utilizar los artefactos construidos previamente en cada desarrollo nuevo que se realiza. Para que la reutilización del software sea sistemática, los procesos de desarrollo deberían abordar la construcción colectiva de familias de productos relacionados con un dominio [21]. Por otro lado, para estar en consonancia con el modelo y los procesos de negocio, se debe considerar un enfoque de desarrollo que se centre en un dominio de negocio específico. Es por ello, que es evidente en el corto y mediano plazo, la necesidad de un cambio revolucionario en la forma de escribir el software. La modernización puede ser tanto estratégica como técnica y las instituciones deben estar preparadas cuando se presente la oportunidad. La modernización se trata principalmente de una reingeniería para aprovechar los beneficios de las arquitecturas y plataformas modernas. El objetivo de la línea de investigación que se presenta en este trabajo es diseñar un marco de trabajo (*framework*) que permita el desarrollo de software de dominio específico para Gobierno Digital. El mismo consiste en investigar y desarrollar herramientas informáticas para facilitar el rápido desarrollo de aplicaciones en Gobierno Digital.

Palabras clave: *Domain Driven Design*, Línea de Productos de Software, *Model-Driven Engineering*, Framework, Gobierno Digital.

CONTEXTO

El presente trabajo se enmarca en el proyecto de investigación "Herramientas Informáticas para el Desarrollo de Servicios Digitales Innovadores para Comunidades Urbanas y Rurales en el Marco de Ciudades y Regiones Inteligentes" desarrollado en el Laboratorio de Informática Aplicada, Sede Atlántica, Universidad Nacional del Río Negro (UNRN). La propuesta de investigación se encuentra en proceso de evaluación en la "Convocatoria Ideas Proyectos de Desarrollo y Transferencia de Tecnología (IP-DTT)" llevada a adelante por la Secretaría de Investigación, Creación Artística, Desarrollo y Transferencia de Tecnología de la UNRN, para obtener financiamiento. El proyecto tiene una duración de 18 meses y los resultados serán transferidos a la Subsecretaría de Innovación, Desarrollo Sostenible y Ambiente de la Municipalidad de Viedma.

1. INTRODUCCIÓN

En 1968, en el Congreso sobre Ingeniería del Software que organizó el Comité de Ciencia de la OTAN, se introdujo por primera vez la idea de la reutilización. En ese Congreso el Ing. M.D. McIlroy, de Bell Laboratories, afirmó que "La industria del software se asienta sobre una base débil, y un aspecto importante de esa debilidad es la ausencia de una subindustria de componentes" [1]. En Ingeniería de Software, se entiende por componente, un fragmento reemplazable de un sistema, dentro del cual se encuentran implementadas un conjunto de funcionalidades [2]. Un componente puede desarrollarse aislado.

Uno de los desafíos centrales en el desarrollo de software, justamente, es la reutilización, lo que permite utilizar nuevamente uno o más artefactos implementados como parte del desarrollo de un nuevo producto o sistema. Existen diferentes técnicas que han facilitado, de alguna manera, el uso de artefactos de desarrollo de granularidad cada vez mayor. Al no disponer de contextos suficientemente amplios como para detectar elementos reutilizables y donde pueden utilizarse se desemboca muchas veces en reutilización oportuna. Esto es, se utiliza algún artefacto ya desarrollado con las correspondientes adaptaciones, en el nuevo sistema.

Desde la perspectiva de una institución, en el contexto de Gobierno Digital, es deseable lograr, en

cada desarrollo nuevo, un enfoque de reutilización sistemática. El fin, es obtener los beneficios asociados a utilizar artefactos construidos previamente (por ejemplo, la madurez y calidad). En el proceso de desarrollo se debe plantear la construcción colectiva de familias de productos relacionados por un dominio para lograr que la reutilización sea sistemática.

La industria del software se ha enfrentado a nuevos desafíos en cuanto a complejidad, costos, tiempo de comercialización, estándares de calidad y evolución [3]. Existen diferentes estrategias para enfrentar los desafíos que cuentan con adeptos tanto en la academia como en la industria, entre ellas, el Diseño Impulsado por el Dominio, DDD por sus siglas en inglés *Domain Driven Design* [4], la Ingeniería de Software Dirigida por Modelos, MDE por sus siglas en inglés Model-Driven Engineering [5] y las Líneas de Productos de Software (LPS) [6]. La relevancia de cada una de estas estrategias se explica brevemente a continuación.

En un contexto de Gobierno Digital, a medida que las aplicaciones crecen en complejidad y tamaño, es posible que puedan dividirse en contextos o subdominios independientes. DDD es un enfoque de desarrollo de software destinado específicamente a abordar los sistemas complejos. Su base estratégica es el proceso de descubrir dominios. En el que entran en juego dos conceptos muy importantes, el Lenguaje Ubicuo y los Contextos Delimitados (Bounded Context- BC). El Lenguaje Ubicuo es una colección de términos específicos del dominio, el lenguaje de negocios [4]. Los Contextos Delimitados son los límites dentro de los cuales cada modelo de dominio existe y opera. Este es un aspecto crucial y poderoso de la modelización de dominios, ya que permite que los modelos en diferentes contextos evolucionen independientemente unos de otros. Esto tiene un efecto significativo en el mantenimiento porque se hace mucho más sencillo aplicar y probar los cambios en un modelo que se centra únicamente en su propio dominio. El modelo sólo cambiará si cambian las reglas en su propio contexto. Por este motivo, el enfoque permite combatir la complejidad desde el lado de la "modularización". Al dividir grandes procesos en subprocesos más pequeños, es posible diseñar e implementar módulos separados y combinarlos para abordar diferentes aplicaciones [7].

A lo largo del tiempo se ha buscado crear artefactos de software de forma rápida, con menos errores y poniendo más énfasis en los requisitos y necesidades. Con el objetivo de cumplir con esto, MDE brinda más agilidad al desarrollo de aplicaciones. Es una propuesta que se ha venido trabajando desde hace varios años. Es un paradigma para el desarrollo software que considera a los modelos como el principal elemento del proceso de desarrollo [8]. Gracias al uso de modelos, es posible elevar el nivel de abstracción y de automatización. Esto permite atacar los principales problemas en la creación de software: a) la complejidad, b) el desarrollo de marcos de trabajo y lenguajes específicos para lograr la comprensión del dominio; c) mejorar diferentes aspectos de la calidad del software como la productividad y el mantenimiento; y d) obtener provecho de las transformaciones para automatizar trabajo repetitivo y mejorar la calidad del software.

Las LPS brindan un enfoque diferente para la creación de aplicaciones individuales a partir de la construcción de Familias de Productos. Este proceso se basa en la selección de elementos comunes y las variabilidades entre los miembros de la Familia [5]. La experiencia [9] muestra que se logra una mayor reutilización, menores costos y un menor tiempo de puesta en marcha. Una LPS es un conjunto de sistemas software que comparten un conjunto de características comunes (*features*) que satisfacen las necesidades específicas de un dominio o segmento de mercado particular, y que se desarrollan a partir de una arquitectura base y un conjunto de activos de software comunes (*core assets*), de un modo preestablecido y sistemático mediante un plan de producción [5]. El desarrollo de LPS consta de dos procesos básicos: la Ingeniería del Dominio, donde se establece las partes comunes y las variables y se construye un conjunto de activos, y la Ingeniería de la Aplicación, donde los activos son reutilizados para derivar productos específicos [10].

En la actualidad, los objetivos de la industria de software están puesto en la reducción de costos, el aumento de la calidad, el aumento de la productividad y la reducción del tiempo de desarrollo del software. Enfoques como LPS, MDE y DDD buscan alcanzarlos. Adoptarlos dentro de una organización implica desafíos importantes con respecto a los artefactos de software existentes que deben transformarse (MDE) para respetar una arquitectura de software centrada en BC (DDD) y la reutilización (LPS). Esta propuesta intenta responder las siguientes preguntas:

- ¿Cómo aplicar LPS como mecanismo de reutilización planificado o sistemático, la metodología MDE como mecanismo de aumento de la productividad en el desarrollo de software y el enfoque DDD para facilitar la modularización y el desarrollo de servicios independientes y reutilizables?
- ¿Cuáles son las herramientas, técnicas y *framewoks* existentes que den soporte al desarrollo de Línea de Productos junto con el enfoque DDD y MDE?

En las siguientes secciones se resume el estado del arte:

1. Ingeniería del Software Basada en Componentes y Reutilización. La Ingeniería del Software basada en componentes (CBSE: component-based software engineering) se ocupa del desarrollo de sistemas a partir de componentes reutilizables. El enfoque busca favorecer la reutilización y, de este modo, evitar la

duplicación de componentes, impidiendo que el análisis, el diseño, la implementación y la prueba de la misma solución, se desarrolle una y otra vez en muchos de los productos. Difiere del enfoque tradicional de desarrollo de software que involucra generalmente la construcción de cada una de las piezas desde cero, cada una de ellas implementada a propósito para el producto en curso. Algunas de las piezas o elementos reutilizables provienen de productos anteriores y son reutilizados con ciertos cambios [1]. CBSE, tiende a concentrar el esfuerzo en la composición de componentes de software desarrollados (que por lo general poseen una funcionalidad concreta), trata el desarrollo de sistemas a partir de la composición e integración de componentes reutilizables. Algunos de ellos, pueden requerir una vinculación con otros componentes para proveer la funcionalidad completa. En cuanto a la comunicación, como la integración de los componentes con los elementos del nuevo sistema que se encuentra desarrollando se realiza entendiendo los componentes como cajas negras, esto permite que se utilicen tal y como están disponibles y sin considerar ninguna modificación interna. [1]. A diferencia del enfoque tradicional, CBSE incluye las siguientes actividades: 1) Selección y evaluación de componentes, 2) Adaptación, 3) Composición, y 4) Evolución [1].

Según [12], existen dos factores principales cuando el enfoque que se utiliza está basado en componentes: el primero, la reutilización – la facilidad para reutilizar componentes existentes en la creación de sistemas más complejos – y el segundo, la evolución – crear un sistema altamente formado por componentes pre-existentes facilita su mantenimiento, ya que si el diseño es correcto los cambios estarán muy localizados y los efectos tendrán poca o ninguna propagación en el resto de los componentes del sistema. Estos factores se aplican bajo las siguientes condiciones: en primer lugar, deben existir componentes para su reutilización. En segundo lugar, debe existir un conjunto de componentes bien diseñados y listos para su utilización. Por otro lado, debe existir un modelo de componentes que soporte el enlazado y la interacción de componentes, esto es, debe existir un marco de referencia estándar en el que los componentes puedan existir y comunicarse. Finalmente, es necesario un proceso y la definición y diseño de arquitecturas que den soporte al CBD.

La reutilización favorece el desarrollo de herramientas de software con un alto grado de flexibilidad, haciendo uso de componentes en diversos proyectos, ampliando las capacidades, mejoras, funcionalidades y mantenimiento del software.

Hay varias definiciones del término Reutilización de Software. Algunas de estas definiciones son las siguientes. “La reutilización de software es el proceso de implementar o actualizar sistemas de software usando activos de software existentes” [12]. “Reutilización de software es el proceso de crear sistemas de software a partir de software existente, en lugar de desarrollarlo desde el comienzo” [13].

Existen varias modalidades de reutilización utilizadas, lo mismo que beneficios y costos, según el enfoque que se aplique. En particular, el oportunista, se beneficia de componentes de software construidos en productos anteriores, pero que no fueron especialmente desarrollados para ser reutilizados. En el enfoque más planificado o sistemático, los componentes se construyen pensando en que serán reutilizados, lo que significa que debe plantearse el componente con mayor generalidad, y por ello en el momento de su desarrollo, se requiere mayor inversión de tiempo y recursos.

2. Línea de Producto de Software. La definición más comúnmente aceptada de una LPS procede del autor Clements, quien define “las líneas del producto de software como un conjunto de sistemas software, que comparten un conjunto común de características (*features*), las cuales satisfacen las necesidades específicas de un dominio o segmento particular de mercado, y que se desarrollan a partir de un sistema común de activos base (*core assets*) de una manera preestablecida” [5].

Las LPS tienen como objetivo cambiar el enfoque de la creación de aplicaciones de software individuales a la construcción de familias de productos mediante la identificación de elementos comunes y variabilidades en un conjunto de aplicaciones de software [14]. En este contexto, una LPS permite acotar el tiempo de producción de componentes software, con altos niveles de calidad y una alta capacidad de reutilización.

En cuanto a los beneficios y costos, según diferentes casos de estudios documentados, existe una tendencia un aumento en la productividad que duplica o triplica la producción de software con respecto a los enfoques tradicionales. Las LPS pueden incrementar la productividad del desarrollo de software, dada la reducción del esfuerzo y el costo necesario para desarrollar, mantener y poner en marcha el conjunto de productos software similares [15].

Una LPS no es únicamente una mera reutilización de código, sino que los activos base se benefician también de cada producto desarrollado en la línea dado que toman ventaja del análisis, diseño, implementación, planificación, prueba. Potencian la reutilización estratégica [11], dado que en este contexto es más planificado que oportunista [16]: en desarrollo de software tradicional, se aprecia la posibilidad de reutilizar un componente después de haberlo desarrollado; en LPS, la reutilización es planificada, de manera que se reutilizan la mayor parte de los activos base en todos los productos de la línea [16]. Una LPS comprende esencialmente dos etapas:

a) Ingeniería de dominio- ID (*core asset development*, por Clements [5]) es la encargada de

diseñar los componentes de software que, de acuerdo a su trazabilidad, hacen parte del dominio común y que son sujetos a reutilización. En esta etapa se definen las similitudes y variantes del modelo de dominio en cada una de las fases. Como producto final en esta etapa se considera un componente altamente reusable y flexible que permite a partir de su definición, la consecución de nuevos modelos y artefactos.

b) Ingeniería de Aplicación - IA (*product development*, por Clements [15]) se encarga de articular los artefactos definidos en la etapa ID a partir de una plataforma específica. El objetivo de esta etapa consiste en lograr la integración de artefactos bajo uno o varios sistemas, logrando con ello un alto grado de reusabilidad de sus componentes a partir de las especificaciones, la definición y desarrollo, documentación e interoperabilidad entre componentes.

3. Ingeniería de Software dirigido por Modelos. La Ingeniería Dirigida por Modelos es una metodología de desarrollo de software que se centra en la creación de modelos, o abstracciones, es decir, identificar conceptos del dominio en lugar de concentrarse en el modelo de cómputo subyacente. A lo largo del tiempo se ha buscado crear artefactos de software de forma rápida, con menos errores y poniendo más énfasis en los requisitos y necesidades. Con el objetivo de cumplir con esto, la metodología MDE brinda más agilidad al desarrollo de aplicaciones [19].

A partir del uso sistemático de los modelos en las diferentes etapas del ciclo de vida, han surgido un conjunto de paradigmas de desarrollo de software que conforman lo que se ha denominado "Ingeniería de Software dirigida por Modelos", MDE por sus siglas en inglés *Model-Driven Engineering* o "Desarrollo de Software Dirigido por Modelos" MDSD por sus sigleas en inglés *Model-Driven Software Development*. El uso de modelos promueve elevar el nivel de abstracción y de automatización y con ello combatir la complejidad en la creación de software, además de mejorar la productividad y mantenimiento del software y la interoperabilidad entre sistemas [18].

En particular, la programación dirigida por modelos es un paradigma para el desarrollo software cuyo principal elemento del proceso son los modelos. A partir de estos modelos el código puede ser generado semiautomáticamente. [17], es decir, como "dibujar un sistema y pasar directamente a ejecutarlo" [1].

4. Domain Driven Design (DDD). El diseño impulsado por el dominio es un enfoque de desarrollo de software destinado específicamente a abordar los sistemas complejos, el cual establece lo siguiente: 1-. Centrarse en el dominio. 2.- Explorar modelos de dominio en colaboración con los expertos y desarrolladores. 3- . Hablar un lenguaje común (Lenguaje Ubicuo) dentro de un contexto delimitado. Plantea un enfoque diferente para el diseño y construcción de sistemas y tiene como objetivo crear aplicaciones a partir de una comunicación entre contextos. Es por ello que dichos sistemas son flexibles y adaptables dentro de cada contexto [4].

DDD ofrece un enfoque sistemático y completo para el diseño y desarrollo de software. Proporciona un conjunto de principios, herramientas y técnicas que ayudan a combatir la complejidad, manteniendo el modelo de negocio como pieza central del enfoque. Se adapta bien a una implementación de microservicios, dado que permite la división de un sistema en diferentes contextos delimitados. El diseño esta dado de tal manera que sólo el contexto tiene que ser modificado para implementar cambios o introducir nuevas características. Por tal motivo, se obtendrá el máximo beneficio del desarrollo independiente en diferentes equipos (de desarrollo), ya que varias características pueden implementarse en paralelo sin necesidad de una coordinación. La reutilizando de los microservicios como componentes, permite lograr el aumento de la calidad y la reducción del tiempo de implementación de procesos de negocio de una forma ágil [20]. DDD apunta a centrarse fuertemente en el dominio y el desarrollo de servicios independientes que forman aplicaciones de software más grandes cuando se combinan [7].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo principal de la línea de investigación aquí presentada es:

- Especificar y diseñar un marco de trabajo (*framework*) que permita el Desarrollo de software de dominio específico en el contexto de Gobierno Digital.

Sus ejes específicos son: (a) Estudiar el estado del arte de los enfoques Líneas de Productos de Software (LPS), paradigma de Ingeniería de Software dirigida por Modelos (MDE) y Diseño dirigido por el Dominio (DDD), (b) Evaluar el estado del arte en marcos de trabajo para la aplicación de LPS combinado con enfoques DDD y MDE, (c) Identificar e integrar herramientas existentes que soporten el desarrollo de LPS, MDE y DDD, (d) Analizar el dominio de aplicación, en Plataformas de servicios con intermediación, se estudiará el contexto del gobierno de la Ciudad de Viedma - Río Negro (e) Llevar adelante un Caso de Estudio donde se especificarán requerimientos e identificarán contextos delimitados bajo el enfoque DDD y con la metodología LPS los activos base y variables por cada uno (f) Diseñar una arquitectura de referencia que permita su instanciación para la aplicación de los enfoques propuestos (g) Construir un marco de trabajo (*framework*) que en base a la arquitectura y los requerimientos que considere un

modelo metodológico y enfoque de construcción acorde para que permita la reutilización sistemática con el fin de lograr beneficios asociados a utilizar los artefactos (componentes) construidos previamente y la implementación de un modelo en constante evolución. (h) Desarrollar y validar prototipos de software, pasibles de ser implementados y transferidos al medio.

3. RESULTADOS OBTENIDOS/ ESPERADOS

Como resultado de este proyecto, se espera identificar herramientas, metodologías y modelos conceptuales para el desarrollo de software de dominios complejos, que permita la integración de herramientas de construcción y sea considerado un modelo metodológico acorde para la reutilización sistemática del software construido previamente y la implementación de un modelo y/o reglas de negocio en constante evolución. En cuanto al diseño de la arquitectura de software se espera construir regiones o capas, para desacoplar las mismas y que evolucionen de manera aislada, independientes de la tecnología. También, se espera que el proyecto signifique un cambio revolucionario en la forma de escribir el software y una solución a la modernización, entre otras, para aquellos sistemas que se encuentran obsoletos por la tecnología que utilizan. Se obtendrá mediante un Caso de estudio, una Plataforma de servicios de empleo con intermediación, esta será implementada y transferida a la Subsecretaría de Innovación, Desarrollo Sostenible y Ambiente de la Municipalidad de Viedma. La elaboración de este prototipo funcional y estable se espera que produzca, un sistema flexible, con capacidad de reutilización de los componentes que lo conforman, mantenible, independiente de la tecnología y tolerante a los cambios (reglas de negocio).

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por tres investigadores formados en la temática, dos investigadores en formación y tres alumnos avanzados de la carrera Licenciatura en Sistemas. En su marco se desarrolla una tesis de Maestría en Ciencias de la Computación y se producirán tres trabajos finales de carrera de grado.

5. BIBLIOGRAFÍA

[1] Usaola, M.: Desarrollo de software basado en reutilización. UOC. Formación de Postgrado. 2013.

[2] Montiliva, J et al: Desarrollo de software basado en Componentes. V Congreso de Automatización y Control Mérida. Noviembre 2013.

[3] Markus Voelter ; Iris Groher. Product Line Implementation using Aspect-Oriented and Model-Driven Software Development. 2007.

[4] E. Evans, Domain-Driven Design: Tackling Complexity in the Heart of Software. Addison-Wesley, 2003

[5] Clements,P., Northrop,L.: Software Product Lines: Practices and Patterns. Addison-Wesley (2001).

[6] Oquendo, F.: p-Method: A Model-Driven Formal Method for Architecture-Centric Software Engineering. ACM SIGSOFT Software Engineering Notes. Volume 31 Number 3. (2006).

[7] Domain Driven Design Reference. Definitions and Pattern Summaries. Eric Evans.2015.

[8] Cabot, J. http://es.slideshare.net/jcabot/mdd-desarrollo-de-software-dirigido-por-modelos-que-funciona-de-verdad.

[9] C. W. Krueger:Homeaway's transition to software product line practice:Engineering and business results in 60 days. In Proc. of the 12th International Software Product Line Conference (SPLC'08), Ireland, pages 297–306. IEEE, September 2008.

[10] J. Gonzalez-Huerta, S. Abrahão, E. Insfran: Un enfoque Multi-modelo para la Introducción de Atributos de Calidad en el Desarrollo de Líneas de Producto Software. XVI Jornadas de Ingeniería del Software y Bases de Datos. Septiembre de 2011

[11] Peñalvo, F., et al.: Líneas de Productos, Componentes, Frameworks y Mecanos. Informe Técnico, Departamento de Informática y Automática - Universidad de Salamanca (2002).

[12] Sodhi, J., Sodhi,P: Software reuse: Domain analysis and design process. McGraw-Hill. (1999)

[13] Sametinger, J.: Software engineering with reusable components. Springer Verlag, (1997)..

[14] C.Parra, D. Joya, L. Giral, A.Infante: An SOA approach for Automating Software Product Line. Proceeding SAC '14 Proceedings of the 29th Annual ACM Symposium on Applied Computing Pages 1231-1238. Republic of Korea. March2014

[15] Díaz O., Trujillo, S.: Fábricas de Software: experiencias, tecnologías y organización" (2º edición). Editorial Ra-Ma, (2010).

[16] Trujillo, S., Batory, D. y Díaz, O: Feature Oriented Model Driven Development: A Case Study of Portlets. International Conference on Software Engineering (ICSE) (2007).

[17] H. Escobar, H.Triana, S.Silveira Netto: Conceptualización de arquitectura de gobierno electrónico y plataforma de interoperabilidad para América Latina y el Caribe. Naciones Unidas, Santiago de Chile. julio de 2007.

[18] J. García Molina, García Rubio, V. Pelechano, A. Vallecillo, JM. Vara, C.Vicente-Chicote: Desarrollo de Software Dirigido por Modelos: Conceptos, Métodos y Herramientas. Edición Ra-Ma 2012.

[19] H. Ed-douibi et al. EMF-REST Generation of RESTful APIs from Models. Proceeding SAC '16 Proceedings of the 31st Annual ACM Symposium on Applied Computing Pages 1446-1453.Italy.2016

[20] Microservices: Flexible Software Architecture. Addison-Wesley Professional.2016

[21] Desarrollo de líneas de producto software mediante un enfoque generativo, disponible en: http://portal.uned.es/portal/page?_pageid=93,56233773&_dad=portal&_schema=PORTAL&idAsignatura=31105043.

Avances en Métodos y Técnicas para la construcción de Aplicaciones basadas en Computación Ubicuas

Únzaga Silvina; Durán Elena, Álvarez Margarita, Salazar Nevelyn, Fernández Reuter Beatriz, González Gabriela, Montoto Sebastián, Sanchez Daniel, Diaz Fátima y Quintana Cancinos, Fernando

Instituto de Investigaciones en Informática y Sistemas de Información (IIISI)
Universidad Nacional de Santiago del Estero (UNSE)
e-mail: {sunzaga; eduran, alvarez, nsalazar, bfreuter, ggonzalez, }@unse.edu.ar,
sebastian.montoto@gmail.com, licprofsanchezdaniel@gmail.com, fatimadiaz91@gmail.com,
f.quintana88@yahoo.com.ar

CONTEXTO

En la Universidad Nacional de Santiago del Estero se está llevando a cabo el desarrollo del proyecto de investigación titulado "Métodos y Técnicas para desarrollos de Aplicaciones Ubicuas"; proyecto acreditado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SICYT - UNSE). El período de ejecución del proyecto es 2017-2020. este proyecto tiene como objetivo favorecer el desarrollo de conocimiento científico-tecnológico de relevancia sobre Computación Ubicua, realizando propuestas de técnicas, métodos y estrategias para el diseño y construcción de aplicaciones ubicuas.

RESUMEN

Los nuevos desarrollos de las Ciencias de la Computación/Informática se orientan hacia la Computación Ubicua, en la que los ordenadores están incluidos en nuestros movimientos naturales y en las interacciones con nuestro entorno, tanto físico como social. Este nuevo paradigma de la computación involucra dos características fundamentales: la movilidad y la pervasividad. Los principales desafíos en la Computación Ubicua se originan, entonces, en la integración de ambas características en la búsqueda de que todo dispositivo computacional, mientras se mueve con nosotros, pueda ir construyendo de forma incremental modelos dinámicos de sus diversos entornos y configurar sus servicios en consecuencia.

Nuevos retos emergen en términos de cómo se debe diseñar y desarrollar aplicaciones ubicuas. Esto requiere el replanteo de las arquitecturas de software posibles, el diseño de ontologías y de modelos de dominio, nuevos escenarios de interacciones y el análisis de nuevas familias de requisitos no funcionales (como configurabilidad y la adaptabilidad).

Con el proyecto de investigación presentado en el contexto, se busca desarrollar nuevas técnicas, métodos y estrategias para el diseño y construcción de aplicaciones ubicuas, atendiendo a los nuevos retos que la computación Ubicua impone.

En este artículo se describe el proyecto, los resultados obtenidos a lo largo de tres años de ejecución del mismo, en especial la formación de recursos humanos concretada en el marco del proyecto.

Palabras clave: Computación Ubicua, Técnicas de Diseño de Software, Razonamiento Automático, Sistemas Adaptativos, Ontologías, Evaluación de Sistemas Ubicuos

1- INTRODUCCIÓN

El término "computación ubicua" se atribuye a Mark Weiser (Weiser, 1991) quien manifestaba que las tecnologías más profundas son aquellas que desaparecen, quedando inmersas en la vida cotidiana, de forma tal que no se pueden distinguir de ella.

Este nuevo campo de la computación ubicua, pretende incorporar a los objetos de la vida cotidiana capacidad de cómputo, de comunicaciones inalámbricas y de interacción entre ellos, para crear un nuevo modelo de la realidad en la que estos objetos interoperan entre sí para facilitar la realización de las tareas a las personas. Uno de los principales objetivos de la Computación Ubicua es volver invisibles a los dispositivos computacionales, ubicándolos en un segundo plano. Este objetivo supone una potencial

revolución que puede hacer cambiar el modo de vida diario.

Las personas se centrarán en las tareas que deben hacer, no en las herramientas que utilizan, porque se pretende que esas herramientas pasen desapercibidas. El significado de enviar la computación a un "segundo plano", se refiere a dos conceptos. El primero es el significado literal de que la tecnología de la computación se debe integrar en los objetos, cosas, tareas y entornos cotidianos. Y la segunda, es que esta integración se debe realizar de forma que la introducción de computación en estas cosas u objetos no interfieran con las actividades para las que son usadas, y que proporcionen un uso más cómodo, sencillo y útil de esos objetos (Chaparro-González, 2003).

La Computación Ubicua revoluciona el paradigma actual de interacción hombre-computador, cuyo objetivo es dar una ayuda personalizada a los sujetos.

Por lo tanto, el desarrollo de aplicaciones de software ubicuas presenta para los desarrolladores un inmenso desafío. Son muchos los problemas a resolver a la hora de construir una aplicación ubicua: conocer el contexto, razonar en base al mismo y actuar en consecuencia. Atendiendo a ello, en Durán et al. (2014) hemos propuesto una arquitectura para desarrollar aplicaciones de aprendizaje ubicuo, pero que perfectamente puede ser adaptada para aplicaciones de computación ubicua en general. Sobre la base de esta arquitectura es posible el desarrollo de aplicaciones que operen en entornos dinámicos de computación ubicua y se adapten a los cambios de contexto. Esta arquitectura ofrece, además, un enfoque de desarrollo de aplicaciones basado en modelos ontológicos que facilita la adaptación dinámica y automática de los servicios disponibles para el usuario.

Considerando los diversos componentes de esta arquitectura (módulos de software que gestionan servicios, modelos, ontologías y repositorios) y que existen distintas técnicas y métodos que se pueden aplicar para el modelado y creación de estos componentes, se propuso en el proyecto que aquí se presenta, investigar las técnicas y métodos para el modelado y construcción de aplicaciones ubicuas.

2- DESCRIPCIÓN DEL PROYECTO

En el Proyecto "Métodos y Técnicas para desarrollos de Aplicaciones Ubicuas"; se han definido los siguientes objetivos:

1. Revisar y analizar las técnicas y métodos existentes para el modelado y la construcción de aplicaciones ubicuas.
2. Construir y evaluar el modelo de contexto ubicuo.
3. Evaluar y proponer técnicas y métodos para el razonamiento y la adaptación en aplicaciones ubicuas.
4. Aplicar las técnicas y métodos propuestos en la construcción de aplicaciones ubicuas de impacto local y regional.
5. Evaluar, en contextos reales, el nivel de satisfacción del usuario y el desempeño de las aplicaciones ubicuas construidas.
6. Formar recursos humanos mediante la dirección de tesis de posgrado, trabajos finales de grado y la dirección de becas de investigación.
7. Transferir y ofrecer servicios al medio a través del asesoramiento y capacitación de estudiantes de grado y posgrado por medio de cursos y seminarios; realizar publicaciones en revistas especializadas, y presentar comunicaciones en congresos y jornadas.

Para cumplir con estos objetivos, se comenzó con una exploración e investigación bibliográfica sobre técnicas y métodos utilizados hasta el momento para el modelado y construcción de aplicaciones ubicuas; se evaluaron las técnicas y métodos estudiados y se diseñaron nuevas técnicas y métodos para mejorar las falencias encontradas.

Se identificaron y analizaron problemáticas que demandaban la construcción de aplicaciones ubicuas. Se estudiaron los requerimientos de estas aplicaciones y la posibilidad de aplicar las nuevas técnicas, métodos y estrategias desarrolladas, para su construcción.

Para la identificación de las problemáticas se aplicó la Metodología de los Sistemas Blandos y sus técnicas asociadas. Para el estudio de los requerimientos de estas aplicaciones se usaron técnicas y métodos de la Ingeniería de Requisitos

Se analizaron los repositorios existentes evaluando la posibilidad de ser utilizados en las aplicaciones ubicuas a diseñar. Se construyó el modelado de requisitos y el modelo de análisis y diseño de las aplicaciones ubicuas. Para ello se usaron los modelos diseñados en etapas previas; y se utilizaron técnicas provenientes de la Ingeniería de Software, de la Ingeniería Web, Ingeniería ontológica, Inteligencia Artificial y las nuevas técnicas, métodos y estrategias propuestos en esta investigación.

Se construyó el software diseñado, aplicando técnicas de programación de dispositivos móviles y de programación web. Para las pruebas de unidad se aplicaron técnicas de prueba caja blanca y de caja negra.

En el último año del proyecto se trabajará para el cumplimiento del objetivo 5, Evaluar, en contextos reales el nivel de satisfacción del usuario y el desempeño de las aplicaciones ubicuas construidas.

Para ello se evaluará el nivel de satisfacción del usuario y el desempeño de la aplicación ubicua, se identificarán indicadores, presentes en el producto y en el uso del sistema. Se investigarán estándares referidos a la usabilidad del software, métricas del producto y antecedentes de otros trabajos en los que se evalúen estas variables.

3- RESULTADOS OBTENIDOS

Los principales resultados obtenidos para cumplir con los objetivo planteados son los siguientes:

a) *Estado del arte sobre sobre creación y gestión de repositorios para aplicaciones ubicuas.* Este resultado se documenta en Zachman et al. 2017, donde se presenta estudio exploratorio y bibliométrico del estado del arte de aplicaciones ubicuas en el escenario de los repositorios institucionales. Estudio realizado con el fin de establecer las bases para la futura creación de un modelo de la arquitectura de un repositorio institucional, procesable desde la computación ubicua.

b) *Método para generar recomendaciones personalizadas para integrar grupos de aprendizaje ubicuo y colaborativo*. Se generaron dos versiones de este método. En la primera (Alvarez et. al. 2017) se generan recomendaciones de compañeros con los que un estudiante puede conformar un grupo de trabajo para desarrollar una tarea colaborativamente en un ambiente de aprendizaje ubicuo. El método considera condiciones temporales, académicas, disponibilidad horaria, y cercanía física. Se presenta, además, un ejemplo demostrativo de la aplicación del método. En la segunda versión del método (Quintana Cancino et. al., 2018; Quintana Cancino et. al., 2019) se recomienda en base a la ubicación de los estudiantes, las condiciones académicas que deben cumplir los mismos y características que aseguren la heterogeneidad de integrantes en el grupo.

c) *Estrategia para la personalización de servicios para aplicaciones ubicuas* (Duran et. al 2017a; Duran et. al, 2017b). La estrategia propuesta permite personalizar servicios en contextos ubicuos y fue implementada en una aplicación para apoyar el aprendizaje de los estudiantes en el curso de ingreso universitario de la FCEyT-UNSE (Montoto et. al. 2018).

d) *Método para recomendar expertos que apoyen el aprendizaje en entornos ubicuos* (Duran et. al; 2017c; Duran et. al; 2018). El método genera recomendaciones automáticas de expertos para asesorar a los estudiantes, considerando las experiencias previas de los expertos con otros estudiantes. Recomienda expertos que estén disponibles y físicamente cerca del estudiante; o expertos que estén disponibles en línea. Sobre la base de este modelo se desarrolló también un sistema multiagente que implementa el método

e) *Método para la generación de caminos de aprendizaje en entornos ubicuos* (Fernández Reuter et. al., 2017)*;* aquí se propone un método híbrido para la generación de caminos de aprendizaje ubicuos personalizados utilizando técnicas de planificación, ontologías y agentes inteligentes. Este método utiliza el modo de generación fuera de línea, lo que significa que no hay demora para determinar la siguiente mejor acción que se debe tomar en la ruta de aprendizaje.

f) *Método para generar un modelo de contexto ambiental* (Gonzalez et. al., 2018); este método permite generar en tiempo real, un modelo de contexto ambiental que incorpora información de eventos locales y de patrones de ocurrencia de eventos, que podrá ser usado en sistemas ubicuos (Fernandez Reuter et. al., 2018).

4- FORMACIÓN DE RECURSOS HUMANOS

En las tablas 1 y 2 se sintetiza la formación de recursos humanos realizada en el marco del proyecto. Específicamente, en Tabla 1 se consignan las tesis de posgrado y los trabajos finales de grado que se encuentran en desarrollo y/o finalizados; y en la Tabla 2, lo becarios con sus correspondientes trabajos y tipos de becas.

Tabla 1. Formación de Recursos Humanos: Tesis de posgrado y Trabajos Finales de Grado

Título del Trabajo	Alumno	Nivel	Estado
Personalización en Sistemas de Aprendizaje Ubicuo	Beatriz Fernández Reuter	Tesis de Doctorado en Ciencias de la Computación, UNICEN	Tesis en ejecución
Representación del Contexto en Aplicaciones de Aprendizaje Ubicuo.	Gabriela Gonzalez	Tesis de Doctorado en Ciencias de la Computación, UNICEN	Tesis en ejecución
Ontología para búsquedas personalizadas de trabajo académicos y científicos en Informática en bibliotecas y repositorios digitales universitarios	Daniel Sanchez	Tesis de Maestría en Informática Educativa, Universidad Nacional de Santiago del Estero.	Tesis en ejecución
Objetos de Aprendizaje para Simulación basado en Pedagogías Emergentes	Nevelyn Salazar	Trabajo Final de la Especialización en Enseñanza de la Tecnología de la UNSE.	Trabajo en Evaluación
Estado del Arte del uso de Ontologías en Bibliotecas Digitales: Revisión sistemática de bibliografía	Daniel Sanchez	Trabajo Final de la Especialización en Informática Educativa, Universidad Nacional de Santiago del Estero.	Trabajo en ejecución
Desarrollo de una Aplicación Ubicua Personalizada de apoyo al curso de ingreso de la FCEyT - UNSE	Sebastían Montoto	Trabajo Final de Grado para optar por el título de Licenciado en Sistemas de Información, UNSE	Trabajo Final Aprobado
Sistema de recomendación personalizado de documentos para repositorios digitales institucionales.	Fátima Diaz	Trabajo Final de Grado para optar por el título de Licenciado en Sistemas de Información, UNSE	Trabajo en ejecución
Aplicación móvil para sugerir establecimientos de venta de alimentos para celiacos.	Paola Romero	Trabajo Final de Grado para optar por el título de Licenciado en Sistemas de Información, UNSE	Trabajo en ejecución

Tabla 2. Formación de Recursos Humanos: Becas

Tipo Beca	Organismo Otorgante	Título del Trabajo	Alumno	Período
Beca interna de Postgrado tipo I	Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET),	Representación del contexto en aplicaciones de aprendizaje ubicuo	Gabriela Gonzalez	2014-2019
Becas UNSE de investigación para estudiantes 2019.	Universidad Nacional de Santiago del Estero	Sistema de recomendación personalizado aplicando minería de texto	Fatima Diaz	2018-2019
Becas UNSE de investigación para estudiantes 2019	Universidad Nacional de Santiago del Estero	Sistema de recomendación de tutores personalizado para entornos de aprendizaje	Luciano Gastón Juárez	2019
Becas UNSE de investigación para estudiantes 2019	Universidad Nacional de Santiago del Estero	Integración de aplicaciones u-learning como servicios web	Nicolás Ezequiel Coali	2019
Becas de Estímulo a las Vocaciones Científicas (Beca CIN)	Universidad Nacional de Santiago del Estero	Prototipo para recomendar grupos de aprendizaje ubicuo y colaborativo	Fernando Quintana Cancino	Mayo 2018 a Abril 2019

5- BIBLIOGRAFÍA

Álvarez, M., Únzaga, S. y Durán, E. (2017). "Método para generar recomendaciones personalizadas para integrar grupos de aprendizaje ubicuo y colaborativo". XXII Congreso Argentino de Ciencias de la Computación. CACIC 2017. Junín.

Chaparro-González, D. (2003) Computación ubicua. (Tesina de grado, Universidad Rey Juan Carlos, Madrid, España). Disponible en http://docplayer.es/6315315-Universidad-rey-juan-carlos.

Durán, E.; Álvarez, M. y Únzaga, S. (2014). "Ontological model-driven architecture for ubiquitous learning applications". EATIS 2014 - 7th Euro American Association on Telematics and Information Systems. Valparaíso (Chile). 2 al 4 de Abril de 2014.

Durán, E.; Unzaga, S. y Alvarez, M. (2017a). "Personalización de servicios para aplicaciones de aprendizaje ubicuo". (Cava 2017)

Durán, E.; Unzaga, S. y Alvarez, M. (2017b). "Personalización de servicios para aplicaciones de aprendizaje ubicuo". RevistaVirtu@lmente, 5(2), 7-25.

Durán, E. y Álvarez, M. (2017c). "Method for Generating Expert Recommendations to advise students on ubiquitous learning experiences". Handbook of Research on Human Interaction and the Impact of Information Technologies/IGI Global. En edición.

Durán, Elena y Alvarez, Margarita (2018). Recommendations of experts to advise students on ubiquitous learning experience: method and validation. IEEE Latin America Transaction. ISSN: 1548-0992. Vol 16 No 9. pp: 2314-2320 (2018): Special Issue on New Trends in Electronics. Disponible en: https://www.inaoep.mx/~IEEElat/index.php/transactions/article/view/ 5

Fernandez Reuter, B., Duran, E. y Amandi (2017)." Designing a Hybrid Method for Personalized Ubiquitous Learning Paths Generation". Jornadas Chilenas de Computación 2017, Arica. Chile.

Fernandez Reuter, B.; Alvarez, M; Gonzalez, G y Durán, E. (2018). "Multi-agent system model for tutor recommendation in ubiquitous learning environments". I Workshop on Advanced Virtual Environments and Education (WAVE2).

Gonzalez, G., Durán, E. y Amandi, A., (2018), "Generación de modelo de contexto ambiental con técnicas de aprendizaje automático", Simposio Argentino de Inteligencia Artificial (ASAI).

Montoto, Sebastían; "Desarrollo de una Aplicación Ubicua Personalizada de Apoyo al Curso de Ingreso de la FCEyT-UNSE"; 47JAIIO - EST - ISSN: 2451-7615 - Página 374; EST, Concurso de Trabajos Estudiantiles, 2018

Quintana Cancino, F. y Alvarez, M. (2018). "Método para formación de grupos de estudiantes en un ambiente de aprendizaje ubicuo", XIII Jornadas de Ciencia y Tecnologías de Facultades de Ingeniería del NOA.

Quintana Cancino, F. y Alvarez, M. (2019). "Método de formación de grupos colaborativos en ambientes de aprendizaje ubicuos", III Congreso Internacional Virtual en Investigación e Innovación Educativa (CIVINEDU).

Weiser, M. (1991). The Computer for the 21st Century. Scientific American, 265(3), 66–75. Retrieved from http://www.syssoft.uni-trier.de/systemsoftware/Download/Fruehere_Veranstaltungen/Ubiquitous_Computing/2004/02 Paradigm Print.pdf

Zachman, P y Duran E., (2017). "Repositorios Institucionales y Computación Ubicua - Estudio bibliométrico"; XII Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2017), Universidad Nacional de Chilecito, La Rioja, ISBN 978-987-24611-1-9, pp. 298-305.

Uso de la tecnología Blockchain Federal (BFA) para dejar pistas de auditoría y trazabilidad a sentencias y acordadas de la Suprema Corte de la provincia de Mendoza

Rotella Carina[1], Ontiveros Patricia[1], Bianchini Germán[2], Caymes-Scutari Paola[2,3], Tagarelli Sandra[4], Salinas Sergio[4], Chirino Pamela[2], Galdamez Mariela[2]

[1]Laboratorio de Gobierno Electrónico, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza
[2]Laboratorio de Investigación en Cómputo Paralelo/Distribuido
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579
[3]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
[4]Laboratorio de Analítica de Datos, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza

crotella@frm.utn.edu.ar, pontiveros@frm.utn.edu.ar, gbianchini@frm.utn.edu.ar, pcaymesscutari@frm.utn.edu.ar, stagarelli@gmail.com, s4salinas@gmail.com, pamelaachirino@gmail.com, mariela.galdamez.16@gmail.com

RESUMEN

El conocimiento y uso de las tecnologías innovadoras es fundamental para la formación de docentes y alumnos de grado e Investigadores Científicos en las áreas de Informática y Ciencias de la Computación.

Este proyecto pretende generar ese espacio de enlace y formación entre el interés que despierta trabajar sobre nuevas tecnologías, como *Blockchain*, y la necesidad de formar investigadores. De esta manera, es posible lograr la permanente actualización, en las nuevas tendencias, ya que se trata de ciencias que evolucionan vertiginosamente

Blockchain (o cadena de bloques) es una base de datos compartida que funciona como un libro para el registro de operaciones o transacciones. Es una tecnología disruptiva que merece espacio de estudio en las universidades por los distintos casos de uso que tiene no sólo en el ámbito financiero. Disruptiva porque se trata de una innovación tecnológica que deja obsoleta a la tecnología anterior produciendo una ruptura brusca y en algunas ocasiones causando cambios profundos.

Debido a la falta de personal calificado en el tema y la alta demanda, se pretende abordarla en el amplio sentido de estudiarla, entenderla y diversificar su aplicación en distintos ámbitos: académico y profesional. Como así también promover y motivar la innovación tecnológica.

Palabras clave: Blockchain, Tecnología, Disruptiva, Innovación, Notarización, Seguridad, Integridad, Cadena, Nodo, Bloque.

CONTEXTO

En noviembre de 2019 el laboratorio "GE-Lab" de la UTN Facultad Regional Mendoza, inicia acciones conjuntas con el Laboratorio "UTN Blockchain Lab" de la UTN Facultad Regional de La Plata [1] siendo la primera actividad conjunta una charla abierta sobre ***BLOCKCHAIN & SMART CONTRACTS*** a la cual asistieron 130 personas interesadas en constituir la comunidad de Blockchain en Mendoza aprobada por Resolución 880 con fecha 12 de noviembre 2019-

En el contexto de esa charla abierta a cargo de Pedro Rey Puma de "UTN Blockchain Lab" cofundador del mismo, se abordaron los principales conceptos y características de *Blockchain*, tipos de *Blockchain* y el paradigma de la descentralización. Otro tema abordado fue la Notarización de documentos *PDF* en *Blockchain*.

De este último tema desarrollado surge el interés de un docente de la UTN - Facultad Regional Mendoza que además se desempeña en la Suprema Corte de la Provincia de Mendoza. En esta institución se plantea la necesidad de modernización del proceso de auditoría y trazabilidad de sentencias y acordadas emitidas por los tribunales utilizando las nuevas tecnologías (BFA específicamente para este caso). La Notarización tiene que ver con la posibilidad de garantizar con la tecnología el valor legal de un documento, asegurar su integridad, incrementar la transparencia y facilitar la auditoría del mismo.

Es de gran importancia destacar que, en este caso, los distintos actores: Universidades (UTN-FRLP y UTN-FRM) y Sector Público (Poder Judicial) trabajan de manera integrada, uniendo y coordinando sus esfuerzos y conocimientos para lograr innovaciones tecnológicas aplicables.

El laboratorio (GE-Lab) perteneciente al Departamento de Ingeniería en Sistemas de Información de la UTN-FRM, también está trabajando en forma colaborativa con los docentes de los laboratorios LICPaD (Laboratorio de Investigación en Cómputo Paralelo/Distribuido) y ADA-Lab (Laboratorio de Analítica de Datos), ambos del mismo Departamento, en el marco del proyecto PID TEUTIME0007658TC "Formación de docentes y alumnos de grado como Investigadores Científicos Iniciales en las áreas de Informática y Ciencias de la Computación".

1. INTRODUCCIÓN

La finalidad de este trabajo es analizar el uso de la tecnología *Blockchain* Federal Argentina (BFA) para dejar pistas de auditoría y trazabilidad a sentencias y acordadas de la Suprema Corte de la provincia de Mendoza.

Según la Enciclopedia Jurídica [2], "Acordada" es una comunicación de un tribunal a otro inferior, para ordenarle la ejecución de algo o para advertirle o proponerle algo reservadamente. Y "Sentencia" es una Resolución judicial que decide definitivamente un proceso o una causa o recurso o cuando la legislación procesal lo establezca.

A continuación, se da una pequeña introducción a los principales conceptos de *Blockchain* y BFA.

Blockchain es una tecnología diseñada para administrar un registro de datos online, que se caracteriza por ser transparente e incorruptible.

De acuerdo a Laurence [3], es una estructura de datos que permite la creación de un registro digital de datos. Este registro es compartido con participantes independientes mediante una red.

El concepto de bloques se refiere a una lista de movimientos que se guardan en un registro digital. Estos bloques pueden tener distinto tamaño y se pueden crear en diferentes periodos del tiempo.

La cadena se refiere a un *hash* que enlaza un bloque con otro, un *hash* es un algoritmo que transforma o resume el contenido del bloque en una salida [4]. Su principal uso es para la protección de los datos, mediante el ocultamiento o disfraz de los mismos con el propósito de que éstos no puedan ser alterados ni eliminados. Estos enlaces creados por la función *hash* permiten crear la fuerte cadena de todos los bloques, que ahora se encuentra encadenados entre sí y cuya misma función permite asegurar los datos que pasan de un bloque a otro. [3]. La siguiente figura muestra el comportamiento de una cadena de bloques.

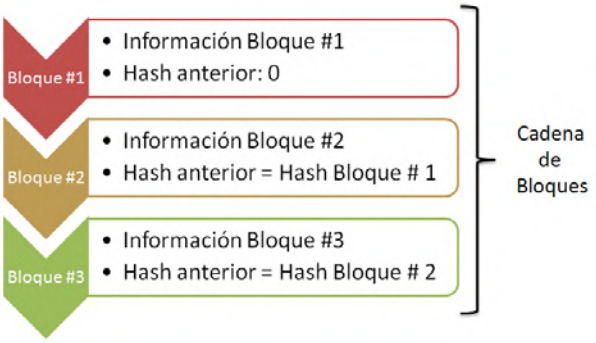


Figura 1: Comportamiento de una Cadena de Bloques.

En la última parte de la estructura de la cadena de bloques, o *blockchain*, se encuentra la red, una de las partes más importantes para que toda la cadena de bloques funcione correcta y eficazmente. [3]

La red está compuesta por "nodos completos", donde cada nodo contiene el registro de datos de todos los movimientos realizados en la cadena de bloques. Los nodos pueden localizarse en cualquier parte del mundo y ser operados por quien sea. A estos operadores se les llama mineros. [3]

Lo más importante de una cadena de bloques es su funcionamiento, ya que cada nodo posee toda la información de toda la cadena de bloques y toda la cadena se actualiza inmediatamente después de un cambio. Este sistema no es un sistema centralizado sino más bien distribuido o descentralizado. Esto implica que, si algún nodo se pierde o es atacado, la información se mantendrá intacta en los demás nodos, por este motivo es considerado como uno de los mejores en el tema de seguridad, ya que es muy complejo realizar un ataque hacía todos los puntos de la cadena. Además, cuando un nodo presenta un comportamiento distinto o sospechoso (a este se le llama nodo bizantino [5]) se pone en alerta a toda la cadena de bloques. Sin embargo, como se explicó anteriormente, la cadena de bloques no se ve afectada por sólo un nodo que falle. [5]

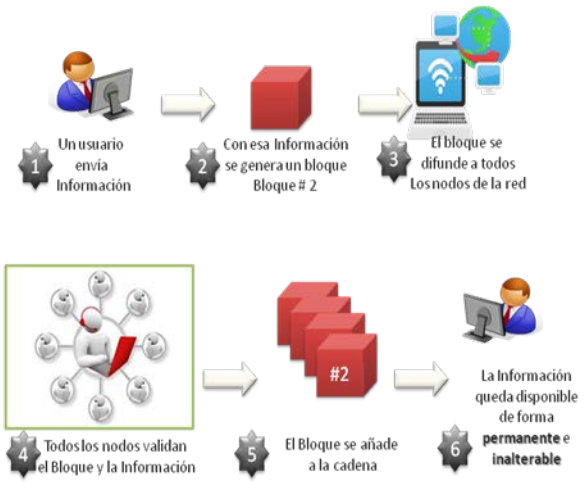


Figura 2: Funcionamiento del *Blockchain*

En el caso bajo estudio se utilizará *Blockchain* Federal Argentina, ya que es una plataforma multiservicios abierta y participativa pensada para integrar servicios y aplicaciones sobre *blockchain* para el sector público.

Blockchain Federal Argentina [6] fue diseñada para potenciarse a través de los aportes de sectores públicos, privados, académicos y de la sociedad civil. BFA cuenta con una estrategia donde la participación de toda la comunidad es esencial, desde la ingeniería organizacional hasta el despliegue de la infraestructura.

En BFA no se almacenan documentos o archivos, sólo los *hashes* de los mismos. No posee criptomoneda [6] asociada y es gratuita.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Como ya se mencionó en un punto anterior, esta línea de trabajo está enmarcada dentro de un proyecto general cuyo objetivo es la formación de investigadores docentes y alumnos, para que, dentro de un ámbito de investigación real, puedan llevar adelante el proceso y las tareas necesarias para la formación teórica y práctica.

Se investigará en distintas temáticas relacionadas con los Laboratorios involucrados en el proyecto, especialmente en lo que hace referencia a las tecnologías disruptivas como *Blockchain* y sus distintos casos de uso en todos los ámbitos con el fin último de desarrollar capacidades, promover su uso y aplicación, especialmente en el ámbito académico.

3. RESULTADOS ESPERADOS

El resultado principal que se pretende alcanzar aplicar esta tecnología actual e innovadora en la Corte Suprema de la provincia de Mendoza para dejar pistas de auditoría y trazabilidad en las acordadas y dictámenes.

Como resultados secundarios se pretende lograr modernización del estado, en el ámbito legislativo, transparencia en la gestión, seguridad de los datos, garantía de integridad de la información y control de la información.

La implementación de esta tecnología implicaría un ahorro de tiempo debido a que no hay necesidad de procesos burocráticos regidos bajo intermediarios o terceras partes. Por ejemplo, la aprobación de una transacción normalmente puede llegar a tardar alrededor de dos días, sin embargo, al utilizar *blockchain*, el proceso puede llegar a durar sólo minutos. Del mismo modo esta nueva tecnología podría reducir los costos de infraestructura para la institución.

La trazabilidad consistirá en hacer un seguimiento de las modificaciones del *hash* de los dictámenes y acordadas, de manera que si han sufrido algún cambio se modificaría y se generaría un nuevo *hash*, lo que nos aseguraría la incorruptibilidad.

Esto se complementaría con la pista de auditoría que permitirá seguir el rastro de las modificaciones que haya sufrido el documento, a partir de la generación del nuevo hash, y que nos podría dar información acerca de cuándo y quién realizó la modificación.

La tecnología *blockchain* facilita la auditoría de la información al asegurar el registro de todas las transacciones, las cuales generan una cadena de bloques que no se puede borrar o modificar sin dejar una huella. Esta característica además permite auditar procesos confiando en la información de la cadena y sin la necesidad de que terceros brinden la información, lo que quita los incentivos a manipularla con fines particulares. [7]

En resumen, esto nos brindará garantía, resguardo, seguridad y control sobre los dictámenes y acordadas de la Corte Suprema de Justicia de la provincia de Mendoza.

4. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto pretende contribuir a la formación y crecimiento de investigadores en el ámbito académico. La idea es motivar e incentivar a alumnos y docentes a investigar sobre temáticas novedosas y en un ámbito donde puedan desarrollar el pensamiento crítico acompañados y guiados por investigadores formados.

Actualmente, participan del proyecto seis docentes con distintos grados y temas de especialización, y dos alumnos de grado de la carrera Ingeniería en Sistemas de Información. Además, se prevé la participación y colaboración con investigadores de otras facultades regionales vinculadas a la misma temática, y se contará

con la participación de personal del Poder Judicial de la Provincia de Mendoza.

5. BIBLIOGRAFÍA

[1] LINSI Laboratorio de Innovaciones en Sistemas de Información. http://www.linsi.edu.ar/utnblockchainlab.html (Accedido el 16/03/2020)

[2] Enciclopedia Jurídica http://www.enciclopedia-juridica.com/inicio-enciclopedia-diccionario-juridico.html / (Accedido el 05/03/2020)

[3] Laurence, Tiana (2019) Blockchain for Dummies. Editorial: For Dummies. ISBN-13: 978-1119555018

[4] Bit2me Academy https://academy.bit2me.com/que-es-hash/ (Accedido el 19/03/2020)

[5] Bushir, Imran (2017) Mastering Blockchain: Deeper insights into decentralization, cryptography, Bitcoin, and popular Blockchain frameworks. Editorial: Packt Publishing. ISBN-13: 978-1787125445

[6] Blockchain Federal Argentina https://bfa.ar/ (Accedido el 05/03/2020)

[7] Florencia Serale, Christoph Redl, Arturo Muente Kunigami (2019) Blockchain en la Administración Pública ¿Mucho ruido y pocos bloques? Disponible en: https://publications.iadb.org/publications/spanish/document/Blockchain_en_la_administraci%C3%B3n_p%C3%BAblica_Mucho_ruido_y_pocos_bloques_es.pdf (Accedido el 19/03/2020)

Herramienta Web para Interoperabilidad Conceptual entre UML, EER y ORM 2

Emiliano Rios Gavagnin[1] Germán Braun[1,2,3] Laura Cecchi[1]
Pablo Fillottrani[2,4]

email: emiliano.rios@est.fi.uncoma.edu.ar,
{german.braun,lcecchi}@fi.uncoma.edu.ar, prf@cs.uns.edu.ar

[1]*Grupo de Investigación en Lenguajes e Inteligencia Artificial*
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE
[2]*Laboratorio de I&D en Ingeniería de Software y Sistemas de Información*
Departamento de Ciencias e Ingeniería de la Computación
UNIVERSIDAD NACIONAL DEL SUR
[3]*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*
[4]*Comisión de Investigaciones Científicas de la provincia de Buenos Aires (CIC)*

Resumen

La presente investigación se desarrolla mediante el trabajo colaborativo de docentes investigadores de la Universidad Nacional del Comahue (UNCo) y de la Universidad Nacional del Sur (UNS), en el contexto de proyectos de investigación financiados por las universidades indicadas.

El objetivo general de esta línea de investigación y desarrollo es el diseño e implementación de una herramienta Web que permita facilitar la interoperabilidad entre los lenguajes de modelado UML, EER y ORM 2. Para esto, se considera un metamodelo integrador, llamado KF, el cual formaliza las bases de los distintos lenguajes, permitiendo así, identificar las similitudes entre los lenguajes previamente mencionados. El resultado de esta implementación será integrado a `crowd`, la cual es una herramienta para modelado visual ontológico utilizando lenguajes de modelado conceptual, desarrollada por nuestros grupos de investigación.

Palabras Clave: Ingeniería de Software basada en Conocimiento, Lógicas Descriptivas, Ontologías, Interoperabilidad de Lenguajes de Modelado Conceptual.

Contexto

Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue, en el marco del proyecto de investigación *Agentes Inteligentes y Web Semántica (04/F014)*, por la Universidad Nacional del Sur a través del proyecto de investigación *Integración de Información y Servicios en la Web (24/N027)*, por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), en el contexto de una beca posdoctoral y por el Consejo Interuniversitario Nacional (CIN) con una Beca de Estímulo a las Vocaciones Científicas para Estudiantes 2019. Los proyectos de investigación tienen una duración de cuatro años, la beca posdoctoral una duración de 2 años, finalizando en mayo de 2021, y la beca CIN una duración de 1 año, finalizando esta última en mayo de 2020.

1. Introducción

En la actualidad, la información es un elemento esencial en instituciones y compañías. Esto requiere de software cada vez más complejo y, por consiguiente, de procesos de diseño y desarrollo que escalen hacia soluciones interoperables y considerando fuentes de datos altamente heterogéneas [2, 8, 13, 17]. Analistas, ingenieros y administradores de bases de datos necesitan crear modelos conceptuales que les permitan representar los datos con los que trabajan, y así, poder hacer una evaluación más acertada del dominio. Existen múltiples lenguajes de modelado de datos que permiten a los diseñadores crear modelos, como por ejemplo UML [4], EER [12] y ORM 2 [15]. Si bien esta diversidad de lenguajes permite un mayor nivel de flexibilidad a la hora de diseñar soluciones de software para un sinnúmero de dominios, y en términos generales tratan de cumplir el mismo objetivo, estos lenguajes tienen diferentes formas de representar la información. Asimismo, pueden ser utilizados en diversas etapas del diseño, por ejemplo, los diagramas ORM 2 para representar información conceptual del sistema, los Diagramas de Clases UML para esquemas lógicos para diseño de objetos y, por último, diagramas EER para esquemas de bases de datos [9, 14, 24].

Teniendo en cuenta el contexto mencionado previamente, en [11, 19] se introduce el diseño de un modelo que abstrae y formaliza los conceptos representados por los lenguajes UML, EER y ORM. Este modelo, conocido como el *Metamodelo KF de unificación de modelado conceptual*, tiene como objetivo reducir la heterogeneidad que existe entre estos lenguajes de modelado, para poder mejorar la interoperabilidad de los mismos. En esencia, es un modelo que integra las entidades, relaciones y restricciones de los tres lenguajes.

El Metamodelo KF permite crear un modelo conceptual sobre cualquiera de los lenguajes de modelado, y luego generar el equivalente en otro lenguaje, teniendo en cuenta las limitaciones semánticas que existen entre los mismos. Sin embargo, en la actualidad, no existe una implementación del Metamodelo KF, y el mismo es un gran candidato para reducir la heterogeneidad entre modelos debido a dos factores: abarca tres de los lenguajes de modelado de datos más populares y relevantes, y además existe una formalización del Metamodelo que nos permite demostrar la correctitud del mismo.

Bajo esta línea de investigación se propone desarrollar una herramienta Web que implemente las bases del Metamodelo KF. La herramienta recibirá como entrada un modelo representado en alguno de los lenguajes de modelado soportados (UML, EER y ORM), o incluso una instancia del Metamodelo, y permitirá convertir dicho modelo a una representación equivalente en otro lenguaje, acorde a la formalización del Metamodelo.

Este trabajo está estructurado como sigue. En la siguiente sección, se describe la línea de investigación actual. En la sección 3, se indican algunos resultados obtenidos y trabajos futuros. Finalmente, en la última sección se comentan aspectos referentes a la formación de recursos humanos en esta temática.

2. Línea de Investigación y Desarrollo

En el proyecto de investigación Agentes Inteligentes y Web Semántica, UNCo, se investigan técnicas de representación de conocimiento y razonamiento, metodologías de modelado conceptual y mecanismos para la interoperabilidad tanto a nivel de proceso como de datos. El empleo efectivo de los conceptos y conocimientos adquiridos da soporte a comunidades de desarrollo de ontologías.

Por otro lado, en el proyecto de investigación Integración de Información y Servicios en la Web, UNS, se desarrollan metodologías y herramientas que asisten la interoperabilidad semántica de información y de servicios en la Web, privilegiando los últimos avances en el área de lenguajes de representación del conocimiento, ontologías y modelado conceptual.

Ambos proyectos comparten el perfil de investigación de este trabajo, en el que se estudian entre otros, sobre temas afines a la representa-

ción del conocimiento, las Lógicas Descriptivas [7], las Ontologías, la Ingeniería de Software basada en Conocimiento y estándares para definir nuevos lenguajes visuales para representar conocimiento.

La OMG (Object Management Group) [22] define varios lenguajes de modelado, dentro de los cuales UML [4] es muy utilizado en la actualidad. Existen otros de amplio uso, como EER [12] para bases de datos y ORM 2 [15, 16] para aquellos expertos en dominios e interacción entre el análisis de requerimientos y datos. Por esta razón, la interoperabilidad de los lenguajes de modelado conceptual se ha vuelto una necesidad, dado que los modeladores podrían requerir vincular entidades entre modelos, representados en diferentes lenguajes de modelado conceptual y comunicarse con otros *stakeholders* mediante lenguajes comunes.

En este sentido, se han obtenido resultados unidireccionales para la transformación de lenguajes como ORM a UML [3], unificación de modelos mediante grafos [5], lógicas descriptivas [20] y transformaciones a través de diccionarios de términos comunes. Incluso, se han desarrollado investigaciones que involucran unificaciones parciales de los distintos modelos mediante metamodelos [25] y procesos automatizados de traducción de lenguajes, que hacen uso de diccionarios de "metaconstrucciones" [1]. Sin embargo, estos trabajos han demostrado tener ciertas limitaciones, como por ejemplo, la omisión de elementos importantes de los lenguajes de modelado, como lo son los roles, restricciones de relaciones y agregaciones.

Así, se optó por el *Metamodelo KF de unificación de modelado conceptual*, dado que considera la integración de la mayoría de los elementos, su desarrollo está basado en un enfoque ontológico [18] y presenta una formalización en Lógica de Primer Orden (FOL).

Una herramienta Web que implemente el metamodelo KF y que permita a los modeladores utilizar diferentes lenguajes de modelado conceptual, particularmente UML, EER y ORM 2, en forma conjunta y simultánea, es el fin de esta línea de investigación, desarrollo y transferencia.

Nuestra propuesta permitirá unificar el uso de los lenguajes de modelado UML, EER y ORM 2. Se espera mejorar la interoperabilidad de estos lenguajes, facilitando la conversión de modelos conceptuales a distintos lenguajes de modelado, manteniendo la semántica de los mismos.

Existen herramientas que permiten el modelado conceptual para cada uno de los lenguajes de modelado mencionados anteriormente. Sin embargo, no hay herramientas que implementen al metamodelo y hagan uso de las ventajas que brinda, tal como se propone en este trabajo.

A su vez, el resultado de la implementación del metamodelo será integrado a una nueva versión de `crowd` [6], plataforma desarrollada también por nuestros grupos de investigación. *crowd* es una plataforma Web para el modelado conceptual visual, en los lenguajes gráficos UML, EER y ORM 2, integrada a sistemas de razonamiento automáticos que permiten validarlos.

3. Resultados Obtenidos y Trabajo Futuro

Inicialmente, se analizó el diseño ontológico del Metamodelo KF [10, 18] para luego profundizar en la formalización FOL descrita en [10].

Con una mejor comprensión del Metamodelo KF y su semántica, se diseñó la implementación del mismo. Para esto, se realizó un relevamiento de patrones de diseño a tener en cuenta, que podían ayudar a representar el Metamodelo KF a un nivel programático. Analizando los trabajos que describen el Metamodelo, fue evidente que la implementación se daría bajo el paradigma orientado a objetos, ya que se ven involucradas una gran cantidad de clases, representadas bajo una jerarquía de las mismas, con un gran número de relaciones entre sí.

Una vez diseñada la herramienta, se decidió desarrollar la implementación del Metamodelo KF como una *API REST*. Esta decisión fue tomada teniendo en cuenta que, la funcionalidad de la misma no debería depender de una aplicación o software preexistente. De esta forma, el

Metamodelo queda como un elemento externo a herramientas de modelado de datos, facilitando la integración de la misma a sistemas enfocados en el diseño de modelos de datos.

Para su implementación se trabajó con el framework *Java* denominado *Spring Boot* [23, 26]. Adicionalmente, se configuró la API dentro de un *Docker* [21] el cual contiene el aplicativo, permitiendo así, una mayor portabilidad y facilitando las futuras integraciones a sistemas de modelado de datos. En este sentido, actualmente se está trabajando en su integración a *crowd*.

Dentro de la formalización del Metamodelo KF, existen tres enfoques para asociar elementos de un lenguaje de modelado con otro. El primer enfoque es el denominado *1:1 Mappings*. Estos son elementos iguales en todos los lenguajes de modelado, y por lo tanto se pueden convertir de manera directa. Las *Transformations* son aquellos elementos que en esencia representan el mismo concepto, pero difieren en aspectos sintácticos y por lo tanto requieren de ciertos pasos de conversión. Las *Approximations* consisten de elementos que para convertirlos a otro lenguaje, requieren de una entrada de parte del usuario que defina ciertos aspectos de la conversión. En esta primer parte del desarrollo de la herramienta, se restringe la implementación del Metamodelo KF a los *1:1 Mappings*, ya que son las conversiones más directas, y son los elementos principales de los lenguajes de modelado de datos.

Como se puede observar, muchas formalizaciones y conceptos del Metamodelo KF quedan aún pendientes a implementar. Trabajos futuros incluyen la implementación de las *Transformations* y *Approximations*. Estos avances implicarían abarcar un mayor número de entidades dentro del Metamodelo KF.

4. Formación de Recursos Humanos

En la Universidad Nacional del Comahue, Facultad de Informática, se otorgaron Becas CIN para estimular la vocación científica. Una de esas becas fue otorgada a uno de los autores de este trabajo, que está desarrollando su tesis de grado de la Licenciatura en Ciencias de la Computación en esta temática.

Por otra parte, otro de los autores de este trabajo es becario posdoctoral CONICET.

Referencias

[1] Paolo Atzeni, Paolo Cappellari, Riccardo Torlone, Philip Bernstein, and Giorgio Gianforme. Model-independent schema translation. *VLDB J.*, 17:1347–1370, 11 2008.

[2] Albert Banal-Estañol. Information-sharing implications of horizontal mergers. *International Journal of Industrial Organization*, 25(1):31 – 49, 2007.

[3] Peter Bollen. A formal orm-to -uml mapping algorithm. *Maastricht : METEOR, Maastricht Research School of Economics of Technology and Organization, Research Memoranda*, 01 2002.

[4] G. Booch, J. Rumbaugh, and I. Jacobson. *Unified Modeling Language User Guide*. Addison-Wesley Professional, 2005.

[5] Michael Boyd and Peter Mcbrien. Comparing and transforming between data models via an intermediate hypergraph data model. *Journal on Data Semantics - JODS*, pages 69–109, 01 2005.

[6] Germán Braun, Elsa Estevez, and Pablo Fillottrani. A Reference Architecture for Ontology Engineering Web Environments. *Journal of Computer Science & Technology*, 2018.

[7] Diego Calvanese, Maurizio Lenzerini, and Daniele Nardi. Description logics for conceptual data modeling. In *Logics for Databases and Information Systems*, pages 229–263. Kluwer, 1998.

[8] Leonardo Candela, Donatella Castelli, and Pasquale Pagano. Data interoperability. *Data Science Journal*, 12, 07 2013.

[9] Brian Dobing and Jeffrey Parsons. How uml is used. *Commun. ACM*, 49:109–113, 05 2006.

[10] Pablo R. Fillottrani and C. Maria Keet. KF metamodel formalization. *CoRR*, abs/1412.6545, 2014.

[11] Pablo Rubén Fillottrani and C. Maria Keet. Conceptual model interoperability: A metamodel-driven approach. In *Rules on the Web. From Theory to Applications - 8th International Symposium, RuleML 2014, Co-located with the 21st European Conference on Artificial Intelligence, ECAI 2014, Prague, Czech Republic, August 18-20, 2014. Proceedings*, pages 52–66, 2014.

[12] M. Gogolla. *Extended Entity-Relationship Model: Fundamentals and Pragmatics*. Springer-Verlag, 1994.

[13] Alon Halevy, Naveen Ashish, Dina Bitton, Michael Carey, Denise Draper, Jeff Pollock, Arnon Rosenthal, and Vishal Sikka. Enterprise information integration: successes, challenges and controversies. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 778–787, 01 2005.

[14] Terry Halpin. Object-role modeling: Principles and benefits. *IJISMD*, 1:33–57, 01 2010.

[15] Terry Halpin and Tony Morgan. *Information Modeling and Relational Databases*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2 edition, 2008.

[16] Terry A. Halpin. ORM 2. In *On the Move to Meaningful Internet Systems: OTM Workshops, OTM Confederated International Workshops and Posters*, 2005.

[17] Eduard Hovy. Data and Knowledge Integration for e-Government. In *Digital Government*, pages 219–231. Springer, 01 2008.

[18] C. Maria Keet and Pablo Fillottrani. An ontology-driven unifying metamodel of UML Class Diagrams, EER, and ORM2. *Data Knowledge Engineering Journal*, 2015.

[19] C. Maria Keet and Pablo Rubén Fillottrani. An ontology-driven unifying metamodel of UML Class Diagrams, EER, and ORM2. *Data Knowl. Eng.*, 98:30–53, 2015.

[20] Catharina Maria Keet. *Ontology-Driven Formal Conceptual Data Modeling for Biological Data Analysis*, chapter 6, pages 129–154. John Wiley & Sons, Ltd, 2013.

[21] Karl Matthias and Sean P. Kane. *Docker: Up & Running*. O'Reilly Media, Inc., 1st edition, 2015.

[22] Object Management Group Homepage. https://www.omg.org/index.htm Accedido última vez marzo 2020.

[23] K. Reddy. *Introduction to Spring Boot*, pages 1–20. 09 2017.

[24] Il-Yeol Song and K. Froehlich. Entity-relationship modeling. *Potentials, IEEE*, 13:29 – 34, 02 1995.

[25] John R. Venable and John C. Grundy. Integrating and supporting entity relationship and object role models. In Michael P. Papazoglou, editor, *OOER '95: Object-Oriented and Entity-Relationship Modeling*, pages 318–328, Berlin, Heidelberg, 1995. Springer Berlin Heidelberg.

[26] Craig Walls. *Spring Boot in Action*. Manning Publications Co., USA, 1st edition, 2016.

Detección de productos Software para la Industria 4.0

Alicia Mon; Horacio René Del Giorgio; Eduardo De María
Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza.
Florencio Varela 1903 - San Justo (CP 1754)
Tel: 4480-8952
alicialmon@gmail.com; hdelgiorgio@unlam.edu.ar; demaria.edu@gmail.com

RESUMEN

El presente artículo expone los resultados parciales de un proyecto de investigación en curso que se propone evaluar el desarrollo tecnológico actual en la industria manufacturera, independientemente de la rama de actividad que se analice, y detectar así los tipos de productos software específicos que componen a la Industria 4.0.

Para ello, se ha desarrollado una aplicación de software que, utilizando los instrumentos de medición desarrollados en un proyecto precedente, permite realizar de manera sistemática el relevamiento, análisis y evaluación de inserción de TICs en la industria. Los resultados de tal evaluación permiten definir los niveles del desarrollo tecnológico en cuanto a productos software, hardware e infraestructura.

Una vez determinados con precisión los parámetros tecnológicos de la industria manufacturera local, se trabajará con la detección de los atributos específicos que contienen los productos software y sus necesidades de implementación en las cadenas de valor, para confluir en la Industria 4.0.

Palabras clave: TICs, Industria 4.0, Usabilidad.

1. CONTEXTO

La línea de investigación que se presenta se está desarrollando en el marco del programa de incentivos del Ministerio de Educación, un Proyecto PICTO financiado por el Ministerio de Ciencia y Tecnología que finaliza en mayo del 2020, sumado a un Proyecto Vincular 2019 financiado por la Secretaría de Políticas Universitarias del Ministerio de Educación, y un proyecto PROINCE en el marco del DIIT (Departamento de Ingeniería e Investigaciones Tecnológicas) de la Universidad Nacional de La Matanza.

Por otra parte, todas estas actividades se están llevando a cabo en colaboración con la Cámara de Industria y Comercio de La Matanza, la Secretaría de Producción del Municipio y el CeDIT-UNLaM (Centro de Desarrollo e Investigaciones Tecnológicas de la UNLaM), de modo tal de realizar un relevamiento en la industria manufacturera del Partido de La Matanza, a efectos de evaluar el nivel de desarrollo tecnológico local y facilitar la detección de necesidades de desarrollo e implementación de productos software en las diferentes ramas de actividad.

Sobre esta línea, el DIIT se propone estudiar los tipos de tecnologías instaladas, analizar el valor agregado del uso de las mismas en la industria y determinar las necesidades de software específicos para implementar en las cadenas de valor que aporten hacia la Industria 4.0.

2. INTRODUCCIÓN

Las Tecnologías de la Información y las Comunicaciones, también conocidas como TICs, constituyen un factor central para el desarrollo económico, social y productivo. La competitividad y la innovación en los sectores productivos requieren del desarrollo tecnológico, el cual se ha transformado en una componente imprescindible en las empresas actuales. Una cantidad creciente de protagonistas de las distintas disciplinas científicas advierten la presencia de una revolución en curso a la que denominan la cuarta revolución industrial [Ministerio de Ciencia, Tecnología e Innovación Productiva,

2015] [Ibáñez Díaz, Cabas Alonso, Cuevas Arce, & Balaguer, 2018] [Kantar Millward Brown, 2017].

La industria del software en general ha crecido exponencialmente en los últimos años, enfocando los desarrollos de productos hacia aplicaciones de uso fuertemente impulsados a las áreas de administración, gestión, servicios financieros y aplicaciones de uso social, recreativos o de servicios, quedando los desarrollos de software destinados a la industria en un reducido porcentaje del total de proyectos.

Teniendo en cuenta que en la estructura productiva del país la producción industrial representa el 28% del PBI, menos del 10% de la industria del software desarrolla productos para este sector productivo. Por otra parte, en Argentina, el 56% de la producción de software se direcciona hacia empresas multinacionales, del sector financiero o comercial, dejando relegada la actualización tecnológica de las pequeñas y medianas empresas manufactureras que representan mayoritariamente la estructura productiva de Argentina.

En este sentido, la industria del software y el desarrollo de TICs en general pareciera no tener definidas estrategias de actualización tecnológica direccionadas hacia las PyMEs industriales, las que, como contracara de esta carencia, parecieran no focalizar la mejora de la competitividad en la innovación y en la actualización tecnológica, dado que no resulta ser un sector demandante del desarrollo de TICs.

Desde esta perspectiva, la incorporación de nuevas productos software en los sectores industriales requiere de un profundo conocimiento sobre la capacidad instalada. Es decir que, sin información relativa a las tecnologías instaladas y utilizadas en los diferentes procesos, no es posible definir necesidades de incorporación tecnológica para generar una reconversión en las cadenas de valor.

Si bien existe diversa bibliografía sobre el desarrollo productivo y los desarrollos tecnológicos, no se ha encontrado aún una forma específica de evaluar los diferentes desarrollos tecnológicos ni el impacto que generan en la productividad, así como en la definición de estrategias de innovación requeridas por la industria manufacturera.

3. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

La posibilidad de conocer las diferentes tecnologías, los tipos de productos software instalados, así como la agregación de valor que aportan en la productividad, resulta una información clave para la toma de decisiones estratégicas tanto en la industria del software como en los diferentes sectores industriales.

El presente proyecto, que se ha iniciado en el Departamento de Ingeniería e Investigaciones Tecnológicas de la UNLaM, se enfoca en las líneas de investigación que el grupo GIS (Grupo de Ingeniería de Software) viene desarrollando desde hace varios años y en red con otras universidades.

El proyecto en curso se propone medir la inserción de TICs (Software, Hardware e Infraestructura) en la industria y definir específicamente los productos software que pueden ser desarrollados o implementados para transformar a una empresa en una Industria 4.0.

Esta línea de investigación se desarrolla en el contexto del DIIT con proyectos sobre realidad virtual, realidad aumentada, impresoras 3D y redes para internet de las cosas.

Asimismo, en los proyectos de investigación mencionados, se ha estudiado la conformación industrial del Partido de La Matanza, siendo la vinculación con el desarrollo local un factor predominante en las líneas de investigación desarrolladas desde el Departamento.

4. RESULTADOS OBTENIDOS/ESPERADOS

Como resultados parciales del proyecto, se pueden diferenciar en un Índice de TICs [Del Giorgio & Mon, 2019] y un Relevamiento en la Industria de La Matanza [Mon & Del Giorgio, 2018].

Respecto al Índice de TICs, se ha creado un método permite evaluar en 3 niveles diferenciados de desarrollo tecnológico según las TICs como básico, medio o avanzado. En el tipo de TICs que se agrupan en el nivel avanzado se encontrarían las industrias más desarrolladas tecnológicamente, sin que necesariamente lleguen a ser reconocidas como Industrias 4.0.

Es dable destacar que se han creado todos los instrumentos necesarios para la aplicación del del índice de TICs y la obtención del cálculo. Dichas herramientas permiten el análisis permanente y sistemático sobre la inserción de TICs y la evaluación del nivel de desarrollo tecnológico de la industria en todas las ramas y en forma independiente de la localización.

En este sentido, los instrumentos creados con el proyecto están disponibles en https://indicetics.unlam.edu.ar/ y son los siguientes:

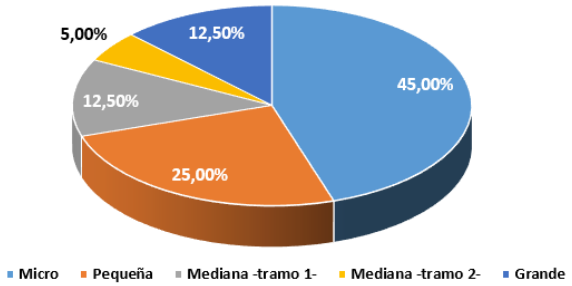
- Diseño de encuesta de relevamiento de TICs en industrias.
- Desarrollo de un software de evaluación automática, que contiene las herramientas y el método de evaluación; soporta una base de datos con la información del relevamiento y está disponible en forma online para realizar la evaluación en cada empresa.
- Desarrollo de una página web donde está disponible el acceso al software de evaluación automática que devuelve el resultado del nivel de TICs a cada empresa, las publicaciones del proyecto, información sobre el grupo de investigación y un video de explicación y difusión del índice.
- Guionado, realización y edición de un video explicativo del índice

Respecto a los resultados del Relevamiento en la Industria de La Matanza se avanzó en el análisis a partir de realizar:

- Relevamiento piloto en 40 industrias de ramas diferentes en el distrito.
- Determinación de niveles de desarrollo tecnológico de las Industrias relevadas.

Del análisis del relevamiento, se puede observar que el 45% de las empresas son Micro, el 25% Pequeñas, en tanto que en la categoría de Mediana tramo 1 y tramo 2 se ubica un 17,5% entre ambas categorías y sólo el 12,5% se ubican en la categoría de Gran empresa, tal como puede se presenta en el siguiente gráfico.

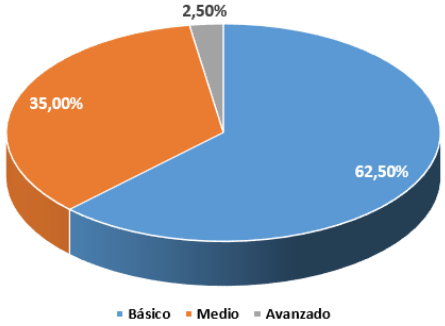
Gráfico 1 - Empresas por tamaño



Fuente: Elaboración Propia.

Del análisis de los resultados sobre el nivel de TICs, aplicando el índice, se puede observar que, de las 40 empresas encuestadas, el 62,5% se encuentra en el Nivel Básico, el 35% en el Nivel Medio y solo el 2,5% en el Nivel Avanzado, tal como se expone en el siguiente gráfico.

Gráfico 2 - Distribución según Puntaje



Fuente: Elaboración Propia.

En tanto que, si se aborda el análisis por rama de actividad para comparar al interior de

una rama, es posible detectar las diferencias sobre las TICs implementadas, y es así como las siguientes tablas presentan los porcentajes obtenido en cada nivel y tipología de productos para 3 empresas.

La siguiente tabla expone los valores para una empresa micro de la rama metalúrgica cuyo resultado de aplicar el índice le asigna una puntuación correspondiente al Nivel Básico.

TIC	Básico	Medio	Avanzado
Software	14,29%	0,00%	0,00%
Hardware	20,00%	0,00%	16,67%
Infraestructura	0,00%	0,00%	0,00%

Tabla 1- Metalúrgica Nivel Básico
Fuente: Elaboración Propia.

La siguiente tabla expone los valores para una empresa pequeña de la rama metalúrgica cuyo resultado de aplicar el índice le asigna una puntuación correspondiente al Nivel Medio.

TIC	Básico	Medio	Avanzado
Software	42,86%	27,27%	16,67%
Hardware	40,00%	66,67%	33,33%
Infraestructura	80,00%	25,00%	0,00%

Tabla 2 - Metalúrgica Nivel Medio
Fuente: Elaboración Propia.

En tanto que la siguiente tabla expone los valores para una gran empresa de la rama metalúrgica cuyo resultado de aplicar el índice le asigna una puntuación correspondiente al Nivel Avanzado.

TIC	Básico	Medio	Avanzado
Software	100,00%	90,91%	83,33%
Hardware	60,00%	100,00%	50,00%
Infraestructura	100,00%	50,00%	100,00%

Tabla 3 - Metalúrgica Nivel Avanzado
Fuente: Elaboración Propia.

Estos resultados, son provisorios y representan una muestra del tipo de análisis que puede realizarse con los instrumentos desarrollados para la industria en general, para un distrito en particular y a su interior por tamaño de empresa, por rama de industria, así como para una empresa en particular.

Tal como se mencionó anteriormente, el desarrollo de las TICs encamina a su vez el desarrollo de la industria hacia la convergencia digital, la conectividad entre objetos y la incorporación de inteligencia artificial en la resolución de problemas de producción, entre otros aspectos que enfocan hacia lo que se conoce como Industria 4.0.

Este tipo de industrias encierran un conjunto de desarrollos tecnológicos sin definiciones precisas, y es por ello por lo que el grupo de investigación GIS ha realizado la evaluación del nivel de TICs en la industria local.

A partir de dichos resultados, se está trabajando sobre los productos software específicos que están implementados en la actualidad y cuáles son los tipos de software que pueden ser incorporados por una industria según su área funcional.

5. FORMACIÓN DE RECURSOS HUMANOS

El grupo de investigación GIS se ha conformado para este proyecto como un grupo interdisciplinario e interuniversitario, integrado por Ingenieros Informáticos, Industriales y Electrónicos, todos docentes-investigadores.

Los proyectos precedentes han generado diversos resultados científicos y académicos publicados en diferentes congresos nacionales e internacionales y la interdisciplinariedad de este ha facilitado el desarrollo de una Tesis del Doctorado en Ciencias Económicas de la UNLaM (*Un modelo de diagnóstico de la competitividad empresarial*) escrita y dirigida por integrantes del grupo de investigación. En el momento de la escritura del presente documento, sólo queda pendiente la asignación de una fecha para su defensa.

Asimismo, un integrante del grupo GIS se encuentra desarrollando una tesis de la Maestría en Dirección Estratégica y

Tecnológica del Instituto Tecnológico Buenos Aires.

6. BIBLIOGRAFÍA

- Del Giorgio, H. & Mon, A. (2019). Las TICs en las Industrias. Editorial Universidad Nacional de La Matanza. Buenos Aires, Argentina. Disponible en https://indicetics.unlam.edu.ar/it/publicaciones-libros.jsp
- Ibáñez Díaz, A., Cabas Alonso, J., Cuevas Arce, S., & Balaguer, C. (2018). Observatorio de la Industria 4.0 - Foro de Profesionales. Obtenido de http://www.observatorioindustria.org/Inicio/
- Kantar Millward Brown. (2017). Tercer Estudio de Competencias Digitales en la Empresa Española. Obtenido de http://www.ticpymes.es/siteresources/files/839/54.pdf
- Ministerio de Ciencia, T. e. (2015). Industria 4.0: Escenarios e impactos para la formulación de políticas tecnológicas en los umbrales de la Cuarta Revolución Industrial. Obtenido de http://www.mincyt.gob.ar/adjuntos/archivos/000/038/0000038319.pdf
- Mon, A. & Del Giorgio, H. (2018). Exploración de la Inserción de las Tecnologías de la Información y la Comunicación en el Desarrollo Industrial de La Matanza. XIº Congreso de Ingeniería Industrial. Universidad de Mendoza. Mendoza. Argentina.
- Mon, A. & Del Giorgio, H. (2019). Modelo de evaluación de Tecnologías de la Información y la Comunicación para la industria 4.0. XXV Congreso Argentino de Ciencias de la Computación. Facultad de Ciencias Exactas, FísicoQuímicas y Naturales – Universidad Nacional de Río Cuarto. Río Cuarto. Córdoba. Argentina.

Inteligencia y tecnologías aplicadas al deporte de alto rendimiento

Laura Fava, Diego Vilches, Alejandro Ferraresso, Ezequiel Boccalari, Javier Díaz
LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.
Facultad de Informática. Universidad Nacional de La Plata, Calle 50 esq. 120, 2do Piso.
Tel: +54 221 4223528
{lfava, dvilches, aferraresso, eboccalari}@linti.unlp.edu.ar, jdiaz@unlp.edu.ar

RESUMEN

En un mundo tan competitivo como es el del deporte, cualquier detalle por mínimo que sea, marca una gran diferencia; los datos que ofrecen los diferentes sensores y otros medios, aportan una ventaja competitiva extra que puede ser diferencial. El uso de *wearables,* cámaras y distintos softwares proveen una variedad de datos que pueden ser procesados y analizados, dando paso a una nueva etapa donde los diferentes actores vinculados al deporte pueden mejorar sus entrenamientos, capacidades de decisión y en consecuencia, los resultados. La tecnología ha alcanzado la industria del deporte, y ha impactado en los resultados de los equipos y deportistas, debido a las mejoras en las técnicas, a la aplicación de nuevos métodos de entrenamiento y al análisis predictivo de los datos.

En este artículo se describen líneas de investigación y desarrollo que forman parte de un proyecto integral destinado a proveer soluciones tecnológicas aplicadas al deporte que abren un abanico de posibilidades para un entrenamiento más controlado y eficiente.

Palabras claves: dispositivos para entrenamiento, tecnología en deporte, aprendizaje automático, Internet of Things (IoT).

CONTEXTO

El Laboratorio de Investigación de Nuevas Tecnologías Informáticas (LINTI) de la Facultad de Informática, viene trabajando en proyectos relacionados con tecnologías aplicadas al deporte acompañando las metodologías clásicas de entrenamiento, así como también, la administración de las nuevas fuentes de datos y mediciones existentes.

La primera línea de trabajo estuvo relacionada con el desarrollo de dispositivos para deporte basados en sensores y LEDs para mejorar el entrenamiento y rendimiento de los deportistas de élite (Fava, L., 2018). El uso de estos dispositivos y de las tecnologías de IoT nos ofrecen una potente fuente de información para planear el progreso del deportista y optimizar sus entrenamientos y competiciones (Alexandre D, 2012). Hoy en día, los profesionales del deporte pueden disponer de datos exactos y en tiempo real, con los que medir velocidad, distancias recorridas, movimientos realizados y una enorme cantidad de datos que abren una nueva línea de trabajo vinculada con el análisis de datos,

aplicación de algoritmos de Machine Learning y la creación de modelos para realizar predicciones.

Las líneas de trabajo que se describen en este artículo se desarrollan en el LINTI y están enmarcadas en el proyecto: *De la Sociedad del Conocimiento a la Sociedad 5.0: un abordaje tecnológico y ético en nuestra región,* acreditado en el marco del Programa de Incentivos, bajo la dirección del Lic. Javier Díaz.

1. Introducción

El término Internet de las Cosas o Internet of Things (IoT) se refiere generalmente a escenarios donde la capacidad de cómputo y la conectividad de las redes se extienden a objetos, sensores y elementos cotidianos, permitiendo que estos dispositivos generen, intercambien y consuman datos. El concepto de combinar computadoras, sensores y redes para monitorear y controlar diferentes dispositivos ha existido durante décadas. Sin embargo, la reciente confluencia de diferentes tendencias del mercado tecnológico está permitiendo que la Internet de las Cosas esté cada vez más cerca de ser una realidad generalizada. Estas tendencias incluyen la conectividad omnipresente, la adopción generalizada de redes basadas en el protocolo IP, la economía en la capacidad de cómputo, los avances en el análisis de datos y el surgimiento de la computación en la nube.

Con todo esto, la implementación a gran escala de dispositivos de la IoT promete transformar muchos aspectos de la forma en que vivimos. Los nuevos productos de IoT para lograr "ciudades inteligentes" son una realidad y como era de esperar IoT también está alcanzando la industria del deporte. La utilización de dispositivos inteligentes o wearables y el análisis de datos que generan en tiempo real están cambiando el mundo del deporte, y los deportistas podrían optimizar su rendimiento y adecuar sus entrenamientos haciendo uso de estas tecnologías.

En este artículo se describen líneas de investigación y desarrollo que forman parte de un proyecto integral destinado a aplicar nuevas tecnologías en deporte y profundizar al análisis de datos que con ellas se generan.

2. Líneas de Investigación, Desarrollo e Innovación

Las líneas de investigación, desarrollo e innovación que se llevan a cabo en este proyecto están vinculadas con la aplicación de tecnologías en el deporte de alto rendimiento.

Los ejes principales de I+D+i son:

- Análisis de tecnologías de vanguardia para la construcción de dispositivos y software que mejore el entrenamiento y rendimiento de los jugadores.
- Construcción de dispositivos *low cost* para estimular el entrenamiento cognitivo y medir cargas externas.
- Diseño e implementación de una plataforma de software para la integración de datos y gestión de historia deportiva. Análisis inteligente de datos.

3. Resultados y Objetivos

Para esta línea de tecnologías aplicadas al deporte se ha trabajado en el desarrollo de dispositivos y software de administración asociado a ellos, se ha construido una plataforma de integración de datos deportivos y se están aplicando diferentes técnicas para analizar los datos recolectados. A continuación se describen los resultados de cada una de las líneas de

I+D+i mencionadas en el inciso anterior:

Análisis de tecnologías y construcción de dispositivos low cost

En el LINTI, desde hace algunos años se viene trabajando en la construcción de dispositivos para entrenamiento. Se han construído un primer prototipo de hardware con luces LEDs, que pueden ser configurados y operados desde dispositivos móviles para crear entrenamiento cognitivo y un primer prototipo con GPS para cuantificar las cargas de entrenamiento. Estos primeros diseños nos permitieron hacer pruebas de campo y evolucionar a dos nuevas versiones.

La Fig. 1 muestra el nuevo dispositivo evolucionado, donde se observa la placa de circuito impreso (PCB) de frente y dorso y el diseño de la carcasa.



Fig. 1: (a) Frente PCB (b)Dorso PCB (c) Carcasa

La ciencia está descubriendo que si bien toda la actividad física tiene un efecto positivo en el cerebro, el ejercicio que combina el uso de múltiples sentidos con movimientos corporales completos, desafía al cerebro a niveles más altos, porque requieren una función cognitiva más compleja para tomar decisiones y ejecutar habilidades. En consecuencia, la práctica de estas actividades mejora la toma de decisiones y el rendimiento físico, sensorial y neurológico de los deportistas (Lamberti, C., 2018). Este nuevo dispositivo, además de contar con una matriz de LEDs independientes que permite armar figuras geométricas de diferentes colores, cuenta con una nueva aplicación móvil para Android y para iOS para la creación de rutinas. La Fig. 2 muestra algunas capturas de pantalla:

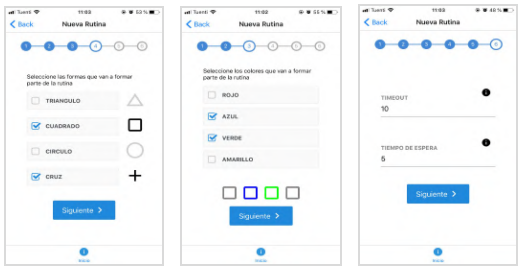


Fig. 2: Generación de rutina

Cabe destacar que esta aplicación está integrada con la plataforma para integración de datos que se describe más adelante en este artículo.

Asimismo se ha avanzado con otro dispositivo, *wereable*, para medir la carga externa de los deportistas. De estos dispositivos se presenta una segunda versión que combina básicamente GPS (los dispositivos con GPS (sistema de posicionamiento global), acelerómetro, magnetómetro y giroscopio. La aplicación de administración de este dispositivo registra los datos en la plataforma para integrarlos con los datos de los deportistas y equipos.

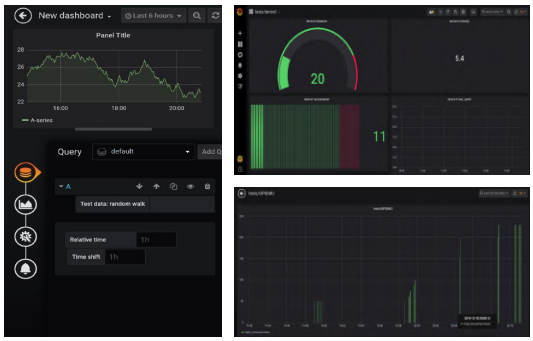


Fig 3.: Prototipo con GPS y acelerómetro

La Fig. 3 muestra algunas capturas de los datos graficados en Grafana, de

registros en una sesión de entrenamiento.

Plataforma de software para la integración de datos y gestión de historia deportiva.

El control de la carga de entrenamiento es una de las herramientas fundamentales para optimizar el rendimiento de los deportistas de alto nivel. Este control incluye no sólo las exigencias biológicas, sino también las psicológicas, y es el que permite individualizar y adaptar los procesos de entrenamiento y recuperación. Uno de los recursos usados para el seguimiento es un test llamado percepción subjetiva del esfuerzo (PSE), que usa una escala sencilla basada en la sensación personal de fatiga e intensidad del esfuerzo que siente el deportista y el otro es el test de bienestar (wellness) en el que se recuperan parámetros como fatiga, calidad del sueño y daño muscular de forma subjetiva.

En este contexto, se está desarrollando una plataforma de integración de datos que se alimenta automáticamente de los dispositivos *wereables* que usan los deportistas y manualmente por los diferentes usuarios (roles) del sistema. La carga manual incluye PSE, wellness, características físicas, aspectos técnicos, habilidades blandas y datos que permiten hacer seguimiento de lesiones.

A partir de este seguimiento, se está analizando establecer la relación que existe entre esta percepción del esfuerzo y cada una de las variables medidas por los dispositivos con GPS.

Los objetivos inmediatos son:

- Realizar nuevas pruebas de campo de los nuevos dispositivos. Analizar los resultados de las pruebas y realizar los ajustes necesarios, en los dispositivos . Mejorar tanto en el hardware de los dispositivos como en las aplicaciones de administración.
- Análisis descriptivo de la información consolidada permitiendo la realización de cruces de información de diferentes fuentes.
- Utilizar modelos de Machine Learning para predecir aspectos físicos/médicos que ayuden a prevenir lesiones en los deportistas.
- Crear herramientas que permitan facilitar el análisis del comportamiento de los equipos durante los partidos (Buldú, J.M. 2019, Kröckel, P. 2017).
- Crear modelos que permitan puntuar el desempeño de los jugadores y realizar scouting (Pappalardo, Luca 2019).

4. Formación de Recursos Humanos

El equipo de trabajo de la línea de I+D+i presentada en esta artículo se encuentra formado por docentes investigadores categorizados del LINTI y alumnos avanzados de la Licenciatura en Informática, Licenciatura en Sistemas e Ingeniería en Computación perteneciente a Facultad de Informática y a la Facultad de Ingeniería.

En relación a las tesinas de grado vinculadas con esta línea de investigación, se está dirigiendo a cuatro tesistas. Se han finalizado cuatro Prácticas Profesional Supervisadas (PPS). También se están ejecutando dos proyectos acreditados de la Secretaría de Políticas Universitarias del Ministerio de Educación y Deportes de la Nación.

Asimismo, este año se está ejecutando el proyecto Tecnologías aplicadas al Deporte: dispositivos de bajo costo y análisis de datos, financiado por la

Facultad de Informática de la UNLP.

5. Referencias

Alexandre D, da Silva CD, Hill-Haas S, et al. *Heart rate monitoring in soccer: interest and limits during competitive match play and training, practical application*. J Strength Cond Res. 2012;26(10): 2890–2906. doi:10.1519/JSC.0b013e3182429ac7

Buldú, J.M., Busquets, J., Echegoyen, I. et al. Defining a historic football team: Using Network Science to analyze Guardiola's F.C. Barcelona. Sci Rep 9, 13602 (2019). https://doi.org/10.1038/s41598-019-49969-2

Fava, L., Vilches, D., Romero Dapozzo, R., Pagano, M., *Tecnología aplicada al deporte de alto rendimiento*, XX Workshop de Investigadores en Ciencias de la Computación, Corrientes, Argentina, 2018.

Kröckel, P., Piazza, A. and Neuhofer, K. *Dynamic Network Analysis of the Euro 2016 Final: Preliminary Results*, 2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Prague, 2017, pp. 114-119.

Lamberti, C., *Brain Training Enhancing Athletic Performance*, SMARTfit Multisensory Fitness. https://www.ibm.com/developerworks/ssa/library/iot-mqtt-why-good-for-iot/index.html, Inc., Febrero 2018.

Pappalardo, Luca & Cintia, Paolo & Rossi, Alessio & Massucco, Emanuele & Ferragina, Paolo & Pedreschi, Dino & Giannotti, Fosca. (2019). *A public data set of spatio-temporal match events in soccer competitions*. Scientific Data. 6. 10.1038/s41597-019-0247-7.

Zeng, J. and Jia J., *The impact of big data on school sports and competitive sports*, 2017 Chinese Automation Congress (CAC), Jinan, 2017, pp. 596-599.

Recursos/Desarrollos Existentes

LongoMatch: análisis de video, accesible en https://longomatch.com/es/pro/, 2020.

Wyscout: professional football platform for football analysis, accesible en https://wyscout.com/, 2020.

Catapult: the tracking system, accesible en https://www.catapultsports.com/, 2020

Instat: sports performance analysis company, accesible en http://www.instatsport.com/, 2020.

Procesamiento de señales y Sistemas de Tiempo Real

MODELO MEJORADO PARA LA ESTIMACION DE PUNTOS DE BORDE EN IMÁGENES SAR POLARIMETRICAS

[1]Daniel Monferrán, [2]Andrés Sartarelli

[1]Instituto de Industria, [2]Instituto del Desarrollo Humano

Universidad Nacional de General Sarmiento

Juan María Gutiérrez 1150, (1613) Los Polvorines, Buenos Aires, Argentina\

[1]dmonferr@campus.ungs.edu.ar, [2]asartare@campus.ungs.edu.ar

RESUMEN

El Radar de Apertura Sintética polarimétrico es un tipo especial de radar ampliamente utilizado en teledetección, permite obtener imágenes de alta resolución a gran distancia. La interpretación automática de estas imágenes es una tarea difícil, contienen gran volumen de información y se encuentran contaminadas con ruido speckle no Gaussiano ni aditivo. Este ruido hace necesario utilizar métodos estadísticos para el procesamiento y análisis de estas imágenes. El modelo de distribución estadística más utilizado en este tipo de imágenes generadas a partir de datos multilook polarimétricos es la distribución Wishart compleja.

El tiempo de procesamiento para estimar la "posición de los puntos de bordes (PPB)" es relevante y existen trabajos de investigación que analizan y comparan los tiempos de procesamiento de diferentes métodos de estimación de PPB.

Este trabajo tiene por objetivo reducir el tiempo de procesamiento en la estimación de PPB. A tal efecto, se analizan técnicas de estimación de PPB que usan distribución Wishart compleja, máxima verosimilitud y distancias estocásticas. Aplicando el modelo de señal en tiempo discreto se obtienen se obtienen expresiones analíticas optimizadas para la estimación de la posición de los puntos de borde y se evalúan los tiempos de procesamiento mediante Montecarlo en imágenes simuladas.

Palabras clave: Imágenes SAR, Verosimilitud, Segmentación PolSAR, Distribución Wishart, Número de Looks

CONTEXTO

Este trabajo, corresponde al 'Doctorado de Ciencia y Tecnología de la Universidad Nacional de General Sarmiento (UNGS)', se inserta en el proyecto: ``Segmentación de imágenes generadas por fuentes coherentes", en curso en la UNGS, cuya finalidad es el estudio de algoritmos para la segmentación de imágenes obtenidas a partir de fuentes coherentes.

I. INTRODUCCIÓN

El Radar de Apertura Sintética polarimétrico (PolSAR - *Polarimetric Synthetic Aperture Radar)* es una herramienta eficaz para la teledetección geofísica. A diferencia de los sensores ópticos, el uso de sensores SAR en diferentes bandas, con distintas alternativas de ángulos de incidencia y polarizaciones múltiples, no dependen de la iluminación solar y no se ven comprometidos por perturbaciones atmosféricas como nubes o humo. Muchos trabajos han demostrado el potencial de las imágenes SAR [1], [2]. Los sistemas PolSAR transmiten y reciben pulsos electromagnéticos con diferentes combinaciones de polarización. Cada pixel de la imagen polarimétrica que se obtiene con esta tecnología representa la información registrada por el sensor en cada celda de resolución. El PolSAR no requiere de una fuente de iluminación externa, hace uso de su propia iluminación coherente y, por ello, se obtienen imágenes contaminadas con ruido speckle, característico de imágenes que se obtienen al iluminar con fuentes coherentes. Este ruido, que no es de carácter aditivo,

confiere un aspecto granular a las imágenes y hace necesario un especial modelado en su procesamiento y análisis. Una de las etapas importantes en el análisis de imágenes es encontrar los bordes de las mismas. Estos bordes delimitan diferentes regiones en la imagen que pueden corresponder por ejemplo a pasturas, zonas urbanas, zonas forestales, etc. Técnicas orientadas a la segmentación en este tipo de imágenes [3], [8] son de gran interés como etapa previa al reconocimiento y la interpretación de objetos. Los trabajos de investigación destinados a la determinación de los puntos de borde utilizan una gran variedad de técnicas, como por ejemplo la de "Contornos Activos", así como de atributos, por ejemplo: máxima verosimilitud, entropía, distancia estocástica, etc. [4], [5]. En [6] se emplean cinco métodos para la detección de bordes, entre ellos máxima verosimilitud. En el artículo [7] se propone un detector de borde para las imágenes SAR que produce un mapa de resistencia de borde sobre píxeles de la imagen, en [8] se desarrolla un proceso jerárquico de optimización paso a paso para la segmentación empleando el método de máxima verosimilitud, utilizan el modelo Gaussiano multivariado, la distribución Wishart y la distribución K. En [9] se propone la combinación de un algoritmo de segmentación basado en campos aleatorios de Markov estructurados en árbol y la clasificación orientada a objetos. Los algoritmos diseñados para determinar las posiciones de los puntos de borde en este tipo de imágenes deben mantener un equilibrio entre la precisión de los resultados obtenidos y el costo computacional. Para identificar puntos de borde en este tipo de imágenes, es necesario modelar los datos con una distribución estadística apropiada y luego estimar los parámetros correspondientes. En el artículo [10], los autores utilizan muestras con un número reducido de datos, con el propósito de reducir el costo computacional.

II. LÍNEA DE INVESTIGACIÓN Y DESARROLLO

El objetivo del plan de trabajo de tesis del doctorado es la segmentación automática de Imágenes SAR polarimétricas utilizando modelos estadísticos y contornos activos

El objetivo de este trabajo es encontrar expresiones analíticas optimizadas para la estimación de PPB, que puedan usarse en un algoritmo para la interpretación automática de imágenes PolSAR. Que además, pueda ejecutarse en un tiempo comparativamente menor con respecto a otros algoritmos tradicionalmente usados, sin perder precisión en la estimación de PPB.

A. Distribución Wishart

La función de densidad de la distribución Wishart compleja está indexada por dos parámetros: la matriz de covarianzas Σ y el número equivalente de looks L de la imagen.

$$f_Z(Z', \Sigma, L) = \frac{L^{mL}|Z'|^{L-m}}{|\Sigma|^L \Gamma_m(L)} exp\left(-L\, tr(\Sigma^{-1}Z')\right) \qquad (1)$$

Donde, Z' representa los posibles resultados de Z, Σ es el valor estimado de la matriz de covarianzas, L es el número equivalente de looks, m es el número de canales de polarización, $\Gamma_m(L)$ es la función Gamma polarimétrica, |.| es el operador determinante y tr(.) es el operador traza.

B. Función de Verosimilitud para Muestra de Datos

Dada una partición *A, B* de los datos sobre un segmento, la función de verosimilitud se expresa:

$$p(j) = \prod_{k=1}^{j} f_Z\left(Z'_{k;}\Sigma_A, L_A\right) \prod_{k=j+1}^{N} f_Z\left(Z'_{k;}\Sigma_B, L_B\right) \quad (2)$$

La matriz Z'_k es hermitiana, definida positiva y sigue una distribución Wishart compleja, j es la posición del pixel que determina la partición dentro del segmento, Σ_A y Σ_B son valores estimados de las matrices de covarianza de los conjuntos A y B que satisfacen la siguiente condición para Z'_k

$$Z_k^{(i)} \sim W\left(\Sigma_A^{(i)}, L_A^{(i)}\right),\ for\ k = 1, \dots, j^{(i)}, \quad (3),$$

$$Z_q^{(i)} \sim W\left(\Sigma_B^{(i)}, L_B^{(i)}\right),\ for\ q = j^{(i)} + 1, \dots, N^{(i)}$$

C. Posición del Punto de Borde

Siendo $\ell(j)$ el logaritmo de la función de verosimilitud, el estimador de la posición del punto de borde se expresa como:

$$\hat{J}_{ML} = \arg\max_j \ell(j) \quad (4)$$

D. Optimización de $\ell(j)$

Siendo que *A, B* pertenecen a una misma imagen, puede considerarse el mismo número equivalente de looks para ambos conjuntos, es decir $L_A = L_B = L$. En un proceso de manipulación algebraica, partiendo de la ec.(2) se obtiene la siguiente expresión analítica:

$$\begin{aligned}\ell(j) = L[-j\ log|\Sigma_A(j)| - (N-j)\ log|\Sigma_B(j)|] \\ - 3NL \\ + (L-m)\left(\sum_{k=1}^{N} log|Z'_k|\right) \\ + N[mL\log L \\ - \log\Gamma_m(L)]\end{aligned} \quad (5)$$

Esta expresión es similar a la que puede encontrarse en algunos trabajos de investigación que utilizan distribución Wishart y Logaritmo de Verosimilitud. En [10] la expresión usada es:

$$\ell(j) = N[-mL(1-\log L) - \log\Gamma_m(L)] + L\left[j log\left|\hat{\Sigma}_A(j)\right| + (N-j) log\left|\hat{\Sigma}_B(j)\right|\right] \quad (6)$$

La ec.(5) es una función de variable independiente *j,* si se la aplica sobre los datos de una muestra o segmento de *N* puntos, los siguientes términos no dependen de *j* y por lo tanto son ctes. para esa función:

$$N, m, L, \Gamma_m(L)\ y\ \sum_{k=1}^{N} log|Z'_k|$$

Se define la constante b y la función $\ell^*(j)$

$$\begin{aligned}b = -3NL + (L-m)\left(\sum_{k=1}^{N} log|Z'_k|\right) \\ + N[mL\log L \\ - \log\Gamma_m(L)]\end{aligned} \quad (7)$$

$$\begin{aligned}\ell^*(j) = -j\ log|\Sigma_A(j)| \\ - (N-j)\ log|\Sigma_B(j)|\end{aligned} \quad (8)$$

Resulta:

$$\ell(j) = L\ell^*(j) + b \quad (9)$$

Teniendo en cuenta que las señales discretas o señales para una variable discreta pueden representarse matemáticamente como funciones de esa variable discreta [12], la función $\ell(j)$ con $j \in N$ se puede interpretar como tal. Para dos señales discretas $f(j)\ y\ f^*(j)$ con $f(j) = a + b\ f^*(j)$ resulta ser b un factor constante que escala la señal $f^*(j)$ y a un término constante que desplaza $f^*(j)$. Vale decir que, $f(j)$ es la señal $f^*(j)$ escalada en b y desplazada en a.

Los valores máximos de $\ell(j)\ y\ \ell^*(j)$ se alcanza para el mismo valor de *j*. Dado que interesa hallar el valor *j* donde la función $\ell(j)$ presenta un máximo y no el valor $\ell(j)$, para estimar la posición de los puntos de borde se empleará la ec. (8) en lugar de la ec. (5).

El estimador de la posición del punto de borde para esta nueva expresión analítica es:

$$\hat{J}_{ML} = \arg\max_j \ell^*(j) \quad (10)$$

III. RESULTADOS OBTENIDOS/ESPERADOS

A. Imagen Sintética

En la Fig. 1 se ilustra en color rojo las posiciones estimadas de 4 puntos de borde en una imagen sintética de 421 x 421 píxeles - Canal HH. Las posiciones estimadas se obtuvieron aplicando las ecuaciones (5), (6) y (8), esas posiciones resultaron ser las mismas

independientemente de la ecuación usada.

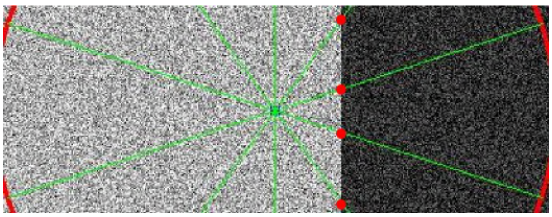


Fig. 1. Puntos de borde en una imagen sintética. Canal HH

B. Imagen SAR Polarimétrica

En las Figuras 2, 3 y 4 se ilustran en color rojo las posiciones estimadas de 11 puntos de borde. Se compararon las posiciones estimadas aplicando a las tres imágenes las ecuaciones (5), (6) y (8). Las posiciones estimadas resultaron ser las mismas independientemente de la ecuación usada

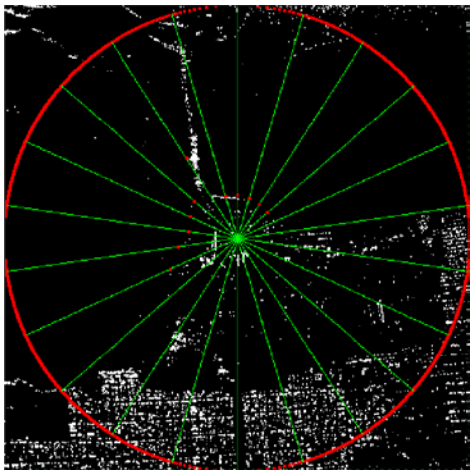


Fig. 2. Puntos de borde sobre una imagen SAR polarimétrica, Imagen de 363 x 363 píxeles. Canal HH. AIRSAR - Bahía de San Francisco, California, EE. UU

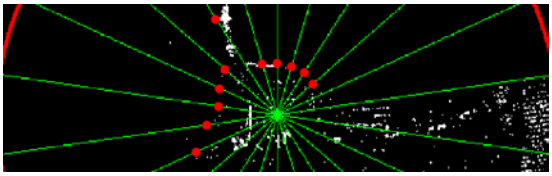


Fig. 3. Puntos de borde sobre una imagen SAR polarimétrica, Imagen de 363 x 363 píxeles. Canal HH. AIRSAR - Bahía de San Francisco, California, EE. UU

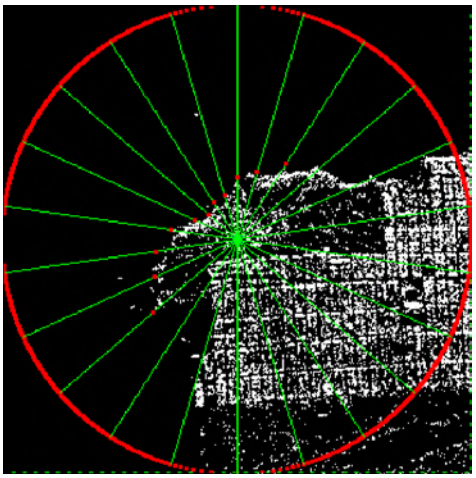


Fig. 4. puntos de borde en una imagen SAR polarimétrica, Imagen de 341 x 341 píxeles. Canal HH. AIRSAR - Bahía de San Francisco, California, EE. UU.

TABLA I
RELACIÓN DE TIEMPOS DE PROCESAMIENTO

Item	Región seleccionada	*Rel_TP*
Fig. 2	363 x 363 píxeles, 11 puntos,	0,7904
Fig. 4	341 x 341 píxeles, 11 puntos,	0,7757

La "relación de tiempos de procesamiento, (*Rel_TP)"* se calcula como el cociente de los tiempos de procesamiento medidos al estimar las PPB.

TiempoOpt es el tiempo medido cuando se usa la ec. (8) correspondiente a la expresión optimizada.

TiempoTrd es el tiempo medido cuando se usa la ec.(6) correspondiente a la expresión tradicional

$$rel_TP = \frac{TiempoOpt}{TiempoTrd}$$

En la Tabla I se observan valores de relación de tiempos menores a la unidad, es decir que el tiempo de procesamiento cuando se usa la ec.(8) es significativamente menor al tiempo de procesamiento que se requiere cuando se usa la ec.(6). Se ha podido alcanzar el objetivo principal de este trabajo de investigación.

Si se usa la ec.(8), aplicando distribución Wishart y el método de máxima verosimilitud para la estimación de la posición de los puntos de borde, es evidente lo innecesario de estimar o

conocer el número de looks *L, igualmente innecesario* calcular la función Gamma polarimétrica. Esto implica ahorro en el tiempo de procesamiento.

IV. FORMACIÓN DE RECURSOS HUMANOS

Tal como se menciona en la Sección I, este trabajo es uno de los trabajos que corresponde al 'Doctorado de Ciencia y Tecnología de la Universidad Nacional de General Sarmiento (UNGS)', se inserta en el proyecto: ``Segmentación de imágenes generadas por fuentes coherentes'', en curso en la UNGS, cuya finalidad es el estudio de algoritmos para la segmentación de imágenes obtenidas a partir de fuentes coherentes.

V. BIBLIOGRAFIA

[1] J. S. Lee and E. Pottier, Polarimetric Radar Imaging: From Basics to Applications, CRC Press: Boca Raton, FL, EE. UU., 2009.

[2] V. Akbari, A. Doulgeris and T. Eltoft, “Monitoring glacier changes using multitemporal multipolarization SAR images”, IEEE Trans. Geosci. Remote Sens., pp. 3729 –3741, Jun. 2014.

[3] K. Ersahin, I. Cumming, and R. K. Ward, “Segmentation and classification of polarimetric SAR data using spectral graph partitioning”, IEEE Trans. Geosci. Remote Sens., vol. 48, pp. 164–174, Sep. 2009.

[4] C. J. Oliver, D. Blacknell, and R. G. White, “Optimum edge detection in SAR”, IEE Proceedings - Radar, Sonar and Navigation vol. 143, pp. 31–40, Feb. 1996.

[5] R. Touzi, A. Lopes, and P. Bousquet, “A statistical and geometrical edge detector for SAR images”, IEEE Trans. Geosci. Remote Sens., vol. 26, pp. 764–773, Nov. 1988.

[6] J. Gambini, M. E. Mejail, J. Jacobo-Berlles and A. C. Frery, Accuracy of edge detection methods with local information in speckled imagery, ”Statistics and Computing”, vol. 18, pp. 15–26, Mar. 2008.

[7] R. Fjortoft, A. Lopes, P. Marthon and E. Cubero-Castan, “An Optimal Multiedge Detector for SAR Image Segmentation”, IEEE Trans. Geosci. Remote Sens., vol. 36, pp. 793–802, May. 1998.

[8] J.-M. Beaulieu and R. Touzi, “Segmentation of textured polarimetric SAR scenes by likelihood approximation”, IEEE Trans. Geosci. Remote Sens., vol. 42, pp. 2063–2072 Oct. 2004.

[9] C. D’Elia, S. Ruscino, M. Abbate, B. Aiazzi, S. Baronti and L. Alparone, “SAR Image Classification Through Information-Theoretic Textural Features, MRF Segmentation, and Object-Oriented Learning Vector Quantization”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, pp. 1116–1126, Apr. 2014.

[10] A. D. C. Nascimento, M. M. Horta, A. C. Frery and R. J. Cintra, “Comparing edge detection methods based on stochastic entropies and distances for Polsar imagery”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, pp. 648–663, Feb. 2014.

Monitoreo de Catástrofes Naturales a partir de la Obtención y Procesamiento de Imágenes Satelitales

Paula B. Olmedo[*], Marco Miretti[†], Emanuel Bernardi[†], Javier Redolfi[†], Gastón Peretti[†], Eduardo J. Adam[‡]
[*]Universidad Tecnológica Nacional, Facultad Regional Córdoba (UTN-FRC)
Córdoba Capital, Argentina.
paulabeatrizolmedo@gmail.com
[†]Universidad Tecnológica Nacional, Facultad Regional San Francisco (UTN-FRSFCO)
San Francisco, Córdoba, Argentina.
marco.miretti@gmail.com, ebernardi@sanfrancisco.utn.edu.ar,
javierredolfi@gmail.com, gastonperetti@gmail.com
[‡]Universidad Nacional del Litoral, Facultad de Ingeniería Química (UNL-FIQ)
Santa Fe Capital, Argentina.

RESUMEN

En la actualidad, el desarrollo de la industria aeroespacial a la par del desarrollo tecnológico, facilita el acceso a la información brindada por los satélites. En este contexto, el trabajo que a continuación se describe tiene por objetivo captar señales satelitales, especialmente imágenes, con el fin de almacenarlas y procesarlas en tiempo real para obtener modelos que auxilien en el análisis y prevención de emergencias ambientales. Es de destacar que este trabajo forma parte de un proyecto de mayor alcance, cuya primera etapa consistió en la construcción de una estación terrena para la adquisición de las imágenes de interés.

Palabras clave: procesamiento de imágenes, emergencias ambientales, comunicación, estación terrena.

CONTEXTO

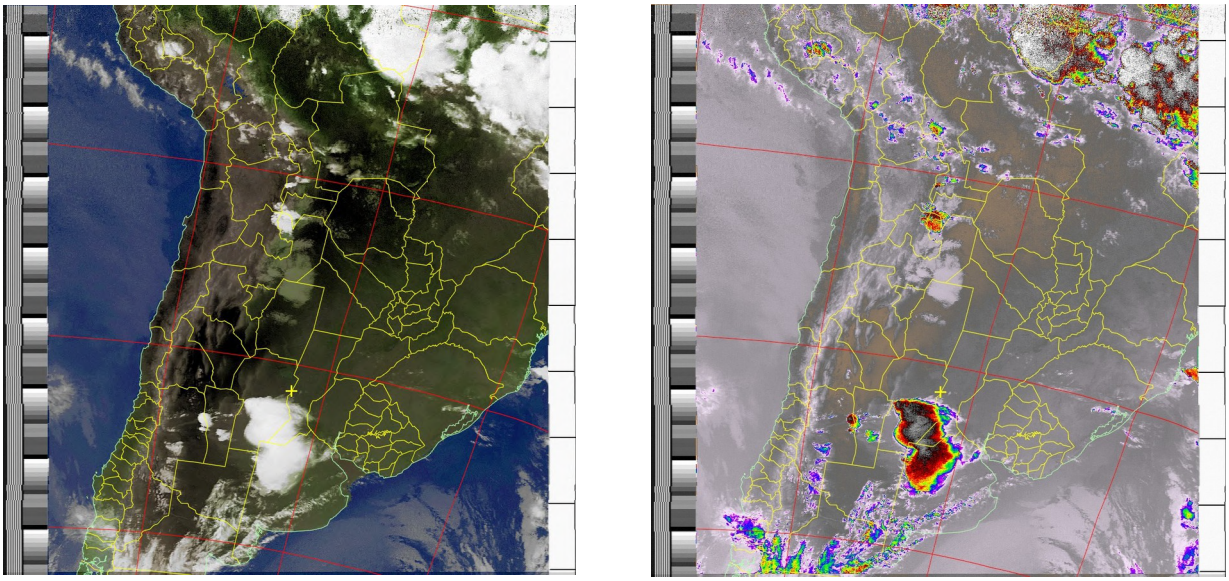
La siguiente propuesta de investigación se enmarca dentro de las áreas Control Automático de Sistemas y Telecomunicaciones del Departamento de Ingeniería Electrónica de la Universidad Tecnológica Nacional, Facultad Regional San Francisco. Específicamente, a través del PID-UTN CCUTNSF0005361. En tanto, las actividades involucradas en el mismo se llevarán a cabo por los integrantes del Grupo de Investigación sobre Control Aplicado y Sistemas Embebidos *(AC&ES-RG)*[1], el cual se compone de becarios de grado, doctorales y docentes investigadores.

1. INTRODUCCIÓN

Debido a su extensión y variedad de biomas, los fenómenos naturales con consecuencias negativas son frecuentes en la Argentina. Por ejemplo, desde hace ya varios años son comunes las inundaciones, las cuales han afectado entre tres y seis millones de hectáreas de superficie [1]. En ciertos casos, dichas áreas son difíciles de dimensionar debido a sus magníficas proporciones y variedades topográficas, sin mencionar que la estimación de daños también resulta una complicación que aún no está del todo resuelta.

Por otro lado, en las sierras cordobesas, los incendios forestales progresivamente se han convertido en un factor cada vez más influyente en la reducción de la riqueza de su flora [2]. En este caso, las dificultades se presentan a la hora de predecirlos y detectarlos con el suficiente tiempo para dar una respuesta adecuada, así como también para el monitoreo de su evolución. Aunque existen herramientas para solventar estas necesidades [3], éstas no se encuentran optimizadas para actuar en nuestro contexto y/o el acceso a muchas de ellas, es restringido.

1 del inglés, *Applied Control & Embedded Systems - Research Group*



(a) Color real (b) Intensidad de precipitaciones

Figura 1: Imágenes de NOAA-19 del 24 de Octubre de 2019.

Diversos autores han calificado a la Argentina como un "País espacial" [4], a causa de una necesidad socio-económica de aprovechar el acceso a este tipo de información, y los múltiples proyectos factibles en esta área. Por esta razón, trabajar con imágenes satelitales brinda la capacidad de proceder con celeridad frente a catástrofes naturales, así como también a realizar una correcta gestión una vez ocurrido el suceso. Es de destacar, que un adecuado procesamiento puede ofrecer tanto información en 2D como en 3D. Ejemplo del primer caso es el análisis del Índice Diferencial de Agua Normalizado *(NDWI)*[2], con el que se obtiene valiosa información para la detección de áreas de sequía [5]. En cambio, una aplicación del segundo caso es la generación de modelos digitales de elevación, los mismos ayudan a estimar deslizamientos de suelo [6], u optimizar la asignación de recursos ante un desastre [7].

Es interesante resaltar la existencia de satélites de acceso abierto, lo que significa que las señales que éstos transmiten no se encuentran cifradas y además realizan una constante emisión de información, posibilitando su recepción desde cualquier parte del mundo, resultando de utilidad para este tipo de proyectos. En la Fig. 1b, se observa la imagen

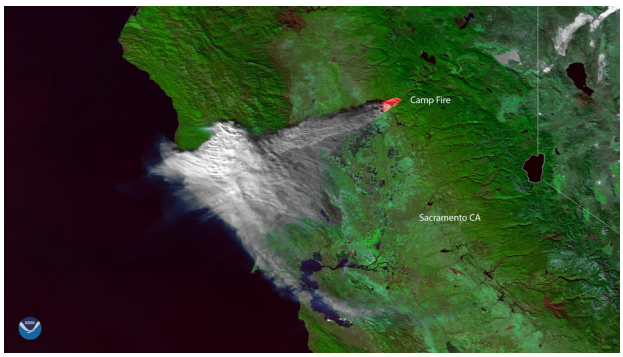


Figura 2: Imagen de incendio desde el NOAA-20.

adquirida desde un satélite NOAA[3] donde se aprecia la intensidad de las precipitaciones sobre nuestro país. Adicionalmente, se muestra en la Fig. 2 un incendio forestal, monitoreado con herramientas de procesamiento de información satelital [8].

Como se mencionó anteriormente, en etapas tempranas de este proyecto [9] se construyó una antena Yagi de alta ganancia, con la capacidad de recibir señales de satélites no-geoestacionarios. Ésta, se probó receptando una imagen proveniente del NOAA-19. Se observan los resultados en la Fig. 1. Posteriormente, se desarrolló un sistema de posicionamiento de antena con motores paso a paso, controlados por el software *GPredict*, ejecutándose en una plataforma *Raspberry-Pi*, y finalmente se logró la recepción de imágenes a través de un

2 del inglés, *Normalized Difference Water Index.*

3 del inglés, *National Oceanic and Atmospheric Administration*.

dispositivo SDR[4] configurado adecuadamente utilizando el software *Gqrx*.

A partir de esto, se propone captar y almacenar las imágenes recuperadas, en una base de datos de tipo *NoSQL*, las cuales permiten no solamente el almacenamiento de grandes volúmenes de datos, sino también de diversos tipos [10]. Luego, se trabajará en el procesamiento de éstas en dos etapas:

- Obtención de parámetros normalizados, como el NDWI, con la motivación de obtener datos relevantes sobre el estado del suelo y así determinar un nivel de riesgo para fenómenos como incendios o inundaciones. Inicialmente se ajustarán los parámetros necesarios para el análisis manualmente, luego se incorporará el uso de redes neuronales convolucionales [11].
- Generación de modelos de elevación, que permiten la medición y estimación de los niveles de riesgo, además de que resultan de utilidad a la hora de medir las consecuencias de las catástrofes naturales [7] por tratarse de una mejor aproximación a la realidad de los terrenos a tratar. Para esto, será necesaria una retroalimentación con la primera etapa.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El presente proyecto se compone de los siguientes ejes de investigación y desarrollo:

- Análisis de las necesidades de hardware y software requeridas para un sistema de almacenamiento y procesamiento de la información obtenida en la estación terrena.
- Estudio de las necesidades de inforomación satelital locales y regionales específicas.
- Evaluación de la estrategia de procesamiento digital de imágenes a emplear, en base a la necesidad propuesta.
- Desarrollo de un software de predicción, modelado y detección de diversos fenómenos.
- Vinculación con otros proyectos de colaboración internacional de estaciones terrenas.
- Escritura de documentación clara que permita un desarrollo continuado y abierto.
- Divulgación de los resultados parciales y finales de la investigación.

3. OBJETIVOS

Objetivo General:

Diseñar y desarrollar un sistema de procesamiento de imágenes satelitales con el fin de realizar un análisis estratégico de determinadas variables de interés en un contexto regional.

Objetivos específicos:

- Configurar un sistema de almacenamiento de imágenes.
- Diseñar y desarrollar un sistema de procesamiento de imágenes provenientes de satélites no-geoestacionarios.
- Determinar las variables de interés para el contexto regional.
- Analizar las imágenes procesadas en función de las variables seleccionadas.
- Examinar la información recopilada para determinar nuevas posibilidades del sistema.
- Estimar la dimensión de catástrofes naturales, a partir de las imágenes obtenidas.

4. RESULTADOS OBTENIDOS/ESPERADOS

Considerando que este trabajo forma parte de un proyecto de mayor dimensión, el cual consiste en el montaje completo de una estación terrena en la Universidad Tecnológica Nacional (Facultad Regional San Francisco) para la adquisición y procesamiento de imágenes, se debe aclarar que el proyecto ya posee algunos resultados obtenidos [9], como los descriptos a continuación:

- Cálculo y construcción de una antena Yagi direccional de alta ganancia.
- Desarrollo de un sistema de posicionamiento automático, utilizando

4 del inglés, *Software Defined Radio*.

información de apuntamiento brindada por el software *Gpredict* [12].

- Construcción de una estructura apta para soportar dicho sistema junto a la antena y el montaje de los mismos.

En cuanto a los objetivos puntuales de este tramo del proyecto:

- Adquisición de imágenes actualizadas de nuestra región.
- Desarrollo de la estrategia de procesamiento de las imágenes recibidas de acuerdo a cada necesidad propuesta.
- Colaboración con los proyectos internacionales pre-existentes.
- Divulgación en el medio de las capacidades del proyecto y sus resultados.
- Formación de recursos humanos capaces de continuar proyectos afines a esta línea de investigación.

5. FORMACIÓN DE RECURSOS HUMANOS

El presente trabajo se encuentra enmarcado en el Proyecto de Investigación y Desarrollo (CCUTNSF0005361) del Departamento de Ingeniería Electrónica de la Universidad Tecnológica Nacional, Facultad Regional San Francisco. El grupo de trabajo esta compuesto por docentes investigadores y estudiantes, entre ellos el Dr. Eduardo J. Adam, el Dr. Javier Redolfi y los integrantes de *AC&ES-RG*.

AC&ES-RG está conformado por dos becarios doctorales, tres ingenieros electrónicos y cinco estudiantes de las carreras Ingeniería Electrónica, Industrial y Sistemas de Información. Por último, es de destacar que el presente proyecto forma parte de la tesina de grado de una estudiante del grupo.

AGRADECIMIENTOS

Los integrantes del artículo desean agradecer al Ing. Emmanuel M. Dovis, integrante del proyecto de investigación y desarrollo en el cual se enmarca este trabajo, por facilitarnos las imágenes que se muestran en la Fig 1.

6. BIBLIOGRAFÍA

[1] Claudia E. Natenzon. «Riesgo, vulnerabilidad e incertidumbre. Desastres por inundaciones en Argentina». En: *Provincia* 2.609 (1998), 69e.

[2] Graciela Verzino y col. «Impacto de los incendios sobre la diversidad vegetal, Sierras de Córdoba, Argentina». En: *Ecología Aplicada* 4.1-2 (2005), págs. 25-34.

[3] Claire Miller y col. «SPARK–A bushfire spread prediction tool». En: *International Symposium on Environmental Software Systems.* Springer. 2015, págs. 262-271.

[4] C. M. Scavuzzo y col. «Satellite image applied to epidemiology, the experience of the Gulich Institute in Argentina». En: *Space Technology for E-health* (2007), pág. 49.

[5] Marco Miretti y col. «Detección y Control de Malezas a través de la Evaluación de Parámetros Normalizados». En: Libro de resúmenes del *IX Congreso de Microelectrónica Aplicada, uEA2018.* San Fernando del Valle de Catamarca, Argentina: Editorial Científica Universitaria de la Universidad Nacional de Catamarca, 2018. ISBN: 978-987-661-325-5.

[6] José Gregorio Roa. «Estimación de áreas susceptibles a deslizamientos mediante datos e imágenes satelitales: cuenca del río Mocotíes, estado Mérida-Venezuela». En: *Revista Geográfica Venezolana* 48.2 (2007), págs. 183-219.

[7] Facundo Ismael Casasola. «Implementación de herramientas de Teledetección e Inteligencia Artificial para la optimización de los recursos del Ejército Argentino en situaciones de emergencias volcánicas». En: *Universidad Nacional de Córdoba* (2016).

[8] National Oceanic y Atmospheric Administration. Plumes of Smoke Cover Portions of Northern California. URL: https://www.nesdis.noaa.gov/content/plumes-smoke-cover-portions-northern-california (visitado 09-11-2018).

[9] Facundo Busano y col. «Estación terrena de adquisición de señales de satélites no geoestacionarios». En: *WICC2018*, 2018, págs. 978-981. ISBN: 978-987-36-1927-4.

[10] Christof Strauch, Ultra-Large Scale Sites y Walter Kriha. «NoSQL databases». En: *Lecture Notes, Stuttgart Media University 20* (2011), pág. 24.

[11] Samer Hijazi, Rishi Kumar y Chris Rowen. «Using convolutional neural networks for image recognition» .En: *Cadence Design Systems Inc*. San Jose, CA, USA (2015), págs. 1-12.

[12] Alexandru Csete.GPredict project.

Modelos Matemáticos y Métodos Computacionales en Ingeniería

Javier Giacomantone - Oscar Bria

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP

Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)
La Plata, Buenos Aires

{jog, onb}@lidi.info.unlp.edu.ar

Resumen

Este trabajo describe una línea de investigación y sus objetivos generales y particulares. El objetivo principal es estudiar y evaluar modelos matemáticos y métodos numéricos que permitan abordar problemas específicos en ingeniería. También se estudian sistemas y procesos, que por su nivel de complejidad, requieren abordajes y trabajos multidisciplinarios para operar en el dominio del problema de interés. Los modelos abordados son dependientes del tipo de sistema estudiado, del fenómeno analizado y del área particular de ingeniería que originó el requerimiento. Determinar el tipo de sistema, el método para evaluar su rendimiento y las soluciones numéricas óptimas o sub-óptimas forma parte de los objetivos en esta línea de investigación.

Palabras Clave: modelos matemáticos aplicados, métodos computacionales, modelos probabilísticos, desempeño de sistemas, confiabilidad de sistemas.

Contexto

Esta línea de investigación y desarrollo (I/D) forma parte del proyecto "Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real". En particular del sub-proyecto "Modelos y métodos computacionales. Procesamiento de señales y reconocimiento de patrones".

1. Introducción

En ingeniería y en ciencias básicas se utilizan distinto tipo de modelo para estudiar y caracterizar objetos y fenómenos de interés. Los modelos matemáticos nos permiten predecir fenómenos naturales y el comportamiento de estructuras, dispositivos, procesos y sistemas diseñados por el hombre. Los métodos computacionales asociados a modelos determinísticos y probabilísticos permiten estudiar el dominio del problema considerando distintas condiciones estructurales y funcionales. El análisis del comportamiento a medida que el objeto o fenómeno real se aleja de las hipótesis fundamentales del modelo permite determinar su alcance, su validez y su utilidad respecto a los objetivos iniciales [1][2]. Este abordaje permite implementaciones en *hardware* o *software* que ofrezcan soluciones viables. Los sistemas estudiados pueden ser lineales o no lineales requiriendo modelos sofisticados [3][4]. Determinar si una solución es viable con fundamento científico, es una tarea compleja y dependiente del problema particular analizado [5]. Involucra estudiar el dominio del problema, los objetos o los fenómenos modelados, y por lo tanto eventualmente requiere soporte multidisciplinar. Este proyecto tiene como primer objetivo analizar y proponer modelos computacionales, métodos y las soluciones

particulares derivadas de los mismos. Alcanzar los objetivos anteriores requiere estudiar los fundamentos que subyacen a cada modelo, evitando soluciones, que por su nivel de encapsulamiento, limiten una verdadera comprensión y abordaje científico de las mismas [6]. El tipo de problema de interés en esta línea de investigación requiere la integración de soluciones de tres áreas, ciencias de la computación, matemáticas aplicadas y un área de ingeniería o ciencia básica en particular. La evaluación de rendimiento es un aspecto fundamental para poder validar las soluciones propuestas o los modelos analizados [7]. Por lo tanto, otro aspecto fundamental es el estudio de las métricas y paradigmas de desempeño en sistemas específicos.

En la sección 2 se presenta un breve resumen de los temas de I/D específicos en el período actual. La sección 3 enumera resultados obtenidos y esperados. Finalmente, la sección 4 resume los objetivos con respecto a la formación de recursos humanos.

2. Líneas de Investigación

2.1 Modelos Probabilísticos Aplicados: Procesamiento y Análisis de Imágenes.

Esta línea de investigación estudia problemas derivados de ciertas modalidades de imágenes. Particularmente en las que la formación de la imagen digital de interés, no es de formación directa. Es necesario, por lo tanto, estudiar el fenómeno físico relacionado, el modelo de formación y los métodos computacionales asociados. El procesamiento y análisis de imágenes de tiempo de vuelo (TOF) requieren este enfoque para obtener soluciones robustas. Las cámaras de TOF permiten obtener simultáneamente una imagen 2D y otra denominada 2 ½ D o imagen de rango. El ruido y los artefactos en este tipo de imagen requieren de los modelos adecuados para caracterizar, filtrar y eventualmente reducir efectos indeseados. Segmentar imágenes de tiempo de vuelo y otras modalidades con estructuras de datos similares puede requerir, dependiendo del objetivo, modelos de segmentación con características específicas [8][9][10]. En particular métodos estadísticamente robustos [11][12]. El objetivo principal al investigar modelos específicos de análisis de imágenes es mejorar la calidad de los descriptores obtenidos en función de su impacto en el sistema de clasificación. Se analizan métodos de clasificación supervisada y semi-supervisada en imágenes TOF, en particular basados en núcleos dispersos [13][14][15].

2.2 Desempeño de Sistemas de Posicionamiento, Navegación y Vigilancia.

En los sistemas de posicionamiento, de navegación y de localización [16][17], el concepto de desempeño excede al habitual que está limitado a la calidad nominal de la estimación de ubicación y eventualmente a la confiabilidad [18][19]. En estos sistemas deben considerarse además los parámetros de integridad y continuidad que le garanticen al usuario que la información proporcionada por el sistema es correcta para que una operación crítica pueda realizarse en forma segura [20][21] .
Un tema relacionado con el desempeño de los sistemas vigilancia es el volumen de transacciones con características aleatorias [22].
Esta línea de trabajo se avoca al estudio de problemas puntuales de desempeño en los sistemas mencionados, utilizando criterios y métodos diversos de modelado, procesamiento y análisis [23][24].

3. Resultados y Objetivos

3.1 Resultados publicados

- Se estudiaron y propusieron métodos para detección en series temporales de fMRI [25][26].
- Se desarrollaron métodos de segmentación de imágenes de rango y supresión del plano de fondo [27][28][29][30].

- Se analizaron y propusieron alternativas para el agrupamiento de objetos de interés en video [31].
- Se estudió el desempeño de un método de exclusión de satélites en un sistema de ayuda a la aeronavegación basado en GNSS [32].
- Se propuso y se presentaron resultados experimentales de un método de aprendizaje basado en problemas para aritmética computacional [33].
- Se propuso un método de segmentación espectral para imágenes de tiempo de vuelo [34].
- Se estudió el comportamiento bajo carga de un algoritmo para programar transacciones de radares aeroportuarios [35].

3.2 Objetivos generales

- Desarrollar modelos y optimizar algoritmos particulares de clasificación supervisada y no supervisada.
- Evaluar métodos de análisis de desempeño y su aplicación sobre los clasificadores y conjuntos de datos particulares.
- Evaluar la monitorización de la integridad de los sistemas de ayuda a la navegación aérea basados en sistemas GNSS.
- Estudiar métodos de selección y extracción de características.
- Investigar modelos y métodos computacionales en procesamiento y análisis de imágenes.
- Promover la interacción con otros grupos y líneas de I/D resultando en un mecanismo de permanente consulta y transferencia.

4. Formación de Recursos Humanos

Se dictan cursos de postgrado que tienen por objetivo formar alumnos en temas específicos y fundamentos relacionados a esta línea de investigación. Los alumnos tienen la posibilidad de realizar trabajos supervisados de investigación. Debido al carácter trans-disciplinar de la línea de I/D, se espera orientar y brindar apoyo a investigadores y alumnos de otras líneas de investigación relacionadas.

BIBLIOGRAFIA

1. Zalizniak V. Essentials of Scientific Computting – Numerical Methods for Science and Engineering. Woodhead Publishing, 2008.
2. Juergen G. Coupled Systems: Theory, Models, and Applications in Engineering. CRC, 2014.
3. Torokhti A., Howlett P. Computational Methods for Modelling of Nonlinear Systems. Elsevier, 2007.
4. Canuto C., et al. Spectral Methods. Evolution of Complex Geometries and Applications to Fluids Dynamics Scientific Computation. Springer, 2007.
5. Ciurpina G. Scientific Computing in Electrical Engineering Springer, 2007.
6. Gustafsson B. Fundamentals of Scientific Computing Springer, 2011.
7. Aslak T., et al. Elements of Scientific Computing. Springer, 2010.
8. Von Luxburg U. A Tutorial on Spectral Clustering. Statistics and Computing, v.17(4), 2007.
9. Kim H.Y., Giacomantone J. O., Cho, Z. H. Robust Anisotropic Diffusion to Produce Enhanced Statistical Parametric Map, Computer Vision and Image Understanding, v.99, pp.435-452, 2005.
10. Han Y., Feng X., Baciu G. Variational and PCA based natural image segmentation. Pattern Recognition 46, pp. 1971-1984, 2013.
11. Fukunaga K. Introduction to Statistical Pattern Recognition. Second Edition. Academic Press, 1990.
12. Devijer P, Kittler, J. Pattern Recognition: theory and applications. Springer, 1986.
13. Corte C, Vapnik V, Support vector networks. Machine Learning v.20, pp.273-297, 1995.

14. Vapnik, V. The Nature of Statistical Learning Theory. N. Y. Springer, 1995.

15. Aytug H. Feature selection for support vector machines using Generalized Benders Decomposition. European Journal of Operational Research, v. 244(1), pp. 210-218, 2015.

16. Partap Misra, Per Enge. Global Positioning System: Signals, Measurements and Performance, Ganga-Jamuna Press, 2010.

17. Hakan Koyuncu, Shuang Hua Yang. A Survey of Indoor Positioning and Object Locating Systems Indoor Positioning System, International Journal of Computer Science, 2010.

18. Petevelo Mark. Quantifying the performance of Navigation Systems and Standars for assisted-GNSS, Inside GNSS, 2008.

19. Morurikis A., Roumeliotis S. Performance Analysis of Multirobot Cooperative Localization, IEEE, 2005.

20. Murphy T., et. al., Fault Modeling for GBAS Airworthiness Assessments, Navigation, 2012.

21. Cosmen-Schortmann J., Azaola-Sáenz, Martínez-Olagüe M. A., Toledo-López M., Integrity in Urban and Road Environments and its use in Liability Critical Applications, IEEE, 2008.

22. Pengfei Duan, Maarten Uijy De Haa. Flight Test Results of a Measurement-Based ADS-B System for Separation Assurance, Navigation, 2013.

23. Shuo-Ju Yeh1, Shau-Shiun, GBAS airport availability simulation tool, GPS Solutions, 2015.

24. Sam Pullen, Todd Walter, Per Enge. SBAS and GBAS Integrity for Non-Aviation Users: Moving Away from "Specific Risk," International Technical Meeting of The Institute of Navigation, 2011.

25. Giacomantone J., Tarutina T. Diffuse Outlier Detection Technique for Functional Magnetic Resonance Imaging. Computer Science and Technology Series. XVI Argentine Congress of Computer Science Selected Papers. pp. 255-265, 2011.

26. Giacomantone J., De Giusti A. Detección de áreas de interés bajo la hipótesis de relación espacial de voxels activados en fMRI. XX Congreso Argentino de Ciencias de la Computación. San Justo. Argentina, 2014.

27. Lorenti L., Giacomantone J. Segmentación espectral de imágenes utilizando cámaras de tiempo de vuelo. XI Workshop Computación Gráfica, Imágenes y Visualización. pp. 430-439. Mar del Plata, Argentina, 2013.

28. Lorenti L., Giacomantone J. Time of flight image segmentation through co-regularized spectral clustering. Computer Science & Technology Series. XX Argentine Congress of Computer Science. Selected papers. La Plata, Editorial de la Universidad Nacional de La Plata, 2015.

29. Giacomantone J., et al. Supresión del plano de fondo en imágenes de tiempo de vuelo. VII Workshop Procesamiento de Señales y Sistemas de Tiempo Real, 2016.

30. Lorenti, L., Giacomantone, J., Bria, O. N., De Giusti, A. E. (2017). Fusión de información de geometría e intensidad para segmentación de imágenes TOF. XXIII CACIC. La Plata, Argentina, 2017.

31. Lorenti L., Giacomantone J., De Giusti A. Agrupamiento de trayectorias vía clustering espectral incremental. XXII CACIC, pp. 222-231, 2016.

32. Bria, O., Giacomantone, J., Lorenti, L., Excluding Ionospherically Unsafe Satellite Geometries in GBAS CAT-I. XXII CACIC. CCIS 790: Revised CACIC Selected Papers, 790, pp. 243–252, 2018.

33. Giacomantone, J., Bria, O., Proactive Independent Learning Approach: A case study in computer arithmetic. XXII CACIC. La Plata, Argentina, 2017.

34. Lorenti L., Giacomantone J., Bria O., Unsupervised TOF Image Segmentation through Spectral Clustering and Region Merging. Journal of Computer Science & Technology, v.18(2), pp. 97-104, 2018.

35. Bria O., Giacomantone J., Villagarcía H., Compound Interleaving Scheduling Algorithm for SLM Transactions in Mode S Surveillance Radar. Comunications in Computer and Information Science 995, pp. 297-312. Springer, 2019.

Identificación y compensación del sesgo en sensores inerciales MEMS de muy bajo costo mediante el uso de algoritmos de aprendizaje automático

Rodrigo Gonzalez [1,2], Carlos A. Catania [1,3], Javier Rosenstein [1] y Fernando O. Pinciroli [1]

[1] Universidad Champagnat, Facultad de Informática y Diseño, Mendoza, Argentina.
{gonzalezrodrigo, rosensteinjavier, pincirolifernando}@uch.edu.ar

[2] Universidad Tecnológica Nacional, Facultad Regional Mendoza, GridTICs, Mendoza, Argentina.

[3] Universidad Nacional de Cuyo, Facultad de Ingeniería, LABSIN Campus Universitario, Mendoza, Argentina.
harpo@ingenieria.uncuyo.edu.ar

RESUMEN

En los últimos años, el uso de sensores inerciales basados en tecnología MEMS (MicroElectroMechanical Systems) se ha consolidado en diferentes áreas que requieren sensores baratos, livianos y de bajo consumo de energía. Algunas de estas áreas son la robótica móvil aérea y la robótica móvil terrestre. Una desventaja que presentan este tipo de sensores es su bajo desempeño comparado con sensores fabricados con otras tecnologías. Los dos errores más gravitantes en el desempeño de un sensor inercial MEMS son el ruido blanco y el sesgo presentes en la señal de salida del sensor. En particular, el sesgo presenta una variación no lineal muy notoria respecto a la temperatura, debido a que los sensores MEMS están construidos mayoritariamente con silicio. En la literatura existente, si bien se encuentran varios trabajos que han tratado de corregir el sesgo en sensores inerciales MEMS utilizando técnicas basadas en aprendizaje automático (machine learning), debido a la naturaleza no lineal del fenómeno que se trata de predecir, todos estos trabajos han utilizado sensores MEMS cuyo costo varía entre los USD 300 y los USD 1.000. En los últimos años han surgido sensores inerciales MEMS de muy bajo costo, los cuales se encuentran en una cantidad importante de sistemas robóticos. El costo de estos sensores ronda los USD 10. En este trabajo se propone analizar y compensar las variaciones del sesgo por temperatura en sensores inerciales MEMS de muy bajo costo, usando diferentes técnicas de aprendizaje automático. La hipótesis central de este proyecto se basa en mejorar el rendimiento de sensores inerciales de muy bajo costos aplicando técnicas de aprendizaje automático del estado del arte, las cuales permitirán identificar la relación no lineal que existe entre el sesgo y la temperatura.

Palabras Claves: Sensores inerciales, MEMS, Sesgo, Aprendizaje automático.

CONTEXTO

El presente proyecto se desarrolla en el marco de la Facultad de Informática y Diseño, junto con el grupo de investigación GridTICs de la Universidad Tecnológica Nacional, Facultad Regional Mendoza. Este trabajo es parte de las líneas de investigación llevadas adelante por el Dr. Rodrigo Gonzalez junto al Dr. Carlos A. Catania desde el año 2016.

1. INTRODUCCIÓN

El uso de sensores inerciales basados en tecnología MEMS (MicroElectroMechanical Systems) se ha consolidado últimamente en ciertas actividades, como por ejemplo en la

fabricación de sistemas de navegación integrados (INS/GNSS) de bajo costo tanto para el control de multicópteros, en la localización en tiempo real de diferentes agentes (bomberos, deportistas, etc.) o para la evaluación de su desempeño físico. Los sensores MEMS poseen tres características fundamentales para ser usados en las aplicaciones mencionadas: bajo peso, bajo consumo de potencia y relativo bajo costo. Como contrapartida, presentan mayores errores cuando se los compara con sensores fabricados con otras tecnologías, como los realizados con fibra óptica. Los errores que tienen mayor influencia en el desempeño de sensores inerciales MEMS son el ruido blanco y el sesgo (*bias*) (Groves, 2008, Sec. 4.4.5).

Debido a que los sensores MEMS están construidos mayoritariamente con silicio, el sesgo presente en la salida es muy dependiente de la temperatura (Fontanella et al., 2018). Comúnmente las variaciones del sesgo con la temperatura se corrigen a través de tablas. Dichas tablas se generan en laboratorios para una unidad en particular. Se coloca el sensor MEMS dentro de una cámara térmica y se varía la temperatura. De esta forma se puede mapear la relación que existe entre temperatura y sesgo para un determinado sensor inercial. Posteriormente, se confecciona una tabla que podrá ser usada para compensar la variaciones del sesgo durante la operación del sensor. Debido a la fuerte relación no lineal que existe entre sesgo y temperatura en sensores inerciales MEMS, el uso de tablas es sub óptimo.

En la literatura existente se encuentran diversos trabajos que han propuestas diferentes técnicas para la compensación no lineal del sesgo en sensores inerciales MEMS. Un método conveniente se basa en ajustar mediante el uso de un polinomio (fitting) la relación temperatura/sesgo usando los datos obtenidos en laboratorio (Günhan and Ünsal, 2014). El problema que presenta esta técnica es que no puede mapear con precisión cambios abruptos en el sesgo.

En los últimos años se han propuesto métodos basados en aprendizaje automático (machine learning) para abordar la compensación del sesgo en sensores inerciales MEMS (Xia et al., 2009) (Zhang et al., 2009) (Shiau, 2011) (Grigorie et al., 2012) (Musavi and Keighobadi, 2015) (Fontanella et al., 2018). Estos trabajos han demostrado ser una mejora respecto a los métodos basados en fitting. Una limitación que presentan estos trabajos es que solo se han centrado en utilizar sensores inerciales MEMS que pueden ser considerados de calidad media, cuyo costo varía entre USD 300 y USD 1.000.

Existen en el mercado sensores inercial MEMS de muy bajo costo, de alrededor de USD 10. Estos dispositivos tienen un tamaño de unos cuantos milímetros cuadrados y son muy livianos. Pesan alrededor de 1 gramo. Ejemplos de estos dispositivos son la unidad de medición inercial (IMU, por sus siglas en inglés) MPU-6000 de Invensense (Invensense, 2013) o la IMU BMI055 de Bosch (Bosh, 2014). No se han encontrado en la literatura existente trabajos que hagan un análisis profundo sobre el uso de diferentes técnicas para compensar los errores presentes en sensores inerciales MEMS de muy bajo costo.

Comúnmente este tipo de IMUs se encuentra en la gran mayoría de sistemas de navegación y orientación en el campo de la robótica, y en productos de electrónica de consumo, como celulares y tablets. Poder lograr una mejora sustancial en el desempeño de estos sensores inerciales puede tener un impacto muy grande en estas industrias. Además, se podría extender su uso a aplicaciones que requieren mayor precisión.

Un área donde los sensores inerciales MEMS de muy bajo costo tienen mucho potencial es en aplicaciones geológicas, por ejemplo en el monitoreo de sismos y la detección temprana de avalanchas. Es sabido que la región conocida como Gran Mendoza está asentada sobre una zona sísmica. Sería muy útil contar con una red de alerta sísmica distribuida por todo el Gran Mendoza usando los acelerómetros MEMS que ya vienen incorporados en teléfonos celulares. Se podría crear una aplicación que sería descargada por

usuarios voluntarios que deseen ser parte de la red de alerta sísmica. Para detectar sismos de baja intensidad, es necesario aumentar la sensibilidad de los acelerómetros. Esto se podría lograr al corregir el sesgo presente en estos sensores. Una vez detectado un sismo, cada celular enviaría un mensaje a través de Internet reportando la ubicación del celular y el nivel de vibración registrado.

Por otra parte, en los últimos años, se han producido en Mendoza varios aludes de barro sobre la ruta nacional 7 camino a Chile, debido a las intensas lluvias que se producen durante el verano (Diario Los Andes, 2016) (El Sol Online, 2018). Un detector de aludes o de avalanchas de nieve podría ser un dispositivo compuesto por una triada de acelerómetros, una radio, una batería y un pequeño panel solar. Este dispositivo estaría instalado en las laderas de las montañas contiguas a la ruta 7. Debido a que el equipo puede ser destruido o perdido ante un alud, es necesario contar con un sensor de muy bajo costo. Usar acelerómetros MEMS de muy bajo costo compensados en sesgo podría aumentar la sensibilidad de estos equipos y permitir emitir una alerta más rápida de avalanchas.

Por lo expuesto, este proyecto considera relevante compensar las variaciones del sesgo en sensores inerciales de muy bajo costo con técnicas del estado del arte en aprendizaje automático.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Para el desarrollo del presente proyecto pueden diferenciarse 4 etapas principales.

1. Generar datos estáticos que representen la relación que existe entre temperatura y sesgo para diferentes IMUs de muy bajo costo. Será central en el proyecto propuesto la generación de datos provenientes de sensores MEMS de muy bajo costo. Se colocarán diferentes sensores en una cámara térmica nivelados y en estado estático, se variará la temperatura ambiente y se registrarán las señales de salida. Con estos datos se podrá analizar en una etapa posterior qué tipo de relación existe entre el sesgo y la temperatura en estos dispositivos. Se presupone que la relación entre ambas variables será no lineal. Adicionalmente, se desarrollará un datalogger ad-hoc para esta tarea usando una computadora Arduino Blue Pill. De esta forma, se podrán leer diferentes tipos de IMUs, las cuales generalmente vienen provistas de un puerto de comunicación I2C o SPI.
2. Evaluar los métodos propuestos en la literatura previa para la compensación del sesgo en IMUs MEMS. Se realizará un estudio del estado del arte sobre compensación del sesgo en IMUs MEMS de diferentes costos. Se analizará si los métodos ya propuestos en la literatura, como confección de tablas, redes neuronales o fitting, pueden efectivamente lidiar con las variaciones del sesgo en sensores inerciales de muy bajo costo.
3. Proponer nuevos modelos para la compensación del sesgo basados en técnicas de aprendizaje automático del estado del arte. En una tercera etapa se propondrán nuevas técnicas basadas en aprendizaje automático del estado del arte para la corrección de la variación del sesgo en sensores inerciales MEMS de muy bajo costo. Algunos de los métodos que se analizarán es el Long-Short Term Memory Recurrent Neural Network, la cual es un tipo de red neuronal que ha sido utilizada con buenos resultados en el modelado de series de tiempo en áreas como el reconocimiento del habla y la detección de anomalías. También se analizarán técnicas de aprendizaje automático utilizada en otras áreas del conocimiento que también se basen en la identificación de señales en el dominio del tiempo, como puede ser estudios sobre el estado del arte sobre reconocimiento del habla (speech recognition) o procesamiento de señales médicas.
4. Generar datos dinámicos para evaluar y comparar las técnicas propuestas en el contexto de un sistema de navegación integrado. En una cuarta etapa de cuantificará el impacto que tendrá el uso de las nuevas técnicas propuestas para mejorar el rendimiento de los sensores inerciales bajo

estudio, en el contexto de un sistema de navegación integrado. En otras palabras, se medirá si efectivamente se produce una mejora en la estimación de posición, velocidad y orientación de un vehículo al aplicar una compensación al sesgo de los sensores inerciales de muy bajo costo que componen el sistema de navegación. Para este propósito, se deberán generar datos montando los sensores en un vehículo junto a un receptor GNSS y registrando las mediciones de estos sensores al trazar una trayectoria predeterminada. Se utilizará el datalogger desarrollado con Arduino Blue Pill. Luego, estos datos se integrarán bajo el entorno MATLAB usando una biblioteca o toolbox para simular sistemas de navegación integrado llamado NaveGo (NaveGo, 2019), el cual fue desarrollado y es mantenido por el Dr. Gonzalez.

3. RESULTADOS ESPERADOS

Al término del tiempo de duración del plan de trabajo, se espera obtener los siguientes resultados:

1. Conocimiento de los modelos matemáticos que relacionan temperatura y sesgo en diferentes sensores inerciales MEMS de muy bajo costo.
2. Creación de conjuntos de datos dinámicos compuestos por sensores inerciales MEMS de muy bajo costo y receptores GNSS. Estos podrán ser usados tanto en el marco de este proyecto como en futuras líneas de investigación en el área de sistemas de navegación.
3. Sólida comprensión de los beneficios y limitaciones de las técnicas basadas en aprendizaje automático orientadas a la identificación de series de tiempo en general, y a la corrección del sesgo en sensores MEMS de muy bajo costo en el contexto de un sistema de navegación, en particular.
4. Un mejor entendimiento sobre los alcances y limitaciones del uso de sensores inerciales MEMS de muy bajo costo en sistemas de navegación integrada.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de este proyecto está integrado por:

1. Dr. Rodrigo Gonzalez, propuesto como Investigador Responsable de este proyecto. Tiene basta experiencia en el tratamiento de señales provenientes de sensores inerciales en el contexto de un sistema de navegación integrado (Gonzalez et al., 2014), (Gonzalez et al., 2015), (Gonzalez and Catania, 2016), (Gonzalez et al., 2017), (Gonzalez and Catania, 2019), (Dabove and Gonzalez, 2019). 2. Dr. Carlos Catania, especialista en el área de Ciencia de los Datos (Data Science). Cuenta con experiencia en la resolución de diferentes problemas aplicando algoritmos específicos del área de aprendizaje automático (Gonzalez and Catania, 2016), (Gonzalez et al., 2017), (Catania et al., 2018), (Gonzalez and Catania, 2019).
3. Lic. Fernando Pinciroli, docente, investigador y decano de la Universidad Champagnat. Tiene basta experiencia en programación y en ingeniería de software, como así también en gestión de proyectos de desarrollo de software.
4. Mg. Javier Rosenstein, quien tiene profundos conocimientos en programación usando diferentes lenguajes y experiencia en el desarrollo de firmware para sistemas embebidos. Adicionalmente, cuenta con conocimientos en el procesamiento digital de señales (Rosenstein, 2019).

En el campo de la formación de recursos humanos, se espera:

1. Iniciar en investigación a profesionales que estén interesados en postularse a diferentes becas doctorales, en el marco del proyecto en referencia.
2. Iniciar en investigación a alumnos, tanto de la UCH, UTN o UNCuyo, que deseen sumarse al proyecto como becarios.
3. Iniciar en investigación a alumnos de los últimos años de carrera que deseen desarrollar su proyecto final de carrera en el marco de este proyecto.

5. BIBLIOGRAFÍA

(Bosh, 2014) Bosh. BMI055. Small, versatile 6DoF sensor module. Rev. 1,2, July 2014.
(Catania et al., 2018) Carlos A. Catania; Sebastian Garcia; Pablo Torres. An Analysis of Convolutional Neural Networks for detecting DGA. Argentina. La plata. 2018. Libro. Artículo Completo. Congreso. XXIV Congreso Argentino de Ciencias de la Computación 2018. Red de Universidades con Carreras en Informática.
(Dabove and Gonzalez, 2019) Dabove, P. and Gonzalez, R. Inertial Measurement Units inside Smartphones: Performances and Future Steps, book chapter in Smartphones: Recent Innovations and Applications, Nova Publisher, 2019. ISBN: 978-1-53615-830-4. URL: https://novapublishers.com/shop/smartphones-recent-innovations-and-applications/
(Diario Los Andes, 2016) Diario Los Andes. Un alud tiró un puente de la ruta 7 en Alta Montaña y en Uspallata quedaron 200 vehículos varados. URL: https://www.losandes.com.ar/article/se-cayo-un-puente-de-la-ruta-7-en-alta-montana. Febrero de 2016.
(El Sol Online, 2018) El Sol Online. Un alud en alta montaña obligó a suspender el tránsito en la ruta 7. URL: https://www.elsol.com.ar/un-alud-en-alta-montana-obligo-a-suspender-el-transito-en-la-ruta-7. Diciembre de 2018.
(Fontanella et al., 2018) Rita Fontanella, Domenico Accardo, Rosario Schiano Lo Moriello, Leopoldo Angrisani, Domenico De Simone, MEMS gyros temperature calibration through artificial neural networks, Sensors and Actuators A: Physical, Volume 279, 2018, Pages 553-565, ISSN 0924-4247.
(Gonzalez et al., 2014) Gonzalez, R., Giribet, J.I., and Patiño, H.D. An approach to benchmarking of loosely-coupled low-cost navigation systems. Journal of Mathematical and Computer Modelling of Dynamical Systems. ISSN 13873954.
(Gonzalez et al., 2015) Gonzalez, R., Giribet, J.I., and Patiño, H.D. NaveGo: a simulation framework for low-cost integrated navigation systems. Journal of Control Engineering and Applied Informatics. ISSN 14548658.
(Gonzalez and Catania, 2016). Rodrigo Gonzalez and Carlos A. Catania. Time-delayed Multiple Linear Regression for Increasing MEMS Inertial Sensor Performance by Using Observations From a Navigation-grade IMU. VI IEEE/ION Position Location and Navigation Symposium (PLANS 2016). ISSN 2153-3598.DOI:10.1109/PLANS.2016.7479677. Savannah, Georgia, USA. April 2016.
(Gonzalez et al., 2017). Gonzalez, R., Catania, C., Dabove, P., Taffernaberry, J.C. and Piras, M. Model validation of an open-source framework for post-processing INS/GNSS systems. In III International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2017). Porto, Portugal. April 2017.
(Gonzalez and Catania, 2019). Rodrigo Gonzalez and Carlos A. Catania. Time-delayed multiple linear regression for de-noising MEMS inertial sensors. Computers & Electrical Engineering, Volume 76, 2019, Pages 1-12, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2019.02.023.
(Grigorie et al., 2012) Grigorie, T. L., Botez, R. M., Lungu, M., Edu, R. I., & Obreja, R. Microelectromechanical systems gyro performance improvement through bias correction over temperature using an adaptive neural network-trained fuzzy inference system. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 226(9), 1121–1138, 2012.
(Groves, 2008) Paul D. Groves. Principles of GNSS, inertial, and multisensor integrated navigation systems, first edition. Artech House, 2008.
(Günhan and Ünsal, 2014) Günhan, Y., and Ünsal, D., "Polynomial Degree Determination for Temperature Dependent Error Compensation of Inertial Sensors", IEEE/ION Position, Location and Navigation Symposium (PLANS), May 2014,

pp.1209-1212, doi: 10.1109/PLANS.2014.6851494.
(Invensense, 2013) Invensense. MPU-6000 and MPU-6050 Product Specification, Revision 3.4, 2013.
(Musavi and Keighobadi, 2015) Musavi, N., and Keighobadi, J., "Adaptive fuzzy neuro-observer applied S to low cost INS/GPS," Applied Soft Computing, Vol. 29, 2015.

(NaveGo, 2019) R. Gonzalez, C. Catania, and P. Dabove. NaveGo: An Open-Source MATLAB/GNU-Octave Toolbox for Processing Integrated Navigation Systems and Performing Inertial Sensors Profiling Analysis. Version 1.2.
URL: https://github.com/rodralez/NaveGo. DOI: 10.5281/zenodo.2536950. June 2019.
(Rosenstein, 2019) Javier Rosenstein. Uso de VRPN para la implementación de BCI partiendo de señales EEG/EOG. Tesis de Maestría. Maestría en Teleinformática, Universidad de Mendoza. Ciudad de Mendoza, Argentina. 2019.
(Shiau, 2011) Shiau, J. K., "MEMS Gyroscope Null Drift and Compensation Based on Neural Network", Advanced Materials Research, Vols. 255-260, 2011, pp. 2077-2081, 2011.
(Xia et al., 2009) Xia, D., Chen, S., Wang, S., and Li, H., "Microgyroscope Temperature Effects and Compensation-Control Methods", Sensors, Vol. 9, No. 10, October 2009, pp. 8349-8376, doi: 10.3390/s91008349.
(Zhang et al., 2009) Zhang, Q., Tan, Z., and Guo, L., "Compensation of Temperature Drift of MEMS Gyroscope Using BP Neural Network," 2009 International Conference on Information Engineering and Computer Science, Wuhan, 2009.

Redes de Sensores, Vehículos móviles y Simulación en Sistemas de Tiempo Real

Fernando Romero[1], Diego Encinas[1], Armando De Giusti[1,2], Santiago Medina[1], Martín Pi Puig[1], Horacio Villagarcía[1,3], Juan Manuel Paniego[1], Fernando G. Tinetti[1,3]

[1]Instituto de Investigación en Informática LIDI (III-LIDI)[3]
Facultad de Informática – Universidad Nacional de La Plata – Centro Asociado CIC

[2] CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas
[3] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

{fromero, dencinas, degiusti, smedina, mpipuig, hvw, jmpaniego, fernando}@lidi.info.unlp.edu.ar

Resumen

El objetivo de esta línea de investigación es el estudio, desarrollo y aplicación de Sistemas de Tiempo Real (STR), en particular los que incumben a robots y redes de sensores, tanto inalámbricas como cableadas. Se llevan a cabo experimentos con diferentes tipos de robots y con plataformas basadas en microcontroladores, diseñados y armados en el laboratorio de Tiempo Real. Una parte importante de los trabajos se enfoca en las comparaciones de escenarios reales y simulados.

Contexto

Esta línea de Investigación forma parte del proyecto 11/F024-Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real SubProyecto CAD-3. Procesamiento para problemas de Tiempo Real / Robótica del Instituto de Investigación en Informática LIDI acreditado por la UNLP.

Palabras Claves: Tiempo Real, Simulación, Sistemas Embebidos, Comunicaciones, Redes de Sensores, Robots, drones, Microcontroladores, Cloud Computing.

1. Introducción

Es una característica primordial de los Sistemas de Tiempo Real (STR) la existencia de plazos de tiempo para llevar a cabo sus acciones [5] [6] [11] [12] [16] [17]. Es debido a que siempre interactúan con el mundo físico y sus acciones deben ocurrir dentro de los límites temporales fijados por el medio al cual controlan. Para ejercer dicho control se utilizan sensores para la adquisición de datos y actuadores para las respuestas. Estos sensores pueden adquirir configuraciones complejas en red cuando la cantidad de variables a controlar sea grande y sobre todo estén situadas en forma remota. A menudo se recurre en estos sistemas a la utilización de robots móviles, terrestres y aéreos, que se combinan con los sensores y actuadores. Los sensores cumplen una doble funcionalidad: a) los que están sobre el móvil adquirir datos en el área que recorren, y b) proveer información sobre la posición de los móviles. Para este propósito se recurre a sistemas GPS en exteriores y técnicas de posicionamiento en interiores (*indoor*) utilizando para ello diferentes tipos de sensores de tecnología

de ultrasonido, infrarrojo como también los sistemas de comunicaciones wifi y bluetooth. Se utilizan robots móviles y drones [22] [23] armados por alumnos y los autores del trabajo, como también robots Khepera [18], y drones Parrot. En el desarrollo de robots propios y redes de sensores se utilizan placas de desarrollo basadas en microcontroladores (como, Arduino, NodeMCU, CIAA [19] [27]) y Computadoras de Placa Simple (como, Raspberry Pi), utilizando diferentes SOTR (Linux RT-Preempt, Free RTOS, MQX, OSEK-OS, etc.) [8]. Se ha experimentado con redes de redes de sensores basadas en los protocolos cableados RS485, SPI, I2C, CANBUS [15] [20] y MODBUS [1]. También se realizan pruebas de alcance, integridad y funcionalidad de redes de sensores e inalámbricas [28] [29] [30] principalmente utilizando módulos WiFi y LoRa [26]. Por otra parte, se trabaja en el campo del modelado y simulación con el fin de obtener simuladores que permitan predecir el comportamiento y la eficiencia de distintos sistemas ante diferentes escenarios. En particular, los sistemas analizados son los: redes de sensores [31], arquitecturas de cómputo en la nube y evacuación de edificios. Las herramientas utilizadas para los desarrollos son: Proteus [25], CloudSim [2], ABMS [14] y SystemC [24].

2. Líneas de Investigación y Desarrollo

Se plantean como temas de estudio:

- Sistemas robóticos autónomos y con intervención humana [4] en el lazo de control [21]. En particular en lo relativo a seguir trayectorias predefinidas, evitando colisiones. Es de gran importancia saber la ubicación del móvil a fin de asociar los datos relevados con la posición en el entorno.
- Análisis y desarrollo de modelos de sistemas de tiempo real.
- Herramientas para implementación de modelado y simuladores.
- Utilización de simulaciones para comprobación de defectos de diseño en etapas tempranas de desarrollo [7].
- Armado de redes inalámbricas específicamente orientadas al control de variables físicas con sensores. En principio, se implementarán experimentos para caracterizar estas redes en términos de métricas como latencia y ancho de banda para el caso de rendimiento, distancia, confiabilidad (pérdida de paquetes), etc. [26] [9].
- Evaluación de redes específicamente diseñadas para distancias mayores a los estándares de WiFi (ej: LoRa) [26]
- Odometría a través de robots Khepera [13] [3] y otros de producción propia.
- Sistemas de posicionamiento en interiores utilizando técnicas que involucran señales WiFi, bluetooth y ultrasonido. Se enfoca principalmente en obtener vehículos autónomos [10] que puedan circular tanto en depósitos utilizados para tareas de logística, como en un entorno de autopistas inteligentes y con capacidad de estacionamiento.

3. Resultados y Objetivos

Se han desarrollado tareas sobre los temas antes expuestos tales como:

- Desarrollo de un robot con codificadores (*encoders*) para odometría.
- Medición de consumo energético de diferentes microcontroladores en distintas condiciones de uso y corriendo diferentes SOTR.
- Estudio de plataformas de hardware: Arduino, Intel Galileo, CIAA, Freescale Kinetis, Raspberry Pi, NodeMCU.
- Construcción y estudio de redes de sensores cableadas, empleando CANBUS, MODBUS y RS485.

- Construcción y estudio de redes de sensores inalámbricas basadas en WiFi y LoRa.
- Desarrollo de interfaces para el control de drones.
- Modelado y simulación de arquitecturas de cómputo en la nube para comparar con arquitecturas de HPC.
- Modelado y simulación de sistemas orientado a aplicaciones sociales.

4. Formación de Recursos Humanos

Se desarrollan trabajos de alumnos en la Convocatoria a Proyectos de Desarrollo e Innovación de la Facultad de Informática de la UNLP. Además, se encuentran en desarrollo y concluidas tesinas de grado de alumnos de Licenciaturas de Informática y Sistemas, como así también Prácticas Profesionales Supervisadas (PPS) con las que concluyen sus estudios los alumnos de Ingeniería en Computación y Analista en TICs. De postgrado, investigadores del grupo están desarrollando un trabajo final de especialización, tres tesis de Maestría y una tesis de Doctorado.

5. Referencias

[1] Jordi Bartolomé "El protocolo MODBUS", 2011. En http://www. tolaemon.com/docs/modbus.htm

[2] R. Calheiros, R. Ranjan, A. Beloglazov, C. De Rose and R. Buyya “CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms” Published online 24 August 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.995.

[3] Azizi, F., N. Houshangi. "Mobile robot position determination using data from gyro and odometry.", Canadian Conf. On Electrical and Computer Engineering, 2004, vol. 2, pp. 719-722. IEEE, 2004.

[4] Bekey, George A. Robotics: state of the art and future challenges. Imperial College Press, 2008.

[5] Burns, A, A, Wellings. "Real-Time Systems and Programming Languages: Ada, Real-Time Java and C/Real-Time POSIX", Addison-Wesley Educational Publishers Inc., 2009.

[6] Buttazzo, G. C., "Hard RealTime Computing Systems", Third edition, Springer, 2011.

[7] Eickhoff, J., Simulating Spacecraft Systems, Springer, 2009.

[8] "FreeRTOS - market leading RTOS (real time operating system) for embedded systems supporting 34 microcontroller architectures". http:// www.freertos.org/.

[9] Jenkins, T., I. Bogost. "Designing for the internet of things: prototyping material interactions." In CHI'14 Extended Abstracts on Human Factors in Computing Systems, pp. 731-740. ACM, 2014.

[10] Jones, J. L., A. M. Flynn, Bruce A. Seiger. Mobile robots: inspiration to implementation. Vol. 2. Wellesley MA: AK peters, 1999.

[11] Kopetz. H., "Real-Time Systems, Design Principles for Distributed Embedded Applications". Second Edition. Springer. 2011.

[12] Liu, J. W. S. Liu, "Real Time Systems", Integre Technical Publishing Co., Inc., 2000

[13] Rekleitis, I. M., G. Dudek, E. E. Milios. "Multi-robot exploration of an unknown environment, efficiently reducing the odometry error". In International Joint Conference on Artificial Intelligence, vol. 15, pp. 1340-1345. Lawrence Erlbaum Associates Ltd, 1997.

[14] C. Macal, M. North, Tutorial on agent-based modeling and simulation part 2: how to model with agents, in: Proceedings of the Winter Simulation Conference, 2006.

[15] Introduction to the Controller Area Network (CAN) Texas Instrument Appli-

cation Report SLOA101A–August 2002–Revised July 2008.
[16] Silberschatz, A., P. B. Galvin, G. Gagne, Operating System Concepts, 8th Edition, ISBN: 978-0-470-12872-5, Wiley, 2009.
[17] PHILLIP A. LAPLANTE, SEPPO J. OVASKA. REAL-TIME SYSTEMS DESIGN AND ANALYSIS Tools for the PractitionerFourth Edition. A JOHN WILEY & SONS, INC., PUBLICATION. IEEE PRESS. 2012.
[18] http://ftp.k-team.com/khepera/documentation/Kh2IRAN.pdf
[19] http://www.proyecto-ciaa.com.ar/devwiki/doku.php?id=desarrollo:edu-ciaa:edu-ciaa-nxp
[20]http://www.can-cia.de/can-knowledge/can/can-fd/
[21] Chi-Pang and Shankar S, A POMDP Framework for Human in the Loop System, Univ. of California at Berkeley.
[22]F. G. Tinetti and O. C. Valderrama Riveros, "Unmanned Vehicles: Towards Heterogeneous Hardware Approaches," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2018, pp. 919-924.
[23]Fernando G. Tinetti, Oscar C. Valderrama Riveros, Fernando L. Romero, "Unmanned Vehicles: Real Time Problems in Drone Receivers", Conf. on Computational Science & Computational Intelligence (CSCI'19), Las Vegas, Nevada, USA , 2019, pp. 1081-1085.
[24] D. Black, SystemC: From the Ground Up. Second Edition, Springer, 2010.
[25] Proteus. https://www.labcenter.com. 2017
[26] LoRa https://www.lora-alliance.org/ 2017
[27] NodeMcu http://www.nodemcu.com/ 2017
[28] Akyildiz, Ian F., and Mehmet Can Vuran. “Wireless sensor networks” Vol. 4. John Wiley & Sons, 2010.
[29]Lewis, Franck L."Wireless sensor networks." Smart environments: technologies, protocols, and applications 11 (2004): 46.
[30] Raghavendra, Cauligi S., Krishna M. Sivalingam, and Taieb Znati, eds. “Wireless sensor networks” Springer, 2006.
[31] M. Pi Puig; S. Medina; A. Batista; D. Encinas; F. Romero; F. Tinetti; A. De Giusti. Design of a CAN Simulation Device for Communications in Sensor Networks. Computer Science & Technology Series - XXII Argentine Congress of Computer Science. Selected Papers. 2017.

ORCID autores:
Fernando Romero: 0000-0002-1498-3752
Diego Encinas: 0000-0002-6948-9786
A. De Giusti: 0000-0002-6459-3592
Santiago Medina: 0000-0001-6852-7165
Martín Pi Puig: 0000-0002-7202-7638
Horacio Villagarcía:
Juan Manuel Paniego: 0000-0001-6721-9822

SÍNTESIS DE SONIDO EN DISPOSITIVOS MÓVILES

Antonio Zimmerman[1], Claudio Aciti[1,2], Pablo Cosso[1], Esteban Sebastiani[1], Fabricio Chungo[1], Juan Amusategui[1], Francisco García[1], Catriel Nievas[1], Romina Gomez[1], Alan Achaval[1], Germán Torós[1] y Diego Leguizamón[1]

[1] Departamento de Ciencia y Tecnología
Universidad Nacional de Tres de Febrero (UNTREF)
Valentín Gómez 4752 (+54-011-47593528) - Caseros - Buenos Aires

[2] Instituto INTIA - Facultad de Ciencias Exactas
Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA)
Pinto 399 (+54-249-4439650) - Tandil - Buenos Aires
caciti@exa.unicen.edu.ar

Resumen

Este proyecto se propone investigar temas relacionados a síntesis de sonido en dispositivos móviles. Las aplicaciones de sintetizadores digitales permiten emular el sonido analógico. Aunque es necesario ahondar en temas como la velocidad de reproducción, el jitter, la latencia "touch to sound", el timbre, entre otros. El equipo de trabajo, con base en UNTREF, está integrado por investigadores y alumnos de las especiales Computación, Música, Sonido y Artes Electrónicas. Hasta el momento, se cuenta con un desarrollo principal de un sintetizador, y varias versiones que permiten comparar entre diferentes arquitecturas de hardware, modelos de dispositivos móviles y lenguajes de programación.

Palabras claves: procesamiento digital de sonido - dispositivos móviles - sintetizador modular analógico

Contexto

Un sintetizador analógico modular es un sistema electrónico desarrollado para la creación y composición del sonido mediante la manipulación de sus parámetros básicos. Utiliza circuitos y señales analógicas para su generación. De esta forma el timbre de un sonido es modelado utilizando diferentes tipos de subsistemas, conformados en tres grandes grupos: Generación, Procesamiento y Sistemas de Control [1,2]. El primer sintetizador fue creado por Robert Moog en el año 1965 y este era capaz de utilizar módulos separados para la generación de una amplia variedad de sonidos mediante técnicas de síntesis como la aditiva, sustractiva y FM, entre otras [3]. Un sintetizador modular es una combinación de módulos separados que el usuario interconecta pero que son en conjunto controlados por tensión o voltaje. Un sistema modular puede variar en tamaño y complejidad en relación a la cantidad de módulos disponibles, y las necesidades del usuario. Los sistemas modulares le brindan al usuario total libertad para realizar las conexiones entre las partes que lo componen. Esta configuración de conexiones es llamada “patch”.

Los sintetizadores digitales, a diferencia de los analógicos, generan y controlan sus

sonidos por medio de señales digitales. Utilizan técnicas de procesamiento digital de señales para su creación [4]. La señal es procesada por algoritmos y finalmente se la convierte en una señal analógica. Algunos ejemplos de sintetizadores digitales son el Synclavier, Yamaha, Roland líneas, etc.

Los avances en la tecnología, lenguajes de programación y técnicas de procesamiento digital de señales hacen posible el desarrollo de un sintetizador modular digital para dispositivos móviles [5][6]. Para ello se utilizan lenguajes de programación de alto nivel como Pure Data [7][8][9], Max-MSP [10], Faust [11], Super Collider [12], entre otros.

En la actualidad, los dispositivos móviles han cambiado la forma en que las personas se relacionan entre sí, ya que son un producto clave de la tecnología de la información [13][14]. Con relación a la música, permiten desarrollar nuevas posibilidades que facilitan al momento de experimentar nuevas formas de expresión de un modo mucho más accesible. Con los sensores adecuados se puede generar nuevas formas de interpretar música, permitiendo que un dispositivo de uso masivo pueda ser utilizado por un público más amplio.

Introducción

En un sintetizador, el usuario puede elegir el tipo de onda de generación del sonido. Los tipos de onda más utilizados en síntesis son: sinusoidal, cuadrada, triangular y diente de sierra. En el dominio digital no pueden existir señales continuas ya que requerirían infinita cantidad de información. Representar una señal continua implica convertirla a una forma discreta; para ello es necesario realizar un muestreo de la información. Según el Teorema de Nyquist, para no perder información en la conversión, la frecuencia de muestreo debe ser igual o mayor que el doble de la frecuencia máxima [15][16][17].

En un sintetizador modular, el usuario determina la configuración de conexiones con un criterio artístico y no meramente por limitaciones técnicas. Esto posibilita y fomenta la libertad de exploración de diversas posibilidades sonoras dentro de los distintos métodos de síntesis [18][19][20].

El objetivo principal de este proyecto es el estudio, diseño y desarrollo de aplicaciones que sintetizen sonido en dispositivos móviles que emulen fielmente un sintetizador modular analógico. Es necesario realizar estudios y comparativas de velocidad, jitter, timbre y latencia "touch to sound", entre diferentes dispositivos, sistemas operativos y lenguajes de programación.

Metodología de trabajo

Se utilizan metodologías ágiles poniendo énfasis en el desarrollo iterativo e incremental.

Captura de requerimientos: Se realiza un relevamiento de sintetizadores analógicos; sus distintos componentes, interacciones y comportamientos. Se releva la resolución del manejo de interfaz de usuario en sintetizadores digitales disponibles para dispositivos móviles.

Sesiones de Brainstorming del equipo interdisciplinario para definir el alcance y las metas de cada etapa.

Líneas de Investigación, Desarrollo e Innovación

Este proyecto se desarrolla en la UNTREF, con un equipo de trabajo multidisciplinario, formado por alumnos y docentes provenientes de las carreras Ingeniería de Sonido, Licenciatura en Artes Electrónicas,

Licenciatura en Música e Ingeniería en Computación.

Hasta el momento se diseñó y se cuenta con un desarrollo avanzado de una aplicación para dispositivos móviles que busca emular fielmente un sintetizador modular analógico. Dicha aplicación funciona en plataformas Android, y fue escrito en lenguaje Java y Pure Data.

Entre las líneas de investigación se destacan:

- La implementación de diferentes aplicaciones que requieren de un estudio previo, puesta en funcionamiento y comparación de diferentes lenguajes de programación. Y las comparativas con app actuales, de licencia libre o privativa.
- El estudio de jitter y de la latencia "touch to sound" de las aplicaciones entre diferentes lenguajes de programación, diferentes marcas y modelos de dispositivos móviles y diferentes versiones de sistemas operativos.
- La creación y estudio de funcionamiento de sonidos preconfigurados en diferentes plataformas y dispositivos móviles.
- El diseño de interfaz de usuario que permita a los usuarios utilizar aplicaciones de este estilo de forma intuitiva, fácil de comprender, que se adapte a diferentes tamaños de pantallas, y que sea capaz de responder a velocidades razonables independientemente del dispositivo donde esté corriendo.
- El diseño y desarrollo de los módulos generadores de audio, de tal manera que puedan combinarse entre sí y funcionen en diferentes dispositivos móviles.

Resultados

Como se dijo anteriormente, se cuenta con un diseño y desarrollo avanzado de un sintentizador modular llamado Synth3F[1], que corre en Android y está programado en Java y Puredata.

Se comparó el nivel de aliasing de los generadores de ondas de las dos APIs utilizadas para sintetizar audio en Android. La primera de ellas es la basada en libPd, que permite embeber un código o patch escrito en el lenguaje Pure Data. La segunda es la API provista por el lenguaje Faust, que compila el código en lenguajes Java y C++ para ser implementados dentro de la aplicación. La comparación entre ambos motores de audio muestra un resultado notablemente favorable para el caso de Faust. La síntesis desde Pure Data muestra componentes no deseados incluso en el caso de la onda senoidal. En general, estos componentes se encuentran distribuidos a lo largo del espectro y sus picos más altos tienen una diferencia cercana a los 20 dB respecto a la fundamental.

Se realizaron mediciones de latencia "touch to sound" en diversos dispositivos Android y se comparó su comportamiento en distintas versiones de la aplicación. Primero, se buscó obtener la medición de latencia de manera óptima. Para esto se utilizó un micrófono de medición apuntando hacia el dispositivo bajo prueba. El transductor se conectó con una placa de audio, y luego a una Digital Audio Workstation (DAW) para su análisis posterior. De esta forma, el transductor logra captar dos impulsos diferentes: el primero referido al golpe del dedo sobre la pantalla del dispositivo, y el segundo relacionado con el sonido emitido por el parlante del mismo, una vez que el proceso de la señal fue realizado. Los parámetros de la aplicación en cuestión se ajustaron para que el sonido emitido posea un valor de ataque bajo (idealmente, lo más cercano posible a 0 ms). De las mediciones obtenidas de los tiempos de latencia "touch to

[1] https://github.com/UntrefSintetizador

sound" obtenidas de la aplicación propia contra aplicaciones comerciales, se puede observar que efectivamente los valores son similares. Es notable destacar esta comparación, ya que coloca a la aplicación (en referencia a tiempos de latencia "touch to sound") al nivel de aplicaciones comerciales actuales.

Además, se realizó la compilación de archivos de Faust en Android Studio con el fin de utilizar módulos de Faust en lugar de los de Pure Data ya que el mismo cuenta con antialiasing y una mejor calidad de audio.

Formación de Recursos Humanos

El equipo de trabajo es dirigido por un profesor con maestría de la carrera Ingeniería en Computación, quien se especializa en procesamiento de señales y espera terminar su doctorado en Matemática Computacional e Industrial de la UNCPBA en 2020. La codirección está a cargo de un Licenciado en Música, quien se especializa en aplicaciones para dispositivos móviles. El resto del equipo de docentes está compuesto por un Lic. en Análisis de Sistemas, un Lic. en Artes Electrónicas, un Lic. en Música y un Ing. de Sonido.

También lo integran 8 alumnos avanzados de la UNTREF, dos de la carrera Ing. en Computación, dos de Lic. en Artes Electrónicas, dos de Ing. de Sonido y dos de Lic. en Música. Tres de ellos están realizando su trabajo final de grado en temas relacionados y se espera que continúen con sus posgrados en la temática.

Referencias

[1] Austin, Kevin. *A Generalized Introduction to Modular Analogue Synthesis Concepts.* Montréal,Canada. 2016

[2] Rudi Esch. *Electri_City: The Düsseldorf School of Electronic Music.* Omnibus Press. Germany. 2016

[3] Russ, M. *Sound Synthesis and Sampling*. 3ed. Focal Press. 2009

[4] J G Proakis, D G Manolakis. *Digital signal processing - Principles, algorithms and applications.* Pearson. 2007

[5] Miller Puckette. *Pure Data: Another Integrated Computer Music Environment.* Proceedings of the Second Intercollege Computer Music Concerts, 37-41, 1996.

[6] Brinkman, Peter. *Making Musical Apps*. Sebastopol, CA: O'Reilly Media, Inc. 2012.

[7] Pure Data (Pd). Pd Community Site. [online] Disponible en: https://puredata.info/.

[8] P. Brinkmann, P. Kirn, R. Lawler, C. McCormick, M. Roth, and H.-C. Steiner. *Embedding Pure Data with libpd*. Pure Data Convention Weimar. 2011.

[9] LibPD [online] Disponible en: http://libpd.cc/about/

[10] Sitio web de MAX-MSP https://cycling74.com/ accedido en Marzo de 2020.

[11] Sitio web de FAUST. https://faust.grame.fr accedido en Marzo de 2020.

[12] Sitio web de Supercollider. https://supercollider.github.io/ accedido en Marzo de 2020.

[13] The Economist [online] Disponible en: https://www.economist.com/news/briefing/21645131-smartphone-defining-technology-age-truly-personal-computer

[14] GartnerInc.: Technology Research [online] Disponible en: http://www.gartner.com/newsroom/id/360987

[15] Alan Oppenheim. *Señales y Sistemas*. Prentince Hall.1998.

[16] Lyons. *Understanding Digital Signal Processing*. Prentince Hall. 2001

[17] Oppenheim, Schafer. Discrete time signal processing. Prentince Hall. 2009

[18] Alan Bloom. *The ARRL Handbook for Radio Communications*, Chapter 8: Modulation. American Radio Relay League. 2010

[19] Simon Haykin. *Communication Systems*. 4th Ed. John Wiley & Sons. New York. 2001

[20] Puckette, Miller. *Theory and Techniques of Electronic Music.* WSP. 2007

Detección de patologías en señales biomédicas mediante técnicas de machine learning.

Bergamini, María Lorena, Liberczuk, Sergio Javier

Sergio.Liberczuk@uai.edu.ar - Maria.Bergamini@uai.edu.ar

Universidad Abierta Interamericana –Centro de Altos Estudios en Tecnología Informática (CAETI)

RESUMEN

El procesamiento de señales biomédicas tiene una importancia relevante en el diagnóstico temprano y prevención de enfermedades. El electrocardiograma (ECG) es un estudio no invasivo, de bajo costo, que brinda información valiosa sobre la actividad eléctrica cardíaca. El análisis de esta señal estudia patrones que se asocian con condiciones anormales de funcionamiento. El objetivo principal de este proyecto es desarrollar técnicas y algoritmos para el análisis, modelado, clasificación y segmentación de señales electrocardiográficas, a fin de que puedan ser aplicados en tiempo real; y poder así dar soporte a la detección temprana de eventos patológicos. Específicamente, nos proponemos diseñar algoritmos de procesamiento de ECG con un enfoque Bayesiano, con el objetivo de sintonizar los parámetros de un modelo dinámico que permitan la síntesis de señales de ECG registrables durante procesos de isquemia e infarto. Asimismo, se aplicarán técnicas de machine learning para procesar los parámetros y configurar un sistema de asistencia al médico en el diagnóstico automático de patologías.

CONTEXTO

La investigación aquí descrita surge de la colaboración entre el Centro de Altos Estudios en Tecnología Informática (CAETI), de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI), conjuntamente con el Instituto Argentino de Matemática Alberto P. Calderón, (IAM) dependiente de CONICET y el Instituto de Ingeniería Biomédica (IIBM) de la Universidad de Buenos Aires.

El CAETI concentra proyectos de investigación básica y aplicada, que pretenden contribuir a la generación de herramientas informáticas y tecnológicas para dar solución a problemáticas sociales y humanas. Uno de los objetivos del centro es promover la inserción de la Tecnología Informática en asuntos multidisciplinarios de impacto social.

1. INTRODUCCIÓN

La señal de electrocardiograma registra la actividad eléctrica del corazón, mostrando la variación temporal del potencial eléctrico que se registra a partir de electrodos dispuestos convenientemente en la superficie del tórax. La morfología de este registro y su interpretación a partir de la detección de sus ondas características, así como diversos cálculos que surgen de la detección de dichos puntos característicos, permiten el diagnóstico de distintas patologías: arritmias cardíacas, cardiopatías isquémicas, infarto y/o alteraciones de la conducción auriculo-ventricular. Dichas patologías pueden ser detectadas con antelación mediante un adecuado análisis de la señal de ECG (Martis et al., 2014).

El estudio del ECG, sus amplitudes de onda, sus intervalos temporales y patrones ha sido objeto de intensas investigaciones, ya que proporciona información sustancial de la

funcionalidad del corazón. La extracción de señales cardíacas de alta resolución de un electrocardiograma ruidoso sigue siendo un problema de interés para la comunidad de ingeniería biomédica. A pesar de la rica literatura en este campo, todavía hay muchas aplicaciones clínicas que carecen de herramientas confiables de procesamiento para extraer las componentes ricas en información del ECG, que se encuentran inmersas en potencias de ruido considerable por lo que poseen baja relación señal ruido (SNR).

Para señales estacionarias, el filtro de Wiener es la técnica de filtrado lineal en el sentido del mínimo error cuadrático medio, aplicado de modo causal en el dominio temporal o de modo no causal en el dominio frecuencial. Sin embargo, no se espera que el filtrado de Wiener dé buenos resultados para señales no estacionarias.

Las técnicas estadísticas como el Análisis de Componentes Principales (PCA), Análisis de Componentes Independientes (ICA) y las Redes Neuronales (NNs) también se han utilizado para construir un modelo estadístico de la señal y del ruido, lo que permite la eliminación de éste en banda descartando las componentes correspondientes exclusivamente a dicho ruido. Aunque estos son esquemas poderosos de filtrado, el modelo utilizado es bastante arbitrario y pueden ser extremadamente sensibles a pequeños cambios tanto en la señal como en el ruido.

En algunos trabajos recientes se ha propuesto un paradigma de Filtrado Bayesiano para la compresión de señales biológicas en general y de ECG en particular; y el filtrado del ruido en la misma (Sayadi et al., 2010, Sameni et al., 2007). Este marco también puede ser utilizado eficazmente para la segmentación y la extracción de características.

La falta de una adecuada comprensión de la dinámica eléctrica cardíaca hace difícil optimizar las técnicas computacionales empleadas, obteniendo como consecuencia un aumento en el número de falsos positivos y por lo tanto una baja especificidad al momento de emitir un diagnóstico asistido por computadora. Una solución a este tipo de problema ha sido la creación de modelos matemáticos de la actividad eléctrica cardíaca, que permitan recrear diversas y complejas situaciones electrofisiológicas. De esta manera, se pueden sintetizar conjuntos de señales de ECG para entrenamiento y validación de los sistemas de diagnóstico asistido (Kundu & Gupta, 2015)

McSharry et al. (2003), han desarrollado un modelo para simular con gran precisión diversos trazados de ECG correspondientes a condiciones fisiológicas y algunas patologías cardíacas.

La tendencia actual en procesamiento de señales biomédicas se dirige a la aplicación de técnicas de inteligencia artificial como redes neuronales, algoritmos genéticos y técnicas basadas en lógica difusa (fuzzy logic) (Engin, 2004; Ahsan et al., 2011; Tamil et al., 2008)

Los métodos de aprendizaje profundo, como las redes neuronales convolucionales (CNN), pueden extraer automáticamente características distintivas y han demostrado su eficacia para reconocimiento y clasificación de señales biomédicas. Recientemente, Rivera & Gonzalez (2019) aplicaron redes neuronales artificiales a la tarea de clasificación de ECG, transformando las señales en imágenes, que usaron de entrada para el aprendizaje. Por otro lado, Labati et al. (2019), desarollaron un enfoque biométrico basado en redes neuronales convolucionales (CNN) para señales de ECG con muy buenos rendimientos usando datos controlados y no controlados.

Muchas tareas de análisis de datos en problemas reales implican estimar cantidades desconocidas a partir de algunas observaciones. Frecuentemente, se cuenta con el conocimiento previo sobre el fenómeno que se modela, y es posible formular distribuciones previas para las cantidades desconocidas y funciones de verosimilitud que relacionan estas cantidades con las observaciones.

Dentro de este contexto, toda inferencia sobre las cantidades desconocidas se basa en la distribución posterior obtenida del teorema

de Bayes. Bajo ciertas hipótesis, es posible derivar una expresión analítica exacta para calcular las distribuciones posteriores. Esta recursión es el conocido y extendido filtro de Kalman.

Las suposiciones necesarias para aplicar un filtro de Kalman podrían resultar muy restrictivas. De acuerdo con el campo de interés, el problema aparece bajo muchos nombres diferentes, incluido el filtrado bayesiano, el filtrado óptimo (no lineal), el filtrado estocástico y la inferencia y el aprendizaje en línea.

El filtrado bayesiano es un marco para tratar problemas dinámicos, donde las variables y los datos responden a dos procesos estocásticos. Su implementación numérica, en el caso de los modelos no lineales generales, requiere el uso de técnicas de integración numérica como Filtros de Partículas (PF), que son esencialmente una clase de métodos secuenciales de Monte Carlo donde los puntos de soporte, llamados partículas, evolucionan con el tiempo de acuerdo con el núcleo de transición del proceso desconocido.

El filtrado estocástico es un problema inverso: dadas las observaciones recopiladas (mediciones u observaciones) en pasos de tiempo discretos, es necesario deducir el estado real que produce esas medidas. En otra perspectiva, este problema puede interpretarse como un problema de aprendizaje de mapeo inverso: encontrar las entradas desconocidas, con una función de mapeo que produce los datos de salida medidos.

Este esquema conduce a un algoritmo para la restauración de la señal de ECG sin ruido a partir de un conjunto de observaciones ruidosas y / o ambiguas.

En Liberczuk et al. (2019) proponemos un nuevo método para encontrar los parámetros que mejor representan un determinado latido. Y en Liberczuk & Bergamini (2018) clasificamos estos latidos en el espacio de parámetros entre fisiológicos y patológicos, con altas tasas performance en la clasificación. Además, hemos desarrollado un método para eliminar ruidos de señales de ECG basado en técnicas de filtro de Monte Carlo que estiman el estado (valor de señal filtrada) a partir de observaciones ruidosas, simuladas con diferentes SNR (Bergamini & Liberczuk, 2019).

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El desarrollo de esta investigación está orientado en principio a sintonizar y validar un modelo computacional del comportamiento electrofisiológico del corazón basado en el modelo de McSharry et al. [6], reformulado por nuestro grupo en la incorporación de modificaciones que consideren y permitan recrear situaciones asociadas a cambios durante eventos cardíacos, como procesos isquémicos y de infarto.

Las técnicas de procesamiento que utilizamos se encuentran dentro del paradigma de filtrado bayesiano. Se están implementando técnicas de filtrado lineal y no lineal para la correcta extracción de información y detección de patrones de los eventos buscados. Estos métodos de procesamiento tienen alto rendimiento en señales con baja SNR, pero requieren un modelo adecuado de la dinámica del sistema que permita inferir el estado del sistema a partir de las observaciones medidas secuencialmente.

Se utiliza en principio el mencionado modelo de McSharry que se basa en suponer que la señal consiste en una suma de funciones gaussianas, cada una caracterizada por su ubicación, su amplitud y su ancho. Así, el modelo queda dependiendo de un conjunto bien determinado de parámetros. El conjunto de parámetros puede utilizarse no solo en esquemas de filtrado sino en aplicaciones de compresión, clustering y/o clasificación de patrones en señales (Clifford et al., 2005).

Se analizará la posibilidad de plantear un nuevo espacio de estado, basado en la descomposición de la señal en distintas ondas senoidales (modos empíricos), que capturen, en diferentes frecuencias, distintos niveles de información [Rilling et al., 2003, Huang et al., 1998]. Este enfoque, al ser empírico, es independiente de cualquier modelo

matemático considerado, y casi no requiere ninguna hipótesis previa. Las características de los distintos modos empíricos podrían evidenciar patrones característicos de eventos cardíacos.

Se experimentará con redes neuronales, redes neuronales convolucionales o redes neuronales recurrentes a fin de lograr la automática clasificación de latidos. Los inputs de dichas redes estarán conformados por las características obtenidas por cada latido, o por cada señal ECG, mediante las técnicas de procesamiento y análisis descritas anteriormente.

Las señales son tomadas de bases de datos validadas internacionalmente como la base de datos de ECG Physikalisch-Technische Bundesanstalt (PTB) disponible en Physio-Bank [10] que aporta gran cantidad de pacientes sanos, con isquemia e infarto. Los ECG están digitalizados a 1Khz, con 16 bits de resolución sobre un rango de 16.384mV. Cada registro incluye las 12 derivaciones simultáneas y las derivaciones ortogonales de Franz. Se dispone de la historia clínica de los pacientes.

3. RESULTADOS OBTENIDOS/ESPERADOS

En nuestra investigación hemos obtenido los siguientes resultados:

- Algoritmo eficiente para la estimación de parámetros del modelo de la actividad eléctrica del corazón. Como resultado del algoritmo se obtienen los parámetros que mejor ajustan el modelo a latidos reales dados, y que reflejan la morfología del latido. El algoritmo ejecuta una optimización global heurística, basado en método de simulación de Monte Carlo.
- Clasificación de latidos provenientes de sujetos sanos y sujetos que han sufrido un evento cardíaco, basado en separación del espacio de parámetros obtenidos, aplicando support vector machine (SVM).
- Implementación de un algoritmo de filtrado con un enfoque bayesiano, a fin de separar las componentes de ruido en una señal de ECG. Se ha logrado muy buena perfomance con distintos niveles de SNR (Signal/Noise Ratio)

Los resultados esperados son:

- Caracterizar señales de electrocardiograma (ECG) con diversas patologías en el espacio de los parámetros del modelo de la señal utilizado para representarla de un modo alternativo a fin de facilitar su clasificación.
- Plantear y testear un nuevo sistema de representación de señales que permita detectar patrones ocultos, indicativos de posibles eventos patológicos.
- Diseñar, entrenar y testear una red neuronal capaz de identificar señales electrocardiográficas patológicas, a partir del aprendizaje de los features que las identifican.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigación llevará adelante las ideas aquí expuestas está conformado por los autores de este trabajo.

El Bioing. Sergio Liberczuk se encuentra desarrollando su tesis de Doctorado, bajo la dirección del Doctor Silvano Zanutto (Director del Instituto de Ingeniería Biomédica de la UBA) y la codirección del Dr. Pedro Arini en la Facultad de Ingeniería de la UBA.

5. BIBLIOGRAFÍA

- Ahsan M.R., M. I. Ibrahimy , O. Khalifa. (2011) " Electromygraphy (EMG) signal based hand gesture recognition using artificial neural network (ANN)". In 2011 4th International Conference on Mechatronics (ICOM), pp 1-6, IEEE.
- Bergamini M.L**.,** S**.** J**.** Liberczuk (2019). "Bayesian Inversion Approach for ECG Denoising". VII Congreso de Matemática Aplicada, Computacional e Industrial VII MACI 2019. Río Cuarto, Córdoba, Argentina.
- G. Clifford, A. Shoeb, P. McSharry, B. Janz. (2005) "Model-based filtering, compression and classification of the ECG". International Journal of Bioelectromagnetism, vol. 7, no. 1, pp. 158–161.

- M. Engin. (2004) “ECG beat classification using neuro-fuzzy network”. Pattern Recognitnion Letters 25(15), pp 1715-1722.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Liu, H. H. (1998). “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”. Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences, 454(1971), 903-995 Proc. R. Soc. Lond. A. 1998
- P. Kundu, R. Gupta. (2015) "Electrocardiogram Synthesis using Gaussian and Fourier Models", IEEE Int Conf on Research in Computational Intelligence and Communication Networks, Nov 20-22, Kolkata, India, pp.312-317.
- Labati R.D., E. Muñoz, V. Piuri, R. Sassi, F. Scotti (2019) “Deep-ECG: Convolutional Neural Networks for ECG biometric recognition”. Pattern recognition letter 126,1, pp. 78-85.
- Liberczuk S.J., M.L. Bergamini, P. D. Arini (2019). “Heart Beat Parametric Modeling Based on Monte Carlo Fitting Techniques”. Revista Argentina de Bioingeniería, Vol. 23 (1), pp. 63-68. ISSN: 2591-376X
- Liberczuk S., M. L. Bergamini, (2018). “Model Based Feature Extraction Method for Myocardial Infarction Detection”. XXI Congreso Argentino de Mecánica Computacional, 2018.
- P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith. (2003), “A dynamical model for generating synthetic electrocardiogram signals”. IEEE Transactions on Biomedical Engineering, vol. 50, no. 3, pp. 289–294.
- R. J. Martis, U. R. Acharya, and H. Adeli. (2014) “Current methods in electrocardiogram characterization”. Computers in biology and medicine, vol. 48, pp. 133–149.
- R. Sameni, M. Shamsollahi, C. Jutten, G. Clifford, (2007) “A nonlinear Bayesian filtering framework for ECG denoising”. IEEE Trans. Biomed. Eng., vol. 54, no. 12, pp. 2172–2185.
- O. Sayadi, M. Shamsollahi, G. Clifford. (2010) “Robust detection of premature ventricular contractions using a wave-based Bayesian framework”. IEEE Trans. Biomed. Eng., vol. 57, no. 2, pp. 353–362.
- Tamil E.B.M., Kamarudin N.H., Salleh R., Tamil A.M. (2008) “A Review on Feature Extraction & Classification Techniques for Biosignal Processing (Part I: Electrocardiogram)”. In: Abu Osman N.A., Ibrahim F., Wan Abas W.A.B., Abdul Rahman H.S., Ting HN. (eds) 4th Kuala Lumpur International Conference on Biomedical Engineering 2008. IFMBE Proceedings, vol 21. Springer, Berlin, Heidelberg.
- Physionet: https://www.physionet.org/physiobank/database
- Ranjan, R., Awasthi, A., Aggarawal, N., Gulati, J. (2006). “Applications of fuzzy and neuro-fuzzy in biomedical health sciences”. In 2006 IEEE International Conference on Electro/Information Technology, pp. 60-65. IEEE.
- Rilling, G., Flandrin, P., & Goncalves, P. (2003). “On empirical mode decomposition and its algorithms”. In IEEE-EURASIP Workshop on nonlinear signal and image processing (Vol. 3, No. 3, pp. 8-11.
- Sánchez, F. R., Cervera, J. G. (2019). “ECG Classification Using Artificial Neural Networks”. In Journal of Physics: Conference Series, Vol. 1221, No. 1, p. 012062. IOP Publishing.

Procesamiento Distribuido y Paralelo

Hacia la Optimización de un Sistema de Recuperación de Información

Osvaldo Spositto[1], Viviana Ledesma[1], Gastón Procopio[1], Hugo Ryckeboer[1]
[1]Departamento de Ingeniería e Investigaciones Tecnológicas,
Universidad Nacional de La Matanza
{spositto, vledesma, gprocopio, hugor}@unlam.edu.ar

RESUMEN

Con el desarrollo de los repositorios digitales cada vez ha cobrado mayor interés el estudio de los Sistemas de Recuperación de Información. El volumen de la información contenida en dichos repositorios crece de forma exponencial con lo cual búsqueda de los documentos que respondan a la necesidad de los usuarios se torna una tarea difícil. En este contexto este grupo de investigación ha estado trabajando por más de ocho años en la construcción de sus propios motores de búsqueda y recuperación orientados a corpus estáticos. Una vez construidos dichos motores las líneas de investigación se han orientado a distintos enfoques que pretenden acelerar los mismos tanto en la búsqueda como en los preprocesos. En particular, en esta etapa se tiene por objetivo la investigación, desarrollo e implementación de algoritmos paralelos, principalmente para resolver el proceso de la de Descomposición en Valores Singulares en arquitecturas basadas en Unidades de Procesamiento Gráfico y su comparación con clústeres de multicores, así como el empleo de soluciones híbridas que combinen ambos enfoques.

Palabras clave: Sistemas de Recuperación de Información, Descomposición en Valores Singulares, Bidiagonalización, Indexación Semántica Latente.

CONTEXTO

La línea de investigación que se presenta se encuentra inmersa en el proyecto de investigación C225 "*Resolución Eficiente de la Descomposición en Valores Singulares en una Arquitectura Híbrida y su Posterior Inserción en un Sistema de Recuperación de Información*" llevado a cabo en el marco del programa PROINCE de la Universidad Nacional de La Matanza (UNLaM). El mismo se desarrolla en el Polo Tecnológico dependiente del Departamento de Ingeniería e Investigaciones Tecnológicas de la UNLaM.

Este trabajo continúa con la línea de investigación de los proyectos PROINCE C151, C177 y C205 cuya temática se orientó: primero al estudio del tema y posteriormente a la realización de un prototipo de un sistema de recuperación de la información, la optimización de la recuperación de documentos usando como técnica base el LSI (Lematización Semántica Latente), y el uso de Minería de Datos para acelerar la recuperación de documentos.

1. INTRODUCCIÓN

A partir de la década del 90 los avances tecnológicos de la informática dieron lugar a un incremento exponencial en la generación y almacenamiento de información que continúa hasta la actualidad. La enorme cantidad de información almacenada hace que su búsqueda y recuperación sea cada vez más dificultosa, lo cual impulsó el estudio de la Recuperación de la Información (RI) como disciplina. Existen dos tendencias principales en el desarrollo de Sistemas de RI (SRI) según el contexto y el ámbito de la fuente documental [1]: La RI vertical que se enfoca en la indexación de fuentes documentales específicas, por ejemplo, una biblioteca de Ciencias Jurídicas. La RI horizontal que se enfoca en fuentes documentales generales, por ejemplo, la Web. Esta línea de investigación se relaciona con el primer grupo dado que se trata la implementación y optimización de un SRI de fuentes documentales específicas y la evaluación de su rendimiento.

En la literatura existen diversas propuestas sobre la organización interna de un SRI [2, 3]. La Figura 1 muestra una representación simplificada de un SRI, tal como se visualiza, los procesos más importantes que intervienen

en dicho sistema son los siguientes:

Indexación – los documentos que alimentan el sistema se representan como objetos indexados.

Búsqueda – se analiza la consulta del usuario y se compara con los objetos indexados, de tal modo se pueden obtener los objetos recuperados que se le presentarán al usuario.

Ranking – se determina la relevancia de cada documento recuperado para dar solución a la consulta que haya ingresado el usuario, finalmente los documentos se ordenan en base a los valores obtenidos en este proceso.

Este trabajo está enmarcado dentro del proceso de indexación.

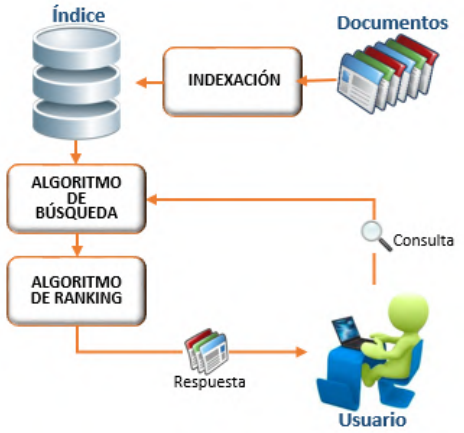


Figura 1. Sistema de RI.

Se han ideado diferentes modelos basados en distintos paradigmas para representar tanto documentos como consultas en SRI y comparar la similitud de esas representaciones [4]. Entre estos se encuentran el modelo booleano, el modelo vectorial y el modelo probabilístico, denominados clásicos. Esta línea de investigación se enmarca en una variante del método de RI vectorial, la Indexación Semántica Latente (ISL) [5].

La ISL es un método para la búsqueda de información en documentos a través de la indexación de términos [5]. Con dicho método se pretende resolver perturbaciones en la RI causados por problemas de sinonimia y polisemia o equivocidad del habla corriente. Para ejemplificar, si se desea buscar la palabra "estación", la cual tiene múltiples significados (polisemia) una búsqueda literal de la palabra produciría muchos resultados posibles (estación de tren, estación del año, etc.). Si lo que se desea buscar es "estación del año", podrían interesar resultados de palabras distintas, pero con un significado igual o similar, por ejemplo "temporada", "época" y así por el estilo (sinonimia). La ISL permite la búsqueda por conceptos o definiciones en contraposición a la búsqueda literal.

La aplicación de la ISL implica la utilización de algoritmos matemáticos especializados, que como resultado simulan el análisis que realizaría una persona. Una técnica ampliamente utilizada a tal fin es la Descomposición en Valores Singulares (DVS), luego la recuperación se realiza utilizando como punto de partida los valores singulares y vectores obtenidos a partir de la aplicación de dicha técnica [6].

La DVS [5,7] consiste en descomponer una matriz en varias matrices que exhiben las propiedades más importantes de la matriz original. Así, una matriz A de tamaño t × d descompuesta con DVS (ver Figura 2) produce tres matrices de la forma:

$$A = T_0 S_0 D_0$$

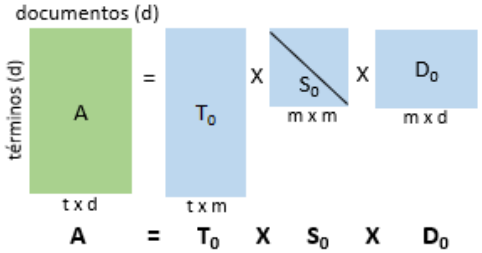


Figura 2. Reducción de dimensiones en DVS. Fuente: [5]

T_0 y D_0 tienen columnas ortonormales (ortogonales y de tamaño uno) y son las matrices izquierda y derecha, respectivamente, de vectores singulares y S_0 es una matriz diagonal compuesta de los valores singulares de A.

Disponer de modelos de orden reducido tiene como ventaja el simplificar la comprensión del sistema, reducir el coste computacional en los problemas de simulación, lo cual a su vez implica menor esfuerzo computacional en el diseño de controladores numéricamente más eficientes y se obtienen leyes de control más simples [8]. De ahí la necesidad de buscar modelos

matemáticos simplificados que aproximen al máximo el comportamiento del sistema original. El modelo resultante, que poseerá un número menor de estados que el sistema original, se denomina modelo reducido o modelo de orden reducido y al procedimiento utilizado para conseguirlo se lo conoce como reducción de modelo.

Existen dos tipos principales de algoritmos que se aplican al cálculo computacional de la DVS de una matriz real, el método unilateral de Jacobi y aquellos algoritmos que se basan en la bidiagonalización [6]. El número de operaciones para los distintos algoritmos se encuentra en el orden de On^3, las diversas propuestas y mejoras que han surgido buscan disminuir operaciones costosas en tiempo.

Mediante este proyecto se pretende optimizar la resolución de la DVS, en especial mediante implementar algoritmos para resolver la primera fase de este proceso, a través de la bidiagonalización. Los métodos más tradicionales de bidiagonalización utilizan las transformaciones de Householder por la izquierda y por la derecha de la matriz [9, 10]. Como desventaja, cuando las matrices son de grandes dimensiones requieren tiempos de computación elevados y además repercuten negativamente en los costos de comunicación de una implementación paralela del algoritmo en sistemas de memoria distribuida [11,12]. Así es que se han realizado diversos trabajos, entre estos se encuentran la propuesta de Ralha [13], mejorada más adelante por Barlow [14], orientada a conseguir un método más sencillo aplicando las transformaciones de Householder solamente por el lado derecho de la matriz. Posteriormente, Da Silva Sanches de Campos [12] presenta una mejora al método de Barlow con el objetivo de reducir el número de comunicaciones necesarias para una implementación paralela destinada a sistemas de memoria distribuida.

Para este proyecto se pondrá especial interés en los algoritmos alternativos de bidiagonalización propuestos por Ralha y Barlow [13,14], dado que están pensados para soportar el paralelismo. La decisión se sustenta en que el resultado final de esta etapa de la investigación se orienta hacia algoritmos que puedan ser implementados en plataformas paralelas y, en particular aprovechando la capacidad de las unidades de procesamiento gráfico (GPU, por sus siglas en inglés).

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

A lo largo de varios proyectos de investigación, se ha profundizado la investigación de la RI. La meta inicial del equipo fue construir íntegramente un prototipo de un sistema de organización de documentos para su posterior recuperación mediante un buscador. Habiendo concluido esa primera etapa se buscó optimizar el trabajo de RI lo cual implica dar atención a distintas líneas de investigación:

a) Mejorar la lematización disponible, en particular para este caso del idioma español;
b) Subdividir el corpus en forma inteligente de modo tal que sin gran pérdida de exhaustividad se pueda resolver la consulta examinando una o más partes de la subdivisión, excluyendo a muchas de ellas; y
c) Acelerar la velocidad de cómputo.

En particular, esta etapa de la investigación se relaciona a la línea de investigación mencionada en el ítem c. El objetivo principal de este trabajo apunta a diseñar, implementar y evaluar algoritmos secuenciales y paralelos para resolver eficientemente cada uno de los algoritmos utilizados en la DVS. En función a tal objetivo se han trazado los siguientes objetivos específicos:

a) Estudiar el problema matemático de la DVS y las variantes algorítmicas existentes para una mejor implantación en GPU.
b) Estudiar a bajo nivel las arquitecturas de los equipos CPU y GPU, sobre los cuales se realizarán los desarrollos, así como de las herramientas de software necesarias para su máximo aprovechamiento.
c) Estudiar las principales librerías disponibles que resuelven problemas relacionados con el álgebra lineal, especialmente aquellas que lo realicen en

arquitectura de cálculo paralelo, por ejemplo, CUDA, CUBLAS, OPENCL, etc.

d) Desdoblamiento de los algoritmos para resolverlos sobre una configuración híbrida.
e) Realizar un estudio comparativo en cuanto al rendimiento al bidiagonalizar matrices de variados tamaños cuando se utilizan distintas implementaciones variando la arquitectura. Determinando que algoritmo e implementación resulta más eficiente.
f) Calcular la DVS utilizando el algoritmo identificado en el punto anterior, y finalmente implementarlo en el SRI desarrollado por el equipo.

3. RESULTADOS OBTENIDOS/ESPERADOS

A continuación, se enumeran los resultados ya alcanzados:

i. Se construyeron dos prototipos de SRI, uno basado en lo que ahora se pueden llamar métodos clásicos y otro según el método LSI. Estos se diseñaron de forma modular de tal manera que permitan mejoras locales con su consiguiente experimentación. Los mismos se almacenaron con código abierto a fin de facilitar a nuevos grupos interesados en esta tecnología para su iniciación en el tema y la base sobre la cual efectuar experimentos propios.
ii. Se extendió la selección de documentos a corpus voluminosos a través de la utilización de DVS y clustering. La DVS proporciona una salida donde las diferencias de las distancias entre los documentos son muy cercanas, mientras que clustering utiliza esas distancias para poder agrupar los documentos, parte de este trabajo ha sido publicado en [15].
iii. Se estudió como línea alternativa fraccionar el corpus. Esto requiere dos algoritmos preparatorios, uno que particione el corpus utilizando una noción de vecindad o similitud y el entrenamiento de un algoritmo de clasificación que direccione la consulta hacia la parte más promisoria. Para ambos servicios se aplicaron técnicas de minería de datos y de la selección de la parte usando redes neuronales. Los resultados se han reflejado parcialmente en [16].
iv. Adicionalmente se desarrolló un generador de corpus, para ser utilizado en esta línea de investigación y pretendiendo, además, con esto, colaborar con la propagación de la Lingüística de Corpus como metodología para investigaciones en RI, el trabajo realizado se describe en [17].

En las próximas etapas de investigación, se profundizará el análisis de la posibilidad de optimizar la DVS, dando atención en particular a la primera fase, la bidiagonalización. Para ello se realizará un estudio comparativo en cuanto al rendimiento al bidiagonalizar matrices de variados tamaños cuando se utilizan distintos algoritmos y distintas implementaciones variando la arquitectura. De esta manera se intentará determinar qué algoritmo e implementación resulta más eficiente. La investigación se centrará en el desarrollo e implementación de algoritmos paralelos, principalmente el de DVS, en arquitecturas basadas en GPU y su comparación con clústeres de multicores, así como el empleo combinado de GPU y multicores. El alto grado de paralelismo de las GPU que sin lugar a duda disminuye el tiempo de cálculo, sufre una mengua en la misma a causa de la lentitud de sus comunicaciones transversales. De allí surge el interés del equipo en explorar si soluciones híbridas pudieran aportar una aceleración en los cómputos, delegando en cada parte, CPU o GPU, aquellas tareas en las cuales mejor se desempeñan.

Por otro lado, se llevará a cabo la puesta a prueba de dicha optimización en el SRI desarrollado por el equipo con el fin de comprobar el nivel de impacto alcanzado en la productividad del proceso. No es objetivo de este proyecto obtener una resolución en forma abstracta y genérica como para enriquecer las bibliotecas del cálculo matricial, sino resolverlo para ciertas arquitecturas concretas disponibles en la sede del proyecto. Esto marcaría el camino para que otros, con más recursos económicos y tamaño del equipo

humano, puedan extenderlo y parametrizarlo para que funcione en otras configuraciones.

4. FORMACIÓN DE RECURSOS HUMANOS

En el proyecto participan seis investigadores, uno de ellos en formación y dos son alumnos de grado. La línea de investigación presentada aquí es parte directa de la tesis "*Estudio comparativo de DBSCAN, KMEANS con redes neuronales en un Sistema de Recuperación de Información*", correspondiente a la Maestría en Informática que está desarrollando el Ing. Casuscelli Marcos en UNLaM.

Durante el último año la Ing. Viviana Ledesma presentó su tesis de maestría y su posterior defensa, la cual desarrolló en la UNLaM, y por su parte, el alumno Gastón Procopio, finalizó su carrera de Ingeniería en Informática con la aprobación del proyecto final de carrera.

Parte de los resultados de esta investigación son divulgados en la cátedra de Diseño de Sistemas que se dicta para la carrera de Ingeniería en Informática de la UNLaM. Se espera además que esta investigación contribuya a la formación de recursos humanos en RI y que el sistema desarrollado pueda servir de base para una transferencia de tecnología a las PYMEs de la región.

5. BIBLIOGRAFÍA

[1] Cleverdon, C.W. "Progress in documentation. Evaluation of information retrieval systems", Journal of Documentation, 26, 55-67, 1970.

[2] Kowalski, G. "Information Retrieval Systems: Theory and Implementation", 1st ed. Norwell, MA, USA: Kluwer Academic Publishers, 1997.

[3] Kuna, H., Rey, M., Martini, E., Solonezen, L. & Podkowa, L. "Desarrollo de un Sistema de Recuperación de Información para Publicaciones Científicas del Área de Ciencias de la Computación", Revista Latinoamericana de Ingeniería de Software, 2014. 2(2): 107-114.

[4] Tolosa G. & Bordignon, F. "Introducción a la Recuperación de Información: Conceptos, modelos y algoritmos básicos". Universidad Nacional de Luján, Argentina, 2008. Recuperado el 01/08/2019 de: http://eprints.rclis.org/12243/1/Introduccion-RI-v9f.pdf

[5] Deerwester, S., Dumais, S., Furnas, G., Landauer, T. & Harshman, R. "Indexing by latent semantic analysis". Journal of the American Society for Information Science (SIAM),1990. 41(6):391–407.

[6] Lahabar, S. & Narayanan, P. "Singular Value Decomposition on GPU using CUDA". IEEE International Symposium on Parallel & Distributed Processing, 2009. 1-10.

[7] Berry, M., Dumais, S. & O'Brien, G. "Using Linear Algebra For Intelligent Information Retrieval". Society for Industrial and Applied Mathematics, Review 37(4): 573-595. Philadelphia, USA, 1995.

[8] L. Fortuna, G. Nunnari & A. Gallo. "Model order reduction techniques with applications in electrical engineering". Springer-Verlag, 1992.

[9] J. Demmel, M. Gu, S. Eisenstat, et al. "Computing the Singular Value Descomposition with Hight Relative Accuracy". Linear Algebra and its Application, 299, 21-80, 1999.

[10] T. Chan. "An Improved Algorithm for Computing the Singular Value Decomposition". ACM Transactions on Mathematical Software, 8(1): 72-83, 1982.

[11] Sangwine, S. & Le Bihan, N. "Quaternion Singular Value Decomposition based on Bidiagonalization to a Real Matrix using Quaternion Householder Transformations" Applied Mathematics and Computation, ELSEVIER, 182(1): 727-738, 2006.

[12] Da Silva Sanches de Campos, C. "Algoritmos de Altas Prestaciones para el Cálculo de la Descomposición en Valores Singulares y su Aplicación a la Reducción de Modelos de Sistemas Lineales de Control". Tesis Doctoral. Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, España, 2014.

[13] Ralha, R. "One-sided reduction to bidiagonal form". Linear Algebra and Its Applications, ELSEVIER, 358(1-3): 219-238, 2003.

[14] Barlow, J., Bosner, N., Drmač, Z. "A new stable bidiagonal reduction algorithm". Linear Algebra and Its Applications, ELSEVIER, 397: 35-84, 2005.

[15] Spositto, O., Procopio, G., Quintana, F.& Ryckeboer H. "Una paralelización del método de Householder", CACIC 2016, pp. 1291-1300. Universidad Nacional de San Luis San Luis, 2016.

[16] Spositto, O., Casuscelli, M., Bossero, J., Matteo, L., Ryckeboer, H. "Aceleración en la Recuperación de Información utilizando Algoritmos de Minería de Datos de R". CACIC 2018, pp.491-500.Universidad Nacional del Centro, Tandil, 2018.

[17] Spositto, O., Procopio, G. Bossero, J. "Método para la Construcción de un Corpus Periodístico mediante Expresiones Regulares". CONAIISI 2018, pp. 491-500. Universidad CAECE, Mar del Plata, 2018.

Métodos y Algoritmos para Procesamiento Distribuido de Grafos Masivos y Evolutivos

Tomás Delvechio[1], Andrés Giordano[1], Gabriel H. Tolosa[1,2]
{tdelvechio, agiordano, tolosoft}@unlu.edu.ar
[1]Departamento de Ciencias Básicas, Universidad Nacional de Luján
[2]CIDETIC, Universidad Nacional de Luján

Resumen

El desafío de representar datos en entornos digitales posee gran interés frente a la masiva adopción de Tecnologías de la Información por parte de las poblaciones a escala mundial. Estas representaciones cobran importancia central y las estructuras elegidas condicionan los algoritmos y métodos que se usan para su procesamiento. Los grafos (o redes) son una estructura de datos que permite gran expresividad de representación, han sido ampliamente estudiados y existen algoritmos bien conocidos para su procesamiento. Sin embargo, en un entorno de datos masivos aparecen retos que en muchos casos no admiten soluciones triviales. La escalabilidad y optimización de los algoritmos de procesamiento de redes son motivo de incesante trabajo. Decisiones como particionar la estructura en varios subgrafos, generar índices que resuman la información para realizar estimaciones o procesar cambios de la red a través del tiempo afectan de formas diversas los tiempos de cáculo de algunas métricas (ejemplo, de centralidad, distancias, etc.) o la necesidad de almacenamiento. En este plan se propone analizar cómo es posible combinar/rediseñar varias de estas técnicas para obtener ventajas en entornos distribuidos o en el contexto nativo en que se ejecutan estos procesos, estudiar cómo son afectados los recursos según las decisiones de diseño a adoptar y realizar pruebas para conjuntos de datos de composición y tamaños heterogéneos.

Palabras clave: grafos, procesamiento distribuido, grafos dinámicos, estimación.

Contexto

Esta presentación se encuentra enmarcada en el proyecto de investigación "Estrategias y Algoritmos para Problemas de Búsquedas a Gran Escala" (Disposición CD-CB Nº 350/19) del Departamento de Ciencias Básicas (UNLu).

Introducción

La disponibilidad de información digital en escala global genera un constante interés en estrategias para su representación y procesamiento. Sin mucho esfuerzo resulta trivial obtener acceso a ingentes cantidades de datos a costos bajos o nulos. Un caso de interés son los datos que forman redes (*linked-data*) y que se modelan utilizando una de las representaciones más versátiles y utilizadas en computación: los grafos o redes [15].

Los nodos de un grafo[1] pueden representar infinidad de conceptos y las aristas expresan relaciones entre los nodos. De aquí, diferentes áreas en las ciencias y la tecnología trabajan cotidianamente con algoritmos de grafos, dada esta versatilidad y poder de abstracción.

Un cambio ocurrido en los últimos años es el crecimiento en el tamaño de los grafos que representan diferentes objetos de estudio (por ejemplo, Facebook reportaba que su grafo social contaba con millones de nodos y miles de millones de aristas para el 2015 [7]). A este escenario se le agrega que algunas aplicaciones requieren realizar consultas con severas restricciones de tiempo. Aparece entonces un problema cuando se consideran grafos masivos y es que ciertos algoritmos *'clasicos'* dejan de ser competitivos por no poder responder en tiempos adecuados.

Otra característica que complejiza el tratamiento de estas estructuras es su dinámica. Hoy en día, existen grafos provenientes de sistemas que evolucionan constantemente, agregando nuevos nodos y

[1]En general, se suele definir un grafo como $G = (V, E, w)$, donde V es un conjunto de nodos, E es un conjunto de aristas que vinculan 2 nodos (v_i, v_j) donde $v_i, v_j \in V$ y w es un vector de pesos para cada arista $e_i \in E$.

aristas en tiempo real, generando grafos dinámicos o evolutivos [13]. Por ello, su versatilidad para representar relaciones y el constante crecimiento en cuanto a tamaño y complejidad genera constante interés en el desarrollo de estrategias para su procesamiento eficiente.

Uno de los aspectos más interesantes de los grafos es que se pueden estudiar muchas propiedades a partir de las conexiones, sin considerar el significado intrínseco de sus componentes. Esto es, no importa si los nodos representan personas en una red social o routers en una red de datos; se pueden analizar y estudiar algunos comportamientos a partir de los nodos y sus enlaces. Un caso de estudio donde este enfoque es aplicado con notable éxito es en el estudio de la web y los algoritmos de análisis de enlaces utilizados para el ranking de las páginas en los motores de búsqueda [3]. En estos casos, una propiedad ampliamente estudiada (como el grado de los nodos) es utilizada como medida de reputación a la hora de reordenar los resultados [16].

Además de la distribución de grado, otras métricas y algoritmos son utilizados en las ciencias de la computación para el estudio de propiedades de los grafos. Por ejemplo, el camino más corto entre dos nodos. De estos algoritmos, cuyas propuestas conocidas suelen ser de orden de complejidad no lineales, surgen métricas importantes, como por ejemplo medidas de centralidad, utilizadas ampliamente en el análisis de redes sociales [12], donde se establece algún tipo de importancia a diferentes vértices en relación a la cantidad de caminos más cortos que pasan por éste [5]. En muchos casos, estos requerimientos obligan a realizar un procesamiento distribuido del grafo, requiriendo versiones de los algoritmos clásicos adaptadas a ejecutarse en un cluster de computadoras siendo, en general, problemas que no son trivialmente paralelizables dada la naturaleza linkeada de la estructura de datos. Un ejemplo de esto es Betweenness Centrality, una medida de centralidad a partir de SSSP[2], que asigna un ranking de importancia a los nodos en función de la cantidad de caminos que lo atraviesan [12]. Una opción al procesamiento distribuido para el cálculo exacto de ciertas métricas, es la generación de un índice que resume la información del grafo para estimar estos valores [17] [3] [2]. El objetivo principal de este enfoque es lograr la mejor calidad posible en este 'resumen' de tal manera que el error de estimación sea el menor posible. En estos casos siempre existe un trade-off entre el costo de almacenamiento de esta estructura y la calidad de la estimación de las métricas. Luego, como método adicional para reducir aún mas el error, las estimaciones pueden ser corregidas si existe alguna función que modele el comportamiento dependiendo de la técnica que se use para estimar.

Como se mencionó, una línea de trabajo desarrolla la problemática de estudiar grafos dinámicos o evolutivos. La idea de que el grafo evoluciona en el tiempo en cuanto a sus componentes agrega una complejidad desafiante que es tema de estudio en la actualidad [9][10]. Dado que existen métricas que se calculan para un versión inicial del grafo en un instante, se considera de interés estudiar cómo es posible calcular las métricas para una nueva versión de la estructura sin tener que procesar todo el grafo completamente. Esta idea resulta relevante principalmente en grafos masivos y métricas computacionalmente costosas, en los cuales el cómputo eficiente proviene solo de procesar las novedades [14] [1].

Líneas de I+D

En este trabajo se describen líneas de I+D del grupo, las cuales se basan, principalmente, en desarrollar estrategias y analizar *trade-offs* existentes entre diferentes metodos de procesamiento (exactos y aproximados), sobre estrategias distribuidas en clusters de *commodity hardware*, con las consiguientes ventajas y desventajas que estos escenarios presentan. De forma continua aparecen nuevos desafíos en estas áreas y oportunidades de investigación en temas ya explorados por la comunidad científica. En particular, las líneas de I+D principales son:

a. Cálculo Distribuido en Grafos Evolutivos

Diversos algoritmos que se aplican sobre grafos son ampliamente utilizados en la industria y la academia para abordar un gran abanico de soluciones. Algunos de estos algoritmos son los de camino mas corto (Single Source Shortest Path o SSSP y All-Pairs Shortest Path o APSP), y suelen constituir la base de muchos procesos en ámbitos como la recuperación de información [6] o análisis de redes sociales [4], por nombrar un par de ejemplos.

Este tipo de procesos suelen tener soluciones aceptadas desde hace mucho tiempo, las cuales no

[2]Single Source Shortest Path

escalan cuando el tamaño de los datos supera un umbral, dando lugar a soluciones en tiempos no triviales [8]. Por otro lado, es todo un desafío la implementación de estrategias que aprovechen las características distribuidas de las plataformas existentes [10]. Otro problema que presentan los enfoques secuenciales tradicionales y no escalables es que muchas aplicaciones exigen de forma creciente respuestas cada vez mas instantáneas, lo que introduce la posibilidad de construir soluciones aproximadas en el resultado pero eficientes en el tiempo insumido. También resulta importante considerar que los grafos, al representar relaciones, pueden cambiar a lo largo del tiempo, que se conoce como grafos dinámicos [11], y admiten soluciones que tengan en cuenta procesamientos previos para no tener que recalcular completamente las métricas ante cada modificación del mismo.

Otro aspecto para lograr esta escalabilidad es tener esquemas de procesamiento parciales de la estructura de datos. Una estrategia conocida es el particionado del grafo (*divide-and-conquer*), que consiste en dividirlo en porciones más pequeñas (subgrafos) para poder procesarlo de forma distribuida. Esta estrategia ha sido ampliamente adoptada en los últimos años dado los tamaños de los grafos provenientes de servicios digitales [7] y las restricciones en cuanto a tiempos de respuesta. Los algoritmos de partición de grafos persiguen diversos en objetivos, aplicaciones y complejidad lo que lleva a esquemas que utilizan diferentes heurísticas, suposiciones y restricciones [1]. Esto dificulta su comparación por lo que la selección de la técnica adecuada para un caso puntual es no trivial.

Dentro del marco del proyecto, el objetivo de esta linea de trabajo es diseñar y analizar algoritmos distribuidos sobre grafos dinámicos masivos considerando el *tradeoff* entre diferentes modelos de actualización de los datos (snapshots vs stream) considerando ademas esquemas de particionamiento adecuados.

b. Estimación de Distancia entre Nodos

El problema de la distancia entre dos nodos tiene múltiple aplicaciones prácticas, por ejemplo, para el ranking en búsquedas en redes sociales o profesionales[3] o, como se mencionó, el ranking de resultados de un motor de búsqueda[4]. A esta métrica se la define como la longitud del camino más corto entre dos nodos y se vuelve inviable si se requiere responder en pocos milisegundos en un grafo masivo. Esta distancia es usada en numerosos algoritmos que apuntan a resolver problemas como la recomendación de links [18] y agrupamiento de usuarios [9], entre otros.

El cálculo exacto de esta métrica entre dos nodos arbitrarios se ve afectado en cuanto a la eficiencia del proceso debido al tamaño de los grafos actuales resultando prohibitivo para aplicaciones prácticas. Por ejemplo, en redes sociales digitales el número de relaciones (aristas) supera ampliamente la cantidad de usuarios (nodos)[5].

La estimación de este valor es una alternativa válida y la reducción del error asociado (a un costo computacional bajo) conforma una de las líneas de investigación de este proyecto. En este enfoque se utiliza un conjunto de nodos, llamados *landmarks* [17], que se toman como referencia para estimar luego la distancia entre dos nodos arbitrarios. El problema de la selección de *buenos landmarks*, es decir, aquellos que permitan minimizar el error de estimación, es una pregunta abierta ya que existen diversos criterios a aplicar que consideran grafos de diferente tamaño, densidad y dinámica. Algunos resultados preliminares sobre nuevos métodos de selección de *landmarks* indican que usar métricas tradicionales en conjunto (*Closeness Centrality* o *Degree*, entre otras) , sobre ciertos grafos, potencian la estimación de la distancia logrando mejores resultados respecto del uso individual de éstas.

El desarrollo de métodos de corrección de la distancia estimada no ha sido abordado hasta el momento y es otra de las lineas de investigación de este proyecto. La idea consiste en lograr un algoritmo que, basado en características generales del grafo (como la densidad, diámetro, tamaño de comunidades, entre otras) permita corregir el valor de la estimación. Algunos resultados preliminares obtenidos por nuestro grupo sobre los mismos grafos estudiados en otros trabajos de referencia indican que es posible reducir el error de forma significativa mediante esta técnica.

[3] Un caso es la red de contactos profesionales Linkedin.

[4] `https://patents.google.com/patent/US9165040B1/en`

[5] Twitter: 321 millones de usuarios activos, 226.947 millones de vínculos. `http://investor.twitterinc.com/home/default.aspx`

Resultados y objetivos

El objetivo principal de esta propuesta es desarrollar, evaluar y transferir algoritmos y estrategias para abordar algunas de las problemáticas relacionadas con el procesamiento de grafos masivos en entornos de clusters. En general, se proponen mejoras que apuntan a la eficiencia en la distribución de trabajo y la escalabilidad. En particular, se espera alcanzar los siguientes objetivos:

- Definir y evaluar estructuras y modelos de cómputo distribuido sobre clusters de *hardware commodity* para problemas de escalabilidad en el cálculo exacto de métricas sobre grafos masivos.
- Considerar y analizar el impacto en el rendimiento y escalabilidad de utilizar estrategias de procesamiento para grafos masivos evolutivos. De forma similar, analizar el efecto de utilizar técnicas de particionado en estos entornos distribuidos.
- Diseñar y estudiar estrategias de estimación de distancias entre nodos de un grafo masivo para problemas de búsqueda.
- Diseñar y estudiar estrategias de corrección de la estimación de distancias entre nodos de un grafo masivo.

Formación de Recursos Humanos

En el marco de estas líneas de investigación se está dirigiendo una tesis de Licenciatura en Sistemas de Información (UNLu). Además, asociados al proyecto de investigación hay una estancia de investigación de la Secretaría de CyT (UNLu) y se espera contar con un becario de doctorado UNLu y al menos un pasante en 2020.

Referencias

[1] Z. Abbas, V. Kalavri, P. Carbone, and V. Vlassov. Streaming graph partitioning: an experimental study. *Proceedings of the VLDB Endowment*, 11(11):1590–1603, 2018.

[2] T. Akiba, T. Hayashi, N. Nori, Y. Iwata, and Y. Yoshida. Efficient top-k shortest-path distance queries on large networks by pruned landmark labeling. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[3] T. Akiba, Y. Iwata, and Y. Yoshida. Dynamic and historical shortest-path distance queries on large evolving networks by pruned landmark labeling. In *Proceedings of the 23rd international conference on World wide web*, pages 237–248, 2014.

[4] D. A. Bader, S. Kintali, K. Madduri, and M. Mihail. Approximating betweenness centrality. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 124–137. Springer, 2007.

[5] U. Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.

[6] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. 1998.

[7] A. Ching, S. Edunov, M. Kabiljo, D. Logothetis, and S. Muthukrishnan. One trillion edges: Graph processing at facebook-scale. *Proceedings of the VLDB Endowment*, 8(12):1804–1815, 2015.

[8] A. Crauser, K. Mehlhorn, U. Meyer, and P. Sanders. A parallelization of dijkstra's shortest path algorithm. In *International Symposium on Mathematical Foundations of Computer Science*, pages 722–731. Springer, 1998.

[9] J. Edachery, A. Sen, and F. J. Brandenburg. Graph clustering using distance-k cliques. In *International Symposium on Graph Drawing*, pages 98–106. Springer, 1999.

[10] N. Edmonds, T. Hoefler, and A. Lumsdaine. A space-efficient parallel algorithm for computing betweenness centrality in distributed memory. In *2010 International Conference on High Performance Computing*, pages 1–10. IEEE, 2010.

[11] D. Ferone, P. Festa, A. Napoletano, and T. Pastore. Shortest paths on dynamic graphs: a survey. *Pesquisa Operacional*, 37(3):487–508, 2017.

[12] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[13] A. Kosmatopoulos, K. Giannakopoulou, A. N. Papadopoulos, and K. Tsichlas. An overview of methods for handling evolving graph sequences. In *International Workshop on Algorithmic Aspects of Cloud Computing*, pages 181–192. Springer, 2015.

[14] A. McGregor. Graph stream algorithms: a survey. *ACM SIGMOD Record*, 43(1):9–20, 2014.

[15] M. Newman. *Networks*. Oxford university press, 2018.

[16] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[17] M. Potamias, F. Bonchi, C. Castillo, and A. Gionis. Fast shortest path distance estimation in large networks. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 867–876, 2009.

[18] Y. Zhang and J. Pang. Distance and friendship: A distance-based model for link prediction in social networks. In *Asia-Pacific Web Conference*, pages 55–66. Springer, 2015.

Sintonización de Aplicaciones científico/ingenieriles: un proceso de medición-mejora para incrementar la eficiencia

Caymes-Scutari Paola[1,2], Bianchini Germán[1], Tardivo M. Laura[1,3], Méndez-Garabetti Miguel[1]

[1]Laboratorio de Investigación en Cómputo Paralelo/Distribuido
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579
[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
[3]Departamento de Computación Paralelo/Distribuido
Facultad de Ciencias Exactas, Físico-Químicas y Naturales
Universidad Nacional de Río Cuarto
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579

pcaymesscutari@frm.utn.edu.ar, gbianchini@frm.utn.edu.ar, lauratardivo@dc.exa.unrc.edu.ar, miguelmendezgarabetti@gmail.com

RESUMEN

La medición constituye una disciplina fundamental para guiar la toma de decisiones. La motivación por medir ha estado históricamente ligada a la necesidad de seleccionar la alternativa más prometedora para mejorar la calidad de los productos y/o procesos. El campo de la informática/procesamiento no escapa a esto, y es allí donde emerge la Sintonización como un proceso de medición-mejora permanente de las aplicaciones, con la finalidad de ajustar y/o adecuar el comportamiento dinámico de la aplicación y lograr una ejecución globalmente más eficiente. En este proyecto se propone abordar la sintonización de aplicaciones de interés científico/ingenieril a través de la utilización de diferentes técnicas y modelos de rendimiento, como así también del estudio y definición de nuevos modelos. Aplicaciones de diversa naturaleza podrán valerse de los beneficios de este proceso: aplicaciones paralelas, de predicción, de generación foto-realista o renderización, *big data*, métodos de reducción de incertidumbre, etc. La calidad de la ejecución, la reducción del tiempo de ejecución, y la eficiencia en la utilización de los recursos computacionales, son los tres objetivos principales al incluir capacidades para la sintonización, a través de modelos y métricas que detectan la aparición de problemas y activan las acciones correctivas o de mejora.

Palabras clave: Sintonización Automática, Modelo de Rendimiento, Sintonización Dinámica, Aplicaciones Paralelas.

CONTEXTO

El proyecto SIUTIME0007840ME inició su ejecución en enero de 2020, en el

marco del LICPaD (Laboratorio de Investigación en Cómputo Paralelo/Distribuido) dentro del ámbito de la UTN-FRM. El mismo continúa la línea de investigación de proyectos previos de nuestro laboratorio, pero en esta oportunidad haciendo foco en la necesidad del proceso de medición-mejora. Tal como expresara a finales del siglo XIX el físico Lord Kelvin, '*cuando puedes medir sobre lo que estás diciendo y expresarlo en números, sabes algo sobre ello, pero cuando no puedes medirlo ni expresarlo en números entonces tu conocimiento es precario e insatisfactorio*'. Y así es que continuamente realizamos mediciones que nos guían a la hora de tomar decisiones y seleccionar la alternativa que creemos mejor, tanto en nuestra vida personal a la hora de comprar un producto, seleccionar una ruta, o elegir un médico, como en lo que respecta a diversas actividades profesionales. La necesidad y motivación por medir está ligada fundamentalmente a la mejora de calidad continua que requieren los procesos y productos. Para poder asegurar que un proceso o sus productos resultantes son de calidad, o bien para poder realizar comparaciones, es necesario asignar valores, descriptores, indicadores o algún otro mecanismo mediante el cual se pueda llevar a cabo dicha comparación [1]. El campo de la informática y del procesamiento no escapan a esta necesidad de medición, evaluación y mejora, y es allí donde emerge la Sintonización como un proceso de medición-mejora permanente de las aplicaciones, con la finalidad de ajustar y/o adecuar el comportamiento dinámico de las mismas y lograr una ejecución globalmente más eficiente [2]. Ésta es precisamente la temática abordada en este proyecto.

1. INTRODUCCIÓN

Frecuentemente, y sobre todo en sistemas críticos, es necesario realizar mediciones para comparar y contrastar diferentes aspectos que hacen poner en relieve la aproximación, herramienta o aplicación más adecuada para el problema que hemos de resolver. La sintonización automática es un paradigma tecnológico que permite que los programas sean adaptables a una variedad de condiciones computacionales. Su mayor auge se ha originado en estrecha relación a la computación de alto rendimiento, dados los requerimientos de rendimiento inherentes a todo sistema paralelo/distribuido [3,4]. El proceso de Sintonización involucra varias **fases** sucesivas que se relacionan en pos del objetivo de mejorar la calidad de la aplicación y de su ejecución. La primera etapa es la **Instrumentación**, en la cual la aplicación es anotada a fin de habilitar la medición de valores propios de la aplicación y/o su ejecución (capturar valores de variables, medir tiempos de ocio o ejecución, etc.). La segunda etapa es la **Monitorización**, durante la cual se recolectan, se clasifican y se almacenan los datos obtenidos por medio de la instrumentación. La siguiente etapa es la de **Análisis**, en la que las mediciones y la información monitorizadas son evaluadas y valoradas a fin de detectar posibles cuellos de botella que actúen en detrimento de la calidad de la ejecución. La etapa final del proceso se denomina **Sintonización** en sí misma, pues es la que introduce y materializa los cambios necesarios en la aplicación para adaptar su comportamiento y ajustar su rendimiento. Cabe mencionar que la fase de análisis constituye el corazón del proceso, y es la fase más compleja de tratar y programar, ya que requiere contar con algún tipo de conocimiento experto

sobre el problema a resolver a fin de evaluar, valorar y tomar decisiones sobre cuál es la mejor acción a tomar frente a la aparición de un problema de rendimiento. Para la toma de decisiones, en cada una de las distintas aproximaciones se suelen utilizar modelos matemáticos, lógica difusa, ciertos algoritmos, y/o métodos basados en la historia. Cada uno de ellos tiene características diferentes en cuanto a la calidad de los resultados obtenidos, la sobrecarga que ocasionan sobre la aplicación original, y la velocidad de respuesta con la que son arrojadas las decisiones [2, 4].

En particular, el proyecto propone dar tratamiento al estancamiento y convergencia prematura que aquejan a diversos algoritmos evolutivos poblacionales [5], a fin de modelar, detectar y actuar en consecuencia para aplicar acciones de sintonización que permitan paliar el efecto negativo que ocasionan en los resultados. Existen diversos algoritmos que utilizan una población de soluciones candidatas en el proceso de búsqueda de una solución aceptable para un problema dado [6, 7]. Dichos algoritmos se basan en conceptos estadísticos y evolutivos a partir de los que consideran que trabajar con una muestra suficientemente representativa del universo de búsqueda es suficiente para alcanzar un resultado aproximado a la solución óptima. Dicha muestra está compuesta por elementos del dominio de búsqueda (también denominados **individuos**, con diferentes características: cada individuo, usualmente, representa un conjunto de parámetros con los que ha de interpretarse un modelo de simulación o ejecutarse un algoritmo, y es en el valor de tales parámetros de entrada en donde difieren unos individuos de otros, con el consecuente efecto en la salida que arrojará el programa), los cuales son sometidos a alguna transformación y/o combinación a lo largo de las iteraciones, hasta que se converge a un óptimo, que en el mejor de los casos será el óptimo global. La computación evolutiva se enmarca en estos principios, y en la taxonomía reúne una variedad de algoritmos cuyos principales puntos de encuentro son la utilización de poblaciones, alguna forma de medir la calidad de los individuos, el constante proceso de selección de los supervivientes a lo largo de las iteraciones (o también denominadas generaciones), y en la utilización de ciertos criterios de parada que permiten decidir, respectivamente, cuándo finalizar una generación para pasar a la siguiente, y cuándo terminar el algoritmo [6].

Cada algoritmo se caracteriza por los operadores que utiliza, los cuales definen la forma en la que el algoritmo transforma, combina o recombina los individuos de la población, ya sea individualmente, entre sí, o creando nuevos individuos. En este grupo de algoritmos cabe mencionar los algoritmos genéticos, los algoritmos evolutivos, la evolución diferencial, colonia de hormigas o abejas, entre otros [6]. Asimismo, existen diversas herramientas –entre ellas las de reducción de incertidumbre [7, 8, 9]– que se valen de los algoritmos evolutivos para guiar su proceso de búsqueda. Un punto de interés común a todos ellos se refiere al criterio de parada o finalización de cada generación. La decisión de frenado, usualmente, se basa en alcanzar cierto umbral en cuanto a la calidad de los resultados obtenidos hasta el momento, y/o a iterar una cierta cantidad de generaciones. No obstante, dichos criterios no siempre son representativos del comportamiento global de la población corriente ni de las características del problema particular que

se esté tratando, por lo que frecuentemente el proceso evolutivo se ve degradado por la convergencia prematura o el estancamiento, dos situaciones que van en detrimento de la calidad de los resultados y que sobrecargan y retrasan la toma de decisiones incrementando innecesariamente el tiempo de cómputo [5].

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

En este proyecto, el cual continúa la línea de trabajo de otros cinco proyectos previos vinculados a la sintonización y al desarrollo automático, se propone la definición de modelos de rendimiento para su utilización en la sintonización automática y dinámica basada en modelos matemáticos de rendimiento. Que la sintonización sea automática significa que el propio algoritmo o aplicación será dotado con la capacidad de realizar todas las etapas del proceso de sintonización, sin necesidad de que el usuario intervenga. Que sea dinámica significa que el proceso de mejora se llevará a cabo a la vez que el programa o aplicación esté ejecutándose, lo cual permitirá reflejar de modo inmediato las decisiones que se tomen en pos de la mejora del rendimiento, y así lograr ejecuciones más eficientes y adaptadas a las características de los datos particulares de entrada y al estado corriente del entorno de ejecución. [2, 4]. Que el proceso de sintonización esté basado en modelos matemáticos de rendimiento significa que el conocimiento experto que se utilizará en la etapa de Análisis estará centrado en la utilización de un conjunto de expresiones que modelan algún problema de rendimiento particular, lo cual, a la hora de tomar decisiones se reduce a la evaluación de tales expresiones matemáticas a partir de los datos recopilados durante la monitorización, lo que permite una toma de decisiones concisa en un tiempo de cómputo razonable, muchas veces despreciable en relación a la ejecución global del programa [2].

Dicha temática puede clasificarse dentro del área del Cómputo Paralelo, la Computación Evolutiva, la Sintonización automática y dinámica, y el Modelado.

3. RESULTADOS ESPERADOS

Se espera obtener dos tipos de resultados: por un lado, lograr el objetivo de sintonización en sí mismo, y de esa forma hacer un aporte a la mejora en el desempeño de los sistemas o métodos bajo estudio. Por otro lado, se espera mejorar los modelos de rendimiento existentes y a la vez definir nuevos modelos que superen las posibilidades de mejora.

4. FORMACIÓN DE RECURSOS HUMANOS

La temática propuesta por este proyecto permite continuar con la formación de los distintos integrantes del grupo de trabajo, de forma complementaria a la formación adquirida hasta el momento. En el caso particular de la Lic. Tardivo, su participación en el proyecto resulta natural dado que representa una temática estrechamente relacionada con su trabajo doctoral, el cual ha sido depositado para su defensa a lo largo del corriente año. En el caso del Mg. Ing. Méndez Garabetti la temática también resulta transversal a su plan de tesis doctoral (la cual acaba de

depositarse) y naturalmente seguirá avanzando y desarrollándose también en esta línea de investigación. Tanto la Lic. Tardivo como el Mg. Méndez Garabetti realizaron el Doctorado en Ciencias de la Computación de la Universidad Nacional de San Luis. Asimismo, el grupo de trabajo siempre está abierto a la incorporación de nuevos integrantes (de grado o postgrado) que deseen familiarizarse con las temáticas que aquí se describen.

5. BIBLIOGRAFÍA

[1] Piattini Velthuis, M., García Rubio, F., Garzás Parra, J., Genero Bocco, M. (2008) Medición y Estimación del Software – Técnicas y métodos para mejorar la calidad y la productividad. Alfaomega Ra-Ma.

[2] Caymes-Scutari, P., Bianchini, G., Sikora, A., Margalef, T., (2016) "Environment for Automatic Development and Tuning of Parallel Applications" HPCS 2016, International Workshop on Parallel Optimization using/for Multi and Many-core High Performance Computing, Innsbruck, Austria. pp. 743-750. IEEE.

[3] Wilkinson, B., Allen, M. (2005) Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers. Pearson Prentice Hall.

[4] Naono, K., Teranishi, K., Cavazos, J. y Suda, R. (2010) Software Automatic Tuning: From Concepts to State-of-the-Art Results, Springer, New York.

[5] Lampinen, J. y Zelinka, I. (2000), "On the Stagnation of the Differential Evolution algorithm", I.C. Soft Computing, pp. 76-83.

[6] Talbi, E. (2009) Metaheuristics: From Design to Implementation. John Wiley & Sons, Hoboken, New Jersey

[7] Bianchini, G., Denham, M., Cortés, A., Margalef, T., Luque, E. (2010) "Wildland Fire Growth Prediction Method Based on Multiple Overlapping Solution", Journal of Computational Science Vol 1 Issue 4, pp. 229-237. Elsevier.

[8] Tardivo, M.L., Caymes-Scutari, P., Méndez-Garabetti, M. y Bianchini, G. (2017) "Optimization for an Uncertainty Reduction Method Applied to Forest Fires Spread Prediction", Computer Science – Cacic 2017, pp. 13-23. Springer.

[9] Mendez Garabetti, M. Bianchini, G., Caymes Scutari, P., Tardivo M.L. (2016). "Increase in the quality of the prediction of a computational wildfire behavior method through the improvement of the internal metaheuristic", Fire Safety Journal, pp. 49-62. Elsevier.

Computación Serverless para tratamiento de datos provenientes de dispositivos de IoT

Nelson Rodríguez[1], María Murazzo[1], Diego Medel[1], Lorena Parra[1], Ana Laura Molina[1], Federico Sánchez[1], Fabiana Píccoli[2], Miguel Méndez Garabetti[3], Daniel Arias Figueroa[4], Martín Tello[5], Cynthia Aguilera[6], Hernán Atencio[6]

[1] Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.
[2] Departamento de Informática - F.C.F.M. y N – UNSL
[3] Instituto de Investigaciones, Facultad de Informática y Diseño, Universidad Champagnat
[4] Consejo de Investigación, FCE, Universidad Nacional De Salta
[5] Alumno con Beca de CIN – F.C.E.F. y N. – U.N.S.J.
[6] Alumno Avanzado Licenciatura en Sistemas de Información y Cs. de la Computación - F.C.E.F. y N. - U.N.S.J.

Complejo Islas Malvinas. Cereceto y Meglioli. 5400. Rivadavia. San Juan, 0264 4234129

nelson@iinfo.unsj.edu.ar, marite@unsj-cuim.edu.ar, vdiego.unsj@hotmail.com, lorenaparra152@yahoo.com.ar, almm95@gmail.com, fgsanchez@unsj-cuim.edu.ar, mpiccoli@unsl.edu.ar, mendez-garabettimiguel@uch.edu.ar, daaf@cidia.unsa.edu.ar, martinl.tello@gmail.com, cynaguilera95@gmail.com, hernan.atencio.98@gmail.com

Resumen

Cloud Computing se ha consolidado como una arquitectura que permite la provisión de recursos bajo demanda y de forma elástica; es ampliamente utilizada y ha demostrado su efectividad. Por otro lado Internet de las cosas ha comenzado a tener una gran adopción fruto de la emergente industria 4.0 y de las smart cities entre otras aplicaciones Serverless en un nuevo paradigma de desarrollo de software en el Cloud, que surge como una evolución del mismo, partiendo de una arquitectura monolítica, pasando por máquinas virtuales, luego contenedores y por último serverless, que en algunos casos se conoce como función como servicios. La misma está enfocada en proveer una arquitectura que permita la ejecución de funciones arbitrarias con mínima sobrecarga en la administración del servidor y soportada bajo la programación orientada a eventos. La cantidad y variedad de datos provenientes del Edge, necesitan ser procesados y en algunos casos en tiempo real. Estrategias como CloudIoT o Fog computing se han propuestos pero debido a los costos de los dispositivos fog enabled o a la latencia hacia el Cloud, estas soluciones son aplicables solo en determinados casos. Serverless provee la ejecución individual de funciones Cloud y es adecuada para el procesamiento de datos (ETL), procesamiento de datos provenientes de dispositivos de IoT (donde claramente favorece la escalabilidad) y computación científica. Esta unión de tecnologías presenta varios desafíos y oportunidades a resolver, entre ellas la redefinición de DevOps para serverless (NoOps), testing e interoperabilidad, performance, seguridad y análisis de costo entre otras. La evaluación de la eficiencia de las arquitecturas serverless y el estudio detallado de las mismas, permitirá plantear nuevos casos de uso, proponer estrategias para el procesamiento IoT y redefinir el proceso de desarrollo y operación

Palabras clave: *Serverless Computing, Edge Computing, IoT, Cloud Computing, Distributed Computing,*

Contexto

El presente trabajo se encuadra dentro del área de I/D Procesamiento Distribuido y Paralelo y es un resultado del proyecto de investigación: Computación Serverless para el tratamiento de datos provenientes de dispositivos de IoT, que está en proceso de evaluación para el período 2020-2021. Asimismo el grupo de investigación viene trabajando en proyectos relacionados con la computación distribuida y de alta performance desde hace más de 19 años. Como continuación del proyecto anterior: Orquestación de Servicios para la Continuidad Edge al Cloud, se sigue trabajando con los investigadores de otras universidades, lo cual favorece notablemente a todos las instituciones participantes.

Introducción

La computación serverless es un paradigma emergente en el que las aplicaciones de software se descomponen en múltiples funciones independientes sin estado [1] [2]. Las funciones solo se ejecutan en respuesta a acciones desencadenantes (como interacciones de usuario, eventos de mensajería o cambios en la base de datos), y se pueden escalar de forma independiente.

Por lo tanto, la computación serverless, también es a veces denominada función como servicio (FaaS). En este enfoque, casi todas las preocupaciones operativas son abstraídas lejos de los desarrolladores. De hecho, los desarrolladores simplemente escriben código e implementan sus funciones en una plataforma sin servidor. La plataforma se encarga de la ejecución de la función, el almacenamiento y la infraestructura de contenedor, redes y tolerancia a fallas. Adicionalmente, la plataforma serverless se encarga de escalar las funciones según la demanda real.

La computación serverless ha sido identificada como un enfoque prometedor para varias aplicaciones, como el análisis de datos en el borde de la red [3]. En consecuencia, una plataforma sin servidor maneja el ciclo de vida, la ejecución y escalada de las funciones reales; estas necesitan correr solo cuando son invocadas o activadas por un evento. Por lo tanto, el mayor beneficio de la computación serverless son las pocas preocupaciones operativas y de gestión y la utilización eficiente de los recursos [4].

Un modelo basado en funciones es particularmente adecuado para ráfagas, uso de CPU intensivo, cargas de trabajo granulares. Actualmente, los casos de uso de FaaS varían ampliamente, incluido el procesamiento de datos, el procesamiento de flujo, la computación de borde (IoT) y la computación científica [5] [6]. Con la continua experimentación en torno a FaaS, es probable que otros casos de uso surjan en un futuro.

En los últimos años, el paradigma de la función como servicio ha revolucionado la forma en que resuelve el procesamiento de eventos distribuidos. Los desarrolladores pueden implementar y ejecutar funciones controladas por eventos en el Cloud que se escalan bajo demanda. La ventaja de la alta disponibilidad y poca administración, son muy valoradas por la conducción de las empresas.

El término "sin servidor" es confuso ya que hay tanto hardware de servidor como procesos de servidor que se ejecutan, pero la diferencia en comparación con los enfoques normales es que la organización que construye y admite una aplicación "sin servidor" no se ocupa de ese hardware o esos procesos, debido a que están subcontratando esta responsabilidad.

Los primeros usos del término aparecieron en 2012, incluido el artículo de Ken Fromm [7]. El término se hizo más popular en 2015, luego del lanzamiento de AWS Lambda, y creció en popularidad después del lanzamiento de API Gateway de Amazon en julio de 2015. En octubre de ese año hubo una charla titulada "The Serverless Company using AWS Lambda", en referencia a PlayOn! Deportes. Hacia fines de año, el proyecto de código abierto "JavaScript Amazon Web Services (JAWS)" se renombró a Serverless Framework, continuando la tendencia.

A mediados de 2016, Serverless se había convertido en un nombre dominante para esta área, dando paso al nacimiento de la serie de conferencias Serverless, y varios proveedores Cloud adoptaron el término en todo, desde marketing de productos hasta descripciones de puestos de trabajo.

La Computación Serverless

Existen varias definiciones y todas son muy similares. Se toma de referencia la del grupo de investigación The SPEC Cloud:

La computación serverless es una forma de computación en el cloud que permite a los usuarios ejecutar eventos y aplicaciones facturadas de forma granular, sin tener que abordar la lógica operativa [8].

Esta definición coloca al servidor como una abstracción, que se superpone parcialmente con la plataforma como servicio (PaaS). Los desarrolladores se centran en abstracciones de alto nivel (por ejemplo, funciones, consultas y eventos) y en crear aplicaciones que los operadores de infraestructura asignan a recursos concretos y servicios de soporte. Esto efectivamente separa las preocupaciones, con los desarrolladores enfocándose en la lógica empresarial y en las formas de interconectar elementos de la lógica empresarial en flujos de trabajo complejos. Mientras tanto, los proveedores de servicios se aseguran de que las aplicaciones están alojadas en contenedores, desplegadas, aprovisionadas y disponibles bajo demanda, mientras se factura al usuario solo por los recursos utilizados [9].

La computación serverless se puede identificar como resultado de la unión de Cloud y Microservices Architecture. Pero la evolución de Cloud ha pasado por varias etapas.

Surgió como una evolución de todas las tecnologías Cloud. Se investigaron muchas vías diferentes para encontrar formas de simplificar la gestión de aplicaciones a escala, entornos virtuales, hosts virtuales dentro de máquinas virtuales, contenedores, por nombrar algunos. Con la aparición de Docker, los contenedores comenzaron a ser tenidos en cuenta y comenzaron a ofrecerse plataformas que permitían desplegar contenedores.

De estos microservicios, surgieron nuevas ofertas de "nivel de servicio" de alto nivel como Backend-as-a-Service (BaaS) y Function-as-a-Service (FaaS). Estos representaban los albores de la computación serverless.

La determinación exacta de lo que es "sin servidor" está en debate, pero se puede definir esto como: cualquier aplicación o servicio para el cual la empresa solo proporciona lógica de aplicación, que se ejecuta en una infraestructura que no administra. Los paradigmas claves detrás de la arquitectura empresarial y las aplicaciones que admite están cambiando rápidamente, presentando nuevas oportunidades de escalabilidad y agilidad.

Adoptar nuevos paradigmas será un desafío en el contexto de arquitecturas heredadas monolíticas y prácticas de TI arraigadas. Pero con la computación serverless puede mejorar enormemente la capacidad de su organización para reaccionar ante los cambios inevitables pero imprevisibles que aún están por venir.

La computación serverless es adecuada para aplicaciones de corta duración sin estado dirigidas por eventos, por ejemplo microservicios, backends IoT móviles, procesamiento de flujo modesto, bots e integración de servicios. No son indicados para cálculo numéricos de larga duración con estado como son: bases de datos, entrenamiento de aprendizaje profundo, análisis de flujo de servicio pesado, simulación numérica y vídeo transmitido en vivo [9].

La computación serverless se está utilizando para admitir una amplia gama de aplicaciones. Desde una perspectiva funcional, serverless y las arquitecturas más tradicionales pueden usarse indistintamente. La determinación de cuándo usar serverless estará influenciada por otros requisitos no funcionales como la cantidad de control sobre las operaciones requeridas y el costo y las características de la carga de trabajo de la aplicación, por ejemplo.

Para los proveedores, la computación serverless promueve el crecimiento empresarial, ya que facilita la programación, ayuda a atraer nuevos clientes y ayuda a los clientes a hacer un mayor uso de lo que ofrece el cloud. Por ejemplo, encuestas realizadas sobre el uso de tecnología indicaron que: el 41% de los encuestados usa tecnología serverless, mientras que el 31% no. Un 28% adicional tiene planes de usarlo dentro del próximo año. De los encuestados indicaron que el 70% utiliza AWS Lambda, seguido por Apache OpenWhisk (12%), Azure Functions (12%) y Google Cloud Functions (13%) [10].

La Arquitectura Serverless está estructurada de forma análoga al cloud tradicional, pero incluye una capa intermedia que se encuentra entre las aplicaciones y la plataforma base del cloud, lo que simplifica la programación. Las FasS en el cloud proporcionan información general, computan y se complementan con un ecosistema de ofertas especializadas de BaaS como almacenamiento de objetos, bases de datos o mensajes, entre otros. Por ejemplo, una aplicación serverless en AWS podría usar Lambda con S3 (almacenamiento de objetos) y DynamoDB (base de datos NoSQL), mientras que una aplicación en el Cloud de Google podría usar Cloud Functions con Cloud Firestore (base de datos de back-end móvil) y Cloud Pub / Sub (mensajería). Serverless también incluye servicios de big data como AWS Athena y Google BigQuery. La plataforma base del cloud subyacente incluye máquinas virtuales, privadas, redes privadas, almacenamiento virtualizado, gestión de identidad y acceso, así como facturación y seguimiento [11].

En el modelo tradicional, las aplicaciones consisten en máquinas virtuales grandes y multifuncionales con minutos de tiempos de aprovisionamiento. Estas máquinas virtuales actúan como cajas negras y son difíciles de modelar y predecir para los operadores. Aunque las aplicaciones suelen estar bloqueadas por uno de los recursos en una parte de la aplicación, los operadores sólo pueden escalar la aplicación completa para resolver el cuello de botella. Es posible eliminar algunos de estos problemas, pero requiere que el usuario vuelva a diseñar la aplicación, típicamente como microservicios. Desde la perspectiva de un proveedor, la computación serverless brinda una oportunidad para controlar toda la pila de desarrollo, reducir los costos operativos mediante una optimización eficiente y gestión de recursos en el cloud y la habilitación de un ecosistema sin servidor que fomenta el despliegue de servicios adicionales [12].

Si bien las primeras ventajas de serverless eran la económica y la sencillez en el desarrollo de aplicaciones, posteriormente surgieron nuevos beneficios como son: la escalabilidad, el código puede ejecutarse desde cualquier lugar y se adecúa al desarrollo de productos ágiles.

Por supuesto que existen desventajas como son: la depuración puede resultar más difícil de llevar a cabo, existe pérdida del control operativo, surgen nuevos problemas al utilizar APIs de terceros, nuevos riesgos de seguridad aparecen y los desarrolladores necesitan de un cambio de mentalidad, entre otras.

Líneas de Investigación, Desarrollo e Innovación

El tipo de investigación que se usará para llevar a cabo este proyecto comenzará con una investigación documental que permita definir las bases de conocimiento sobre la "Computación Serverless" y su convergencia al IoT. Esto será posible mediante la confección de revisiones sistemáticas y mapeos sistemáticos de la literatura a fin de definir el problema científico a solucionar.

Posteriormente la investigación se conducirá de forma hipotética mixta (experimental y deductiva). Esta forma de trabajo permitirá la selección de una plataforma de despliegue serverless sobre la cual se realizará el tratamiento de datos provenientes de los dispositivos IoT. Realizada esta selección, se fijará como principal variable el grado de escalabilidad horizontal lograda con la solución desarrollada.

Para lograr esto, se usará un método empírico analítico, que permita determinar si mediante el uso de la arquitectura serverless, se logra mejor escalabilidad en el tratamiento de datos provenientes de los dispositivos IoT.

Resultados y Objetivos

Resultados Obtenidos

Durante los últimos doce años se trabajó en el área de Computación de Altas Prestaciones y distribuidas, en particular sobre análisis de diversas arquitecturas paralelas y distribuidas, tales como: Cloud Computing, Cluster de commodity, arquitecturas distribuidas y paralelas de bajo costo y fog computing. Dicha experiencia impulsó esta línea de investigación. El grupo ha realizado varias

publicaciones durante el último año: doce trabajos de investigación en Congresos y Jornadas, se realizaron tres publicaciones en revistas científicas y se transfirieron los resultados mediante conferencias en eventos científicos.
Se han aprobado siete tesinas de grado, se incorporó un becario de investigación categoría alumno y otra beca está en evaluación.

Objetivos
El objetivo del grupo de investigación es Evaluar la eficiencia de las arquitecturas serverless para el tratamiento de datos proveniente de dispositivos de IoT, analizando variadas estrategias y evaluando cómo se comportan los parámetros de desempeño comparado con el modelo de cloud tradicional.

Formación de Recursos Humanos

El equipo de trabajo está compuesto por los nueve investigadores que figuran en este trabajo de las universidades Nacional de San Luis, Champagnat, Nacional de Salta y San Juan, un becario CIN y cinco alumnos.
Se está desarrollando una tesis doctoral sobre paralelismo híbrido y Big Data, seis tesinas de grado en el área de Serverless computing, Concurrencia y Computación distribuida y una tesis de maestría en áreas afines. Además se espera aumentar el número de publicaciones. Por otro lado también se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización o publicaciones de divulgación y asesoramiento a empresas y otros organismos del estado.

Referencias

[1] Adzic, G., & Chatley, R. (2017). Serverless computing: economic and architectural impact. In Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering (pp. 884-889). ACM.
[2] Baldini, I., Castro, P., Chang, K., Cheng, P., Fink, S., Ishakian, V. & Suter, P. (2017). Serverless computing: Current trends and open problems. In Research Advances in Cloud Computing (pp. 1-20). Springer, Singapore.
[3] Nastic, S., Rausch, T., Scekic, O., Dustdar, S., Gusev, M., Koteska, B. & Prodan, R. (2017). A serverless real-time data analytics platform for edge computing. IEEE Internet Computing, 21(4), 64-71.
[4] Mohanty, S. K., Premsankar, G., & Di Francesco, M. (2018). An Evaluation of Open Source Serverless Computing Frameworks. In CloudCom (pp. 115-120).
[5] Gottlieb, N. (2016). State of the Serverless Community Survey Results.
https: //serverless.com/blog/state-of-serverless-community/.
[6] Jonas, E., Pu, Q., Venkataraman, S., Stoica, I., & Recht, B. (2017). Occupy the cloud: Distributed computing for the 99%. In Proceedings of the 2017 Symposium on Cloud Computing (pp. 445-451). ACM.
[7] Fromm, K. (2012). https://readwrite.com/2012/10/15/why-the-future-of-software-and-apps-is-serverless/
[8] Van Eyk, E., Iosup, A., Seif, S., & Thömmes, M. (2017). The SPEC cloud group's research vision on FaaS and serverless architectures. In Proceedings of the 2nd International Workshop on Serverless Computing (pp. 1-4). ACM.
[9] Van Eyk, E., Toader, L., Talluri, S., Versluis, L., Uță, A., Iosup, A. (2018). Serverless is more: From paas to present cloud computing. IEEE Internet Computing, 22(5), 8-17.
[10] Conway, S. (2017) Cloud Native Technologies Are Scaling Production Applications. CNCF website: https://www. cncf. io/blog/2017/12/06/cloud-native-technologies-scaling-production-applications..
[11] Jonas, E., Schleier-Smith, J., Sreekanti, V., Tsai, C. C., Khandelwal, A., Pu, Q., ... & Gonzalez, J. E. (2019). Cloud programming simplified: a Berkeley view on serverless computing. arXiv preprint arXiv:1902.03383.
[12] Castro, P., Ishakian, V., Muthusamy, V., & Slominski, A. (2017). Serverless programming (function as a service). In 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS) (pp. 2658-2659). IEEE. DOI 10.1109/ICDCS.2017.305.

Software y aplicaciones en Computación de Altas Prestaciones para el contexto de la UNdeC

Fernando Emmanuel FRATI, Jose TEXIER, Paula Cecilia RIVERA, Jonathan ALVAREZ, Fernanda CARMONA, Patricia FIGUEROLA, Francisco FRATI, Sebastián GUIDET, Roberto MILLON, Raul MORALEJO, Matías PEREZ, Emmanuel PORTUGAL, Donna RATTALINO, Alberto RIBA, Daniel ROBINS, Mara ROVERO, Javier RUITTI, Jorge TEJADA, Jusmeidy ZAMBRANO, Carlos Esteban GRAFFIGNA

Universidad Nacional de Chilecito
9 de julio 22, Chilecito, La Rioja, Argentina
{fefrati, jtexier, privera, jalvarez, fbcarmona, pfiguerola, flfrati, sguidet, rmillon, rmoralejo, mperez, eportugal, drattalino, ariba, drobins, mrovero, jruitti, jtejada, jzambrano, cgraffigna}@undec.edu.ar

RESUMEN

En la actualidad, para abordar problemas de mayor tamaño y complejidad, los estudios de ciencia básica y aplicada utilizan Computación de Altas Prestaciones (HPC - High Performance Computing). El HPC permite mejorar la capacidad, velocidad y precisión en el procesamiento de datos. Con el proyecto que da origen a este trabajo se abordan seis estudios desde la perspectiva del HPC, para explorar los aspectos centrales del paralelismo aplicado desde las Ciencias de la Computación en otras disciplinas.

Algunas de estos estudios se realizan exclusivamente en la Universidad Nacional de Chilecito, mientras que otros son en cooperación con otras instituciones nacionales y extranjeras. Entre estos, tres formalizan trabajos finales de postgrado. En todos los casos, el HPC será abordado a través de un proceso metodológico organizado para:

- Consolidar una infraestructura de experimentación, desarrollo y producción de soluciones a problemas de HPC
- Desarrollar las capacidades científico-tecnológicas del equipo
- Fomentar la vinculación y transferencia con los sectores académico, social y productivo

Cada problema abordado reúne entre sus integrantes investigadores especialistas en la disciplina del estudio, investigadores de Ciencias de la Computación y estudiantes en sus últimos años de formación de grado. Con esto, se está consolidando un grupo de investigación, desarrollo y transferencia que generará oportunidades de formación de recursos humanos, proveerá de servicios a la comunidad en el área de estudio y potenciará los vínculos de cooperación con otras instituciones.

Palabras clave: HPC, cómputo paralelo, aplicaciones, interdisciplinariedad.

CONTEXTO

La línea de investigación presentada es parte del proyecto “Software y aplicaciones en Computación de Altas Prestaciones” fue aprobado en la convocatoria a proyectos de Investigación y Desarrollo 2018 de la Secretaría de Ciencia y Tecnología de la UNdeC, y se encuentra en ejecución desde junio de 2019 con una duración estimada de 18 meses. Además, en 2018 la UNdeC destinó fondos de PROMINF para la adquisición de 12 PC con procesadores i7 y 8GB RAM, 5 de las cuales están equipadas con placas de video NVIDIA GTX 1060 para el “laboratorio de sistemas paralelos”. También el “Plan de mejoramiento de la función de I+D+i” (Ministerio de Ciencia, Tecnología e Innovación Productiva de la Nación) con el objeto de desarrollar las capacidades en HPC de la UNdeC, a fines de 2019 financió la adquisición de un servidor Dell PowerEdge R740, equipado con 2 Xeon Platinum 8176 (56 núcleos físicos, 112 threads en total), 256 GB de RAM y 2 GPGPU NVIDIA Quadro P4000, actualmente en proceso de instalación y configuración. Estas iniciativas permitirán consolidar una infraestructura de experimentación, desarrollo y producción de soluciones a problemas de HPC.

1. INTRODUCCIÓN

La informática tiene su origen en la necesidad de los distintos sectores de la sociedad de conseguir mayor velocidad, confiabilidad y precisión para resolver sus problemas. Sin embargo, la capacidad de solución a un problema dado encuentra su límite en los tiempos requeridos por sus algoritmos. Superar ese límite requiere que el problema sea abordado mediante cómputo paralelo. Normalmente, esto implica estudiar tres aspectos clave: hardware, aplicaciones y software.

Durante décadas, la industria respondió a la creciente demanda de mayor poder computacional incrementando exponencialmente el rendimiento de los procesadores[1]. Sin embargo, esta forma de obtener mayor poder de cómputo encontró barreras físicas, limitando el rendimiento de los microprocesadores y los sistemas en general [2]. Desde el año 2005, el escalado tecnológico se ha venido aprovechando para aumentar el número de cores dentro del chip, dando lugar a una importante variedad de arquitecturas (multicores, commodity clusters, GPGPU y Cloud) [3], [4].

No obstante, reducir los tiempos de procesamiento y obtener la mayor eficiencia de ese hardware requiere el diseño y desarrollo de algoritmos paralelos [5]. Transformar un algoritmo secuencial en uno paralelo no es trivial. En general procesos concurrentes necesitan algún mecanismo para comunicar resultados parciales entre sí. Dependiendo de la arquitectura de cómputo, se consigue a través del uso de variables compartidas o del paso de mensajes entre procesos. Una transformación 'implícita' o transparente es deseable, pero el costo es una pérdida importante de rendimiento [6]. En su lugar, el programador recurre a librerías estándares para expresar explícitamente el paralelismo: OpenMP, Pthreads, CUDA, OpenCL, MPI [7].

La secuencia temporal en que se ejecutan las comunicaciones entre los procesos se conoce como historia del programa. Las suposiciones de orden de ejecución entre instrucciones heredadas del modelo de programación secuencial ya no son válidas, obligando al programador a utilizar algún mecanismo de sincronización para garantizar estados consistentes del programa. En este contexto, la correctitud de los algoritmos es más difícil de garantizar que en la computación serial. Frecuentemente el programador se equivoca al sincronizar los procesos, dando lugar a nuevos errores de programación: deadlocks, condiciones de carrera, violaciones de orden, violaciones de atomicidad simple y violaciones de atomicidad multivariable, requiriendo el uso de herramientas de depuración específicas [8]. Al diseñar una solución serial, existe un modelo teórico que permite estimar el desempeño de los programas antes de escribirlos. La evaluación del sistema paralelo (software y hardware) se realiza a través de distintas métricas: tiempo de ejecución, speedup, eficiencia y overhead. Las soluciones paralelas están tan estrechamente vinculadas con el hardware subyacente que dificultan enormemente conseguir portabilidad de rendimiento [9].

Existen muchos aspectos que requieren ser tomados en cuenta al diseñar la solución paralela: tamaño del problema, división de datos o tareas, balance de carga, requerimientos de memoria, precisión de los cálculos, comunicaciones y sincronización entre procesos, errores de concurrencia, detección y tolerancia a fallos entre los más relevantes. La complejidad de los problemas requiere habilidades especiales de los desarrolladores: dominio de múltiples paradigmas de programación y frecuentemente múltiples lenguajes, conocimientos de redes y comprensión de la concurrencia y sus consecuencias. Por todo esto, se considera de gran interés el estudio de estos temas para el desarrollo de capacidades científico-tecnológicas en la UNdeC que favorezcan el trabajo interdisciplinario en la institución.

2. LÍNEAS DE I+D

Se abordaron las siguientes líneas de I+D desde la perspectiva del HPC como eje central:

- Análisis de la diversidad molecular de microorganismos del suelo [10]–[13]. Estudio e implementación de algoritmos que contribuyan a reducir los tiempos de procesamiento y aumentar la capacidad de

análisis referidos a este campo de la bioinformática, a fin de profundizar en el estudio de la diversidad molecular de microorganismos del suelo asociados a cultivos regionales.

- Evaluación de enfoques de desarrollo HDL y HLL en FPGA para aplicaciones de procesamiento de imágenes [14]–[16]. Estudio de lenguajes de desarrollo HDL y HLL en FPGA para implementar soluciones de procesamiento de imágenes eficientes.
- Identificación biométrica masiva mediante venas del dedo usando redes de aprendizaje extremo (ELM) [17]–[19]. Estudio de técnicas de computación paralela para mejorar la eficiencia de identificación biométrica masiva basada en venas de dedo, para la aceleración del preprocesamiento y extracción de características biométricas, y el diseño de algoritmos de ELM mejorados que manejen eficientemente lotes de datos de gran tamaño.
- Servicios basados en lingüística computacional para análisis de texto [20]–[26]. Estudio sobre modelos computacionales que reproduzcan aspectos del lenguaje humano, con el fin de realizar análisis lingüísticos como servicios para el Centro de Escritura en la UNdeC.
- Documentos inteligentes a través del Blockchain [27]–[30]. Estudio de la tecnología blockchain para garantizar la integridad de documentos universitarios.
- Nodo de información meteorológica [31], [32]. Estudio, diseño e implementación de algoritmos para reducir los tiempos de procesamiento, aumentar la capacidad de análisis y favorecer la escalabilidad de aplicaciones de análisis y proyección de datos climáticos.

3. RESULTADOS

Durante los primeros meses de ejecución del proyecto, se alcanzaron los siguientes resultados:

- Una publicación en revista[33].
- Seis presentaciones en congresos y/o workshops [34]–[39].
- Seis charlas científicas con invitados externos expertos en los distintos temas del proyecto:
  - Computación Paralela en Biometría.
  - Nuevos retos del análisis inteligente de datos en entornos Big Data.
  - SWIFOLD: Aceleración del Alineamiento de Secuencias Largas de ADN combinando FPGA+OpenCL.
  - Tendencias en Investigación y Formación de Recursos Humanos en Informática.
  - Deep Learning. Estado actual en aplicaciones de Visión por Computadora.
- Cinco capacitaciones en otros centros de formación del país y del exterior en temas relacionados con el proyecto.
- Tres tesis de maestría en desarrollo, dos codirigidas por doctores especialistas en los temas de la UCM (Chile) y la tercera por un doctor de la UNLP (Argentina).
- Dos becarios CIN (en proceso de evaluación).
- Un becario interno (en proceso de selección)
- Se instaló y configuró el cluster del laboratorio de Sistemas Paralelos, con 12 equipos en red 5 de los cuales están equipados con Nvidia GTX 1060.

Actualmente se está configurando el servidor de altas prestaciones recientemente adquirido con fondos del Plan de Mejoramiento de la función de I+D+i del MINCyT (ver contexto).

4. FORMACIÓN DE RECURSOS HUMANOS

Seis miembros del equipo poseen formación de postgrado a nivel de doctorado, uno de ellos es especialista en Cómputo de Altas Prestaciones. Cinco miembros se encuentran en su etapa final para obtener el grado de maestría en Informática, tres de los cuales desarrollan como tesis temas abordados por esta propuesta. Dos de estas tesis de maestría están siendo codirigidas por docentes de la Universidad Católica de Maule (Chile) y una está siendo codirigida con un docente de la Universidad Nacional de La Plata. Cada línea I+D propuesta integra al menos un docente investigador

experto en el campo de cada estudio específico. Todos los temas propuestos se trabajan con estudiantes de grado de las carreras Ingeniería en Sistemas y Licenciatura en Sistemas de la UNdeC (ambas acreditadas por CONEAU). Los docentes forman parte de los equipos de diversas asignaturas de estas carreras, entre las que se encuentran programación, arquitecturas de computadoras y arquitecturas paralelas. Nueve docentes se encuentran categorizados en el programa de incentivos.

5. BIBLIOGRAFÍA

[1] S. Borkar y A. A. Chien, «The Future of Microprocessors», *Commun ACM*, vol. 54, n.º 5, pp. 67–77, may 2011.

[2] K. Ahmed y K. Schuegraf, «Transistor wars», *IEEE Spectr.*, vol. 48, n.º 11, pp. 50-66, nov. 2011.

[3] V. V. Kindratenko *et al.*, «GPU clusters for high-performance computing», en *2009 IEEE International Conference on Cluster Computing and Workshops*, 2009, pp. 1–8.

[4] J. Jeffers, J. Reinders, y A. Sodani, *Intel Xeon Phi Processor High Performance Programming: Knights Landing Edition*. Morgan Kaufmann, 2016.

[5] A. Grama, A. Gupta, G. Karypis, y V. Kumar, *Introduction to Parallel Computing - Second Edition*. Pearson Education and Addison Wesley, 2003.

[6] G. Hager y G. Wellein, *Introduction to high performance computing for scientists and engineers*. Boca Raton, FL: CRC Press, 2011.

[7] J. Dongarra *et al.*, *Sourcebook of parallel computing*, vol. 3003. Morgan Kaufmann Publishers San Francisco, 2003.

[8] F. E. Frati, «Software para arquitecturas basadas en procesadores de múltiples núcleos», Tesis, Facultad de Informática, 2015.

[9] S. Ghosh, *Distributed Systems: An Algorithmic Approach*, 1.ª ed. University of Iowa, Iowa City, USA: Chapman and Hall/ CRC, 2006.

[10] M. S. De, M. Prager, R. E. Naranjo, y O. E. Sanclemente, «El suelo, su metabolismo, ciclaje de nutrientes y prácticas agroecológicas», *Agroecología*, vol. 7, n.º 1, pp. 19–34, 2012.

[11] J. P. Hulsenbeck y F. Ronquist, «MrBayes: Bayesian inference of phylogeny», *Bioinformatics*, vol. 17, pp. 754–755, 2001.

[12] A. Stamatakis, «RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies», *Bioinformatics*, vol. 30, n.º 9, pp. 1312-1313, may 2014.

[13] M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, y A. Rambaut, «Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10», *Virus Evol.*, vol. 4, n.º 1, ene. 2018.

[14] R. Nane *et al.*, «A Survey and Evaluation of FPGA High-Level Synthesis Tools», *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 35, n.º 10, pp. 1591-1604, oct. 2016.

[15] S. Windh *et al.*, «High-Level Language Tools for Reconfigurable Computing», *Proc. IEEE*, vol. 103, n.º 3, pp. 390-408, mar. 2015.

[16] R. Tessier, K. Pocek, y A. DeHon, «Reconfigurable Computing Architectures», *Proc. IEEE*, vol. 103, n.º 3, pp. 332-354, mar. 2015.

[17] K. Wang, H. Ma, O. P. Popoola, y J. Liu, «Finger vein recognition», en *Biometrics*, InTech, 2011.

[18] D. Ezhilmaran y P. R. B. Joseph, «A study of feature extraction techniques and image enhancement algorithms for finger vein recognition», p. 8, 2015.

[19] A. Akusok, K. Björk, Y. Miche, y A. Lendasse, «High-Performance Extreme Learning Machines: A Complete Toolbox for Big Data Applications», *IEEE Access*, vol. 3, pp. 1011-1025, 2015.

[20] AMPLN, «Asociación Mexicana para el Procesamiento del Lenguaje Natural Main/ Home Page». https://www.ampln.org/ (accedido ago. 30, 2018).

[21] M. Vallez y R. Pedraza, «El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines», *Hipertext Net*, 2007.

[22] J. Texier, F. E. Frati, F. B. Carmona, A. E. Riba, M. Pérez, y J. Zambrano, «La gestión de la información en abierto, vehículo importante para maximizar la visibilidad web», presentado en XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina), may 2016.

[23] P. Gamallo Otero, J. C. Pichel Campos,

M. García González, J. M. Abuín Mosquera, y T. Fernández Pena, «Análisis morfosintáctico y clasificación de entidades nombradas en un entorno Big Data», 2014.
[24] S. W. D. Chien, C. P. Sishtla, S. Markidis, J. Zhang, I. B. Peng, y E. Laure, «An Evaluation of the TensorFlow Programming Model for Solving Traditional HPC Problems», en *International Conference on Exascale Applications and Software*, 2018, p. 34.
[25] H. Guan, X. Shen, y H. Krim, «Egeria: A Framework for Automatic Synthesis of HPC Advising Tools Through Multi-layered Natural Language Processing», en *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, New York, NY, USA, 2017, pp. 10:1–10:14.
[26] P. Suber, *Ensuring open access for publicly funded research*. British Medical Journal Publishing Group, 2012.
[27] M. U. Wasim, A. A. Ibrahim, P. Bouvry, y T. Limba, «Law as a service (LaaS): Enabling legal protection over a blockchain network», en *Smart Cities: Improving Quality of Life Using ICT & IoT (HONET-ICT), 2017 14th International Conference on*, 2017, pp. 110–114.
[28] A. Preukschat, *Blockchain: la revolución industrial de internet*. Gestión 2000, 2017.
[29] L.-Y. Yeh, P. J. Lu, y J.-W. Hu, «NCHC blockchain construction platform (NBCP): rapidly constructing blockchain nodes around Taiwan», en *Digital Libraries (JCDL), 2017 ACM/IEEE Joint Conference on*, 2017, pp. 1–2.
[30] H. Dai *et al.*, «TrialChain: A Blockchain-Based Platform to Validate Data Integrity in Large, Biomedical Research Studies», *ArXiv Prepr. ArXiv180703662*, 2018.
[31] A. Botta, W. De Donato, V. Persico, y A. Pescapé, «On the integration of cloud computing and internet of things», en *Future internet of things and cloud (FiCloud), 2014 international conference on*, 2014, pp. 23–30.
[32] P. Yue, H. Zhou, J. Gong, y L. Hu, «Geoprocessing in cloud computing platforms–a comparative analysis», *Int. J. Digit. Earth*, vol. 6, n.º 4, pp. 404–425, 2013.
[33] R. Hernández-García *et al.*, «Fast Finger Vein Recognition Based on Sparse Matching Algorithm under a Multicore Platform for Real-Time Individuals Identification», *Symmetry*, vol. 11, n.º 9, p. 1167, sep. 2019, doi: 10.3390/sym11091167.
[34] F. E. Frati *et al.*, «Software y aplicaciones en Computación de Altas Prestaciones para el contexto de la UNdeC», en *Actas del XXI Workshop de Investigadores en Ciencias de la Computación*, San Juan, Argentina, abr. 2019, p. 5.
[35] S. Guidet y E. Frati, «Desarrollo de un método para identificación de personas por venas de dedo en grandes bases de datos.», en *Actas de las VIII Jornadas Científicas de Estudiantes Investigadores*, Chilecito, La Rioja, Argentina, oct. 2019.
[36] A. Riperto y J. Texier, «Smart Contract en la generación de título de grado bajo infraestructura HPC», en *Actas de las VIII Jornadas Científicas de Estudiantes Investigadores*, Chilecito, La Rioja, Argentina, oct. 2019.
[37] A. Ortiz y I. Flores, «Sistema para el análisis de textos científicos a través de HPC», en *Actas de las VIII Jornadas Científicas de Estudiantes Investigadores*, Chilecito, La Rioja, Argentina, oct. 2019.
[38] J. Zambrano y J. Texier, «Laboratorio de escritura: dispositivos de acompañamiento en escritura para estudiantes de la UNdeC.», en *Actas del 1er Congreso Internacional de Ingeniería Aplicada de Ibero-American Sicence & Technology Education Consortium*, 2019.
[39] R. Hernández-García, S. Guidet, R. J. Barrientos, y F. E. Frati, «Massive Finger-vein Identification based on Local Line Binary Pattern under Parallel and Distributed Systems», en *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*, 2019, pp. 1–7.

FUNDAMENTOS, ALGORITMOS Y EVALUACION DE RENDIMIENTO EN DIFERENTES PLATAFORMAS DE HPC

Marcelo Naiouf[1], Armando De Giusti[1][2], Laura De Giusti[1][3], Franco Chichizola[1], Victoria Sanz[1][2][3], Adrián Pousa[1], Enzo Rucci[1][2], Silvana Gallo[1][2], Erica Montes de Oca[1], Emmanuel Frati[1], Mariano Sánchez[1], María José Basgall[1][2], Adriana Gaudiani[4]

[1]Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 115, La Plata, Buenos Aires
Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires (CIC)
526 e/ 10 y 11 La Plata Buenos Aires

[2] CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas
[3]CIC – Comisión de Investigación Científica de la Provincia de Buenos Aires
[4] Universidad Nacional de General Sarmiento
{mnaiouf, degiusti, ldgiusti, francoch, vsanz, apousa, erucci, sgallo, emontesdeoca,fefrati, msanchez, mjbasgall}@lidi.info.unlp.edu.ar, agaudi@ungs.edu.ar

RESUMEN

El eje central de la línea de I/D son los temas de procesamiento paralelo y distribuido para HPC (fundamentos y aplicaciones). Interesa la construcción, evaluación y optimización de soluciones sobre diferentes plataformas de software y arquitecturas con múltiples procesadores (multicore, clusters, cloud, aceleradores y placas de bajo costo), los lenguajes y paradigmas de programación paralela (puros e híbridos), los modelos de representación de aplicaciones paralelas, los algoritmos de mapping y scheduling, el balance de carga, las métricas de evaluación de complejidad y rendimiento computacional y energético, y la construcción de ambientes para la enseñanza de la programación concurrente y paralela.

Se propone aplicar los conceptos en problemas numéricos y no numéricos de cómputo intensivo y/o sobre grandes volúmenes de datos con el fin de obtener soluciones de alto rendimiento.

En la dirección de tesis de postgrado existe colaboración con el grupo HPC4EAS (High Performance Computing for Efficient Applications and Simulation) del Dpto. de Arquitectura de Computadores y Sistemas Operativos de la Universidad Autónoma de Barcelona; con el Departamento de Arquitectura de Computadores y Automática de la Universidad Complutense de Madrid; y con el grupo Soft Computing and Intelligent Information Systems (SCI2S) de la Universidad de Granada, entre otros.

Palabras clave: Cómputo paralelo y distribuido de altas prestaciones. Algoritmos paralelos y distribuidos. Clusters. Multicore. Aceleradores. Consumo energético. Balance de carga. Aplicaciones. Performance.

CONTEXTO

La línea de I/D que se presenta es parte del Proyecto "Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real." del III-LIDI acreditado por el Ministerio de Educación, y de proyectos acreditados y subsidiados por la Facultad de Informática de la UNLP. Además, existe cooperación con Universidades de Argentina, Latinoamérica y Europa a través de proyectos acreditados por AECID, CyTeD, OEI y CIC y becas de Telefónica de Argentina. Asimismo, el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD).

1. INTRODUCCIÓN

El área de cómputo de altas prestaciones (HPC, High-Performance Computing) es clave dentro de las Ciencias de la Computación, debido al creciente interés por el desarrollo de soluciones a problemas con alta demanda computacional y de almacenamiento, produciendo transformaciones profundas en las líneas de I/D [1].

El rendimiento en este caso está relacionado con dos aspectos: las arquitecturas de soporte y los algoritmos que hacen uso de las mismas, y el desafío se centra en cómo aprovechar las prestaciones obtenidas a partir de la evolución de las arquitecturas físicas. En esta línea la mayor importancia está en los algoritmos paralelos y en los métodos utilizados para su construcción y análisis a fin de optimizarlos.

Uno de los cambios de mayor impacto ha sido el uso de manera masiva de procesadores con más de un núcleo (*multicore*), produciendo plataformas distribuidas híbridas (memoria compartida y distribuida) y generando la necesidad de desarrollar sistemas operativos, lenguajes y algoritmos que las usen adecuadamente. También creció la incorporación de placas aceleradoras a los sistemas multicore constituyendo plataformas paralelas de memoria compartida con paradigma de programación propio asociado como pueden ser las unidades de procesamiento gráfico (GPU, Graphic Processing Unit) de NVIDIA y AMD, los coprocesadores Xeon Phi de Intel [2] o los aceleradores basados en circuitos integrados reconfigurables (FPGAs, Field Programmable Gate Array) [3]. En la actualidad se comercializan placas de bajo costo como Raspberry PI [4] u Odroid [5] que poseen múltiples núcleos de baja complejidad y en algunos casos son procesadores multicore asimétricos

(AMPs) con el mismo repertorio de instrucciones. Es de interés estudiar como explotar el paralelismo en estos dispositivos para mejorar el rendimiento y/o consumo energético de las aplicaciones [6], así como las características de scheduling en los mismos [7]. Asimismo, los entornos de computación cloud introducen un nuevo foco desde el punto de vista del HPC, brindando un soporte "a medida" sin la necesidad de adquirir el hardware.

La creación de algoritmos paralelos en arquitecturas multiprocesador no es un proceso directo [8]. El costo puede ser alto en términos del esfuerzo de programación y el manejo de la concurrencia adquiere un rol central en el desarrollo. Si bien en las primeras etapas el diseñador de una aplicación paralela puede abstraerse de la máquina sobre la que ejecutará el algoritmo, para obtener buen rendimiento debe tenerse en cuenta la plataforma de destino. En las máquinas multiprocesador, se deben identificar las capacidades de procesamiento, interconexión, sincronización y escalabilidad. La caracterización y estudio de rendimiento del sistema de comunicaciones es de interés para la predicción y optimización de performance, así como la homogeneidad o heterogeneidad de los procesadores [9].

Muchos problemas algorítmicos se vieron impactados por los multicore y clusters de multicore. A partir de incorporar varios chips multicore dentro de un nodo y conectar múltiples nodos vía red, se puede crear una arquitectura NUMA, de modo que los cores en un chip compartan memoria principal, y puedan acceder remotamente a la memoria dedicada de otro chip, aunque ese acceso sea más costoso, surgiendo así varios niveles de comunicación. Esto impacta sobre el desarrollo de algoritmos que aprovechen adecuadamente las arquitecturas, y motiva el estudio de performance en sistemas híbridos. Además, es necesario estudiar la utilización de lenguajes y bibliotecas ya que aún no se cuenta con un standard, aunque puede mencionarse el uso de los tradicionales MPI, OpenMP y Pthreads [10][11] o los más recientemente explorados UPC, Chapel y Titanium del modelo PGAS [12].

La combinación de arquitecturas con múltiples núcleos con aceleradores dio lugar a plataformas híbridas con diferentes características. Más allá del acelerador utilizado, la programación de estas plataformas representa un desafío. Para lograr aplicaciones de alto rendimiento, los programadores enfrentan dificultades como: estudiar características específicas de cada arquitectura y aplicar técnicas de programación y optimización particulares de cada una, lograr un balance de carga adecuado entre los dispositivos de procesamiento y afrontar la ausencia de estándares para este tipo de sistemas.

Por otra parte, los avances en las tecnologías de virtualización han llevado a que Cloud Computing sea una alternativa a los tradicionales sistemas de cluster [13]. El uso de cloud para HPC presenta desafíos atractivos, brindando un entorno reconfigurable dinámicamente sin la necesidad de adquirir hardware, y es una excelente plataforma para testear escalabilidad de algoritmos aunque queda mucho por hacer en cuanto al diseño, lenguajes y programación

Métricas de evaluación del rendimiento y balance de carga

La diversidad de opciones vuelve complejo el análisis de performance de los Sistemas Paralelos, ya que los ejes sobre los cuales pueden compararse dos sistemas son varios. Existe un gran número de métricas para evaluar el rendimiento, siendo las tradicionales: tiempo de ejecución, speedup, eficiencia. Por su parte, la *escalabilidad* permite capturar características de un algoritmo paralelo y la arquitectura en que se lo implementa. Posibilita testear la performance de un programa sobre pocos procesadores y predecirla en un número mayor, así como caracterizar la cantidad de paralelismo inherente en un algoritmo.

Un aspecto de interés que se ha sumado como métrica, a partir de las plataformas con gran cantidad de procesadores, es el del *consumo* y la *eficiencia energética* [14]. Muchos esfuerzos están orientados a tratar el consumo como eje de I/D, como métrica de evaluación, y también a la necesidad de metodologías para medirlo.

El objetivo principal del cómputo paralelo es reducir el tiempo de ejecución haciendo uso eficiente de los recursos. El *balance de carga* es un aspecto central y consiste en, dado un conjunto de tareas que comprenden un algoritmo y un conjunto de procesadores, encontrar el mapeo (asignación) de tareas a procesadores tal que cada una tenga una cantidad de trabajo que demande aproximadamente el mismo tiempo, y esto es más complejo si hay heterogeneidad. Dado que el problema general de mapping es *NP*-completo, pueden usarse enfoques que dan soluciones subóptimas aceptables. Las técnicas de planificación a nivel micro (dentro de cada procesador) y macro (en un cluster) deben ser capaces de obtener buen balance de carga. Existen técnicas estáticas y dinámicas cuyo uso depende del conocimiento que se tenga sobre las tareas de la aplicación.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

- Investigar en temas de cómputo paralelo y distribuido de alto desempeño, en lo referido a los fundamentos y a la construcción y evaluación de las aplicaciones. Esto incluye los problemas de software asociados con el uso de arquitecturas multiprocesador:
  - Lenguajes, modelos y paradigmas de programación paralela (puros e híbridos a distintos niveles).
  - Asignación de procesos a procesadores optimizando el balance de la carga de procesamiento.
  - Métricas de evaluación de complejidad y rendimiento: speedup, eficiencia, escalabilidad, consumo energético, costo de programación.

- Construir, evaluar y optimizar soluciones utilizando algoritmos concurrentes, paralelos y distribuidos sobre diferentes plataformas de software y arquitecturas con múltiples procesadores:
  - Arquitecturas de trabajo homogéneas, heterogéneas e híbridas: multicores, clusters, GPU, Xeon Phi, FPGA, placas de bajo costo y cloud.
  - Se propone aplicar los conceptos en problemas numéricos y no numéricos de cómputo intensivo y/o sobre grandes volúmenes de datos (aplicaciones científicas, búsquedas, simulaciones, imágenes, realidad virtual y aumentada, bioinformática, big data, n-body).
- Analizar y desarrollar ambientes para la enseñanza de programación concurrente y paralela.
  - Caracterizar diferentes modelos de arquitecturas paralelas.
  - Representar distintos modelos de comunicación/sincronización.
  - Definir métricas de evaluación de rendimiento y eficiencia energética.

3. RESULTADOS OBTENIDOS/ESPERADOS

▪ Desarrollar y optimizar algoritmos paralelos sobre diferentes modelos de arquitectura. En particular, en aplicaciones numéricas y no numéricas de cómputo intensivo y tratamiento de grandes volúmenes de datos (big data).
▪ Utilizar arquitecturas híbridas que combinan memoria compartida y pasaje de mensajes, evaluando performance para distintos modelos de comunicación.
▪ Estudiar y comparar los lenguajes sobre las plataformas multiprocesador para diferentes modelos de interacción entre procesos.
▪ Investigar la paralelización en plataformas que combinan clusters, multicore y aceleradores. Comparar estrategias de distribución de trabajo teniendo en cuenta las diferencias en potencias de cómputo y comunicación, dependencia de datos y memoria requerida.
▪ Evaluar la performance (speedup, eficiencia, escalabilidad, consumo energético) de las soluciones propuestas. Analizar el rendimiento de soluciones paralelas a problemas con diferentes características (dependencia de datos, relación cómputo / comunicación, memoria requerida).
▪ Mejorar y adecuar las técnicas disponibles para el balance de carga (estático y dinámico) entre procesos a las arquitecturas consideradas.

En este marco, pueden mencionarse los siguientes resultados:
▪ Para la experimentación se han utilizado y analizado diferentes arquitecturas homogéneas o heterogéneas, incluyendo multicores, cluster de multicores (con 128 núcleos), GPU y cluster de GPU, Xeon Phi y FPGA.
▪ Se experimentó la paralelización en arquitecturas híbridas, con el objetivo de estudiar el impacto del mapeo de datos y procesos, así como de los lenguajes y librerías.
▪ Respecto de las aplicaciones y temas estudiados, se trabajó fundamentalmente con los siguientes problemas:

➢ **Aceleración de aplicaciones con cómputo colaborativo CPU-GPU.** Las computadoras comerciales actuales incluyen decenas de cores y al menos una GPU. El uso de ambas unidades de procesamiento de forma colaborativa puede mejorar significativamente el rendimiento de una aplicación. Sin embargo, esto supone un desafío para los programadores ya que dichas unidades difieren en arquitectura, modelo de programación y rendimiento. En [15] se propuso un modelo híbrido para estructurar código a ser ejecutado sobre un sistema heterogéneo con múltiples cores y 1 GPU (utilizando todos los recursos disponibles). Utilizando este modelo se desarrolló un algoritmo paralelo de pattern matching para sistemas heterogéneos CPU-GPU [16]. Los resultados revelaron que este algoritmo supera en rendimiento a trabajos previos, desarrollados para sistemas multicore y GPUs, para datos de tamaño considerable.

➢ **Alineamiento de secuencias biológicas.** Esta operación consiste en comparar dos o más secuencias biológicas, como pueden ser las de ADN o las de proteínas, y resulta fundamental en investigaciones de la bioinformática y la biología molecular. El algoritmo de Smith-Waterman es considerado el método de alineamiento más preciso. Desafortunadamente, este algoritmo resulta costoso debido a su complejidad computacional cuadrática mientras que la situación se agrava aún más a causa del crecimiento exponencial de datos biológicos en los últimos años [17]. El reciente surgimiento de aceleradores en HPC (GPU, Xeon Phi, FPGA, entre otros) da la oportunidad de acelerar los alineamientos sobre hardware comúnmente disponible a un costo accesible. En primer lugar, se exploró el empleo del modelo de programación OpenCL sobre FPGAs para acelerar el alineamiento de secuencias largas de ADN [18]. Luego, se desarrolló una herramienta capaz de procesar alineamientos de secuencias de ADN sin restricciones de tamaño y se analizó su rendimiento comparándolo con el de otras implementaciones basadas en multicores, Xeon Phi y GPUs [19]. Por otra parte, la salida al mercado de las nuevas generaciones de procesadores Xeon (Skylake) y Xeon Phi (Knights Landing) brindó la oportunidad de evaluar su uso para búsquedas de similitud en bases de datos de proteínas [20][21]. A futuro, interesa explorar el uso de próximas generaciones de procesadores y aceleradores para esta aplicación.

➢ **Cálculo de los caminos mínimos**. es uno de los problemas básicos y de mayor antigüedad de la teoría de grafos teniendo aplicación en el dominio de las comunicaciones, del ruteo de tráfico, de la

bioinformática, entre otros. El algoritmo de Floyd-Warshall (FW) permite computar la distancia mínima entre todos los pares de un grafo. Además de poseer una alta demanda de ancho de banda, FW resulta costoso computacionalmente al ser $O(n^3)$. Empezando por una implementación secuencial de base, se estudió y cuantificó cómo diferentes optimizaciones a nivel de datos, hilos y compilador permiten mejorar su rendimiento sobre los nuevos procesadores Xeon Phi Knights Landing [22]. Como trabajo futuro, interesa realizar un estudio comparativo con otros aceleradores (como GPUs), no sólo desde el punto de vista del rendimiento sino también de la eficiencia energética.

- **Simulaciones utilizando modelos basados en agentes.** El objetivo de esta línea son las simulaciones basadas en agentes sobre infraestructuras de altas prestaciones con modelos más cercanos a la realidad que ayuden a la toma de decisiones a científicos de otras disciplinas, y no expertos en el área de la la informática, (biología, ecología, física, etc). Para ello, y por ser modelos complejos que necesitan mucha potencia de cómputo, resulta imprescindible el uso de soluciones de HPC con el fin de lograr tiempos de respuesta reducidos para entornos complejos. Se busca desarrollar algoritmos y simuladores como soluciones HPC que sean eficientes y escalables y por ello una cuestión importante, tratada en esta línea de trabajo, es lograr que además de estas premisas, la simulación sea realizada con el menor consumo possible de energía. Dado que son simulaciones con un gran número de ejecuciones por escenario, no solo es necesario que la solución tenga buenas prestaciones sino que sea posible predecir, mediante un modelo energético, la energía necesaria para llevar a cabo diferentes escenarios de simulación.

- **Simulación de enfermedades infecciosas transmitida por vectores utilizando modelos basados en agentes.** El Dengue, Zika y Chikungunya, son las enfermedades reemergentes de mayor preocupación a nivel mundial. La carencia de tratamientos médicos obliga a los agentes de salud a abordar la contención de los focos infecciosos desde la identificación y eliminación de los criaderos del vector transmisor. Una herramienta tecnológica que ayude en la toma de decisiones representa la solución a los costos asociados al tiempo de recolección de muestras y el análisis de datos, además de la consiguiente reducción de los gastos económicos. Se implementó un modelo basado en agentes en GPU para la evaluación de la reproducción del vector Aedes aegypti, orientado a la toma de decisiones. El modelo ha sido validado con datos de Santo Tomé, Corrientes, dando excelentes resultados [23]. Se realizará un estudio y análisis de performance en otras arquitecturas paralelas basadas en GPU. Además, se evaluará el consumo energético para determinar las relaciones entre escalabilidad, consumo y eficiencia.

- **Simulación de N cuerpos computacionales con atracción gravitacional.** Su propósito es aproximar en forma numérica la evolución de un sistema de cuerpos en el que cada uno interactúa con todos los restantes. El uso más conocido de esta simulación quizás sea en la astrofísica, donde cada cuerpo representa una galaxia o una estrella particular que se atraen entre sí debido a la fuerza gravitacional. Si bien existen diferentes métodos para procesar la simulación de los N cuerpos, en todos los casos se requiere alta demanda computacional. Se ha utilizado este problema para analizar diferentes aspectos y/o arquitecturas como: cluster de GPUs homogéneas [24]; análisis de diferentes distribuciones de trabajo en cluster de GPUs heterogéneas [25]; desarrollo de modelos de estimación de tiempo de ejecución y de energía para GPU, cluster de GPU y multiGPU [26].
Además, se estudió la paralelización de la versión directa de esta simulación sobre un acelerador Xeon Phi KNL [27], focalizando en cómo diferentes técnicas de optimización logran mejorar el rendimiento final. A futuro, interesa continuar el desarrollo de soluciones siguiendo métodos avanzados, además de extender el estudio a otras arquitecturas.

- **Problemas de optimización de simulación de sistemas dinámicos complejos mediante heurísticas**. La búsqueda de un conjunto de parámetros de entrada que optimicen el funcionamiento de un simulador de un sistema físico es un proceso de alto costo computacional que puede considerarse intratable y requiere de heurísticas que permitan disminuir el tiempo de ejecución. En los nuevos avances realizados en esta línea se aprovechó la continuidad en los valores de los parámetros físicos distribuidos sobre el dominio del sistema, de manera de realizar búsquedas de parámetros ajustados sobre espacios de búsquedas de tamaño mucho más reducidos que los utilizados en una primera etapa del trabajo. Esta metodología mucho más eficiente se puede extender a otros simuladores de fenómenos físicos y en particular con simuladores de inundaciones de ríos con las que se llevaron adelante las experiencias. Las experiencias se pueden correr de manera colaborativa beneficiándose ampliamente del uso de plataformas de clusters de procesadores [28] [29].

- **Ambientes para la enseñanza de concurrencia.** Se desarrolló el entorno CMRE para la enseñanza de programación concurrente y paralela a partir de cursos iniciales en carreras de Informática. Incluye un entorno visual que representa una ciudad en la que pueden definirse varios robots que interactúan. Combina aspectos de memoria compartida y distribuida mediante instrucciones para bloquear y liberar esquinas de la ciudad y el concepto de pasaje de mensajes a través de primitivas de envío y recepción. Además, se incluyen los conceptos de heterogeneidad (diferentes velocidades de los robots) y consumo energético [30]. Se ha

integrado con el uso de robots físicos (Lego Mindstorm 3.0) que ejecutan en tiempo real las mismas instrucciones que los robots virtuales y se comunican con el entorno mediante bluetooth [31]. Se ha ampliado para incorporar conceptos básicos de computación en la nube (Cloud Computing) [31]. Actualmente, se está desarrollando una nueva herramienta para la enseñanza de programación concurrente en cursos avanzados. Su objetivo principal es visualizar los conceptos de sincronización y comunicación entre procesos.

- **Aplicaciones en Big Data.** Durante los últimos años, los escenarios de Big Data son cada vez más comunes y esto se debe a la enorme cantidad de datos que continuamente se está generando a partir de distintas fuentes tales como sensores, redes sociales, dispositivos móviles, entre otras. Para tratar con problemas de esta magnitud, se requiere de cómputos de altas prestaciones y de herramientas específicas que colaboren con el proceso de adaptación de los algoritmos tradicionales o permitan desarrollar nuevos para este fin específico, y que puedan ser ejecutados de manera paralela y distribuida. Existen distintos tipos de herramientas para trabajar con Big Data utilizando HPC, y la mayoría de ellas se basan en el paradigma de programación llamado MapReduce, propuesto por Google. Actualmente, Apache Spark es una de las más utilizadas y entre sus principales características se destaca utiliza intensivamente la memoria RAM evitando así escritura en disco de grandes volúmenes de información, presenta mecanismos eficientes de tolerancia a fallos, y se dispone una amplia variedad de librerías que permiten orientar su uso a Machine Learning, a Streaming Processing, etc. En esta línea de trabajo, se están desarrollando nuevas técnicas para el preprocesamiento de datos para problemas de clasificación no balanceada en Big Data [32], utilizando el framework Apache Spark sobre Cómputo de Altas Prestaciones. Este trabajo se está llevando a cabo en colaboración con los Drs. Alberto Fernández Hilario y Francisco Herrera, del grupo de investigación SCI2S.

4. FORMACIÓN DE RECURSOS HUMANOS

Dentro de la temática de la línea de I/D se concluyó 1 tesis doctoral, 1 Trabajo Final de Especialización y 1 Tesina de Grado de Licenciatura. Se encuentran en curso en el marco del proyecto 3 tesis doctorales, 2 de maestría, 3 trabajos de Especialización y 3 Tesinas de grado.

Se participa en el dictado de las carreras de Doctorado en Cs. Informáticas y Magíster y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática UNLP, por lo que potencialmente pueden generarse más Tesis y Trabajos Finales.

Hay cooperación con grupos de otras Universidades del país y del exterior, y tesistas de diferentes Universidades realizan su trabajo con el equipo del proyecto.

5. BIBLIOGRAFÍA

[1]. Giles MB, Reguly I. “Trends in high-performance computing for engineering calculations”. Phil.Trans.R.Soc.A 372: 20130319. 2014. http://dx.doi.org/10.1098/rsta.2013.0319

[2]. Jeffers, James; Reinders, James. “Intel Xeon Phi Coprocessor High Performance Programming”. Morgan Kaufmann. 2013.

[3]. Sean Settle. “High-performance Dynamic Programming on FPGAs with OpenCL”. IEEE High Performance Extreme Computing Conf (HPEC ’13). 2013. https://doi.org/10.1016/j.physa.2017.04.159.

[4]. Raspberry PI. https://www.raspberrypi.org/ Accedido 21 de Marzo de 2016.

[5]. Odroid http://www.hardkernel.com Accedido 21 de Marzo de 2016.

[6]. Annamalai A., Rodrigues R., Koren I., Kundu S., "Dynamic Thread Scheduling in Asymmetric Multicores to Maximize Performance-per-Watt," 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, pp. 964-971, 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, 2012.

[7]. Juan Carlos Saez, Adrian Pousa, Daniel Chaver, Fernando Castro, Manuel Prieto Matias. “ACFS: A Completely Fair Scheduler for Asymmetric Single-ISA Multicore Systems”. ACM SAC 2015 (The 30TH ACM/SIGAPP Symposium on applied computing). 2015.

[8]. McCool, Michael. “Structured Parallel Programming: Patterns for Efficient Computation”, Morgan Kaufmann, 2012

[9]. De Giusti L, Naiouf M., Chichizola F., Luque E., De Giusti A. “Dynamic Scheduling in Heterogeneous Multiprocessor Architectures. Efficiency Analysis”. Computer Science and Technology Series – XV Argentine Congress of Computer Science Selected Papers. La Plata (Buenos Aires): Editorial de la Universidad de La Plata (edulp). 2010. p85 - 95. isbn 978-950-34-0684-7.

[10]. Chapman, B., Jost, G. & Van der Pas. “Using OpenMP – Portable Shared Memory Parallel Programming”. (2008). UK: MIT Press.

[11]. Hager, G. & Wellein, G. “Introduction to HPC for Scientists and Engineers”. (2011) EEUU: CRC Press.

[12]. De Wael, M; Marr, S; De Fraine, B; Van Cutsem, T; De Meuter, W. “Partitioned Global Addess Space Languages”. ACM Computing Surveys (CSUR) 47 (4), 2015.

[13]. Amazon Elastic Compute Cloud (Amazon EC2). http://aws.amazon.com/es/ec2/. Febrero 2013.

[14]. Balladini J., Rucci E., De Giusti A., Naiouf M., Suppi R., Rexachs D., Luque E. “Power Characterisation of Shared-Memory HPC Systems”. Computer Science & Technology Series – XVIII Argentine Congress of Computer Science Selected Papers. ISBN 978-987-1985-20-3. Pp. 53-65. EDULP, La Plata (Argentina), 2013

[15]. V. Sanz, A. Pousa, M. Naiouf, and A. De Giusti, "Accelerating Pattern Matching with CPU-GPU Collaborative Computing", Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2018), ISBN: 978-3-030-05051-1, págs. 310-322, doi. https://doi.org/10.1007/978-3-030-05051-1_22, 2018.
[16]. Sanz V., Pousa A., Naiouf M., De Giusti A. (2020) Efficient Pattern Matching on CPU-GPU Heterogeneous Systems. In: Wen S., Zomaya A., Yang L. (eds) Algorithms and Architectures for Parallel Processing. ICA3PP 2019. Lecture Notes in Computer Science, vol 11944. Springer, Cham.
[17]. Enzo Rucci, Armando De Giusti, Marcelo Naiouf, Carlos García Sanchez, Guillermo Botella Juan, Manuel Prieto-Matías. "State-of-the-art in Smith-Waterman Protein Database Search". Big Data Analytics in Genomics. Ka-Chun Wong (Editor). ISBN: 978-3-319-41278-8 (print) 978-3-319-41279-5 (online), Springer, págs. 197-223, 2016.
[18]. E. Rucci, C. Garcia, G. Botella, A. De Giusti, M. Naiouf, and M. Prieto-Matias. "Accelerating Smith-Waterman Alignment of Long DNA Sequences with OpenCL on FPGA". En: Bioinformatics and Biomedical Engineering. IWBBIO 2017. Lecture Notes in Computer Science, vol 10209., ISBN: 978-3-319-56154-7, Springer, Cham, págs. 500-511, 2017.
[19]. E. Rucci, C. Garcia, G. Botella, A. De Giusti, M. Naiouf, and M. Prieto-Matias. "SWIFOLD:Smith-Waterman Implementation on FPGA with OpenCL for Long DNA Sequences". BMC Systems Biology. ISSN: 1752-0509. In press. 2018.
[20]. E. Rucci, C. Garcia, G. Botella, A. De Giusti, M. Naiouf, and M. Prieto-Matias. "First Experiences Accelerating Smith-Waterman on Intel's Knights Landing Processor". Algorithms and Architectures for Parallel Processing. ICA3pp 2017. Lecture Notes in Computer Science, vol 10393, ISBN: 978-3-319-65482-9, Springer International Publishing, págs. 569-579, 2017.
[21]. Enzo Rucci, Carlos García, Guillermo Botella, Armando De Giusti, Marcelo Naiouf and Manuel Prieto-Matías. "SWIMM 2.0: enhanced Smith-Waterman on Intel's Multicore and Manycore architectures based on AVX-512 vector extensions". International Journal of Parallel Programming, Springer US, 2018. DOI: https://doi.org/10.1007/s10766-018-0585-7
[22]. E. Rucci, A. De Giusti, and M. Naiouf, "Blocked All-Pairs Shortest Paths Algorithm on Intel Xeon Phi KNL Processor: A Case Study". En: Computer Science – CACIC 2017. Communications in Computer and Information Science, vol 790, ISBN: 978-3-319-75213-6 978-3-319-75214-3, Springer, Cham, págs. 47-57, 2018.
[23]. Erica Montes de Oca, Remo Suppi, Laura De Giusti, Marina Leporace, María Victoria Micieli, María Soledad Santini, Marcelo Naiouf. "Simulación de altas prestaciones (GPU) para la reproducción del mosquito Aedes aegypti en el cementerio de Santo Tomé, Corrientes". XXV Congreso Argentino de Ciencias de la Computación 2019 (CACIC 2019). Universidad Nacional de Río Cuarto, Argentina. pp. 183-193. ISBN 978-987-688-377-1.
[24]. E. Montes de Oca, L. De Giusti, F. Chichizola, A. De Giusti, M. Naiouf. "Utilización de Cluster de GPU en HPC. Un caso de estudio". Proceedings del XX Congreso Argentino de Ciencias de la Computación (CACIC 2014), pp. 1220-1227, 2014.
[25]. E. Montes de Oca, L. C. De Giusti, F. Chichizola, A. E. De Giusti, M. Naiouf. "Análisis de uso de un algoritmo de balanceo de carga estático en un Cluster Multi-GPU Heterogéneo". Proceedings del XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016), pp. 159-168, 2016.
[26]. Montes de Oca E. S. "Análisis de consumo energético en Cluster de GPU y MultiGPU en un problema de alta demanda computacional". Trabajo Final Carrera de Especialista en Cómputo de Altas Prestaciones y Tecnología GRID. Facultad de Informática, Universidad Nacional de La Plata. (2018).
[27]. Enzo Rucci, Ezequiel Moreno, Malek Camilo, Adrián Pousa, Franco Chichizola. "Simulación de N Cuerpos Computacionales sobre Intel Xeon Phi KNL". XXV Congreso Argentino de Ciencias de la Computación 2019 (CACIC2019). Universidad Nacional de Río Cuarto, Argentina. pp. 194-204. ISBN 978-987-688-377-1.
[28]. Trigila M., Gaudiani A., Luque E. (2018) "Agile Tuning Method in Successive Steps for a River Flow Simulator". In: Shi Y. et al. (eds) Computational Science – ICCS 2018. ICCS 2018. Lecture Notes in Computer Science, vol 10862. Springer, Cham
[29]. Trigila, Mariano, Gaudiani, Adriana, Luque, Emilio. "Adjustment of a simulator of a complex dynamic system with emphasis on the reduction of computational resources". Actas del Workshop. p. 142-147. 2018.
[30]. J. Castro, L. D. Giusti, G. Gorga, M. Sánchez, and M. Naiouf, "ECMRE: Extended Concurrent Multi Robot Environment". Computer Science – CACIC 2017. Communications in Computer and Information Science, vol 790, ISBN: 978-3-319-75213-6 978-3-319-75214-3, Springer, Cham, págs. 285-294. 2018.
[31]. L. C. De Giusti, F. Chichizola, S. Rodriguez Eguren, M. Sanchez, J. M. Paniego, A. E. De Giusti. "Introduciendo conceptos de Cloud Computing utilizando el entorno CMRE". Proceedings del XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016) – Workshop de Innovación en Educación en Informática. Octubre 2016. Pp 1357-1365.
[32]. Basgall, M. J., Hasperué, W., Naiouf, M., Fernández, A., & Herrera, F. (2019). An Analysis of Local and Global Solutions to Address Big Data Imbalanced Classification: A Case Study with SMOTE Preprocessing. En M. Naiouf, F. Chichizola, & E. Rucci (Eds.), Cloud Computing and Big Data (pp. 75–85). Cham: Springer International Publishing.

Hardware, Software, Modelos, Métricas y Tendencias en Arquitecturas Multiprocesador

De Giusti Armando[(1)(2)], Marcelo Naiouf[(1)], Tinetti Fernando[(1)(3)], Villagarcía Horacio[(1)(3)], Franco Chichizola[(1)], Laura De Giusti[(1)(3)], Enzo Rucci[(1)], Adrián Pousa[(1)], Victoria Sanz [(1)(3)], Montezanti Diego[(1)], Encinas Diego[(1)], Ismael Rodríguez[(1)], Sebastián Rodríguez Eguren[(1)], Erica Montes de Oca[(1)], Juan Manuel Paniego[(1)], Martín Pi Puig[(1)], César Estrebou[(1)], Leandro Libutti[(1)], Costanzo Manuel[(1)], Boggia Marcos[(1)], Joaquín De Antueno[(1)], Julieta Lanciotti[(1)], Javier Balladini[(4)]

[1]Instituto de Investigación en Informática LIDI (III-LIDI),
Facultad de Informática, Universidad Nacional de La Plata – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
[2]CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas
[3]CICPBA – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
[4]Universidad Nacional del Comahue

{adegiusti,mnaiouf,fernando,hvw,francoch,ldgiusti,erucci,apousa,vsanz,dmontezanti,dencinas,ismael,seguren,emontesdeoca,jmpaniego,mpipuig,cesarest,llibutti, mcostanzo,mboggia,jdeantueno,jlanciotti}@lidi.info.unlp.edu.ar;
javier.balladini@gmail.com

Resumen

El eje de esta línea de I/D lo constituye el estudio de las arquitecturas multiprocesador que integran sistemas distribuidos y paralelos. Incluye como temas centrales:

- Arquitecturas many-core (GPU, procesadores MIC, TPUs), FPGAs, híbridas (diferentes combinaciones de multicores y aceleradores), y asimétricas.
- Cloud Computing para HPC (especialmente para aplicaciones de Big Data) y sistemas distribuidos de tiempo real (Cloud Robotics).
- Desarrollo y evaluación de algoritmos paralelos sobre nuevas arquitecturas y su evaluación de rendimiento computacional y energético.

Palabras clave: *Sistemas Paralelos. Clusters. Arquitecturas asimétricas. GPU, MIC, FPGA, TPU. Cloud Computing. Cloud robotics. Eficiencia energética. Resiliencia. E/S paralela.*

Contexto

Se presenta una línea de Investigación que es parte del proyecto "Computación de Alto Desempeño: Arquitecturas, Algoritmos, Métricas de rendimiento y Aplicaciones en HPC, Big Data, Robótica, Señales y Tiempo Real." del III-LIDI y de proyectos específicos apoyados por organismos nacionales e internacionales. También del proyecto "Transformación de algoritmos para nuevas arquitecturas multiprocesador" financiado por la Facultad de Informática de la UNLP.

En el tema hay cooperación con varias Universidades de Argentina y se está trabajando con Universidades de América Latina y Europa en proyectos financiados por CyTED, AECID y la OEI (Organización de Estados Iberoamericanos).

Por otra parte, se tiene financiamiento de Telefónica de Argentina en Becas de grado y posgrado y se ha tenido el apoyo de diferentes empresas (IBM, Microsoft, Telecom, Intel) en la temática de Cloud Computing.

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Asimismo, el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD) del Ministerio de Educación, Cultura, Ciencia y Tecnología de la Nación.

Introducción

Una de las áreas de creciente interés lo constituye el cómputo de altas prestaciones, en

el cual el rendimiento está relacionado con dos aspectos: por un lado, las arquitecturas de soporte, y por otro, los algoritmos que hacen uso de estas.

A la aparición de arquitecturas *many-core* (como las GPU o los procesadores MIC), se ha sumado el uso de FPGAs debido a su potencia de cómputo y rendimiento energético. Su combinación en sistemas HPC da lugar a plataformas híbridas con diferentes características [1].

Lógicamente, esto trae aparejado una revisión de los conceptos del diseño de algoritmos paralelos (incluyendo los lenguajes mismos de programación y el software de base), así como la evaluación de las soluciones que éstos implementan. También resulta necesario investigar las estrategias de distribución de datos y procesos a fin de optimizar la performance.

Además, el estudio del consumo y la eficiencia energética de los nuevos sistemas paralelos se vuelve tan importante como el de métricas clásicas (speedup, eficiencia, escalabilidad) debido a los costos económicos y los problemas operativos asociados [2].

Por otra parte, los avances en las tecnologías de virtualización y cómputo distribuido han dado origen al paradigma de Cloud Computing, que se presenta como una alternativa a los tradicionales sistemas de Clusters y Multicluster para ambientes de HPC [3]. A su vez, este concepto se puede ampliar a sistemas distribuidos de tiempo real, en particular sistemas inteligentes como son los robots que pueden trabajar en paralelo utilizando su propia capacidad de procesamiento y al mismo tiempo conectándose con la potencia de un servidor en la nube (*Cloud Robotics*) [4-6].

En esta línea de I/D se trabaja sobre aspectos que marcan tendencias en el área.

GPUs y Cluster de GPUs

Las GPUs son el acelerador dominante en la comunidad de HPC hoy en día por su alto rendimiento y bajo costo de adquisición. En la actualidad, tanto NVIDIA como AMD trabajan especialmente en mejorar la eficiencia energética de sus placas y disminuir el alto costo de programación.

La combinación de GPUs con otras plataformas paralelas como clusters y multicores, brindan un vasto conjunto de posibilidades de investigación en arquitecturas híbridas, a partir de diferentes combinaciones a saber:

- Máquinas multicore con más de una GPU, que combinan herramientas de programación paralela como OpenMP/CUDA o Pthread/CUDA.
- Cluster de máquinas multicore cada una con una o más placas de GPU, lo que permite combinar OpenMP/MPI/CUDA o Pthread/MPI/CUDA.

Los desafíos que se plantean son múltiples, sobre todo en lo referido a distribución de datos y procesos en tales arquitecturas híbridas a fin de optimizar el rendimiento de las soluciones [7,8].

MIC

En forma reciente Intel brinda una alternativa a partir de la arquitectura MIC (*Many Integrated Core Architecture*). Esta arquitectura permite utilizar métodos y herramientas estándar de programación con altas prestaciones (lo que los distingue especialmente de las GPUs). De esta forma, se remueven barreras de entrenamiento y se permite focalizar en el problema más que en la ingeniería del software. Xeon Phi es el nombre elegido por Intel para su serie de procesadores many-core. Recientemente, Intel ha lanzado Knights Landing (KNL), la segunda generación de Xeon Phi. A diferencia de sus predecesores que operaban como co-procesador a través del puerto PCI, los procesadores KNL pueden operar en forma autónoma. Además, integran las nuevas extensiones vectoriales AVX-512 y tecnología de memoria 3D, entre otras características avanzadas [9].

FPGAs

Una FPGA (*Field Programmable Gate Array*) es una clase de acelerador basado en circuitos integrados reconfigurables. La capacidad de adaptar sus instrucciones de

acuerdo a la aplicación objetivo le permite incrementar la productividad de un sistema y mejorar el rendimiento energético para ciertos tipos de aplicaciones. Tradicionalmente fueron utilizadas para el procesamiento digital de señales. Sin embargo, en los últimos años, existen dos tendencias claras para extender su uso a otros dominios. En primer lugar, el establecimiento de alianzas estratégicas entre fabricantes de procesadores y de FPGAs para integrar estos dispositivos en arquitecturas híbridas (Intel con Altera; IBM con Xilinx) [10,11]. En segundo lugar, el desarrollo de nuevas herramientas de programación para FPGAs empleando estándares familiares para HPC, con las cuales se espera reducir los tradicionales tiempos y costos de programación [12,13]. Por último, la incorporación de FPGAs a los servicios de Cloud abre nuevas oportunidades para la explotación de esta clase de aceleradores.

TPUs

Las unidades de procesamiento tensorial (TPU) son una clase de circuitos integrados personalizados específicos de aplicaciones (ASIC) desarrolladas por Google con el propósito de acelerar las cargas de trabajo de aprendizaje automático que requieren las aplicaciones desarrolladas en su framework TensorFlow [14]. Su uso provee una alternativa a otras arquitecturas ya conocidas como CPUs, GPUs y MICs. En ese sentido, interesa analizar las tasas de aceleración y eficiencia energética provistas por esta nueva arquitectura, en comparación con el resto.

Eficiencia energética

La mejora de la eficiencia energética es una de las principales preocupaciones en la informática actual, principalmente a partir de las plataformas con gran cantidad de procesadores. Muchos esfuerzos están orientados a tratar la eficiencia energética como eje de I/D, como métrica de evaluación, y también a la necesidad de metodologías para medirla.

Entre los puntos de interés pueden mencionarse:

- Análisis de metodologías y herramientas para medir y optimizar el consumo energético.
- Estudio de técnicas para reducir el consumo energético en aplicaciones de HPC de acuerdo con las arquitecturas utilizadas.
- Evaluación de eficiencia energética de diferentes algoritmos y plataformas paralelas.
- Optimización de la eficiencia energética. A partir de los valores de energía que brindan los contadores hardware es posible definir estrategias de programación que lleven a reducir el consumo, manteniendo a su vez el rendimiento en valores aceptables [15].

Cloud Computing

Cloud Computing, proporciona grandes conjuntos de recursos físicos y lógicos (como pueden ser infraestructura, plataformas de desarrollo, almacenamiento y/o aplicaciones), fácilmente accesibles y utilizables por medio de una interfaz de administración web, con un modelo de arquitectura "*virtualizada*" [16,17]. Estos recursos son proporcionados como servicios ("*as a service*") y pueden ser dinámicamente reconfigurados para adaptarse a una carga de trabajo variable (escalabilidad), logrando una mejor utilización y evitando el sobre o sub dimensionamiento (elasticidad) [18].

Por otro lado, Cloud Robotics es una de las áreas más prometedoras de la investigación informática actual en la cual se cuenta con "robots" dotados de diferentes sensores y capacidades, conectados a un Cloud vía Internet. Los temas de investigación derivados son múltiples: sensores, redes de sensores e inteligencia distribuida; robótica y sistemas colaborativos de tiempo real basados en robots; aplicaciones críticas (por ej. en ciudades inteligentes o en el ámbito industrial).

Recientemente, *Edge* y *Fog Computing* surgen como una evolución del modelo tradicional de Cloud en busca de disminuir el tiempo de respuesta de las soluciones y mejorar el aprovechamiento de recursos [19].

Resiliencia

En la actualidad, lograr sistemas resilientes resulta un verdadero desafío considerando el creciente número de componentes, la cercanía a los límites físicos en las tecnologías de fabricación y la complejidad incremental del software. La corrección de las aplicaciones y la eficiencia en su ejecución se torna más importante en HPC debido a los extensos tiempos de ejecución. En ese sentido, resulta interesante desarrollar estrategias de detección y recuperación de fallos, especialmente a través de librerías de software.

Entrada/Salida paralela

A pesar de los avances tecnológicos, las operaciones de E/S en los centros de supercómputo siguen siendo un cuello de botella para determinadas aplicaciones HPC. El rendimiento de un sistema dependerá de la carga de trabajo (patrones de E/S de las aplicaciones) y de su configuración (hardware y software). Contar con herramientas que permitan modelar y predecir el comportamiento de este tipo de aplicaciones en HPC resulta fundamental para mejorar su rendimiento.

Dispositivos de bajo costo con capacidades para cómputo paralelo

En la actualidad se comercializan placas de bajo costo como Raspberry PI [20] u Odroid [21] que poseen múltiples núcleos simples. Asimismo, existen diversos dispositivos móviles con capacidades similares. Es de interés estudiar como explotar el paralelismo en estos dispositivos para mejorar el rendimiento y/o consumo energético de las aplicaciones.

Líneas de Investigación, Desarrollo e Innovación

- Arquitecturas many-core (procesadores MIC, GPU y TPU) y FPGA. Análisis de este tipo de máquinas y de técnicas para desarrollar código optimizado.
- Arquitecturas híbridas (diferentes combinaciones de clusters, multicores, manycores y FPGAs). Diseño de algoritmos paralelos sobre las mismas. Técnicas de resiliencia.
- Exploración de nuevos lenguajes y modelos de programación para HPC.
- Cloud Computing para realizar HPC. Evaluación de perfomance en este tipo de arquitectura. Análisis del overhead por el software de administración del Cloud.
- Sistemas inteligentes de tiempo real distribuidos (Cloud Robotics). Edge y Fog Computing.
- Interconexión de Brokers de mensajes MQTT sobre Cloud Públicos y Cloud Privados.
- Consumo energético en las diferentes arquitecturas paralelas, en particular en relación a los algoritmos paralelos y la configuración de la arquitectura. Análisis de metodologías y herramientas de medición. Modelado y estimación del consumo de potencia de arquitecturas HPC.

Resultados y Objetivos

Investigación experimental a realizar

- Desarrollar y evaluar algoritmos paralelos sobre nuevas arquitecturas. Analizar rendimiento, eficiencia energética y costo de programación.
- Analizar el overhead introducido por el sistema gestor del Cloud en un entorno de HPC para aplicaciones científicas de Big Data.
- Estudiar y comparar las diferentes estrategias para interconectar brokers de mensajes MQTT tanto sobre cloud públicos como privados.
- Realizar el desarrollo de nuevos planificadores de tareas para multicores asimétricos sobre diferentes sistemas operativos con el objetivo de maximizar el rendimiento y minimizar el consumo de energía [15,22].
- Optimizar algoritmos paralelos para controlar el comportamiento de múltiples robots que trabajan colaborativamente, considerando la distribución de su capacidad de procesamiento “local” y la coordinación con la potencia de cómputo y capacidad de almacenamiento (datos y conocimiento) de un Cloud.

- Desarrollar técnicas de tolerancia a fallas que permitan aumentar la resiliencia de sistemas paralelos y distribuidos.
- Desarrollar un modelo de E/S en HPC que permita predecir cómo los cambios realizados en los diferentes componentes del mismo afectan a la funcionalidad y al rendimiento del sistema.

Resultados obtenidos

- Se diseñó y desarrolló un sistema inteligente distribuido para monitorización y control del consumo energético [23].
- Se evaluó el rendimiento de diferentes aplicaciones web en entornos virtualizados basados en Cloud [24].
- Se desarrolló una plataforma distribuida, escalable, híbrida y redudante que permite aprovechar recursos de cómputo basados en arquitecturas ARM a través de un servicio de cloud [25].
- Se desarrolló y validó un modelo estadístico para consumo de potencia en placas RPi de diferentes generaciones [26].
- Se propuso una metodología de selección de eventos de contadores de hardware que sirvan como entrada para un modelo de predicción de consumo energético multi-arquitecturas GPUs [27].
- Se optimizó una solución al problema de los N cuerpos computacionales con atracción gravitacional (versión directa) sobre arquitectura Xeon Phi Knights Landings [28].
- Se estudió y evaluó el desempeño de la herramienta SEDAR para detección y recuperación de fallos transitorios [29].
- Se estudiaron los factores que afectan al consumo energético en las operaciones de checkpoint y restart en clusters con el propósito de caracterizarlos [30].
- Se desplegó un sistema multi-robot integrado por un chasis de auto Rover de 4 ruedas y un cuadricóptero Parrot Bebop. A partir de la comunicación mediante un servicio de cloud público, los robots son capaces de alcanzar un punto destino en común [31].
- Se diseñó y desarrolló un modelo predictivo de tiempo y consumo energético de arquitecturas GPUs y cluster de GPUs para un problema de alta demanda computacional [32].
- Se optimizaron soluciones híbridas CPU-GPU para el procesamiento de pattern matching [8].
- Se desarrolló un modelo conceptual del sistema de archivos paralelos PVFS2 y se implementó por medio del paradigma de agentes en un simulador [33].

Organización de Eventos

En el año 2019 se han organizado las VII Jornadas de Cloud Computing y Big Data (JCC&BD 2019) en Argentina, con participación de especialistas académicos del país y del exterior y de empresas con experiencia en Cloud Computing [34]. En junio de 2020 se organizarán las VIII Jornadas de Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2020).

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D el último año se concluyó 1 trabajo de Especialización y 1 Tesis de Doctorado. Al mismo tiempo se encuentran en curso 4 tesis de Doctorado en Ciencias Informáticas y 3 tesis de Maestría.

Además, se participa en el dictado de las carreras de Doctorado en Ciencias Informáticas, y Magíster y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática de la UNLP (acreditadas por la CONEAU con categoría A, B y A, respectivamente), por lo que potencialmente pueden generarse nuevas Tesis de Doctorado y Maestría y Trabajos Finales de Especialización.

Existe cooperación con grupos de otras Universidades del país y del exterior, y hay tesistas de diferentes Universidades realizando su Tesis con el equipo del proyecto.

Respecto a las carreras de grado, se dictan por parte de integrantes de la línea de investigación tres materias directamente relacionadas con los temas de esta: “Taller de Programación sobre GPUs”, “Cloud Computing y Cloud Robotics” y “Conceptos y Aplicaciones en Big Data”.

Referencias

[1] Rucci, Enzo: "Evaluación de rendimiento y eficiencia energética en sistemas heterogéneos para bioinformática". Tesis de Doctorado en Ciencias Informáticas (Facultad de Informática – UNLP). 2016.

[2] Balladini, J., Rucci, E., De Giusti, A., Naiouf, M., Suppi, R., Rexachs, D., Luque, E. "Power characterisation of shared-memory HPC systems". Computer Science & Technology Series – XVIII Argentine Congress of Computer Science Selected Papers. ISBN 978-987-1985-20-3. Págs. 53-65. 2013.

[3] Bertogna, M., Grosclaude, E., Naiouf, M., De Giusti, A., Luque, E.: "Dynamic on Demand Virtual Clusters in Grids". In: 3rd Workshop on Virtualization in High-Performance Cluster and Grid Computing (VHPC 08). España. (2008).

[4] Lorencik D., Sincak P.: "Cloud robotics: Current trends and possible use as a service". In: Applied Machine Intelligence and Informatics (SAMI), 2013 IEEE 11th International Symposium on , vol., no.,

[5] Guoqiang Hu, Wee Peng Tay, Yonggang Wen: "Cloud robotics: architecture, challenges and applications". In: Network, IEEE, vol.26, no.3, pp.21-28. 2012.

[6] Kehoe B., Patil S., Abbeel P., Goldberg K.: "A Survey of Research on Cloud Robotics and Automation". In: IEEE Transactions on Automation Science and Engineering (T-ASE): Special Issue on Cloud Robotics and Automation. Vol. 12, no. 2. 2015.

[7] V. Sanz, A. Pousa, M. Naiouf, and A. De Giusti, "Accelerating Pattern Matching with CPU-GPU Collaborative Computing", Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing (ICA3pp 2018), ISBN: 978-3-030-05051-1, págs. 310-322, doi. https://doi.org/10.1007/978-3-030-05051-1_22, 2018.

[8] Sanz V., Pousa A., Naiouf M., De Giusti A. (2020) Efficient Pattern Matching on CPU-GPU Heterogeneous Systems. In: Wen S., Zomaya A., Yang L. (eds) Algorithms and Architectures for Parallel Processing. ICA3PP 2019. Lecture Notes in Computer Science, vol 11944. Springer, Cham. 2020

[9] Reinders, J., Jeffers, J., Sodani, A. "Intel Xeon Phi Processor High Performance Programming Knights Landing Edition". Morgan Kaufmann Publish-ers Inc., Boston, MA, USA, 2016
pp.85-88. 2013.

[10] IBM. "IBM and Xilinx Announce Strategic Collaboration to Accelerate Data Center Applications". Disponible en https://www-03.ibm.com/press/us/en/pressrelease/48074.wss

[11] Intel. "Intel Acquisition of Altera". Disponible en http://intelacquiresaltera.transactionannouncement.com

[12] Sean Settle: "High-performance Dynamic Programming on FPGAs with OpenCL". In: IEEE High Performance Extreme Computing Conference. 2013.

[13] Xilinx Inc. "SDAccel Development Environment". [Online]. Disponible en http://www.xilinx.com/products/design-tools/software-zone/sdaccel.html

[14] Google Inc. "Cloud Tensor Processing Unit (TPU)" Disponible en https://cloud.google.com/tpu/docs/tpus?hl=es-419

[15] Saez, J.C., Pousa, A., Rodríguez-Rodríguez, R., Castro, F., Prieto-Matias, M. "PMCTrack: Delivering performance monitoring counter support to the OS scheduler". The computer journal Volume 60, Issue 1 January 2017.

[16] Shafer, J.: "I/O virtualization bottlenecks in cloud computing today". In: Proceedings of the 2nd conference on I/O virtualization (VIOV10). USA (2010).

[17] Xing, Y., Zhan, Y.: "Virtualization and Cloud Computing". In: Proceedings pp.305-312, Springer Link. ISBN 978-3-642-27323-0. (2012). Morgan Kaufmann. 2013.

[18] Velte, A.T., Velte, T.J., Elsenpeter, R.: "Cloud Computing: A Practical Approach". McGraw Hill Professional. 2009.

[19] Ashkan Yousefpour, Caleb Fung, Tam Nguyen, Krishna Kadiyala, Fatemeh Jalali, Amirreza Niakanlahiji, Jian Kong, Jason P. Jue. "All one needs to know about fog computing and related edge computing paradigms: A complete survey", Journal of Systems Architecture, Volume 98, 2019, Pages 289-330, ISSN 1383-7621, https://doi.org/10.1016/j.sysarc.2019.02.009.

[20] Raspberry PI. Disponible en https://www.raspberrypi.org/

[21] Odroid Disponible en http://www.hardkernel.com

[22] Juan Carlos Saez, Adrian Pousa, Daniel Chaver, Fernando Castro, Manuel Prieto Matias: "ACFS: A Completely Fair Scheduler for Asymmetric Single-ISA Multicore Systems". In: ACM SAC 2015 (The 30TH ACM/SIGAPP Symposium on applied computing). 2015.

[23] Pi Puig M. et al. (2019) Intelligent Distributed System for Energy Efficient Control. In: Naiouf M., Chichizola F., Rucci E. (eds) Cloud Computing and Big Data. JCC&BD 2019. Communications in Computer and Information Science, vol 1050. Springer, Cham

[24] Tinetti F.G., Rodríguez C. (2019) Evaluating Performance of Web Applications in (Cloud) Virtualized Environments. In: Naiouf M., Chichizola F., Rucci E. (eds) Cloud Computing and Big Data. JCC&BD 2019. Communications in Computer and Information Science, vol 1050. Springer, Cham

[25] Petrocelli D., De Giusti A., Naiouf M. (2019) Hybrid Elastic ARM&Cloud HPC Collaborative Platform for Generic Tasks. In: Naiouf M., Chichizola F., Rucci E. (eds) Cloud Computing and Big Data. JCC&BD 2019. Communications in Computer and Information Science, vol 1050. Springer, Cham

[26] J. M. Paniego, L. Libutti, M. Pi Puig, F. Chichizola, L. De Giusti, M. Naiouf, A. De Giusti. "Modelado de potencia en placas SBC: integración de diferentes generaciones Raspberry Pi". Actas del XXV

Congreso Argentino de Ciencias de la Computación (CACIC 2019), pp. 159-169, octubre de 2019.
[27] Pi Puig, Martín; De Giusti, Laura; Naiouf, Marcelo, De Giusti Armando. “A Study of Hardware Performance Counters Selection for Cross Architectural GPU Power Modeling”. Actas del XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019), pp. 126-139, octubre de 2019.
[28] Rucci, Enzo; Moreno, Ezequiel; Camilo, Malek; Pousa, Adrián; Chichizola, Franco. “Simulación de N Cuerpos Computacionales sobre Intel Xeon Phi KNL”. Actas del XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019), pp. 194-204, octubre de 2019.
[29] Montezanti, Diego; Rucci, Enzo; Rexachs, Dolores; Luque, Emilio; Naiouf, Marcelo; De Giusti, Armando. “SEDAR: detectando y recuperando fallos transitorios en HPC”. Actas del XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019), pp. 170-182, octubre de 2019.
[30] Morán M., Balladini J., Rexachs D., Luque E. (2019) Checkpoint and Restart: An Energy Consumption Characterization in Clusters. In: Pesado P., Aciti C. (eds) Computer Science – CACIC 2018. CACIC 2018. Communications in Computer and Information Science, vol 995. Springer, Cham.
[31] Costanzo, Manuel; Boggia, Marcos; Rodriguez, Ismael; De Giusti, Armando; “Cloud/Edge Robotics: Navegación autónoma de Auto-Robot y Cuadricoptero”. Actas del XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019), pp. 1062-1071, octubre de 2019.
[32] Montes de Oca E., De Giusti L., De Giusti A., Naiouf M. (2019) Energy Consumption Analysis and Time Estimation Model in GPU Cluster and MultiGPU in a High Computational Demand Problem. In: Pesado P., Aciti C. (eds) Computer Science – CACIC 2018. CACIC 2018. Communications in Computer and Information Science, vol 995. Springer, Cham.
[33] Encinas D.; Naiouf M; De Giusti A; Mendez S; Rexachs D; Luque E. “On the Calibration, Verification and Validation of an Agent-Based Model of the HPC Input/Output System”. Proceedings from The Eleventh International Conference on Advances in System Simulation (SIMUL 2019), November 24 - 28, 2019.
[34] M. Naiouf, F. Chichizola, and E. Rucci (Editores). Cloud Computing and Big Data – Revised Selected Papers from JCC&BD 2019. Springer CCIS, ISBN: 978-3-030-27713-0, 2019.

Solución de Grandes Problemas aplicando HPC Multi-tecnología

Mercedes Barrionuevo [1], Julian Escalante[2], Mariela Lopresti[1], Maximiliano Lucero[1], Natalia Miranda[1], M. Antonia Murazzo[3] y Fabiana Piccoli[1 y 2].

[1]LIDIC- Universidad Nacional de San Luis, San Luis

[2]Universidad Autónoma de Entre Ríos, Concepción del Uruguay

[3] Universidad Nacional de San Juan, San Juan

Argentina

{mdbarrio, omlopres, mlucero, ncmiran, mpiccoli}@unsl.edu.ar

{escalantejulian18,maritemurazzo}@gmail.com

RESUMEN

Los problemas a resolver hoy en día mediante soluciones computacionales demandan muchos recursos, esto puede obedecer a dos factores: el tamaño del problema en relación a los datos a tratar, y la complejidad del problema. Puede ser uno o ambos factores, en cualquier caso, siempre los resultados se esperan en un tiempo razonable. Una de las soluciones propuestas es pensar en sistemas heterogéneos: una computadora formada por procesadores many y multicores con software capaz de tomar ventaja de cada una de las componentes subyacentes.

En este trabajo se exponen dos líneas de trabajo orientadas a distintos tipos de problemas, en las cuales se propone desarrollar técnicas de Computación de Alto Desempeño para el escenario descripto.

Palabras clave: Computación de Alto Desempeño. Modelos Híbridos. Problemas Complejos. Big Data.

CONTEXTO

Esta propuesta de trabajo se lleva a cabo dentro de los proyectos de investigación: "Tecnologías Avanzadas aplicadas al Procesamiento de Datos Masivos" (LIDIC, UNSL) y "Cómputo de Altas Prestaciones aplicado a la Solución de Grandes Problemas" (UADER); y del proyecto binacional CAPG-BA 66/13 entre la Universidad Nacional de San Luis y la Universidad de Pernambuco, Recife, Brasil. Además de docentes-investigadores de otras universidades, UNSJ.

1. INTRODUCCIÓN

El crecimiento de los problemas a resolver computacionalmente, respecto a la cantidad de datos y complejidad de las soluciones, hizo intensificar la búsqueda de nuevas opciones. Desde el punto de vista del hardware, los avances se enfocaron en el desarrollo de computadoras de propósito general con múltiples núcleos, dando origen a dos líneas de desarrollo: las arquitecturas multi-core (multi-núcleos) y las arquitecturas many-cores (muchos-núcleos o muchos-cores) [Vaj11]. En los multi-core, los avances se centraron en el desarrollo de mejoras para acelerar, generalmente, las aplicaciones secuenciales incorporando varios núcleos de procesamiento. Surgen así computadoras con al menos 2 procesadores por unidad central (multi). En el caso de los many-cores, los desarrollos se centraron en optimizar el desempeño de aplicaciones paralelas, en este tipo de arquitectura se encuentran las tarjetas gráficas o GPUs [Aam18].

Los multi-core y many-core tienen características propias, una de sus principales diferencias, como ya se enunció, es la de proveer mejor desempeño: soluciones secuenciales vs.

solución paralela. Para lograrlo, las primeras brindan lógica de control más compleja o memorias caché más rápidas y grandes por ejemplo. En el caso de las otras, la idea es optimizar el throughput de muchos threads ejecutando en paralelo, por ello si alguno de ellos está esperando por la finalización de una operación, otro es asignado al procesador para que no permanezca ocioso. En este tipo de arquitecturas, las áreas dedicadas a control y memoria caché son generalmente pequeñas [Hon10].

Ambos tipos de arquitecturas, si bien difieren en varios aspectos, ofrecen a todos los cores acceso a una única memoria compartida. Como la cantidad de memoria disponible es limitada, una alternativa son las arquitecturas con memoria distribuida, quienes permiten incrementar el espacio de almacenamiento incorporando el costo de las comunicaciones a través de la red [Bas16]. Las arquitecturas con memoria distribuida ponen a disposición un gran número de máquinas y sus memorias, eliminando el problema de la limitación de la RAM. Los clusters y cloud computing son ejemplos de ellas[Kau14, Mar17].

Una tercera alternativa son los sistemas híbridos, los cuales permiten combinar las características de los sistemas de memoria compartida y de memoria distribuida, incrementando la capacidad y poder de cómputo de la arquitectura resultante. Esto posibilita aprovechar las ventajas ofrecidas por ambas, múltiples procesos y threads con distintas administraciones de memoria coexistiendo en el sistema.

Desde el punto de vista del software, también existen diferentes clasificaciones, ellas dependen del objeto a considerar, por ejemplo si se considera qué se divide, los modelos de programación estándares, o paradigmas, son: paralelismo de datos y paralelismo de tareas [Pac11]; en cambio si se tiene en cuenta la arquitectura, es posible clasificar a los modelos como de memoria compartida y de memoria distribuida o pasaje de mensajes. Para cada uno de ellos, existen diferentes herramientas, las cuales permiten llevarlos a cabo en las distintas arquitecturas, entre ellos OpenMP[Op19], MPI[Gro14] y CUDA[NVi19].

Teniendo en cuenta la demanda de cómputo de los problemas actuales y la posibilidad que nos brindan las arquitecturas híbridas: arquitecturas distribuidas con componentes multi y many-core, es posible combinar los paradigmas en un modelo de programación híbrido logrando aplicar una estrategia de paralelización más efectiva mediante múltiples niveles de paralelismo y reducción del overhead de comunicación. Existen muchas soluciones a problemas complejos de gran escala y para grandes volúmenes de datos, las cuales aplican técnicas HPC multi-tecnología, tanto de hardware como de software [Car18, Isu19].

Ante todo lo expuesto, nuestra motivación es investigar, verificar y poner en marcha nuevas técnicas y arquitecturas híbridas que ayuden a mejorar el procesamiento y sus tiempos de respuesta. Las técnicas de HPC serán nuestras herramientas para resolver con eficiencia cada uno de los objetivos: aplicar HPC multi-tecnología para resolver problemas complejos o de datos masivos en ambientes paralelos híbridos sobre arquitecturas many y multi-core. En la próxima sección presentamos diferentes casos de estudio para dos líneas de investigación.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Los trabajos desarrollados se encuadran en dos líneas de investigación, una relacionada a la solución de problemas complejos con técnicas HPC y el otro a problemas Big Data. Para cada área nos enfocamos en:

- Problemas Complejos: Un Autómata Celular (AC) es un modelo matemático que

permite trabajar con sistemas dinámicos discretos de comportamiento complejo mediante el desarrollo de un conjunto de reglas simples. Estas reglas permiten especificar el nuevo estado de un componente en función de su estado actual y de su vecindario. Por sus características los AC son naturalmente paralelos, lo cual permite aplicar técnicas HPC en su solución computacional [Kau 84, Wol84].

- Uno de los usos más difundidos hoy en día de los AC, por su simplicidad y posibilidad de constante mejora, es en el ámbito de la simulación de procesos de difusión o dispersión, ya sea de rumores, enfermedades, etc. Analizar, prevenir y/o controlar la difusión y contagio de enfermedades, rumores, virus, etc. constituye una de las grandes preocupaciones del ser humano. Existen casos donde una enfermedad provocó la desaparición de una población entera o cambios demográficos importantes, por ejemplo la peste en (Europa, siglo XIV), fiebre amarilla (Buenos Aires, siglo XIX) o el cólera (Asia, siglo XIX), continuado la situación hasta hoy, por ejemplo con la Gripe A, el dengue, el ébola y el SIDA [Hup13, Wor14] o más recientes como el COVID-19. Ante estos problemas es importante y prioritario el estudio y control de estas enfermedades, como así también el análisis de su modo de transmisión o contagio. Una forma de atacar este problema es estudiando su comportamiento en un determinado medioambiente o realidad [Joh09, Mai13, Rey11, Yir12]. Los AC permiten expresar modelos realistas capaces de capturar con mayor fidelidad las características y el comportamiento del estado del sistema y su evolución, los cuales no pueden ser abordados analíticamente.
- En esta línea de trabajo, proponemos desarrollar un prototipo paralelo portable basado en AC para simular un proceso de difusión, tanto en arquitecturas multi-core, como many-core. En el caso de las primeras, hemos considerado tanto arquitecturas de memoria compartida y de memoria distribuida. Para los many-core, se tienen en cuenta las arquitecturas más conocidas y propuestas por Nvidia y AMD. Los problemas de difusión considerados son, en enfermedades: Gripe A y Covid-19; y, área social: noticias/rumores. Estas propuestas no permiten, por un lado estudiar en menos tiempo, el proceso de propagación teniendo en cuenta diferentes realidades: el tipo de población, su distribución y otras características; para luego tomar decisiones, como las campañas de vacunación, el aislamiento, cuarentena, mecanismo de contención, etc. Por el otro lado, evaluar el prototipo para las distintas arquitecturas y herramientas, analizando su portabilidad a plataformas heterogéneas.
- Grandes Volúmenes de Datos: En las últimas décadas el conjunto de datos a manipular es tan grande y complejo que los medios tradicionales de procesamiento son ineficaces. Existen diferentes áreas o aplicaciones donde se trabaja con gran cantidad de datos, en esta línea nos enfocamos en: seguridad en redes de datos, recuperación de la información y evolución y parentesco de las especies. El trabajo abordado en cada una de ellas es:
  - Seguridad en Redes de Datos: En el campo de la detección de anomalías en el tráfico en redes de datos, el problema consiste en la identificación de patrones no acordes al comportamiento normal del tráfico en la red [Bar16, Bar17]. Detectar un posible ataque requiere contar con tecnologías para la

clasificación del tráfico, asociando flujos de datos con las aplicaciones que los generan, lo que se denomina Sistema de Detección de Intrusos (IDS: Intrusion Detection System). Aquí proponemos la creación de un IDS híbrido, quien mediante técnicas de clasificación supervisada y no supervisada, visualización de datos, procesamiento de imágenes y HPC en GPU, detecta tanto ataques conocidos como potenciales en una red de área local. Para lograrlo, se busca identificar patrones que se desvían del comportamiento normal a partir del tráfico circulantes en la red (gran volumen de datos) en un tiempo cercano a su ocurrencia, a fin de poder tomar decisiones rápidas [Bar18].

- Recuperación de la Información: El incesante crecimiento de los datos, hace que en la actualidad las bases de datos no sólo son utilizadas para almacenar números o palabras, sino también documentos, archivos multimedia, muestras biológicas, huellas digitales entre otros tipos de datos, haciendo imposible realizar búsquedas de manera tradicional. Las consultas por similitud o búsqueda aproximada se presentan como una solución, por lo cual se necesitan métodos de acceso eficientes (índices) para permitir recuperar rápidamente los elementos que satisfacen los criterios de consulta, es decir los elementos son buscados considerando la cercanía de los mismos al elemento consultado. El modelo de espacio métrico [Cha01] resulta adecuado para ello. En él, las evaluaciones de la función de distancia requieren un alto costo computacional, por lo tanto es necesario la construcción de un índice para reducir la cantidad de cálculos. Uno de los índices está basado en permutantes y establece que dado un conjunto de datos se pueden seleccionar elementos como puntos de referencia, donde el grado de cercanía de dos elementos está dado por la similitud de sus permutaciones [Lop13]. Hoy en día los grandes volúmenes de datos almacenados en bases de datos métricas hace que no sea suficiente procesar previamente el conjunto de datos mediante la creación de un índice, también es necesario acelerar las consultas mediante HPC. Aquí proponemos utilizar una arquitectura many-core para resolver consultas utilizando permutantes a fin de optimizar las búsquedas sobre grandes volúmenes de datos.
- Construcción de Árboles Filogenéticos: En el área de biología evolutiva [Ben09, Gar14, Yan14], la elaboración de un árbol filogenético en el cual se muestran las relaciones evolutivas de los seres vivos entre sí, reconocer los grados de cercanía de ancestros comunes, requiere contar con datos moleculares (fundamentalmente ADN) y/o morfológicos, los cuales identificarán la/s especie/s a incorporar en el árbol. El análisis filogenético demanda cálculos más complejos a medida que la cantidad de especies a catalogar crece, la cantidad de árboles a examinar en una búsqueda aumenta exponencialmente. Por esta razón es necesario contar con métodos, herramientas y técnicas HPC para datos masivos con el objeto de obtener información dentro de tiempos razonables. Para este caso, el objetivo es considerar ambientes HPC híbridos, a fin de lograr obtener buenos resultados, sin considerar el crecimiento exponencial de los datos a procesar.

Todas estas líneas de investigación tienen en cuenta la escalabilidad del problema y la

portabilidad de los desarrollos, aplicando técnicas de paralelismo híbrido y arquitecturas heterogéneas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Como objetivos de las líneas de investigación nos planteamos facilitar el desarrollo de soluciones paralelas portables, de costo predecible, capaces de explotar las ventajas de modernos ambientes HPC a través de herramientas y "frameworks de computación" de alto nivel. Los resultados obtenidos hasta el momento son muy satisfactorios.

4. FORMACIÓN DE RECURSOS HUMANOS

Los resultados esperados respecto a la formación de recursos humanos son hasta el momento el desarrollo de 3 tesis de posgrado: 1 de doctor y 2 de maestría; y varias tesinas de grado en las distintas universidades intervinientes.

5. BIBLIOGRAFÍA

[Ben09] Benítez Burraco, A.. "*Genes y lenguaje: aspectos ontogenéticos, filogenéticos y cognitivos*". ISBN 8429110046, 9788429110043. Reverte, 2009.

[Bas16] Basu, S. "*Parallel and Distributed Computing Architectures and Algorithms*". ISBN: 8120352122, 9788120352124. PHI Learning Pvt. Ltd., 2016.

[Bar16] Barrionuevo, M. Lopresti, M., Miranda, N. Piccoli, F.. "*Un enfoque para la detección de anomalías en el tráfico de red usando imágenes y técnicas de Computación de Alto Desempeño*". XXII Congreso Argentino De Ciencias de la Computación. CACIC 2016. Pp. 1166-1175. Octubre 2016, San Luis, Argentina.

[Bar17] Barrionuevo, M. Lopresti, M. Miranda, N. Piccoli, F. "*Un Modelo de Detección de Anomalías en una LAN usando K-NN y Técnicas de Computación de Alto Desempeño*". Communications in Computer and Information Science" (CCIS). ISBN: 978-3-319-75213-6. Editorial: Springer. Edición: Extranjera (inglés). Año: 2017. Volumen 790. Pp 219-228.

[Bar18] Barrionuevo, M., Lopresti, M., Miranda, N. and Piccoli, F. . "*Secure Computer Network: Strategies and Challenges in Big Data Era*". Journal of Computer Science & Technologies. Editorial Red de Universidades Nacionales con Carreras de Informática(RedUNCI) Iberoamerican Science & Technology Education Consortium (ISTEC). Vol 3, pp 248-257. ISSN: 1666-6038 (electrónico), 1666-6046 (impreso).Diciembre 2018.

[Car18] CARLA 2018, *Selected Papers High of Performance Computing* 5th Latin American Conference. Communications in Computer and Information Science, Vol. 979. ISBN: 3030162052, 9783030162054. Colombia, 2018. Springer, 2019.

[Cha01] Chávez, E. Navarro, G., Baeza-Yates, R. and Marroquín, J.. "*Searching in metric spaces*". ACM Comput. Surv. 33, 3 (September 2001), 273-321. 2001.

[Gar14] Garamszegi, L.. "*Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*". ISBN 3662435500, 9783662435502. Springer, 2014.

[Gro14] Gropp, W., Hoefler, T., Thakur, R. and Lusk, E. "*Using Advanced MPI: Modern Features of the Message-Passing Interface*". Computer science & intelligent systems Scientific and Engineering Computation. ISBN: 0262527634 - 9780262527637. MIT Press, 2014.

[Hon10] Hong, S., Kim,H. , Hong, S. and Kim, H.. "*An integrated GPU power and performance model*" in Proceedings of the 37th annual international symposium on Computer architecture - ISCA '10, Vol. 38, Nº 3.Pp. 280-289. 2010.

[Hup13] Huppert, A. and Katriel, G.. "*Mathematical modelling and prediction*" in infectious disease epidemiology. Clinical Microbiology and Infection. Vol 19 Nº11. Pp 999-1005, 2013.

[Isu19] ISUM 2019, *Selected Papers of Supercomputing*: 10th International Conference on Supercomputing. Communications in Computer and Information Science, Vol. 1151. ISBN: 3030380432, 9783030380434. Springer Nature. México 2019.

[Joh09] Johnson, T. and McQuarrie, B.. "*Mathematical modeling of diseases: Susceptible-infected-recovered (sir) model*". In University of Minnesota, Morris, Math 4901 Senior Seminar, 2009.

[Kau14] Kaur, K. and Rai, A.. "*A Comparative Analysis: Grid, Cluster and Cloud Computing*" Int. J. Adv. Res. Comput. Commun. Eng., vol. 3, no. 3, pp. 2278–1021, 2014.

[Kau84] Kaufiman, S.. "*Emergent properties in random complex automata*". Physica D: Nonlinear Phenomena. Vol 10 Nº1 Pp 145-156, 1984.

[Lop13] Lopresti, M., Miranda, N., Piccoli, F., Reyes, N. "*Solving Multiple Queries through the Permutation Index in GPU*". 4th International supercomputing Conference in Mexico. Colima-México. 5-8 Marzo 2013.

[Mai13] Maitanmi, O., Adekunle, Y. and Agbaje, M.. "*Model-Based Cellular Automata on Spread of Rumours*". Computer Science, Babcock University. Nigeria. 2013.

[Mar 17] Marinescu, D.. "*Cloud Computing: Theory and Practice*". 2nd Ed. ISBN: 0128128119, 9780128128114. Morgan Kaufmann, 2017.

[NVi19] Nvidia. "*CUDA C++ Programming Guide, Design Guide*". https://docs.nvidia.com/cuda /pdf/CUDA_C_Programming_Guide.pdf. 2019.

[Op19] OpenMP Architecture Review Board. "*OpenMP Application Programming Interface Specification Version 5.0*". ISBN: 1795759887 - 978-1795759885. 2019.

[Pac11] Pacheco, P. "*An Introduction to Parallel Programming*", 1st ed., San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.

[Rey11] Reyes, D.. "*Descripción y aplicaciones de los autómatas celulares*", 2011. Technical Report, Autónoma de Puebla.

[Vaj11] Vajda, A. "*Multi-core and Many-core Processor Architectures. In: Programming Many-Core Chips*". Springer, Boston, MA. 2011.

[Wol84] Wolfram, S.. "*Universality and complexity in cellular automata*". Physica D: Non-linear Phenomena, Vol 10 Nº 1. Pp 1-35, 1984.

[Wor14] World Health Organization. *Influenza (seasonal)*. Fact sheet N211. 2014.

[Yan14] Yang, Z.. "*Molecular Evolution: A Statistical Approach*". OUP Oxford, 2014.

[Yir12] Yiran, G. and Jinzhu, D. "*Research on Rumors Spread Based on Cellular Automata*", in Proceedings of the 2nd International Conference on Green Communications and Networks. Vol 2. ISBN 978-3-642-35418-2. Pp 235-244 Londres, Inglaterra. Springer Heidelberg. 2012.

Técnicas de modelado y simulación en entornos HPC.

Diego Encinas[1,2], Jimena Jara[1], Daniel Rosatto[1], Román Bond[1], Lucas Maccallini, Mauro Gomez[1] Martin Morales[1,3]

[1]Proyecto de Investigación SimHPC - Programa TICAPPS - Instituto de Ingeniería y Agronomía - UNAJ

[2]Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP – Centro Asociado CIC

[3] Unidad CodApli - Facultad Regional La Plata - UTN

dencinas@unaj.edu.ar, elchejime@gmail.com, danielrosatto@gmail.com, rbond@unaj.edu.ar, lucas.maccallini@gmail.com, mauro.f.gmz@gmail.com, martin.morales@unaj.edu.ar

Resumen

El objetivo de esta línea de investigación es el estudio de diferentes técnicas de modelado y simulación para entornos de Cómputo en Altas Prestaciones (HPC, High Performance Computing). Enfocando a la obtención de herramientas que permitan predecir la eficiencia del sistema ante posibles escenarios y reconfigurar el sistema físico. Analizando los diferentes componentes del sistema que pueden influir en las prestaciones significativamente y pueden llegar a modelarse y/o reconfigurarse.

Por otra parte se ha iniciado el desarrollo de un simulador para entornos en el área de salud, específicamente en el sector de emergencias hospitalarias.

Palabras clave: *Arquitecturas Multiprocesador. Simulación. Sistema E/S paralela. Modelado y Simulación basado en agentes (Agent-BasedModeling and Simulation, ABMS). Cloud Computing. CloudSim. Simulación y Salud.*

Contexto

Se presenta una línea de Investigación que es parte del Proyecto de Investigación “Simulación y tecnología en Cómputo de Altas Prestaciones (High Performance Computing, HPC) para aplicaciones de interés social” - SimHPC de la Universidad Nacional Arturo Jauretche (UNAJ), acreditado por resolución interna 148/18. Además, el proyecto aporta al Programa “Tecnologías de la información y la comunicación (TIC) en aplicaciones de interés social” – TICAPPS de la UNAJ.

En el tema existe un convenio de colaboración en actividades de Investigación y Postgrado con el Instituto de Investigación en Informática – LIDI de la Universidad Nacional de La Plata y el Área de Computación del Instituto de Ciencias de la Universidad Nacional de General Sarmiento.

Introducción

El crecimiento sostenido en la demanda del poder de cómputo remarca la necesidad de sistemas con enfoques de paralelización masiva y cómputo de alta performance (HPC, High Performance Computing) [1]. Los clusters se han convertido en uno de los enfoques principales para lograr paralelismo a bajo costo. Una noción extendida lo constituye la utilización de grid computing y más recientemente cloud computing. Independientemente de la solución, estos sistemas constan de un gran número de componentes incluyendo nodos de procesamiento, bancos de memoria, discos, entre otros.

En cuanto a las herramientas de simulación para Cloud Computing, CloudSim es un framework desarrollado en Java que provee las

APIs necesarias para que el usuario genere una simulación de un Data Center funcionando como servidor de nube capaz de simular la ejecución de CloudLets.

Por otra parte, los servicios de urgencias hospitalarias son considerados como una de las unidades del sistema sanitario de mayor complejidad y fluidez, lo que unido a la variabilidad de su actividad da lugar a que su gestión operativa sea una tarea muy complicada [2]. Es por ello que resultaría de mucha utilidad para sus responsables disponer de un sistema de ayuda a la toma de decisiones (Decision Support System-DSS) tan flexible como un simulador, que permitiría tomar medidas disponiendo de información suficiente sobre las alternativas posibles.

Sistemas de E/S Paralela

Las exigencias en los sistemas de E/S paralelos se han incrementado debido al aumento en número, velocidad y potencia de las unidades de procesamiento en los clusters. También las aplicaciones científicas que utilizan cómputo de altas prestaciones acrecientan estos requerimientos.

En muchos casos, el cuello de botella de los sistemas paralelos es la E/S a causa de las exigencias que debe afrontar [3]. La E/S Paralela es esencial para emparejar el avance de las arquitecturas de los procesadores y el rápido crecimiento de la capacidad computacional. Aunque la arquitectura jerárquica de memoria multinivel puede evitar grandes pérdidas de prestaciones debido a los retardos de acceso a disco, la capacidad de memoria es limitada. Además, como la capacidad computacional aumentará, la disponibilidad de memoria por core decrecerá, especialmente si la escala de los sistemas de HPC se proyecta a millones de cores o más. Varias simulaciones científicas y de ingeniería de áreas críticas de investigación, tales como la nanotecnología, astrofísica, clima y energía física están convirtiéndose en aplicaciones intensivas de datos. Para poder disminuir la brecha entre CPUs-E/S se deben identificar los factores que influyen en las prestaciones y proponer nuevas soluciones [4] [5].

En el área de tolerancia a fallas en sistemas de cómputo de alta prestaciones se puede notar la importancia de la unidad de E/S en las arquitecturas paralelas como un punto a mejorar para lograr cubrir las exigencias de las aplicaciones que utilizan HPC. Una manera de llevar a cabo este trabajo es utilizar técnicas de simulación para evaluar el efecto de los cambios de los factores con mayores influencias en las prestaciones del sistema de E/S paralelo.

Se puede disminuir la complejidad y la probabilidad de errores en la generación de sistemas híbridos desarrollando una simulación específica de éstos utilizando diferentes frameworks [6] [7] [8].

Las aplicaciones científicas con un uso intensivo de datos utilizan software de E/S paralelo para acceder a archivos. Contar con una herramienta que permita predecir el comportamiento de este tipo de aplicaciones en HPC es de gran utilidad para los desarrolladores de aplicaciones paralelas. Por otro lado, ABMS ha sido utilizado para modelar problemas y sistemas complejos en diversas áreas de la ciencia.

Evaluar las prestaciones del subsistema de E/S con diferentes configuraciones y la misma aplicación, permite adaptar la configuración de E/S teniendo en cuenta el patrón de acceso de la aplicación. Pero también puede ser una gran ventaja analizar las necesidades de las aplicaciones antes de configurar el sistema físico. Una manera de predecir el comportamiento de las aplicaciones en el sistema de cómputo, ante distintas configuraciones, es utilizando técnicas de modelado y simulación.

Se está desarrollando modelos e implementando una simulación de la arquitectura de E/S paralela, por medio de técnicas de simulación basadas en agentes o Sistemas Multi-Agente, (MAS-MultiAgentSystems), para evaluar el efecto de dimensionar el sistema de E/S o cambiar componentes como la red de almacenamiento, dispositivos de E/S, entre otros [9].

Simulación de arquitecturas de Cloud Computing

CloudSim [10] [11] es un Framework de simulación generalizado y extensible que permite el modelado, y simulación de

diferentes infraestructuras y servicios de aplicaciones de Cloud Computing. Un ejemplo de utilización es la simulación de muchos centros de datos.

Su arquitectura consiste en entidades específicas que se representan como clases Java que pueden ser heredadas o instanciadas. Estas clases representan centros de datos, hosts físicos, máquinas virtuales, servicios a ejecutar en los centros de datos y servicios en la nube de usuarios [12] [13]. Además, CloudSim soporta la inserción dinámica de los elementos de simulación y proporciona aplicaciones de paso de mensajes y la topología de la red del centro de datos.

La versatilidad de CloudSim es la principal ventaja del sistema. La integración de nuevos parámetros y conceptos de la simulación es implementada desde abstracciones preestablecidas convenientemente por los autores. Las abstracciones principales son SimEvent [14], SimEntity [15], DataCenterCharacteristics y Vm.

El aporte de un desarrollo de nuevos actores al componente Vm (Virtual Machine) posibilita que por medio de simulación se obtengan métricas de entrada/salida. Las estadísticas que aporta la nueva implementación de Vm dan soporte a un espacio de memoria ram en tiempos de simulación. Dicho espacio de memoria principal está controlado por otro componente que es capaz de procesar instrucciones guardadas en el espacio de memoria sintético denominado RamEntity. Si las instrucciones se guardan lógicamente, la nueva versión de Vm es capaz de administrar procesos en la nueva capa de ejecución [16]. La idea general de la implementación consiste en la creación de una nueva SimEntity y sus agregaciones necesarias para manipular el espacio de memoria proporcionada por RamEntity, en tiempos de simulación. De tal forma que cada celda de RamEntity no sólo guarda datos, sino que también es posible que mantenga objetos interpretables como eventos para la nueva SimEntity que se denomina InitEntity.

Lo destacable es que el entorno de la cola de procesos es en el espacio de memoria de RamEntity y que cada instrucción que compone a un código objeto del proceso es en realidad un evento con todos sus parámetros.

Simulación y Salud

Al analizar las necesidades presentes en los servicios de salud, se obtiene que dentro de una sala de urgencias se encuentre una gran diversidad de escenarios posibles donde cada uno de estos puede afectar a resultados sensibles como por ejemplo, la tasa de mortalidad de personas. Para solucionar este problema se desarrolló un simulador con el objetivo de ser una herramienta capaz de recrear una gran cantidad de escenarios y así poder tomar decisiones rápidas.

El simulador desarrollado se centra en la prevalencia puntual de infecciones intrahospitalarias [17] en una sala de urgencias y cómo la afectan distintos factores relacionados con la gestión hospitalaria.

Para llevar a cabo el modelado del simulador se utilizó el paradigma de Modelado y Simulación basado en Agentes (ABMS). El framework utilizado es Repast Simphony [18], una herramienta especializada en ABMS, la cual provee una serie de ventajas respecto al modelado e implementación de agentes, así como también la coordinación concurrente de los agentes.

De esta manera, se clasificaron distintos agentes intervinientes en el ambiente de una sala de urgencias, como los pacientes y médicos, entre otros. Dichos agentes definen su comportamiento mediante máquinas de estado, las cuales determinan las acciones correspondientes tanto a la atención hospitalaria como también al estado de salud respecto de una enfermedad intrahospitalaria específica.

La propagación de la infección intrahospitalaria estará modelada mediante interacciones entre estos agentes, por ejemplo, uno de los focos de infección más importante de las salas de urgencias son las salas de espera. La forma más efectiva para calibrar el simulador se da mediante el grado de interacción de agentes, ajustando así la tasa de transmisión de la enfermedad en cuestión.

Una vez calibrado el simulador, se pueden obtener resultados. El trabajador de la salud dispone de distintos datos para configurar la simulación, como por ejemplo, la cantidad de

pacientes, la cantidad de camas disponibles, etc. Al realizar distintas ejecucioness, puede obtener resultados analizando distintos posibles cuellos de botella, configurando la cantidad de médicos clínicos disponibles, la cantidad de recepcionistas encargados de la admisión, el triage, entre otros. De esta manera, se pueden tomar decisiones respecto a la asignación de recursos y personal para agilizar la estancia hospitalaria de los pacientes y evitar los contagios producidos por sus interacciones.

Líneas de Investigación, Desarrollo e Innovación

Temas de Estudio e Investigación

- Arquitecturas multiprocesador para procesamiento paralelo: multiprocesador de memoria compartida, multiprocesador on-chip de memoria distribuida. Multicore, Clusters, Clusters de multicore. Grid. Cloud.
- Arquitectura de E/S paralela considerando el software, hardware, comunicaciones entre módulos y dispositivos de almacenamiento.
- Nuevos aportes de desarrollos que mejoren los modelos de simulaciones con CloudSim para el análisis de la performance en sistemas de arquitecturas de software de Cloud Computing.
- Modelado y simulación para la administración de sistemas de salud.

Resultados y Objetivos

Investigación experimental

- Diseño y desarrollo de modelos mediante técnicas de Modelado y simulación basada en agentes (ABMS) para analizar el comportamiento de las distintas capas de la pila de software de E/S.
- Utilización de agentes para generar la funcionalidad de los elementos físicos (procesadores, memoria, buses, drivers, entre otros) como así también de las interfaces en las arquitecturas de E/S.
- Análisis, modelado e implementación de las operaciones típicas de E/S: read, write, open, close, flush.
- Análisis y modelado de librerías de archivos para aplicaciones que utilizan cómputo de altas prestaciones-HPC. Se ha utilizado Amazon Web Services para creación de cluster virtuales y obtener métricas de la pila de software de E/S.
- Incorporación de tiempos de entrenamiento, obtenidos en AWS, en el simulador. Con esto se logra una salida más detallada y un método para validar tiempos y métricas del simulador con AWS.
- Implementación de comandos para ejecutar desde command center en NetLogo. Con esto se logró sintetizar el benchmark IOR correspondiente a la capa de aplicación de la pila de E/S, logrando introducir nuevos parámetros como tamaño de archivo y cantidad de nodos de E/S (metadata server y data server). De ejecutar este comando, se obtiene una nueva salida similar a la del benchmark IOR.
- Desarrollo de pruebas de conceptos con el modelo inicial del sistema de archivos paralelos utilizando el framework NetLogo.
- Obtención de un método de desarrollo de nuevos actores genéricos CloudSim que mejoran el modelado y la producción de estadísticas virtuales.
- Implementación de soportes de memoria principal en tiempos de simulación que se activan como componentes de las máquinas virtuales.
- Implementación de la entidad InitEntity que procesa instrucciones en el espacio de memoria de las máquinas virtuales. Se vinculan exitosamente los tiempos de procesamiento de un cloudlet y las instrucciones en las máquinas virtuales.
- El objetivo a mediano plazo es el contraste con las estadísticas de un despliegue de hardware que proporcione estadísticas de un sistema cloud computing privado.
- Modelado de infraestructuras de sistemas de salud.
- Implementación de un simulador para analizar el contagio de enfermedades intrahospitalarias.
- Paralelización en la ejecución de simuladores.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D se participa en el dictado de la carrera de Ingeniería en Informática de la UNAJ. También aportan trabajos de alumnos de las materias Redes de Computadoras 2 y Programación en Tiempo Real.

Durante 2019 se han realizado 6 publicaciones nacionales e internacionales. Además, se encuentran en desarrollo y concluidas varias Prácticas Profesionales Supervisadas (PPS) con las que concluyen sus estudios los alumnos de Ingeniería en Informática.

En esta línea de I/D existe cooperación a nivel nacional e internacional. Hay un investigador realizando su Doctorado, 1 becario de grado (EVC CIN) y 2 alumnos avanzados de grado colaborando en las tareas.

Referencias

1. Grama A, Gupta A, Karypis G, Kumar V. "Introduction to parallel computing". SecondEdition. Pearson Addison Wesley, 2003.
2. R. Galeano, C. Villalba, D. Rexachs, E. Luque. Agent-Based Model to Simulate Outpatient's Consultations at the "Hospital de Clínicas". The Eighth International Conference on Advances in System Simulation (SIMUL 2016). 1:46-51
3. H Hennessy, J. L., Patterson, and D. A., Computer Architecture, Fourth Edition: A Quantitative Approach. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2006.
4. J. M. May, Parallel I/O for high performance computing. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001.
5. V. Balaji, Earth system modelling – Volume 4. IO and Postprocessing. Springer, 2013.
6. D. Encinas, Utilización de un reloj global para el modelado de un ambiente simulado distribuido. XVIII Congreso Argentino de Ciencias de la Computación. 2012
7. D. Encinas, Simulación de una red CAN para dimensionar las comunicaciones de una IMU. VII Congreso Argentino de Tecnología Espacial. 2013.
8. D. Black, SystemC: From the Ground Up. Second Edition, Springer, 2010.
9. D. Encinas et al.,Modeling I/O System in HPC: An ABMS Approach. The Seventh International Conference on Advances in System Simulation (SIMUL), ISBN: 978-1-61208-442-8, 2015.
10. R. Calheiros, R. Ranjan, A. Beloglazov, C. De Rose and R. Buyya "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms" Published online 24 August 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.995.
11. http://www.cloudbus.org/cloudsim 2018.
12. Hamza Ouarnoughi, Jalil Boukhobza, Frank Singhoff, Stephane Rubini, Erwann Kassis. "Considering I/O Processing in CloudSim for Performance and Energy Evaluation". OpenStack Cloud Software: Open source software for building private and public clouds. © Springer International Publishing AG 2016 M. Taufer et al. (Eds.): ISC High Performance Workshops 2016, LNCS 9945, pp. 591–603, 2016. DOI: 10.1007/978-3-319-46079-6 40.
13. Kushang Parikh, Nagesh Hawanna, Haleema. P.K, Jayasubalakshmi.R and N.Ch.S.N.Iyengar. School of Computing Science and engineering Vellore Institute of Technology, Tamil Nadu,"Virtual Machine Allocation Policy in Cloud Computing Using CloudSim in Java." 2015.
14. www.icsa.inf.ed.ac.uk/research/groups/hase/simjava 2018.
15. F. Howell, R Mc Nab. A discrete event simulation library for java. International Conference on Web-Based Modeling and Simulation. 1998.
16. D. Rosatto, R. Bond, M. Belizán, M. Morales, D. Encinas. Modelado y simulación de arquitecturas de Cloud Computing con CloudSim: comunicación entre entidades. XXIII Congreso Argentino de Ciencias de la Computación. ISBN 978-950-34-1539-9. 2017
17. Prevalencia puntual de infección nosocomial, disponible en internet: https://www.researchgate.net/publication/242363760_Prevalencia_puntual_de_infeccion_nosocomial. Fecha: 29/03/2020
18. Repast Simphony Frequently Asked Questions, disponible en internet: https://repast.github.io/docs/RepastFAQ/RepastFAQ.html. Fecha: 29/03/2020

Rendimiento de Cloud Computing para HPC en IaaS privados y públicos.

Brian Galarza[1], Gonzalo Zaccardi[1], Román Bond[1], Federico Montes de Oca[1], Eduardo Maxit, Jorge Osio[1], David Duarte[1], Martin Morales[1,2], Diego Encinas[1,3]

[1]Proyecto de Investigación SimHPC - Programa TICAPPS - Instituto de Ingeniería y Agronomía - UNAJ

[2]Unidad CodApli - Facultad Regional La Plata - UTN

[3]Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP – Centro Asociado CIC

bgalarza@unaj.edu.ar, gzaccardi@unaj.edu.ar, rbond@unaj.edu.ar, federicomdo97@gmail.com, eduardomaxit5@gmail.com, josio@unaj.edu.ar, davito.duarte.22@gmail.com, martin.morales@unaj.edu.ar, dencinas@unaj.edu.ar

Resumen

El objetivo de esta línea de investigación es el estudio del performance de las arquitecturas tipo cloud a través del despliegue de IaaS y utilización de IaaS públicos, en particular en el área de cómputo paralelo de altas prestaciones (HPC). Enfocando a la obtención de herramientas que permitan predecir la eficiencia del sistema ante posibles escenarios. Analizando los diferentes componentes del sistema que pueden influir en las prestaciones significativamente y pueden llegar a modelarse y/o configurarse.

Palabras clave: *Arquitecturas Multiprocesador. Cloud Computing. OpenStack. Sistemas de Archivos en clústers. Redes definidas por Software.*

Contexto

Se presenta una línea de Investigación que es parte del Proyecto de Investigación “Simulación y tecnología en Cómputo de Altas Prestaciones (High Performance Computing, HPC) para aplicaciones de interés social” - SimHPCde la Universidad Nacional Arturo Jauretche (UNAJ), acreditado por resolución interna 148/18. Además, el proyecto aporta al Programa “Tecnologías de la información y la comunicación (TIC) en aplicaciones de interés social” – TICAPPS de la UNAJ.

Introducción

Cloud Computing es un paradigma que ha estado en constante crecimiento, cada vez más compañías y grupos de investigación trabajan en conjunto con el fin de explotar las oportunidades ofrecidas por el mismo [1]. Dicho paradigma ofrece muchas ventajas, tales como el bajo costo de implementación, ya que no se necesitan computadoras de

última tecnología debido a que éstas trabajan conjuntamente (Clustering) con la posibilidad de escalar horizontalmente de manera sencilla. Además, hay software Open Source disponible para los nodos en el clúster como las infraestructuras Eucalyptus, OpenNebula, CloudStack u OpenStack integradas con GNU/Linux y compatibles, por ejemplo, con Amazon WebServices.

Despliegue de IaaS

Las comunicaciones en Cloud Computing son una parte fundamental del paradigma que consisten en utilizar distintos nodos y lograr hacerlos funcionar conjuntamente.

Para lograr una comunicación sincronizada entre estos nodos se propone utilizar OpenStack [2] como también OpenNebula [3].

OpenStack es una plataforma de tecnología open source que utiliza recursos virtuales agrupados para diseñar y gestionar nubes privadas y públicas a través de múltiples servicios que, de manera coordinada, cumplen diferentes propósitos para lograr el correcto funcionamiento de, por ejemplo, una ”Infraestructure as a service” (IaaS). Algunos de los servicios ofrecidos por OpenStack son: hypervisor (Nova), autentificación (Keystone), Imágenes (Glance), Dashboard (Horizon), Networking (Neutron) y block storage (Cinder). Según las necesidades se pueden requerir de ciertos servicios u otros. La Arquitectura básicamente consiste en dos tipos de nodos: “Compute Node” y “Controller Node”. Se llaman Compute Node a todos aquellos que se encargan del procesamiento de servicios específicos mientras que Controller Node es aquel que comunica a cada uno de los anteriores [4] [5] [6].

Fuel es una herramienta open source desarrollada por Mirantis en la cual se ejecuta un script que permite configurar, de manera más amigable respecto a OpenStack, los recursos que se desean otorgar a la infraestructura, como la cantidad de nodos, los núcleos de procesador, la memoria RAM, entre otros [7].

Fuel trabaja con un nodo master el cual es el encargado de controlar a los nodos slaves que contendrán la infraestructura OpenStack. Es decir, desde el nodo Fuel Master se indican qué paquetes se van a instalar en cada nodo slave (Glance, Nova-Compute, Keystone, etc.) para luego en los slaves tener armados los nodos compute y controller, sin necesidad de realizar configuraciones manuales en cada uno de los mismos.

OpenNebula es un software de código abierto que permite el despliegue de IaaS. Busca reducir la complejidad generada por OpenStack y ofrece soporte con hypervisores tales como KVMy VMware vCenter.

OpenNebula clasifica a los nodos en dos tipos, Front – end los cuales entran en contacto con los usuarios y a su vez se comunican con los nodos de la infraestructura en los cuales se lanzaran las instancias y los nodos virtualizados los que a su vez deben contar con los paquetes correspondientes de storage, autenticación y networking para poder funcionar correctamente.

La implementación de estas infraestructuras ofrece ventajas en las cuales los clústeres virtualizados trabajan en conjunto ofreciendo un buen rendimiento a bajos costos y con posibilidad de escalabilidad al poder agregar mayor cantidad de nodos para procesamiento de manera sencilla.

Sistemas de Archivos Paralelos en clústeres

Amazon Webservices [8], mediante el servicio EC2 (Amazon Elastic Compute Cloud) permite desplegar clústeres virtuales mediante instancias de VMs y almacenamiento para las mismas. Mediante este servicio y el correspondiente clúster conformado, se utiliza un sistema de archivos paralelo (PVFS2) que permite la gestión de datos particionados y distribuidos en los distintos nodos, mediante múltiples tareas de una aplicación ejecutada sobre el clúster.

PVFS2 utiliza una estructura cliente-servidor. En función de su rol, existen tres tipos de nodos dentro de un clúster con este sistema de archivos: servidores de datos, servidores de metadatos y clientes, en donde cada uno de los nodos puede cumplir los tres roles.

PVFS2 [9] contiene una herramienta interna que permite volcar información en logs durante la ejecución de tareas dentro del sistema de archivos. Por cada uno de los nodos, el administrador del clúster puede obtener información de depuración (GOSSIP) conformado por registros de debug, de acceso, contadores de rendimiento y errores producidos durante la ejecución. Referido a los contadores de rendimiento, se obtienen mediante el software de monitoreo Atop. Esta herramienta permite obtener reportes de la actividad de los procesos y la utilización de los diferentes recursos del sistema (memoria, disco, CPU, red, etc) [10].

Asimismo, es posible especificar los datos a obtener en función de las distintas capas de PVFS2, del rol específico de cada nodo (cliente-servidor), así como también de otros factores relacionados con el funcionamiento interno y operaciones asociadas a la gestión de archivos.

Teniendo conocimiento acerca del funcionamiento del código fuente correspondiente a PVFS2 y del modo en el que realiza la escritura de registros en los logs, resulta factible realizar modificaciones que permitan obtener otro tipo de parámetros de interés a partir de esta herramienta.

Finalmente se propone el análisis, uso y configuración de distintas herramientas no invasivas para determinar la performance del sistema de archivos en clústeres virtuales.

Redes definidas por Software

Las redes definidas por software (SDN) son un paradigma de gestión y administración de redes por medio de software que permiten tener un control más flexible con respecto al control del tráfico de datos por medio de hardware ya que permite cambiar en tiempo real las normas y políticas establecidas en la red.

Mininet[11] es un emulador de redes SDN open source que permite generar tráfico artificial entre nodos virtuales de la red.

Se ha llevado a cabo el montaje de un laboratorio SDN utilizando Mininet junto con Amazon WebServices con el fin de analizar el comportamiento de este nuevo paradigma de redes de computadoras en un entorno de Cloud Computing.[12]

Líneas de Investigación, Desarrollo e Innovación

Temas de Estudio e Investigación

- Arquitecturas multiprocesador para procesamiento paralelo: multiprocesador de memoria

compartida, multiprocesador on-chip de memoria distribuida. Multicore, Clusters, Clusters de multicore. Grid. Cloud.
- Plataformas de software para implementar y administrar Clouds públicos, privados e híbridos.
- Sistemas de Archivos Paralelos.

Resultados y Objetivos

Investigación experimental

- Implementación de un IaaS encargado de realizar operaciones en procesamiento paralelo aumentando la eficiencia y reduciendo los costes generados.

- Implementación de OpenStackDashboard y de un sistema desarrollado para poder controlar/administrar de manera visual (web) y más básica cada uno de los servicios.

- Implementación de OpenNebula en un sistema con las mismas características que el implementado por OpenStack con el fin de poder realizar pruebas en entornos similares.

- Utilización de Fuel para administrar OpenStack como sistema de administración de nube (Cloud Computing) a partir de la infraestructura de 2 nodos compute y e1 controller [13].

- Lograr escalabilidad agregando nuevos nodos compute a la infraestructura obteniendo un mayor performance en el sistema.

- Análisis del rendimiento de un Cloud privado en la ejecución de instancias personalizadas.

- Ejecutar sobre el modelo de nodos implementado en OpenStack la distribución GNU/Linux Hetnux, desarrollada en la UNAJ en el marco del programa “Universidad, Diseño y Desarrollo Productivo 2014”.

- Ejecutar diferentes benchmarks en la infraestructura desplegada sobre OpenStack y OpenNebula para así comparar los resultados obtenidos de ambas infraestructuras y poder realizar un análisis del rendimiento en cada caso.

- Análisis y configuración de clústeres virtuales.

- Análisis y configuración de herramientas no invasivas para la obtención de métricas en las distintas capas de software de los sistemas de archivos paralelos.

Formación de Recursos Humanos

Dentro de la temática de la línea de I/D se participa en el dictado de la carrera de Ingeniería Informática de la UNAJ. También aportan trabajos de alumnos de las materias Sistemas Operativos 1, Redes de Computadoras 2, Programación en Tiempo Real y Organización y Arquitecturas de Computadoras.

Durante 2019 se han realizado 6 publicaciones nacionales. Además, se encuentran en desarrollo y concluidas varias Prácticas Profesionales Supervisadas (PPS) con las que concluyen sus estudios los alumnos de Ingeniería en Informática.

En esta línea de I/D existe cooperación a nivel nacional. Hay 2 investigadores realizando carreras de postgrado, 1 becario de grado (EVC CIN) y 2 alumnos avanzados de grado colaborando en las tareas.

Referencias

1. Kondo, D., Javadi, B., Malecot, P., Cappello, F., Anderson, D. P.: "Cost-benefit analysis of Cloud Computing versus desktop grids". In: IPDPS '09 Proceedings. IEEE International Symposium on Parallel and Distributed Processing. Washington, USA (2009).
2. OpenStack Cloud Software: Open source software for building private and public clouds. http://www.openstack.org. Febrero 2015.
3. OpenNebula. https://opennebula.org/. Febrero 2019
4. Galarza, B.; Tuamá, C.; Zaccardi, G.; Encinas, D.; Morales, M. "Implementaciones de Cloud Computing y aplicaciones en el ámbito universitario". 1° Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2013). Ciudad de Córdoba, Argentina.
5. Zaccardi, G.; Galarza, B.; Encinas, D.; Morales, M. "Implementación de Cloud Computing utilizando OpenStack". 2° Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2014). Ciudad de San Luis, Argentina.
6. Galarza, B.; Zaccardi, G.; Encinas, D.; Morales, M. "Análisis de despliegue de una IaaS utilizando Openstack". XXI Congreso Argentino de Ciencias de la Computación (CACIC 2015). Ciudad de Junín, Argentina.
7. OpenStack Deployment Fuel. https://www.mirantis.com/products/mirantis-openstack-software/openstack-deployment-fuel/. Febrero 2016
8. Amazon Web Services (AWS)-Cloud Computing Services. https://aws.amazon.com Marzo 2019
9. T. PVFS2, "PVFS 2 File System Semantics Document," tech. rep., PVFS Development Team, 2015
10. Atop Tool. https://www.atoptool.nl/index.php
11. Mininet. http://mininet.org
12. Montes de Oca, F.; Galarza, B.; Morales, M.; Encinas, D. "Redes Definidas por Software en Entorno de Cloud Computing". 6° Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2018). Mar del Plata. Argentina.
13. Zaccardi, G.; Galarza, B.; Morales, M.;Encinas, D. "Despliegue y ejecución de un cloud privado". 4°Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2016). Ciudad de Salta, Argentina.

Procesamiento de datos masivos en tiempo real y consumo energético de sistemas paralelos

Javier Balladini[1], Marina Morán[1], Claudia Rozas[1], Rodrigo Cañibano[1], Rafael Zurita[1], Belén Casanova[1], Cristina Orlandi[2], Armando De Giusti[3], Remo Suppi[4], Dolores Rexachs[4], Emilio Luque[4]

[1] Facultad de Informática, Universidad Nacional del Comahue
{javier.balladini, marina, claudia.rozas, rcanibano, rafa}@fi.uncoma.edu.ar;
{mb.casanova.retamal}@gmail.com
[2] Hospital Francisco Lopez Lima
orlandi.mariacristina@gmail.com
[3] Instituto de Investigación en Informática LIDI, Universidad Nacional de La Plata
degiusti@lidi.info.unlp.edu.ar
[4] Departamento de Arquitectura de Computadores y Sistemas Operativos,
Universidad Autónoma de Barcelona
{remo.suppi, dolores.rexachs, emilio.luque}@uab.es

Resumen

Los avances tecnológicos de los sistemas de cómputo paralelo y distribuido permiten el desarrollo de aplicaciones antes impensadas. Una de nuestras líneas de investigación se enfoca en aplicar estas tecnologías a Unidades de Cuidados Intensivos y Unidades de Vigilancia Intermedia. Buscamos mejorar sustancialmente el rendimiento de ellas con el procesamiento en tiempo real de datos masivos generados por el equipamiento médico y otras fuentes. Adicionalmente, trabajamos en la reducción del consumo energético de los sistemas de computación de altas prestaciones, con especial atención en los mecanismos de tolerancia a fallos. Todas nuestras investigaciones se centran en desarrollar metodologías, modelos y soluciones informáticas para colaborar en la resolución de problemas que tengan una alta demanda computacional e impacto social. Los trabajos se desarrollan en colaboración con otras universidades, y un hospital público de Argentina. La formación de recursos humanos en estas líneas está orientada al nivel de grado, maestría y doctoral.

Palabras claves: computación de altas prestaciones, eficiencia energética, big data, aplicaciones para la salud.

1. Contexto

Las líneas de investigación aquí presentadas están enmarcadas dentro del proyecto de investigación 04/F013 "Aplicaciones de Cómputo Intensivo con Impacto Social", financiado por la Universidad Nacional del Comahue (UNComa), con inicio el 01/01/2017 y finalización el 31/12/2020, y acreditado por el Ministerio de Educación de Argentina.

Uno de los dos ejes de nuestra investigación, las aplicaciones para la salud, se desarrolla en colaboración con la Unidad de Terapia Intensiva, perteneciente al "Hospital Francisco López Lima" de la ciudad de General Roca, provincia de Río Negro. Integramos el comité de Inteligencia Artificial y Big Data, creado en el presente año, de la Sociedad Argentina de Terapia Intensiva (SATI).

Respecto al eje relacionado con el consumo energético de los sistemas de cómputo paralelo, se desarrolla en colaboración con el Instituto de Investigación en Informática LIDI de la

Universidad Nacional de La Plata (UNLP), y el grupo de investigación HPC4EAS (High Performance Computing for Efficient Applications and Simulation) de la Universidad Autónoma de Barcelona (UAB) de España.

2. Introducción

La masificación de las tecnologías de cómputo paralelo hacen que ellas sean cada vez más accesibles, y se pueda pensar en el desarrollo de nuevas aplicaciones. El sector de la salud es un campo que puede ser altamente beneficiado de estos avances tecnológicos. En especial, las Unidades de Cuidados Intensivos (UCIs) pueden mejorar sustancialmente su rendimiento con el uso de Big Data. A su vez, una de las mayores problemáticas que afectan a los sistemas de cómputo paralelo de altas prestaciones (sobre todo a los de grandes dimensiones) es el elevado consumo energético y potencia eléctrica que demandan. Es por esta razón que también enfocamos nuestra investigación en la gestión del consumo energético de estos sistemas, para reducir su impacto económico, medioambiental y social. A continuación se introduce la motivación y problemática que direccionan ambas temáticas de investigación:

Procesamiento en tiempo real de datos masivos en salud

En los centros de salud, las UCI proveen cuidados continuos y rigurosos para personas adultas que están críticamente enfermas y que pueden beneficiarse de tratamiento, y dan un buen morir a los pacientes irrecuperables. En una UCI, los datos de los pacientes involucran datos clínicos de baja frecuencia, y flujos de datos fisiológicos de alta frecuencia generados por el equipamiento médico (como monitores médicos de signos vitales, respiradores artificiales, etc.). En una UCI típica, los enfermeros completan manualmente datos en formularios, registrando datos clínicos y fisiológicos. Los datos fisiológicos se obtienen por observación de las pantallas del equipamiento médico (ubicados a los lados de las camas) a intervalos de tiempo (horas o fracciones) que definen los médicos para cada paciente. El equipamiento médico emite alertas cuando hay riesgo en la salud del paciente basándose en las mediciones de ciertos parámetros. Luego, los médicos analizan exhaustivamente los datos de los formularios y dan a los enfermeros las indicaciones de tratamientos a realizar a los pacientes.

Los principales problemas que ocurren en las UCI típicas son:

1. La metodología actual de gestión de la información es propensa a errores humanos.
   Causas:
   - Podrían cometerse errores en el registro manual, que realizan los enfermeros, de datos observados en los monitores de signos vitales u otro equipamiento médico.
   - El análisis realizado por los médicos, que interrelaciona una gran cantidad de variables, provenientes de enormes volúmenes de información de diferente tipo, origen y soporte (datos fisiológicos, radiografías, análisis de laboratorio, datos de observación del paciente, etc.) es una tarea sumamente compleja. Ella, sin asistencia informática alguna, es difícil de concretar sin cometer errores.

   Efectos: Posibles diagnósticos erróneos o información inconsistente que implica un esfuerzo adicional del personal para detectar su origen.

2. Detección tardía del deterioro de la salud de los pacientes.
   Causas: Desde que los datos se toman por los enfermeros hasta que son analizados por los médicos, puede transcurrir un tiempo prolongado.
   Efectos: Incidencia negativa en la salud de los pacientes, que puede impactar en:
   - Aumento del riesgo de muerte.
   - Aumento de los costos económicos de los tratamientos.
   - Estadías más largas, y por ende la atención de menos pacientes.

3. Pérdida de información de pacientes.

Causas: Los datos observados en el equipamiento médico son registrados en formularios a intervalos de horas (o varios minutos), perdiendo todas las muestras intermedias (por ejemplo, el monitor de signos vitales actualiza la frecuencia cardíaca una vez por segundo).
Efectos:
- Disminución de la precisión de los diagnósticos.
- Imposibilidad de registrar el conjunto completo de todos los datos medidos por el equipamiento médico para uso futuro. Disponer de un gran conjunto de datos históricos permite la extracción de nuevo conocimiento, útil para futuras investigaciones médicas y el descubrimiento de patrones para predicción del patologías. La predicción de patologías, como sepsis (una de las principales causas de muerte en UCIs), permitiría comenzar los tratamientos más temprano, reduciendo el riesgo de muerte, el costo económico de los tratamientos, y las estadías de los pacientes.

4. Ausencia de personal médico especialista: En la mayoría de las UCI del país, no hay médicos intensivistas durante las 24hs. Las Unidades de Vigilancia Intermedia (UVIs) no cuentan con personal médico permanente.

Un gran avance sería disponer de un sistema de cómputo que detecte el deterioro en la salud de los pacientes en tiempo real, basándose en múltiples parámetros de diferentes orígenes. La detección temprana del deterioro permitirá aumentar la efectividad de los tratamientos, y por consiguiente la mejora de la salud de los pacientes, la reducción del costo económico de los tratamientos, y el aumento del número de pacientes atendidos. Ante la ausencia física de médicos en la sala del paciente, el sistema permitiría alertar al personal de enfermería y a médicos intensivistas remotos. Ellos, los médicos intensivistas, podrían observar el estado del paciente de manera remota.

Hay múltiples problemas o desafíos que deben enfrentarse para la construcción de este sistema, y que creemos viable con la aplicación de técnicas de computación paralela y distribuida, y de los sistemas embebidos. No hay muchos sistemas de este tipo, algunos de ellos se encuentran en etapa experimental inicial y otros ya llevan algunos años de investigación. La información disponible de estos sistemas es normalmente escasa por tratarse mayormente de software privativo. En la bibliografía se encuentran algunos trabajos como [9, 2, 1].

Consumo energético de sistemas de cómputo paralelo

Mientras el rendimiento de los sistemas de computación de altas prestaciones (HPC, *High Performance Computing*) continúa creciendo, las máquinas aumentan significativamente la cantidad de unidades de procesamiento. Este aumento en el número de componentes hace disminuir la confiabilidad y aumentar el consumo energético de un sistema de cómputo. El consumo energético se ha identificado como uno de los mayores desafíos a enfrentar en la era exaescala [12, 8].

El consumo energético es hoy en día un gran problema. Para dar una idea de la magnitud del mismo, la máquina de mayores prestaciones de la actualidad, la estadounidense Summit, demanda casi 10 MW de potencia, lo mismo que se requiere para abastecer a los hogares de una ciudad con alrededor de 135.000 habitantes (cálculo realizado en base al consumo de un hogar en Argentina). Además del alto impacto económico, la generación de tanta energía podría tener un alto impacto medioambiental, por ejemplo, represas hidroeléctricas que modifican el ecosistema, y social, por ejemplo, la mayor fuente de energía mundial se obtiene del carbón, cuya extracción minera es altamente peligrosa.

La computación ecológica es el estudio y la práctica de la computación ambientalmente sostenible. Ella se ocupa de diferentes aspectos de los sistemas de cómputo: diseño, manufactura, eliminación, y uso. Este último aspecto, el uso ecológico, se refiere al uso de los sistemas de cómputo con conciencia ambiental. Es

posible reducir el consumo de energía de los sistemas de cómputo utilizando diferentes estrategias que deben ser consideradas a nivel del software, y consisten en realizar cambios en la configuración del sistema (escalado dinámico de frecuencia y tensión, hibernación de recursos) o en las aplicaciones (uso adecuado de la jerarquía de memoria, rediseño de algoritmos, planificación de tareas, y asignación de tareas a recursos hardware).

3. Líneas de investigación

El eje central de nuestra investigación es desarrollar metodologías, modelos y soluciones informáticas para colaborar en la resolución de problemas que tengan una alta demanda computacional e impacto social en los siguientes campos: aplicaciones para la salud, y el consumo energético de los sistemas de HPC.

Aplicaciones para la salud

Esta línea está enfocada en el diseño y desarrollo de un sistema para detección automática y temprana del deterioro de la salud de pacientes en UCIs. Se busca que el sistema, basándose en el conocimiento de expertos y guías clínicas, emita alertas que orienten en la detección de la afección del paciente y en posibles tratamientos a seguir. La construcción del sistema involucra diferentes desafíos tecnológicos relacionados fundamentalmente a: Extracción de datos del equipamiento médico; Procesamiento en tiempo real de grandes volúmenes de datos (Big Data) en tiempo real, generado por el equipamiento médico, especialmente las curvas como el electrocardiograma, la saturación de oxígeno en sangre, entre otros.

Consumo energético de sistemas de HPC

Nos centramos en el desarrollo de metodologías, modelos y construcción de software para administrar y gestionar el consumo de energía y prestaciones de sistemas de cómputo paralelo. Nuestro principal objetivo actual es la gestión energética en mecanismos de tolerancia a fallos. La tolerancia a fallos agrega una carga de trabajo significativa al sistema de cómputo, sobre todo en sistemas que tienen enormes cantidades de unidades de procesamiento [7], haciendo necesario gestionar el consumo energético de los distintos mecanismos.

4. Resultados y objetivos

Aplicaciones para la salud

En [6] presentamos un análisis del estado general de las UTI del hospital Francisco Lopez Lima, y la propuesta del diseño de alto nivel del sistema. En [3] presentamos el diseño de la infraestructura del sistema de procesamiento de reglas y un prototipo, y el desarrollo de un dispositivo embebido adquisidor de datos (actualmente obtiene información de la señal analógica del electrocardiograma desde un monitor médico) que envía datos a la plataforma de procesamiento por medio de una red de Wi-Fi . En [4] proponemos una herramienta para mejorar la prestación de servicios integrados de cuidados intensivos de la salud. Actualmente, estamos trabajando en: desarrollar hardware y software para extraer datos del equipamiento médico, y transmitirlos por WiFi a la plataforma de procesamiento de la información; desarrollar aplicaciones para el procesamiento eficiente de señales; detección de datos fisiológicos erroneos; optimización de la infraestructura de un sistema de Big Data de tiempo real para el procesamiento de reglas médicas; desarrollo de una aplicación para la interacción del sistema con médicos y enfermeros.

Consumo energético de sistemas de HPC

Los métodos de tolerancia a fallos tienen fuerte incidencia en el consumo energético de los sistemas de HPC, y resulta de suma importancia conocer, antes de ejecutar una cierta aplicación, el impacto que pueden producir los diferentes métodos y configuraciones del mismo. En [5], presentamos una metodología para

predecir el consumo energético producido por el método de checkpoint coordinado remoto, y en [11] expusimos un análisis de los factores que afectan el consumo energético de operaciones de checkpoint y restart en clusters. En [10] propusimos un modelo para estimar el consumo energético de operaciones de checkpoint y restart, y un método para su construcción. Actualmente, estamos construyendo un simulador para evaluar mecanismos de gestión de tolerancia a fallos que procuren un uso eficiente del cluster, permitiendo maximizar la productividad y minimizar el consumo energético.

5. Formación de recursos humanos

El equipo de trabajo local a la Universidad Nacional del Comahue, cuenta con un doctor en la temática (recibido en 2008). En Marzo de 2019 se ha presentado la tesis de grado "Una Infraestructura de Big Data de Tiempo Real para un Sistema de Detección del Deterioro de Pacientes en Terapia Intensiva", en la Universidad Nacional del Comahue. Se están iniciando dos trabajos de maestría y finalizando una tesis de grado en temas de aplicaciones para la salud. En el corriente año se espera finalizar una tesis doctoral en el tema de consumo energético de sistemas de HPC.

Referencias

[1] ehcos smarticu, http://www.ehcos.com/productos/eh-cos-smarticu/ (accedido en marzo de 2020).

[2] Excel medical, http://excel-medical.com/ (accedido en marzo de 2020).

[3] Javier Balladini, Pablo Bruno, Rafael Zurita, and Cristina Orlandi. An automatic and early detection of the deterioration of patients in intensive and intermediate care units. *Journal of Computer Science and Technology*, 18(03):e25, Dec. 2018.

[4] Javier Balladini, Pablo Bruno, Rafael Zurita, Cristina Orlandi, Remo Suppi, Dolores Rexachs, and Emilio Luque. A tool for improving the delivery of integrated intensive health care performance. *International Journal of Integrated Care*, 19(4), 07/2019 2019.

[5] Javier Balladini, Marina Morán, Dolores Rexachs, and Emilio Luque. Metodología para predecir el consumo energético de checkpoints en sistemas de hpc. *XX Congreso Argentino de Ciencias de la Computación (CACIC 2014)*, 2014.

[6] Javier Balladini, Claudia Rozas, Emmanuel Frati, Nestor Vicente, and Cristina Orlandi. Big data analytics in intensive care units: challenges and applicability in an argentinian hospital. *Computer Science and Technology (JCST)*, 2015.

[7] Franck Cappello, Al Geist, William Gropp, Sanjay Kale, Bill Kramer, and Marc Snir. Toward exascale resilience: 2014 update. *Supercomputing Frontiers and Innovations*, 1(1), 2014.

[8] Robert Lucas, James Ang, Keren Bergman, Shekhar Borkar, William Carlson, Laura Carrington, George Chiu, Robert Colwell, William Dally, Jack Dongarra, and et al. Doe advanced scientific computing advisory subcommittee (ascac) report: Top ten exascale research challenges. 2 2014.

[9] Carolyn McGregor. Big data in neonatal intensive care. *Computer*, 46(6):54–59, 2013.

[10] M. Morán, J. Balladini, D. Rexachs, and E. Luque. Prediction of energy consumption by checkpoint/restart in hpc. *IEEE Access*, 7:71791–71803, 2019.

[11] Marina Morán, Javier Balladini, Dolores Rexachs, and Emilio Luque. Factores que afectan el consumo energético de operaciones de checkpoint y restart en clusters. *XXIV Congreso Argentino de Ciencias de la Computación (CACIC 2018)*, 2018.

[12] John Shalf, Sudip Dosanjh, and John Morrison. Exascale computing technology challenges. In *Proceedings of the 9th International Conference on High Performance Computing for Computational Science*, VECPAR'10, pages 1–25, Berlin, Heidelberg, 2011. Springer-Verlag.

Modelo lattice-Boltzmann para flujo multifásico con transferencia de calor en GPU

Argañaras Pablo E.[1], Fogliatto Ezequiel O.[12], Coronel Thomás[2]
[1]Departamento Mecánica Computacional / CAB - CNEA
[2]Instituto Balseiro / Universidad Nacional de Cuyo
Av. Bustillo 9500 – San Carlos de Bariloche - Río Negro
banda@cab.cnea.gov.ar, fogliatto@cab.cnea.gov.ar, thomascoroneltuc@gmail.com

RESUMEN

En este trabajo se presentan las tareas de formación de recursos humanos y de investigación y desarrollo en la programación del modelo de lattice-Boltzmann para flujo multifásico con transferencia de calor en unidades de procesamiento gráfico (GPU) en el departamento Mecánica Computacional (Centro Atómico Bariloche) donde un estudiante del Instituto Balseiro lleva a cabo su proyecto integrador de grado para obtener el título de ingeniero mecánico. Desde el año 2019 un grupo de docentes investigadores del departamento Mecánica Computacional busca innovar en la resolución de problemas ingenieriles que emplean herramientas de cálculo numérico para modernizar los métodos empleados y maximizar el aprovechamiento de las técnicas y de las herramientas de hardware que se disponen. Las tareas involucradas permiten incorporar estudiantes de último año de ingeniería del Instituto Balseiro para formarse y generar la tesina de grado como resultado de su actividad en este proyecto. Se discutirán los recursos informáticos empleados, los ejes de investigación, los resultados obtenidos y futuros, y las producciones asociadas. Mostraremos que el uso de entornos de desarrollo colaborativos, en proyectos de investigación y desarrollo para la formación de nuevos profesionales, ayudan a compartir la realización de tareas, permitiendo incorporar nuevas formas de trabajo para docentes investigadores y estudiantes, con importantes consecuencias tales como la valoración del trabajo colaborativo que los alumnos podrán capitalizar luego en su carrera.

Palabras clave: *lattice-Boltzmann, gpu, cuda, Python, transferencia de calor.*

CONTEXTO

Este trabajo se presenta como una línea de investigación que se relaciona con los Proyectos de investigación PICT-201-0937 denominado “Uso de herramientas numéricas para el estudio y diseño de dispositivos termohidráulicos utilizados en la industria nuclear”, y PID SIIP Tipo 1 Bienal 2019 denominado “Métodos numéricos aplicados en problemas de ingeniería y de medicina nuclear”. Este último, cofinanciado entre el Instituto Balseiro (IB) de la Universidad Nacional de Cuyo y la Comisión Nacional de Energía Atómica a través del departamento Mecánica Computacional.

El presente conforma una primera parte del Trabajo Integrador de Thomás Coronel para obtener el grado de Ingeniero Mecánico en el Instituto Balseiro, con tema y director y codirector aprobados según Resolución C.A./IB N.º 109/19. El mismo se lleva a cabo en el departamento Mecánica Computacional (Mecom a partir de ahora) del Centro Atómico Bariloche, y más precisamente en la Gerencia de Investigación Aplicada dependiente de la Gerencia de Área de Aplicaciones de la Tecnología Nuclear.

Tanto la información empleada para redactar la oferta del proyecto integrador como los datos de salida empleados para la validación de la implementación propuesta, son contenidos proporcionados por la investigación del doctorando Ing. Mgter. Ezequiel Fogliatto. La originalidad del presente trabajo se encuentra en las actividades de formación y en el desarrollo de soluciones informáticas compatibles con múltiples plataformas buscando facilitar la realización de nuevos estudios y desarrollos que ayuden a multiplicar la producción científica y tecnológica en el área de la ingeniería nuclear y mecánica, en una primera instancia.

1. INTRODUCCIÓN

El método de lattice-Boltzmann se ha convertido en una estrategia eficiente y confiable para la resolución de un amplio rango de fenómenos y procesos asociados a la mecánica de fluidos computacional. A diferencia de otras técnicas numéricas convencionales, el método de lattice-Boltzmann permite encontrar la solución de ecuaciones de mecánica de fluidos sin involucrar la resolución de sistemas de ecuaciones, usando un operador convectivo lineal, y permitiendo la representación de condiciones de contorno complejas mediante reglas mecánicas elementales.

A pesar de los significativos avances logrados en la resolución precisa y eficiente de flujos multifásicos isotérmicos, la simulación de una ecuación de energía con esquemas lattice-Boltzmann aún continúa en desarrollo. Los principales inconvenientes asociados a la búsqueda de un esquema lattice-Boltzmann han sido documentados detalladamente en trabajos como los de Li y Luo (2014); Huang y Wu (2014); Fogliatto, Teruel y Clausse (2019), vinculados a la recuperación de una ecuación macroscópica para la temperatura con términos adicionales no deseados en las escalas de expansión relevantes. [1] [2]

Gracias a esta sencilla formulación es posible resolver problemas que resultan prohibitivos para otras metodologías, además de que la implementación computacional de modelos lattice-Boltzmann presenta un elevado nivel de escalabilidad en sistemas de procesamiento en paralelo, en particular en las unidades de procesamiento gráfico (GPUs).

Desarrollos recientes de modelos lattice-Boltzmann para la resolución de flujos multifásicos con transferencia de calor y cambio de fase demuestran que con esta metodología es posible resolver adecuadamente problemas asociados a fenómenos de ebullición y condensación con tiempos de cálculo aceptables, incluso con implementaciones sobre CPUs.

En este sentido, la resolución de los métodos antes mencionados usando procesamiento en GPUs permitiría reducir significativamente los tiempos de cálculo involucrados, logrando ampliar (con fines prácticos) la variedad de problemas que pueden ser resueltos con esta técnica.

En nuestra búsqueda como docentes investigadores tratamos de que las TIC conformen redes entre universidad y estudiantes que amplíen la transferencia mutua de conocimientos y la formación integral de ciudadanos. ”En suma, la apuesta por una sociedad del conocimiento en la educación universitaria del siglo XXI procura abrir espacio y tiempo para el desarrollo de aprendizajes y saberes no sólo intelectuales sino personales, sociales y culturales que permitan al estudiante aprender a conocer y conocerse en sus dimensiones personal, colectiva y global (terrestre)” [3], en donde “a partir de la adopción de una perspectiva constructivista del aprendizaje, se

comienza a concebir al alumno como sujeto activo en la conducción del proceso de construcción del conocimiento" [4].

Para llevar a cabo esta propuesta se incorpora el uso del modelo de programación paralela Cuda con el compilador nvcc del lenguaje C, del entorno de desarrollo integrado (IDE) y multiplataforma de código abierto Spyder para lenguaje Python, la herramienta generalizada de automatización de código Cmake, del sistema de gestión de proyectos y control de versiones de código online Github, para trabar en grupo porque "*compartir procesos cognitivos, a través del razonamiento y la argumentación, promotores de la construcción y reconstrucción de conocimientos científicos en grupos (...) puede ser una respuesta a los desafíos de la práctica docente universitaria.*" [5]

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta investigación se enmarca en los proyectos PICT-201-0937 "Uso de herramientas numéricas para el estudio y diseño de dispositivos termohidráulicos utilizados en la industria nuclear", y PID SIIP Tipo 1 Bienal 2019 "Métodos numéricos aplicados en problemas de ingeniería y de medicina nuclear". Este último, cofinanciado entre la Universidad Nacional de Cuyo a través del Instituto Balseiro (IB), y la Comisión Nacional de Energía Atómica a través del departamento Mecánica Computacional. La investigación y desarrollo involucra a estudiantes del Instituto Balseiro y a profesionales de Centro Atómico Bariloche.

Los ejes de la presente investigación son:

- el método de lattice-Boltzmann como simplificación de la resolución de problemas de mecánica de fluidos computacional,
- la paralelización de código para cálculo científico para ser procesado por placas gráficas,
- el uso de entornos de trabajo colaborativo en proyectos universitarios de formación, y en proyectos de investigación y desarrollo. Porque "explorar experiencias con TIC en educación superior, es siempre un aporte rico y prometedor para comprender y encontrar estrategias" [6] que mejoran el desempeño estudiantil.

3. RESULTADOS OBTENIDOS/ESPERADOS

Algunas de las interrogantes que guían esta investigación son: *¿Qué grado de complejidad encierra la capacitación de estudiantes en el área de programación paralela? ¿Cuánto contribuyen los entornos y las herramientas colaborativas para el desarrollo de software de cálculo? ¿Qué nivel de escalabilidad proporciona una implementación en GPU del modelo de lattice-Boltzmann?*

Los objetivos de la presente son:

- desarrollar una herramienta numérica que permita resolver modelos de lattice-Boltzmann para flujo multifásico con transferencia de calor en geometrías sencillas empleando unidades de procesamiento gráfico,
- validar la herramienta desarrollada,
- optimizar la herramienta numérica.

Nos enfocaremos en la capacitación de un estudiante de ingeniería en el uso de herramientas para desarrollo de software colaborativas, en su

incorporación a un grupo de trabajo conformado en investigación y desarrollo en el área de mecánica computacional, y en el desarrollo de un software para cálculo numérico que basado en el modelo de lattice-Boltzmann para la resolución de flujo multifásico con transferencia de calor, permite resolver problemas relacionados con fenómenos de ebullición y condensación.

Actualmente se está validando una implementación en cuda C con resultados de una implementación previa en C, y se están haciendo las mismas implementaciones para Python empleando Pycuda.

Se optimizaron funciones elementales del modelo lattice-Boltzmann en estudio.

En etapas posteriores se continuará con la optimización del software de cálculo numérico apuntando a mejorar el almacenamiento de estructuras de datos en la memoria, la generación de una versión compatible con el sistema operativo Linux y Windows, y se intentará proponer una interfaz gráfica amigable para la aplicación de cálculo que facilite su uso al usuario final.

Producción 2018/2020:

- Fogliatto, E. O., Clausse, A., & Teruel, F. E. (2019). Simulation of phase separation in a Van der Waals fluid under gravitational force with Lattice Boltzmann method. *International Journal of Numerical Methods for Heat & Fluid Flow*.
- Fogliatto, E. O., Teruel, F. E., y Clausse, A.. Transferencia de calor en flujo multifásico mediante el método de Lattice Boltzmann. XXIV Congreso sobre Métodos Numéricos y sus Aplicaciones. Santa Fe, 5-7 Noviembre 2019. In: Mecánica Computacional, vol. 37, no 22, Multiphase Flow and Transport in Porous Media and Microscale (A), p. 875-885 ISSN 2591-3522 (2019).
- Fogliatto, E. O., Teruel, F. E., y Clausse, A.. Modelado y simulación de flujo multifásico mediante el método de Lattice-Boltzmann. XII Congreso Argentino de Mecánica Computacional Numéricos y sus Aplicaciones. Tucumán, 6-9 Noviembre 2018. In: Mecánica Computacional, vol. 36, no 23, Multiphase Flow and Transport in Porous Media and Microscale, p. 875-885 ISSN 2591-3522 (2018).

4. FORMACIÓN DE RECURSOS HUMANOS

- El Mgter. Ezequiel O. Fogliatto se encuentra realizando el doctorado en Ingeniería en el Instituto Balseiro de la Universidad Nacional de Cuyo, posee categoría V como docente investigador por la Universidad Nacional de Cuyo y ejerce la docencia en el área de Termofluidos en Instituto Balseiro.
- El Ing. Pablo E. Argañaras realiza la maestría en Procesos Educativos Mediados por Tecnología del CEA en la Universidad Nacional de Córdoba, posee categoría IV como docente investigador por la Universidad Nacional de Río Negro en el Taller de Informática y TICs de la Sede Andina Bariloche, y es docente de Fundamentos de Informática en Ingeniería Mecánica en UTN FRBA Extensión Áulica Bariloche.
- Thomás Coronel lleva adelante su tesina de grado en el Instituto Balseiro en relación directa con las líneas de investigación presentadas.

- El equipo de trabajo realiza sus tareas de investigación y desarrollo en el departamento Mecom en Centro Atómico Bariloche.

5. BIBLIOGRAFIA

[1]. Fogliatto, E. O., Clausse, A., & Teruel, F. E. (2019). Simulation of phase separation in a Van der Waals fluid under gravitational force with Lattice Boltzmann method. *International Journal of Numerical Methods for Heat & Fluid Flow*.

[2]. Fogliatto, E. O., Teruel, F. E., y Clausse, A.. Transferencia de calor en flujo multifásico mediante el método de Lattice Boltzmann. *XXIV Congreso sobre Métodos Numéricos y sus Aplicaciones*. Santa Fe, 5-7 Noviembr 2019. In: Mecánica Computacional, vol. 37, no 22, Multiphase Flow and Transport in Porous Media and Microscale (A), p. 875-885 ISSN 2591-3522 (2019).

[3]. Fandiño Parra, Y.J.: La educación universitaria en el siglo XXI: de la sociedad de la información a la sociedad del conocimiento. In:Revista Iberoamericana de Educación, vol. 55, no 3, ISSN 1681-5653 (2011).

[4]. Dumrauf, A.G., Colinvaux, D., Cordero, S.: ¿Y si trabajan en grupo...? Interacciones entre alumnos, procesos sociales y cognitivos en clases universitarias de Física. In: Enseñanza de las Ciencias: revista de investigación y experiencias didácticas, vol. 20, no 3, p. 427-442 (2002).

[5]. Kuznik, F., Obrecht, C., Rusaouen, G., & Roux, J. J. (2010). LBM based flow simulation using GPU computing processor. *Computers & Mathematics with Applications*, *59*(7), 2380-2392.

[6]. Argañaras, P. E., Britos, P. V., & Verdun, N. (2014). Experiencias educativas en el uso de TIC en la educación superior en la sede andina de la Universidad Nacional de Río Negro. In *XVI Workshop de Investigadores en Ciencias de la Computación*.

[7]. Li, Q., & Luo, K. H. (2014). Thermodynamic consistency of the pseudopotential lattice Boltzmann model for simulating liquid–vapor flows. Applied Thermal Engineering, 72(1), 56-61.

[8]. Huang, R., & Wu, H. (2014). A modified multiple-relaxation-time lattice Boltzmann model for convection–diffusion equation. Journal of Computational Physics, 274, 50-63.

Seguridad Informática

Aplicación de redes neuronales profundas para la detección automática de Nombres de Dominio generados de manera algorítmica.

Carlos A. Catania, Jorge Guerra, Martin Marchetta, Gabriel Caffaratti, Lucia Cortez, Alfredo Rezinovsky y Franco Palau

Universidad Nacional de Cuyo, Facultad de Ingeniería, LABSIN
Campus Universitario, Mendoza, Argentina.

{harpo, martin.marchetta, gabriel.caffaratti, jorge.guerra,franco.palau, alfredo.rezinovsky, lucia.cortez}@ingenieria.uncuyo.edu.ar

RESUMEN

En el contexto de la seguridad de redes de datos, un nombre de dominio generado de manera algorítmica (DGA, de sus siglas en inglés) es utilizado por el software malicioso (malware) para generar de manera dinámica un gran número de nombres de dominios de manera pseudo aleatoria, y luego utilizar un subconjunto de estos como parte del canal de Comando y Control (C&C). Dada la simplicidad y rapidez con la que los nuevos dominios son generados, las estrategias basadas en listas de dominios estáticas resultan inefectivas. Es por ello que resulta importante el desarrollo técnicas de detección automática que permitan encontrar los patrones comunes en los dominios generados. El presente proyecto propone el desarrollo de algoritmos de detección de DGA mediante la utilización de algoritmos de aprendizaje de máquinas en general y las redes neuronales profundas en particular. Se espera que la aplicación de redes neuronales profundas para el aprendizaje de los patrones comunes a los DGA permita desarrollar herramientas de detección no solo con una baja tasa de falsos positivos sino también con la capacidad de operar en tiempo real. Esto último resulta fundamental para lidiar con las amenazas de seguridad de hoy.

Palabras Claves: Seguridad de Redes, Detección de Anomalías, Aprendizaje Automático

CONTEXTO

El presente proyecto se desarrolla en el marco de Facultad de Ingeniería dentro Laboratorio de sistemas inteligentes (LABSIN). Este trabajo es parte del proyecto de investigación que dio inicio en setiembre de 2019 en el marco de los proyectos bienales de secretaria de Investigación, Internacionales y Posgrados (SIIP) de la Universidad Nacional de Cuyo.

1 INTRODUCCIÓN

La detección temprana de secuencias DGA y su posterior bloqueo por parte de los administradores de sistemas resulta fundamental a fin de evitar o al menos mitigar la propagación de las acciones maliciosas dentro de la red. Es por ello que es de vital importancia desarrollar herramientas de detección, no solo con una baja tasa de falsos positivos sino también con la capacidad de operar en tiempo real. La utilización de técnicas de aprendizaje de máquinas surge como la alternativa más interesante para la construcción de una herramienta de detección de DGA. La metodología más común para la construcción de modelos de detección basados en técnicas de aprendizaje de máquinas consiste en utilizar un conjunto de datos conteniendo información sobre nombres de dominios normales y DGA. Este conjunto de datos es utilizado para entrenar un algoritmo que da como resultado un modelo de detección capaz de discriminar entre

dominios normales y DGA. La principal ventaja de las técnicas de aprendizaje de máquinas es su capacidad de generalizar a casos similares nunca antes vistos. Lo que facilita considerablemente la tarea del administrador de redes, ya que este no tiene que mantener actualizada la base de datos con nombres de dominios maliciosos. Sin embargo estas presentan algunos inconvenientes para su aplicación en escenarios reales. Entre los inconvenientes principales se destacan: (a) la tasa de falsos positivos, (b) el diseño del grupo correcto de atributos de entrada y (c) la necesidad de reentrenar periódicamente.

La capacidad de reconocer casos nunca antes vistos trae como consecuencia que algunos nombres de dominios normales, puedan ser detectados como maliciosos. Esto se conoce como falsos positivos. En algunos dominios de aplicación, el número de falsos positivos puede llegar a ser considerablemente alto sin perjuicio de la viabilidad en la aplicación de la técnica de aprendizaje de máquina elegida. Sin embargo, en el caso de la detección de DGA, un nombre de dominio bloqueado de manera incorrecta puede perjudicar seriamente la usabilidad del servicio de nombres de dominios (DNS). Basta que un usuario que no pueda acceder a un sitio con un nombre de dominio erróneamente clasificado como DGA para que genere inconvenientes en las tareas cotidianas de los administradores de sistemas. Es por esto último que mantener un número de falsos positivos muy bajo, resulta fundamental para una realizar la detección en escenarios reales.

El número de falsos positivos está directamente relacionado no solo con el tipo de algoritmo de aprendizaje de máquina elegido, sino también con los atributos (variables predictoras) utilizados como entrada del algoritmo. Dentro del área de aprendizaje de máquinas, la construcción de los atributos de entrada adecuados al problema se denomina normalmente ingeniería de atributos. La ingeniería de atributos es una de las tareas que muchas veces demanda no solo los mayores recursos computacionales sino también humanos.

Por otro lado, los modelos de detección construidos a partir de técnicas de aprendizaje de máquinas requieren en muchos casos ajustes periódicos a fin de lidiar con casos significativamente diferentes respecto a los vistos durante la construcción del modelo de detección. Este proceso normalmente implica la incorporación al conjunto de entrenamiento de los nuevos casos observados y la posterior re-ejecución del algoritmo de aprendizaje de máquinas. Lamentablemente, el tiempo de entrenamiento de algunas de las técnicas de aprendizaje puede resultar excesivamente alto para los requerimientos de una aplicación en un escenario real. La realidad es que si se quiere entrenar un modelo de detección en un tiempo adecuado se debe considerar una alta demanda de recursos computacionales. Sobre todo cuando el volumen de los datos del conjunto de entrenamiento comienza a ser significativo. Es por ello que hay que considerar aquellas técnicas que sean capaces de minimizar el tiempo requerido para el ajuste de los modelos.

En los últimos 10 años han sido desarrolladas técnicas de aprendizaje de máquinas conocidas bajo el nombre de Aprendizaje Profundo (DL). La utilización de estas tecnicas ha sido la causa detrás de los mayores avances en el reconocimiento automático de imágenes, audio, video y análisis de texto. En particular las redes neuronales profundas ofrecen ventajas que permiten lidiar con los inconvenientes (a), (b) y (c) mencionados anteriormente. La principal ventaja radica en no tener que lidiar con el diseño del grupo correcto de atributos de entrada. A través de sus múltiples capas, las redes neuronales profundas son capaces ir aprendiendo los atributos de entrada más adecuados para el problema. Además, esta selección automática de los atributos de entrada puede mejorar significamente la eficiencia en términos su tasa de detección de casos tanto positivos como negativos. Por último, recientes avances en la tecnología de

procesamiento de los algoritmos de redes neuronales profundas han permitido disminuir considerablemente los tiempos de entrenamiento con grandes conjuntos de datos. En particular, la utilización de la capacidad de cómputo provista por las placas gráficas (GPU, del inglés Graphics Processing Unit) ofrece una ventaja significativa frente al procesamiento en computadoras con CPU (del inglés Central Processing Unit).

El presente proyecto propone analizar la aplicación de redes neuronales profundas para el aprendizaje de los patrones comunes a los DGA de tal manera que permita desarrollar herramientas de detección no solo con una baja tasa de falsos positivos sino también con la capacidad de operar en tiempo real. Esto último resulta fundamental para lidiar con las amenazas de seguridad de hoy.

2 LÍNEAS DE INVESTIGACIÓN Y DESARROLLO.

El presente proyecto se enmarca en una línea de investigación que viene desarrollando en el LABSIN desde 2017, la cual se centra en la aplicación de técnicas de aprendizaje automático a la seguridad informática.

Para el desarrollo del presente proyecto pueden diferenciarse 3 etapas principales:

A) Análisis preliminar del problema. Las tareas asociadas a esta etapa tienen por objetivo conocimiento de los problemas asociados a la detección de DGA como así también los modelos probabilísticos actualmente implementados. Una tarea fundamental consistirá en realización de una breve revisión de la literatura sobre la aplicación de las técnicas de aprendizaje profundo a problemas de detección de DGA. Dicha tarea tiene por objetivo el de tratar de determinar cuáles han sido los beneficios e inconvenientes al aplicar este tipo de algoritmos.

B) Desarrollo de un algoritmo para la detección de DGA basado en técnicas de aprendizaje profundo. Esta etapa tiene por objetivo el desarrollo, evaluación y puesta a punto de un primer prototipo funcional para la detección de anomalías en el tráfico de red. Este primer prototipo será desarrollado utilizando alguna de las bibliotecas disponibles que permitan la implementación de modelos basados en aprendizaje profundo de manera simple y eficiente. En esta etapa se definirán los diferentes aspectos de la red como ser: el tipo de red a utilizar, la topología de la red y la secuencia de entrada entre otras. Luego se evaluarán los resultados utilizando un conjunto de datos conteniendo tráfico etiquetado (DGA y normal).

C) Experimentación. Finalmente en la última etapa se centrará en la evaluación del algoritmo propuesto sobre distintos conjuntos de datos: En particular, se evaluará el algoritmo propuesto con diferentes conjuntos de datos previamente etiquetados. Durante este proceso se consideran las métricas habituales en el área utilizando mecanismos para validar la generalidad del modelo obtenido como validación cruzada. Durante esta etapa se realizaran también estudios comparativos con otros modelos de reconocimiento de anomalías de la literatura.

3 RESULTADOS OBTENIDOS

Durante los primeros 6 meses del proyecto se realizó completaron las actividades de

a) Recopilación de información bibliográfica sobre el tema poniendo especial énfasis en la aplicación de técnicas de aprendizaje profundo a problemas relacionados con la temática de detección de DGA.
b) Construcción de un conjunto de datos de entrada adecuado con información sobre peticiones de nombres de dominio normales y DGA. Las

mismas se obtuvieron a partir de datos disponibles de manera libre en Internet

c) Implementación de un modelo basado en redes recurrentes de tipo LSTM a partir de lo recopilado en la bibliografía.

una recopilación de dominios generados de manera algorítmica por diferentes tipos de software malicioso. Dicho conjunto de datos se utilizó para evaluar un primer algoritmo basado en redes neuronales convolucionales.

3 RESULTADOS ESPERADOS

Al término del plan de trabajo se pretende que se haya logrado:

1. Fortalecer la línea de investigación en la aplicación de aprendizaje de máquinas aplicados a la seguridad informática.
2. Obtener una implementación funcional del modelo para la detección de DGA basado en redes neuronales con aprendizaje profundo.
3. Incrementar la experiencia para la posterior aplicación de modelos probabilísticos basados en aprendizaje profundo a nuevas líneas de investigación relacionadas con problemas de ciencia y tecnología.

4 FORMACIÓN DE RECURSOS HUMANOS

Se espera capacitar en el ámbito de la investigación a profesores y alumnos interesados en participar en un entorno académico y tecnológico innovador y a todos aquellos actores interesados en los resultados del proyecto.

Sobre la temática de este proyecto se está trabajando en:

• La tesis doctoral de Jorge Guerra, en el doctorado en Ciencias Informáticas de la Universidad Nacional del centro de la provincia de Buenos Aires.

• La capacitación del Sr. Franco Palau, alumno de la carrera de Ingeniería en Mecatrónica de la Facultad de Ingeniería.

5 BIBLIOGRAFÍA

[1] Plohmann, D., Yakdan, K., Klatt, M., Bader, J., Gerhards-Padilla, E. (2016): A comprehensive measurement study of domain generating malware. In: 25th USENIX Security Symposium (USENIX Security 16). pp. 263–278. USENIX Association, Austin, TX.

[2] Kührer, M., Rossow, C., Holz, T. (2014): Paint it black: Evaluating the effectiveness of malware blacklists. In: Stavrou, A., Bos, H., Portokalidis, G. (eds.) Research in Attacks, Intrusions and Defenses. Springer International Publishing.

[3] Antonakakis, M., & Perdisci, R. (2012). From throw-away traffic to bots: detecting the rise of DGA-based malware. Proceedings of the 21st USENIX Security Symposium, 16. Retrieved from https://www.usenix.org/system/files/conference/usenixsecurity12/sec12-final127.pdf

[4] Jason Reed, Adam J Aviv, Daniel Wagner, Andreas Haeberlen, Benjamin C Pierce, and Jonathan M Smith (2010). Differential privacy for collaborative security. In Proceedings of the Third European Workshop on System Security, pages 1–7. ACM.

[5] Grill, M., Nikolaev, I., Valeros, V., & Rehak, M. (2015). Detecting DGA malware using NetFlow.

[6] M. J. Erquiaga, C. Catania and S. García, "Detecting DGA malware traffic through behavioral models," 2016 IEEE Biennial Congress of Argentina (ARGENCON), Buenos Aires, Argentina, 2016, pp. 1-6.

[7] Yadav, S., Reddy, A.K.K., Narasimha Reddy, A.L., Ranjan, S.: Detecting algorithmically generated domain-flux attacks with DNS traffic analysis. IEEE/ACM

Transactions on Networking 20(5), 1663–1677 (2012)
[8] Schiavoni, S., Maggi, F., Cavallaro, L., & Zanero, S. (2014). Phoenix: DGA-based botnet tracking and intelligence. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8550 LNCS, 192–211. https://doi.org/10.1007/978-3-319-08509-8_11.
[9] Ahluwalia, A., Traore, I., Ganame, K., Agarwal, N.: Detecting broad length algorithmically generated domains. In: Traore, I., Woungang, I., Awad, A. (eds.) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. pp. 19–34. Springer International Publishing, Cham (2017)

Detección de vulnerabilidades en especificaciones de contratos inteligentes de la plataforma Ethereum

Mauro C. Argañaraz(1), Mario M. Berón(1), Pedro Rangel Henriques(2) &Daniel Riesco(1)
(1)Departamento de Informática - Facultad de Ciencias Física Matemáticas y Naturales (FCFMyN) – Universidad Nacional de San Luis
(2)Universidade do Minho - Braga, Portugal
marganaraz@gmail.com(1), {mberon, driesco}@unsl.edu.ar(1), pedrorangelhenriques@gmail.com(2)

RESUMEN

Ethereum es el principal ecosistema basado en blockchain que proporciona un entorno para codificar y ejecutar contratos inteligentes, propiedad que en estos últimos años recibió gran atención por las nuevas aplicaciones comerciales y, también, por la comunidad científica. El proceso de escritura de contratos seguros y de buen desempeño en la plataforma Ethereum es un gran desafío para los desarrolladores. Implica la aplicación de paradigmas de programación no convencionales debido a las características inherentes de la ejecución de programas de computación distribuida. Además, los errores en los contratos desplegados pueden tener graves consecuencias debido al acoplamiento inmediato del código del contrato y las transacciones financieras. El manejo directo de los activos significa que las fallas tienen más probabilidades de ser relevantes para la seguridad y tienen mayores consecuencias económicas que los errores en las aplicaciones típicas. Una serie de reportes de incidentes muestran que los problemas de seguridad se han utilizado con propósitos fraudulentos.

En este artículo, se describe una línea de investigación que se enfoca en fortalecer los aspectos de seguridad, basado en una base sólida de diseño, patrones de código establecidos y probados que faciliten el proceso de escritura de código funcional y libre de errores, para proporcionar una herramienta que permita analizar y detectar falencias de manera automática.

Palabras Claves: blockchain, Ethereum, smart contract, Solidity, Vyper, static analysis tool, verification, security patterns.

CONTEXTO

La presente línea de investigación se enmarca en el Proyecto (PO/16/93) de *"Fortalecimiento de la Seguridad de los Sistemas de Software mediante el uso de Métodos, Técnicas y Herramientas de Ingeniería Reversa"*. Realizado en conjunto con la Universidade do Minho Braga, Portugal. Recientemente aprobado por el Ministerio de Ciencia Tecnología e Innovación Productiva (Mincyt). Y por el Proyecto (P031516.) de Investigación: *"Ingeniería de Software: Conceptos, Prácticas y Herramientas para el Desarrollo de Software con Calidad"*. De la Facultad de Ciencias Físico Matemáticas y Naturales, Universidad Nacional de San Luis. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el Programa de Incentivos.

1. INTRODUCCIÓN

El año 2017 se caracterizó por la popularidad mundial que alcanzó Bitcoin [1] llegando a su valor de mercado más alto en su breve recorrido desde su creación, hito que coincidió con el reconocimiento de Wall Street al incluirlo en los mercados de futuros. Ante la merma en el precio durante 2018, la cadena de bloques (blockchain), tecnología que subyace a Bitcoin y al resto de las criptomonedas, llegó a la cresta de la ola IT. Sin embargo, de ahora en más, el foco del mercado y de la comunidad

científica inevitablemente se concentrará en los contratos inteligentes, ya que permiten que partes no confiables manifiesten términos de contrato en el código del programa y, por lo tanto, eliminan la necesidad de un tercero confiable.

Blockchain es una tecnología que se basa en una combinación de criptografía, redes y mecanismos de incentivos para respaldar la verificación, ejecución y registro de transacciones entre diferentes partes. En términos simples, las plataformas de blockchain se pueden ver como bases de datos descentralizadas que ofrecen propiedades muy atractivas. Esto incluye la inmutabilidad de las transacciones almacenadas y la creación de confianza entre los participantes sin un tercero. Esto hace que esta arquitectura sea adecuada como un libro mayor abierto y distribuido que pueda almacenar transacciones entre partes de manera verificable y permanente.

Una aplicación destacada es el intercambio de activos digitales, conocidos como criptomonedas. Las criptomonedas más conocidas y transadas son Bitcoin, Ethereum y Litecoin. Ofrecen, más allá de la transferencia de activos digitales, la ejecución de contratos inteligentes.

Los contratos inteligentes son esencialmente dos conceptos combinados. Una es la noción de software. Código frío y austero que hace lo que está escrito y se ejecuta para que el mundo lo vea. La otra es la idea de un acuerdo entre las partes. Son programas informáticos que facilitan, verifican y hacen cumplir la negociación y ejecución de contratos legales. Se ejecutan a través de transacciones de blockchain, interactúan con las criptomonedas y tienen interfaces para manejar la información de los participantes del contrato. Cuando se ejecuta en la cadena de bloques, un contrato inteligente se convierte en una entidad autónoma que ejecuta automáticamente acciones específicas cuando se cumplen ciertas condiciones.

Ethereum es la principal plataforma de computación distribuida pública basada en blockchain que proporciona un entorno de ejecución de contratos inteligentes dentro del contexto de una máquina virtual descentralizada, conocida como Ethereum Virtual Machine (EVM) [2, 3].

La máquina virtual EVM maneja el cómputo y el estado de los contratos y se basa en un lenguaje basado en una estructura de pila con un conjunto predefinido de instrucciones (códigos de operación) [3]. En esencia, un contrato es simplemente una serie de declaraciones de código de operación, que son ejecutadas secuencialmente por la máquina virtual EVM. La máquina EVM puede considerarse como una computadora global descentralizada en la que se ejecutan todos los contratos inteligentes. Aunque se comporta como una computadora gigante, es más bien una red de máquinas discretas más pequeñas en comunicación constante.

Los contratos inteligentes en Ethereum generalmente se escriben en lenguajes de alto nivel y luego se compilan en bytecode EVM. Los lenguajes de alto nivel son LLL (lenguaje tipo Lisp de bajo nivel), Serpent (un lenguaje de estilo Python), Vyper (un lenguaje de tipo Python), Bamboo (lenguaje basado en estados) y Solidity (un lenguaje estilo Javascript) [4]. LLL y Serpent se desarrollaron en las primeras etapas de la plataforma, mientras que Vyper y Bamboo se encuentran actualmente en desarrollo por la fundación Ethereum. El lenguaje más prominente y ampliamente adoptado es Solidity.

Solidity es un lenguaje de programación de alto nivel orientado a contratos. Su sintaxis es similar a JavaScript, está tipado de manera estática y admite herencia y polimorfismo, así como bibliotecas y tipos complejos definidos por el usuario.

El proceso de escritura de contratos seguros y de buen desempeño en Ethereum

es una tarea difícil para los desarrolladores. Implica la aplicación de paradigmas de programación no convencionales, debido a las características inherentes de la ejecución de programas basados en blockchain. Además, los errores en los contratos desplegados pueden tener graves consecuencias, debido al acoplamiento inmediato del código del contrato y las transacciones financieras. Por lo tanto, es beneficioso tener una base sólida de diseño, patrones de código establecidos y probados que faciliten el proceso de escritura de código funcional y libre de errores, y herramientas que permitan analizar y detectar falencias de manera automática.

Un análisis de los contratos inteligentes existentes realizado por Bartoletti y Pompianu [5] muestra que las plataformas Bitcoin y Ethereum se centran principalmente en los contratos financieros. En otras palabras, la mayor parte del código del programa define cómo se mueven los activos (dinero). Por lo tanto, es crucial que la ejecución del contrato se realice correctamente. El manejo directo de los activos significa que las fallas tienen más probabilidades de ser relevantes para la seguridad y tienen mayores consecuencias financieras que los errores en las aplicaciones típicas.

Los incidentes, como el desbordamiento de valor en Bitcoin, o el ataque al proyecto DAO en Ethereum, causaron que una bifurcación dura de la blockchain anulara las transacciones maliciosas. Estos incidentes muestran que los problemas de seguridad se han utilizado con propósitos fraudulentos. Muchas de estas vulnerabilidades se pueden clasificar en tres grupos: lenguaje de programación, máquina virtual (EVM) y las peculiaridades propias de la blockchain, y pueden abordarse siguiendo las mejores prácticas para escribir contratos inteligentes seguros, que se encuentran dispersos en toda la comunidad Ethereum.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En la búsqueda de asegurar los contratos inteligentes en la plataforma Ethereum se han adoptado diferentes enfoques, cubriendo aspectos como la semántica formal, patrones de seguridad y herramientas de verificación. De acuerdo con el tipo de enfoque abordado, se distingue entre enfoques de verificación y de diseño.

El objetivo de los enfoques de verificación es comprobar que los contratos inteligentes escritos en lenguajes de alto nivel existentes (como Solidity por mencionar alguno) cumplan con una política o especificación de seguridad, entre los que se pueden mencionar:

- Herramientas de análisis estático para la búsqueda automática de bugs [6, 7, 8, 9]
- Herramientas de análisis estático para la verificación automática de propiedades genéricas [10, 11, 12, 13]
- Frameworks para pruebas semi-automatizadas de propiedades específicas del contrato [14, 15, 16, 17]
- Monitoreo dinámico de propiedades de seguridad predefinidas [18, 19]

En contraste, los enfoques de diseño apuntan a facilitar la creación de contratos inteligentes seguros al proporcionar frameworks para su desarrollo: abarcan nuevos lenguajes que son más susceptibles de verificación, proporcionan una semántica clara y sencilla que es entendible por los desarrolladores de contratos inteligentes o que permiten una codificación directa de las políticas de seguridad deseadas. Además, se incluyen trabajos que tienen como objetivo proporcionar patrones de diseño para contratos inteligentes seguros.

- Lenguajes de alto nivel [20, 21, 22, 23]
- Lenguajes intermedios [24]
- Patrones de seguridad [25]

- Herramientas [26]
- Ontologías [27, 28]

La línea de investigación presentada en este artículo busca plantear una estrategia para la detección de vulnerabilidades, nutriéndose de ambos campos de estudio. Por un lado, utiliza un enfoque de verificación con las siguientes características:

- Lenguaje destino: lenguajes de alto nivel de la plataforma Ethereum (se selecciona Solidity como caso de estudio de mercado y Vyper como caso de estudio experimental)
- Método de análisis: estático
- Garantías: búsqueda de bugs
- Grado de automatización: verificación automatizada

Por el lado de los enfoques de diseño, se toman como base los patrones de seguridad para lenguajes existentes [25] y la idea de construir una representación intermedia para analizar los aspectos de seguridad.

3. RESULTADOS OBTENIDOS/ESPERADOS

El trabajo de investigación apunta a responder qué especificaciones tienen vulnerabilidades a través de la ejecución de análisis léxicos y sintácticos. El análisis estático garantiza una cobertura completa sin ejecutar el programa y lo suficientemente rápido en un código de tamaño razonable.

Para llevar adelante la propuesta, se creó el proyecto open source denominado OpenBalthazar. Este proyecto consiste en una herramienta web de análisis estático para los contratos inteligentes de la plataforma Ethereum implementada con Microsoft .NET Core.

Actualmente está implementado Solidity y se inició el desarrollo de las reglas de Vyper, si bien es una herramienta extensible y se pueden incorporar nuevos lenguajes como Bamboo a través de los mecanismos previstos.

En el proyecto se utiliza la librería ANTLR 4 y las gramáticas de Solidity y Vyper para generar el árbol de análisis (AST). Este árbol se puede enriquecer con información adicional utilizando algoritmos y técnicas de procesamiento de lenguajes. Las reglas de verificación construidas utilizan un repositorio de patrones que definen los criterios en términos del árbol.

Los investigadores de esta línea continuarán con estudios en este campo con el objetivo de perfeccionarse en el área y realizar un seguimiento de nuevas amenazas, vulnerabilidades y ciberataques en materia de despliegue y ejecución de contratos inteligentes. También se planea seguir con los siguientes trabajos futuros:

- Generalización para otras plataformas de blockchain que soporten contratos inteligentes, tales como NEM, NEO, Cardano o Hyperledger.
- Obtención de código fuente a partir de bytecode EVM para aplicar el análisis que se propone en esta línea de investigación.
- Análisis dinámico de código fuente, lo cual implica correr el contrato y considerar sólo un conjunto de todas las ejecuciones posibles en base a datos de entrada arbitrarios.
- Análisis de los aspectos de seguridad que surjan de la interoperabilidad con otras plataformas blockchain, también conocida como cross-chain smart contracts.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de profesionales de la UNSL que forman parte de la línea de investigación de este trabajo llevan adelante diferentes trabajos finales integradores de *Ingeniería en Informática, Ingeniería en Computación, Licenciatura en Ciencias de la Computación*, y en un futuro próximo trabajos finales de especialización, tesis de

maestría y doctorado. En particular, las investigaciones desarrolladas en este trabajo forman parte del lineamiento inicial como trabajo final de uno de los autores para optar al grado de Doctor en Ingeniería Informática en la UNSL.

5. BIBLIOGRAFÍA

[1] Nakamoto, S., (2008), Bitcoin: Un Sistema de Efectivo Electrónico Usuario-a-Usuario, http://bitcoin.org/bitcoin.pdf

[2] Ethereum White Paper: https://github.com/ethereum/wiki/wiki/White-Paper

[3] Ethereum Yellow Paper: https://ethereum.github.io/yellowpaper/paper.pdf

[4] Solidity — solidity 0.4.18 documentation. [Online]. Available: https://media.readthedocs.org/pdf/solidity/develop/solidity.pdf

[5] Bartoletti, M. & Pompianu, L. (2017) *An empirical analysis of smart contracts: platforms, applications, and design patterns*. arXiv preprint arXiv:1703.06322.

[6] Luu, L., Chu, D.H., Olickel, H., Saxena, P., Hobor, A.: Making smart contracts smarter. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, ACM (2016) 254–269

[7] Grishchenko, I., Maffei, M., Schneidewind, C.: A semantic framework for the security analysis of ethereum smart contracts. In: Proceedings of the 7th International Conference on Principles of Security and Trust (POST), Springer (2018)

[8] Zhou, E., Hua, S., Pi, B., Sun, J., Nomura, Y., Yamashita, K., Kurihara, H.: Security assurance for smart contract. In: New Technologies, Mobility and Security (NTMS), 2018 9th IFIP International Conference on, IEEE (2018) 1–5

[9] Nikolic, I., Kolluri, A., Sergey, I., Saxena, P., Hobor, A.: Finding the greedy, prodigal, and suicidal contracts at scale. arXiv preprint arXiv:1802.06038 (2018)

[10] Kalra, S., Goel, S., Dhawan, M., Sharma, S.: Zeus: Analyzing safety of smart contracts, NDSS (2018)

[11] Buenzli, F., Dan, A., Drachsler-Cohen, D., Gervais, A., Tsankov, P., Vechev, M.: Securify (2017) Available at http://securify.ch.

[12] Mythril Available at https://github.com/ConsenSys/mythril.

[13] Manticore Available at https://github.com/trailofbits/manticore.

[14] Hirai, Y.: Defining the ethereum virtual machine for interactive theorem provers. In: International Conference on Financial Cryptography and Data Security, Springer (2017) 520–535

[15] Amani, S., B´egel, M., Bortin, M., Staples, M.: Towards verifying ethereum smart contract bytecode in isabelle/hol. CPP. ACM. To appear (2018)

[16] Hildenbrandt, E., Saxena, M., Zhu, X., Rodrigues, N., Daian, P., Guth, D., Rosu, G.: Kevm: A complete semantics of the ethereum virtual machine. Technical report (2017)

[17] Bhargavan, K., Delignat-Lavaud, A., Fournet, C., Gollamudi, A., Gonthier, G., Kobeissi, N., Kulatova, N., Rastogi, A., Sibut-Pinote, T., Swamy, N., et al.: Formal verification of smart contracts: Short paper. In: Proceedings of the 2016 ACM Workshop on Programming Languages and Analysis for Security, ACM (2016) 91–96

[18] Grossman, S., Abraham, I., Golan-Gueta, G., Michalevsky, Y., Rinetzky, N., Sagiv, M., Zohar, Y.: Online detection of effectively callback free objects with applications to smart contracts. Proceedings of the ACM on Programming Languages 2(POPL) (2017) 48

[19] Cook, T., Latham, A., Lee, J.H.: Dappguard: Active monitoring and defense for solidity smart contracts

[20] OConnor, R.: Simplicity: A new language for blockchains. arXiv preprint arXiv:1711.03028 (2017)

[21] Pettersson, J., Edström, R.: Safer smart contracts through type-driven development

[22] Coblenz, M.: Obsidian: A safer blockchain programming language. In: Software Engineering Companion (ICSE-C), 2017 IEEE/ACM 39th International Conference on, IEEE (2017) 97–99

[23] Schrans, F., Eisenbach, S., Drossopoulou, S.: Writing safe smart contracts in flint

[24] Sergey, I., Kumar, A., Hobor, A.: Scilla: a smart contract intermediate-level language. arXiv preprint arXiv:1801.00687 (2018)

[25] W¨ohrer, M., Zdun, U.: Smart contracts: Security patterns in the ethereum ecosystem and solidity. (2018)

[26] Mavridou, A., Laszka, A.: Designing secure ethereum smart contracts: A finite state machine based approach. arXiv preprint arXiv:1711.09327 (2017)

[27] Ugarte, H.: BLONDiE https://github.com/hedugaro/Blondie

[28] Sitio oficial de la ontología EthOn: https://github.com/ConsenSys/EthOn

Blockchain para aseguramiento de Evidencia Digital en entornos Forensic Readiness

Javier Díaz [(1),] Mónica D. Tugnarelli [(2)], Mauro F. Fornaroli [(2)], Lucas Barboza [(2)]
[(1)] Facultad de Informática – Universidad Nacional de La Plata
[(2)] Facultad de Ciencias de la Administración – Universidad Nacional de Entre Ríos
e-mail:jdiaz@unlp.edu.ar, montug, maufor, lbarboza[@fcad.uner.edu.ar]

Resumen

La tecnología blockchain tiene múltiples usos cuando se trata de validar la integridad, la transparencia y trazabilidad de datos. En este trabajo se presentan los avances del PID-UNER 7059 que abordará el estudio de esta tecnología focalizando su aplicación para asegurar la preservación, integridad y trazabilidad de evidencia digital, obtenida de activos esenciales, en un entorno preventivo como lo es Forensic Readiness.

Palabras clave: blockchain, evidencia digital, Forensic Readiness, seguridad.

Contexto

El artículo presenta los primeros avances del Proyecto de Investigación y Desarrollo PID-UNER 7059 denominado "*Tecnología Blockchain para aseguramiento de evidencia digital en entornos Forensic Readiness*" que se encuadra en una de las líneas de investigación establecidas como prioritarias para su fomento, "Arquitectura, Sistemas Operativos y Redes", de la carrera Licenciatura en Sistemas de la Facultad de Ciencias de la Administración. Se adecua además, a las prioridades de la Universidad Nacional de Entre Ríos por ser un proyecto aplicado a la investigación sobre Tecnologías de la Información y la Comunicación.

Introducción

Blockchain puede describirse como una base de datos distribuida y organizada bajo una estructura de conjunto de bloques que se van encadenando entre sí mediante dos códigos hash que actúan como enlaces: uno para el bloque de datos creado anteriormente y otro para que se comparta y grabe en el bloque que se cree a continuación, de forma tal de obtener una lista enlazada o cadena de bloques. Cada bloque que se encadena se vuelve inmutable y en eso radica la fortaleza de seguridad y verificabilidad de esta tecnología en la que, cuanto mayor es el grado de replicación de los bloques, más sencillo resulta detectar adulteraciones. [1]. La aplicación más visible de esta tecnología son las criptomonedas, siendo el Bitcoin una de las más reconocidas que fue presentada por Satoshi Nakamoto [2] como un sistema de dinero en efectivo/pago electrónico basado en pruebas criptográficas, en contraposición al sistema financiero tradicional que utiliza un esquema basado en confianza con instituciones financieras que actúan como intermediarias. De esta manera la criptomoneda permite que, sin necesidad de un tercero, dos partes realicen transacciones entre ellas por medio de criptografía, (operaciones computacionalmente imposibles de revertir con las tecnologías actuales [3]), redes pares, una cadena de firmas digitales, un servidor de sellado de

tiempo y los nodos suficientes para lograr el consenso distribuido requerido para validar cada transacción.

Más allá de su aplicación en las criptomonedas, la tecnología fue considerada en todo su potencial de habilitar consensos distribuidos para que cada transacción en línea pueda ser verificada en cualquier momento futuro. Esto es posible porque blockchain contiene un registro determinado y verificable de cada transacción realizada y los datos que se introducen son permanentes. Es decir, una vez escrito un nuevo hecho este no se puede borrar ni modificar, lo cual se consigue replicando el registro de información entre varios nodos, de manera que cualquier alteración requiera modificar el registro de cada uno de los participantes.

Esta tecnología permite tener una red distribuida en la que no existe ninguna entidad central o intermediario que coordine las interacciones, sino que se trata de una red peer-to-peer (P2P) en la que los participantes se comunican entre pares. Debido a ello, cuando se desea introducir un nuevo hecho en el registro compartido, se requiere alcanzar un consenso entre los participantes para determinar en qué bloque se registrará esa información.

Las transacciones en un bloque se consideran que ocurrieron en el mismo momento de tiempo, por ende los bloques se enlazan entre sí en un orden cronológico lineal y, a medida que la cadena de bloques se enlaza, crea un registro público irrefutable soportado por un esquema de encriptación de clave pública y claves hash.

Lo expuesto permite imaginar múltiples prestaciones de la tecnología relacionadas a transacciones, a la seguridad, a la trazabilidad y a la transparencia. Sería posible poner en la cadena de bloques cualquier tipo de archivo/dato que requiera aseguramiento con hash.

Específicamente en este proyecto se propone analizar las prestaciones de la tecnología Blockchain para asegurar la integridad y trazabilidad de la cadena de custodia en un entorno de Forensic Readiness, que como método preventivo, requiere de estrictas respuestas del entorno tecnológico para resguardar los datos considerados como evidencia digital.

Forensic Readiness o Preparación Forense propone que la evidencia digital se recolecte y asegure de manera anticipada, es decir, antes de la ocurrencia de un incidente de seguridad. Este término fue enunciado por John Tan [4] quien lo describió principalmente a través de dos objetivos: maximizar la capacidad del entorno para reunir evidencia digital confiable y minimizar el costo forense durante la respuesta a un incidente.

En este enfoque, que fue analizado en proyectos anteriores [5] [6] [7], los datos que se recolectan pueden ser utilizados como insumo para el análisis de incidentes de seguridad y también como prueba legal, lo que involucra el aseguramiento de la prueba a medida que se realiza la recolección activa de los datos, tarea que fue realizada, en dicho proyecto, utilizando funciones hash para resguardar la integridad de la evidencia digital.

De acuerdo con la ISO/IEC 27037:2012 [8] la evidencia digital es gobernada por tres principios fundamentales:

a. Relevancia: la evidencia digital debe estar relacionada con los hechos investigados,
b. Confiabilidad: la evidencia debe ser repetible y auditable, de tal manera que un tercero que aplique el mismo método utilizado, llegue al mismo resultado y
c. Suficiencia: la evidencia recolectada debe ser suficiente para sustentar los hallazgos obtenidos por el analista forense.

Considerando estos requisitos, una instancia fundamental para garantizar su admisibilidad como elemento de prueba es la preservación de la Cadena de Custodia como aval de la integridad y trazabilidad de la evidencia. Esta cadena de custodia debe estar claramente documentada y con un registro detallado desde su recolección hasta su almacenamiento, por lo que se plantea, con especial interés, la aplicación de la tecnología blockchain para cumplimentar este requisito.

Líneas de Investigación, Desarrollo e Innovación

Siguiendo la línea de investigación mencionada en el contexto de este trabajo, se llevarán a cabo actividades que propicien la conformación de una base de conocimiento sobre la tecnología blockchain y sus aplicaciones en diversos ámbitos, destacando el aseguramiento de la integridad y trazabilidad de cualquier activo digital que se considere evidencia digital.

Resultados y Objetivos

El PID 7059 tiene como objetivo primario analizar el impacto de la utilización de la tecnología blockchain aplicada a la preservación, la integridad y trazabilidad de la evidencia digital.
Como objetivos secundarios se establecen:

- Integrar esquemas de recolección de datos y bases de datos de resguardo de evidencia con una solución de blockchain.
- Analizar la relación entre la escalabilidad de blockchain y los algoritmos de consenso.
- Avanzar en la identificación de incidentes de seguridad y el análisis de aspectos de seguridad informática relacionada con la tecnología blockchain.

Para lograr estos objetivos se analizarán las estructuras y tipos de blockchain, las soluciones disponibles en el mercado y se realizará la instalación y configuración de un entorno de prueba con Hyperledger, sin criptomoneda asociada. Asimismo, se espera obtener un procedimiento automatizado de autenticación y trazabilidad de la evidencia digital.
Como primera etapa del proyecto se relevaron casos de uso a nivel regional y nacional, entre las que se destaca la iniciativa y puesta en funcionamiento de la Blockchain Federal Argentina (BFA) [9] que brinda una plataforma pública para integrar servicios y aplicaciones sobre blockchain. BFA sigue el modelo de Múltiples Partes Interesadas por lo que participan de ella entidades del sector público, privado, académico y de la

sociedad civil que aportan la infraestructura, desarrollo y soporte técnico. Cabe mencionar que integrantes de este proyecto participan como responsables de la implementación y mantenimiento del nodo sellador de la Facultad de Ciencias de la Administración de la UNER, siendo ésta última un miembro Parte de la BFA.

En el siguiente cuadro se presentan los principales usos relevados:

Tipo de organización: Administración pública	
Aplicación	**Aportes del Blockchain**
• Garantizar la integridad de documentación oficial. (Boletín Oficial, Carpeta Ciudadana CABA.) • Certificación de dominios de internet. (NIC Argentina) • Mediciones de altura de los ríos. (PNA)	• Seguridad jurídica y legislativa. • Garantía de integridad y control ciudadano • Transparencia en la publicación de datos • Trazabilidad de datos y mediciones
Tipo de organización: Instituciones	
Aplicaciones	**Aportes de Blockchain**
• Verificación de documentos notariales digitales. (Colegio Escribanos CABA) • Verificación de registros de graduados universitarios. (SIU) • Certificación de recepción ofertas de proveedores. (SIU-Diaguita) • Verificacion de informacion academica en sistema de gestión de alumnos. (UNC)	• Garantía de integridad en documentación. • Valor agregado en la certificación de documentos notariales. • Verificación de información pública. • Auditoría de información por las partes interesadas. • Trazabilidad en la emisión de certificaciones •
Tipo de organización: Industria	
Aplicaciones	**Aportes de Blockchain**
• Trazabilidad Citrícola (KYAS-SENASA) • Comercialización commodities agrícolas (Plataforma Agree Market) • Validación de boletos de compraventas de inmuebles (Bildenlex)	• Verificación de integridad de información en el documento • Permite sortear exitosamente auditorías documentales en los diferentes lugares donde se destina • Agregado de valor al producto • Uso de Smart Contracts para inmutabilidad de las negociaciones y los contratos

Cuadro 1. Aplicaciones de Blockchain a nivel nacional y regional

Formación de Recursos Humanos

Este proyecto propicia la formación en co-dirección de proyectos de un docente, la formación en actividades de investigación de dos docentes de la carrera Licenciatura en Sistemas y de un colaborador estudiante de posgrado de la Maestría en Sistemas de Información que se dicta en la Facultad de Ciencias de la Administración.

Además, se brindará el espacio de participación a aquellos estudiantes que deseen realizar su Trabajo Final de la carrera Licenciatura en Sistemas sobre este tema. Se prevé también, la incorporación de becarios de investigación en el marco del programa de Becas de Iniciación a la Investigación de la UNER.

Referencias

[1] Michael Crosby, et. al. BlockChain Technology: Beyond Bitcoin. Applied Innovation Review (AIR). Issue No. 2 June 2016. Berkeley. http://scet.berkeley.edu/wp-content/uploads/AIR-2016-Final-version-Int.pdf

[2] Satoshi Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf

[3] Kirill Bryanov. Quantum Computing Vs. Blockchain: Impact on Cryptography. https://cointelegraph.com/news/quantum-computing-vs-blockchain-impact-on-cryptography

[4] Tan, John. (2001). Forensic Readiness. http://isis.poly.edu/kulesh/forensics/forensic_readiness.pdf

[5] Tugnarelli, M.; Fornaroli, M.; Santana, S.; Jacobo, E.; Díaz, F.J. Análisis de metodologías de recolección de datos digitales. Workshop de Investigadores en Ciencias de la Computación (WICC 2017). ISBN: 978-987-42-5143-5. http://sedici.unlp.edu.ar/handle/10915/61343

[6] Mónica D. Tugnarelli, Mauro F. Fornaroli, Sonia R. Santana, Eduardo Jacobo, Javier Díaz: Análisis de metodologías de recolección de datos digitales en servidores web. Libro de Actas. XXIII Congreso Argentino de Ciencias de la Computación CACIC 2017. VI Workshop de Seguridad Informática, pp. 1230-1238. ISBN 978-950-34-1539-9.

[7] Tugnarelli, M., Fornaroli,M. Santana,S. Jacobo, E. Díaz, J. Analysis of Methodologies of Digital Data Collection in Web Serves. Communications in Computer and Information Science (Springer), Vol. 790, Pag.265. (2018) https://link.springer.com/content/pdf/bfm%3A978-3-319-75214-3%2F1.pdf

[8] Guidelines for identification, collection, acquisition and preservation of digital evidence ISO/IEC 27037:2012

[9] Blockchain Federal Argentina https://bfa.ar/

[10] Auqib Hamid Lone, Roohie Naaz Mir. Forensic-chain: Blockchain based digital forensics chain of custody with PoC in Hyperledger Composer. https://doi.org/10.1016/j.diin.2019.01.002

[11] Iuon-Chang Lin,Tzu-Chun Liao. A Survey of Blockchain Security Issues and Challenges. https://pdfs.semanticscholar.org/f61e/db500c023c4c4ef665bd7ed2423170773340.pdf

ANÁLISIS DE LA SEGURIDAD DEL PROTOCOLO DE COMUNICACIONES CAN

Ing. Alejandro Fourcade, Mg Jorge Eterovic
Departamento de Ingeniería e Investigaciones Tecnológicas Universidad Nacional de La Matanza
Florencio Varela 1903 (B1754JEC), San Justo, (5411) 4480-8900

afourcade@unlam.edu.ar; eterovic@unlam.edu.ar

RESUMEN

Desde su concepción, el protocolo de comunicaciones CAN (Controller Area Network), trata de dar solución a la transmisión de datos en los exigentes entornos industriales. Creado por Bosch en 1986, ofrece una opción robusta y confiable, por lo cual se ha transformado en el estándar para la industria automotriz [1].

Su arquitectura está pensada para satisfacer eficientemente las necesidades operativas de un vehículo. Pero hoy, el automóvil tal como se lo conocía una década atrás, ya no existe. Sus mecanismos han sido invadidos por dispositivos digitales, sensores, actuadores y tiene hasta 80 ECUs (Engine Control Unit: Unidad de Control de Motor) comunicadas entre sí. A esta complejidad se ha sumado una aún mayor: estar conectado a Internet constantemente.

Como sucedió recientemente con algunos protocolos inalámbricos ante el advenimiento del Internet de las Cosas (IoT), pasar de un dominio cerrado a uno abierto conectado a la nube, propicia la aparición de amenazas, vulnerabilidades y brechas de seguridad.

Este cambio de escenario no estuvo en los planes en el momento de la concepción de la arquitectura, y en el caso de CAN termina transformándose en una bomba de tiempo. La diferencia entre los dos escenarios es que no tiene las mismas consecuencias un ataque informático a una heladera que a un camión de carga circulando por una autopista.

El principal problema es que algunas de las fortalezas del protocolo CAN, se transforman en limitaciones a la hora de adaptarse a entornos abiertos. Este dilema no es nuevo, pero como el riesgo es creciente, es de crucial importancia encontrar soluciones de seguridad.

Palabras Clave: *Seguridad de Datos. Protocolo CAN. Seguridad en Automóviles Conectados.*

CONTEXTO

Este proyecto de investigación se desarrolla en el marco de un Programa de Incentivos a Docentes Investigadores de la Secretaría de Políticas Universitarias (PROINCE) en el Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza.

El proyecto es financiado por el propio Departamento y es del tipo investigación aplicada. El mismo consiste en el desarrollo de un banco didáctico de ensayo de motores térmicos. Los trabajos de campo y relevamientos realizados aportaron información valiosa y sirvieron como base para el presente trabajo.

1. INTRODUCCIÓN

Los automóviles eran hasta hace unos años dispositivos casi exclusivamente electromecánicos, pero en las últimas décadas se ha incorporado masivamente la electrónica digital y ha tomado preponderancia y criticidad en su funcionamiento. También la Tecnología de la Información (IT) tiene un papel fundamental en los diseños automotrices.

El uso de la IT en los vehículos es cada vez mayor y sus objetivos eran brindar conectividad y confort. Si bien se han alcanzado estas metas y se han maximizado las prestaciones, no se han tenido en cuenta las amenazas y vulnerabilidades ante ataques externos que tiene todo dispositivo conectado a una red abierta como Internet.

Es muy importante la gestión de la seguridad en los sistemas de información. Uno de los principales factores a considera cuando se evalúan

los posibles daños que una intrusión maliciosa puede ocasionar, es el tipo y tiempo de respuesta del sistema.

Los automóviles son sistemas que reaccionan en tiempo real y una de las características del protocolo CAN es su respuesta rápida garantizada. Por eso el daño eventual ante un ataque al sistema de control de un automóvil es muy alto, teniendo en cuenta que, por ejemplo, la dirección, los frenos y el acelerador dejaron de tener control mecánico y pasaron a ser comandados electrónicamente.

Las ECU son las que leen los sensores y comandan los actuadores de un vehículo. En las primeras unidades que incluían esta tecnología había exclusivamente una ECU central, pero en los vehículos actuales se pueden encontrar hasta 80 de ellas conectadas a través de buses de diferentes jerarquías. Por ejemplo, el bus CAN destinado al sistema de confort, tiene menor prioridad y velocidad de respuesta que el que controla el airbag.

Este complejo sistema maneja actualmente casi todos los dispositivos de un automóvil, desde su dirección hasta los frenos, pasando por las trabas de las puertas, el tablero de instrumentos, las ópticas y hasta en algunos modelos, las bombas de agua y combustible. En muchos casos no existe una opción mecánica que prevalezca por sobre la electrónica. Si la ECU decide doblar a la derecha, el volante no puede evitar que esto suceda.

El periodista de la revista Forbes Andy Greenberg, pudo constatar que su acompañante podía habilitar o deshabilitar desde su notebook MacBook, los frenos de una camioneta Ford Escape de 1500 kg. de peso [2]. La situación descripta deja ver la gravedad de una intrusión al sistema.

En los años venideros este panorama se hará todavía más complejo, lo que lleva a los investigadores de la seguridad de la información como Eugene Kaspersky, ampliamente conocido por su software antivirus, a agregar al entorno automotriz dentro de los posibles objetivos de un ataque informático [3].

En el futuro tendremos nuevas formas de conectividad, como CAR TO CAR (C2C) o CAR TO INFRASTRUCTURE (C2I), lo que aumentará la integración de los vehículos a las redes de datos, por lo que las amenazas y vulnerabilidades expuestas son solo la punta del iceberg.

Varios trabajos han demostrado la posibilidad de atacar la estructura de datos de un automóvil con los métodos tradicionales (sniffing, spoofing) o generando mensajes falsos. El bus CAN puede accederse desde varios puntos de ingreso, directamente desde el puerto OBD2 (On Board Diagnostics) disponible debajo del tablero, o mediante “wire tapping”, o sea hackeando los conductores internos.

La primera opción de seguridad es el cifrado de los datos [4], no es de fácil aplicación por las características del protocolo CAN, como ser: la velocidad de transmisión, la latencia y su estructura multimaestro/multicast.

En el escenario descripto, las soluciones parecerían provenir del análisis estadístico de los mensajes que ocupan el bus. De esta forma es posible observar desviaciones con respecto a datos históricos y alertar así de una intrusión externa maliciosa. En este sentido se han expuesto trabajos sobre la detección de anomalías que corresponden tanto a problemas técnicos como a ataques a la seguridad. Los procedimientos que se utilizan incluyen análisis de frecuencias, análisis multivariado de series temporales [5] y los basados en la entropía de la información [6].

Gracias al Big Data están disponibles las grabaciones de miles de vehículos que alimentan una gran base de datos que describe modelos de comportamiento usuales. Con estos datos podrían establecerse patrones de comportamiento, cuyas desviaciones podrían ser analizadas para determinar su origen.

A partir de 2006, un nuevo protocolo llamado FlexRay [7] entró al mercado, con mayor ancho de banda (dos canales de 10 Mbps) y la posibilidad de cifrado. Si bien su implementación es más cara y todavía no es un estándar, demuestra el interés de las grandes fábricas de automóviles de migrar a un nuevo protocolo más seguro.

En base al panorama expuesto y al estado del arte, el proyecto de investigación propone el análisis de las amenazas y vulnerabilidades a los automóviles y la búsqueda de soluciones de seguridad para el parque automotor existente.

2. LÍNEAS DE INVESTIGACION Y DESARROLLO

Como sucede frecuentemente, en nuestro caso el interés por la detección de los problemas de

seguridad en el bus CAN fue un efecto colateral de los estudios e investigaciones realizadas en pos de obtener información de los diferentes dispositivos electrónicos y mecánicos de un automóvil.

El proyecto que propició el presente trabajo fue el de la construcción de un banco de pruebas de motores. Para fabricar un banco de ensayo normalmente se utilizan sensores y dispositivos auxiliares (frenos, sistemas de enfriamiento), pero en las ocasiones en que el dispositivo bajo ensayo conserva el sistema CAN, este puede ofrecer interesante información complementaria.

Se realizaron ensayos sobre algunos modelos de automóviles para estudiar los diferentes sistemas de información de un vehículo. Se decidió, luego de conocer las capacidades de los sistemas, trabajar con tableros de instrumentos, que generan y reciben tramas CAN y también con vehículos completos. Para esta tarea se utilizaron tableros de Volkswagen Suran, Volkswagen Polo, una camioneta Toyota Hilux y una Volkswagen Suran.

El primer paso fue contactar especialistas en electrónica automotriz para que aportaran sus experiencias y expandieran así los alcances de la investigación. Se realizaron varias reuniones con conclusiones sumamente interesantes:

- Crecimiento del equipamiento electrónico en los automóviles (principalmente los de alta gama).
- La complejidad y diversidad de implementaciones en vehículos.
- Los cambios en el perfil del especialista que debe resolver los nuevos problemas relacionados con el ambiente digital.
- Al participar de pruebas de conexión al bus CAN en automóviles, ayudaron a comprender la facilidad con que se ingresa, la variedad de herramientas informáticas que existe para ello y lo potencialmente peligroso que resultaría un acceso malicioso.

Una vez finalizados los relevamientos de campo se comenzó con la elección del dispositivo a ensayar. Era necesario que reciba, transmita y que además muestre los resultados en forma gráfica. En este punto la decisión era utilizar como dispositivo a una ECU o a un tablero de instrumentos. Se optó por el tablero, por ajustarse mejor a los requisitos expuestos.

Si bien existen implementaciones estándar de la mensajería CAN, la mayoría de los fabricantes utilizan, además, mensajes exclusivos. Por eso para establecer comunicación con el tablero, se utilizaron varios métodos, diferentes velocidades y diferentes combinaciones de mensajes.

Para contar con la flexibilidad necesaria para hacer estas pruebas, era necesario tener acceso directo a la trama CAN, tanto para leer como para generar mensajes. Se desarrolló para esa tarea una herramienta informática llamada “CAN Sniffer” que permite la lectura y grabación de mensajes CAN y la generación de tramas. Está basada en el MCP2515, un chip que hace de interfaz del CAN SPI y la plataforma Raspberry Pi (3 y Zero). Los dispositivos se probaron en entornos de maestro/ esclavo y de multimaestro con varias ECUs [8].

Una vez que se contó con la posibilidad de capturar mensajes se determinó el procedimiento para leer los mensajes de control de una herramienta especializada de diagnóstico automotriz especial para Volkswagen. De esa forma se pudo descifrar cómo establecía la comunicación con el tablero de instrumentos. Esto permitió determinar el método más eficiente para tratar de que el tablero reaccione a mensajes externos generado por nuestro dispositivo. Luego de probar varias formas, la más eficaz demostró ser el ataque por fuerza bruta aplicado de forma metódica. Se enviaron mensajes generados secuencialmente para luego analizar la reacción del tablero.

Se verificó también, que al cambiar el tablero por uno externamente idéntico, pero del modelo Polo, fue necesario realizar nuevamente todo el procedimiento de detección de mensajes.

Para la grabación y análisis de las tramas recibidas y generadas se utilizó una herramienta de código abierto llamada Wireshark en su versión para Raspbian (sistema operativo de Raspberry Pi)

En paralelo a las pruebas con el tablero, se realizaron ensayos con una Toyota Hilux y un Volkswagen Suran, para verificar que lo que sucedía en las pruebas de laboratorio, se repitiera en vehículos reales.

3. RESULTADOS OBTENIDOS/ESPERADOS

Luego de realizar los ensayos descriptos, los resultados fueron interesantes tanto desde el punto de vista de la lectura de lo que sucede con algunos dispositivos conectados al bus CAN como en lo que respecta a su reacción ante la generación de mensajes. Como se mencionó, si bien existen varios estándares de protocolo CAN (SAE J1708, SAE J1587 SAE J1939), los fabricantes implementan mensajería propietaria y diferentes variantes. Una de ellas es, por ejemplo, el protocolo UDS (Unified Diagnostic Service), un protocolo de capa 7, adaptación del KPW2000 en CAN. Es por eso que la tarea de tratar de encontrar un dispositivo que se comunique con todos los sistemas es casi imposible y explica a su vez la variedad de herramientas, para diferentes marcas y usos que se ofrecen en el mercado.

Dentro de los resultados obtenidos podrían citarse los obtenidos en el desarrollo de dispositivos para llevar a cabo la investigación. En el proceso hacer ingeniería inversa fue necesario desarrollar hardware y programarlo. Los lenguajes de programación utilizados fueron C++ de diferentes plataformas (Arduino, Raspberry, STM32) y Python sobre Raspbian. Los dispositivos desarrollados fueron:

- CAN Sniffer básico
- CAN Sniffer extendido
- Driver de Tablero
- Nodo de relevamiento remoto

En la Figura 1 se pueden ver el tablero con su driver y el Sniffer. Esto fue presentado funcionando en la feria Exproyecto 2019, en la UNLaM.

En cuanto a los resultados de la investigación propiamente dichos fueron, hasta el momento:

- Ingeniería inversa de implementación CAN (estándar y propietaria).
- Inyección de mensajes a tableros y vehículos que interfieren su funcionamiento maliciosamente.
- Lectura de parámetros del automóvil en tiempo real vía ODB2.
- Lectura de parámetros de un automóvil interviniendo su cableado interno.
- Transmisión inalámbrica al nodo de registro de datos operativos de un vehículo.
- La verificación de que la mayoría de los resultados obtenidos en el banco de pruebas, eran consistentes con vehículos reales.

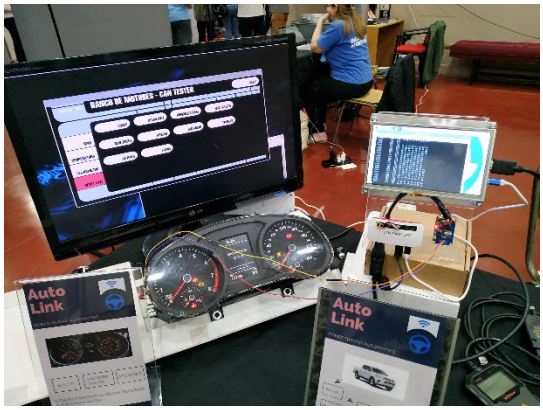


Figura 1: Tablero y Sniffer

El objetivo principal era obtener datos e interactuar con entornos automotrices, pero una de las conclusiones adicionales fue que la seguridad de un vehículo de calle está en riesgo dada la facilidad para ingresar y atacar su lógica de control.

Otra de las conclusiones fue que la electrónica automotriz ya tenía latente estos riesgos, pero las nuevas tecnologías de vehículos "siempre conectados" le suman gravedad. Hace años que la seguridad en los vehículos está en discusión, pero hasta hace poco era necesario el contacto físico para realizar la intrusión. Hoy eso ha comenzado a cambiar. En el horizonte cercano vemos camiones autónomos sin chofer surcando las rutas, por eso es indispensable hacer más robusta la seguridad de los sistemas vehiculares.

Las metas por cumplir para garantizar la seguridad de los datos en sistemas tecnológicos, confidencialidad, integridad, autenticidad, disponibilidad y no repudio; están lejos de ser alcanzadas actualmente por cualquier dispositivo que basa su comunicación en el protocolo CAN [9].

Si bien la industria está trabajando en mitigar estos riesgos en los nuevos modelos, hay un parque automotor existente que está en riesgo. Esta necesidad abre una línea de investigación posible: el desarrollo de un dispositivo conectable al puerto ODB2 para detectar intrusiones maliciosas por comparación estadística. En caso de camiones o vehículos de flota también sería instalable en un bus específico. Por ejemplo, el que

controla el consumo de combustible, el airbag, la dirección o los frenos.

Atacar un dispositivo puede ser tan sencillo como bombardearlo con mensajes y generar lo que se conoce como denegación de servicio (Denial of Service), dejándolo fuera de servicio en forma temporal o definitiva.

Como ejemplo del resultado obtenido en el proceso de ingeniería inversa del tablero de instrumentos de un VW Vento, se puede apagar o encender el testigo del airbag. Si se envía al identificador 80 (0X050 hexadecimal) el dato 0X000001, si X=0 apaga y si X=1 prende. Si se envía al identificador 1136 (0X470) los datos (X0000000) para X=1 prende la luz de giro a la izquierda, X=2 la derecha, X=3 la baliza. Ese mismo identificador maneja la luz del baúl y la luz de batería.

Los códigos para realizar las funciones descriptas se obtuvieron haciendo un barrido de identificadores y enviando datos al azar. También se determinó que, por ejemplo, en el caso del velocímetro, medidor de RPM y de combustible, es necesario enviar una combinación de varias tramas.

En resumen, los resultados obtenidos evidencian una interesante línea de investigación para desarrollar dispositivos no invasivos que aumenten la seguridad de los vehículos.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de este proyecto está formado por un ingeniero mecánico, un ingeniero electrónico y un especialista en seguridad. Como se mencionó anteriormente, este trabajo se desarrolló en el marco del proyecto de investigación: “Desarrollo de un banco didáctico de ensayo de motores térmicos”.

El desarrollo del proyecto de investigación generó varias líneas de trabajo, de múltiples disciplinas. Dada la complejidad, fue necesaria la colaboración de varios expertos con amplia experiencia en la industria y la investigación académica, tanto en la parte de electrónica automotriz, como de bancos de prueba de motores y de seguridad de datos.

Uno de los miembros del equipo de investigación se encuentra desarrollando su trabajo de tesis de posgrado de la Maestría en Informática de la UNLaM titulada: “µFramework: marco de referencia para desarrollos de sistemas embebidos” y su tutor es el Mg. Jorge Eterovic, integrante del proyecto de investigación [10].

5. BIBLIOGRAFIA

[**1**] Protocolo de Comunicaciones CAN; http://www.bosch-semiconductors.com/en/ubk_semiconductors/ip_modules_3/produkttabelle_ip_modules/can_literature_1/can_literature.html. 2019.

[2] Greenberg A., Hackers Reveal Nasty New Car Attacks — with me behind the wheel, https://www.forbes.com/sites/leemathews. 2019.

[3] Kapersky E., Viruses coming aboard? - https://securelist.com. 2005.

[4] Weimerskirch A., Wolf M., Wollinger T., State of the art: embedding security in vehicles, *Horst-Gortz-Institute for IT Security, Ruhr-University Bochum, Universittatsstraße, 44780 Bochum, Germany.* 2007.

[5] Theissler A., Anomaly detection in recordings from in-vehicle networks, IT-Designers GmbH, Esslingen, Germany, 2014.

[6] Marchetti M., Stabili D., Guido A., Colajani M. Evaluation on anomaly detection for in-vehicle networks through information-theoretic algorithms, University of Modena and Reggio Emilia, Italy, 2015.

[7] Hartzell, S., Stubel C. Automobile CAN bus network security and vulnerability, University of Washington, Seattle, USA. 2018.

[8] Miller C., Valasek C. Car Hacking: for poories, http://illmatics.com/car_hacking_poories.pdf. 2018.

[9] Hoppe T., Kilz S., Security threats to automotive can networks – practical examples and selected short-term countermeasures, Otto von Guericke University of Magdeburg, Germany. 2010.

[**10**] Fourcade A., Eterovic J., Pérez A., Rodofile G., µFramework: marco de referencia para desarrollo de sistemas embebidos; CoNaIISI 2019; RIISIC-CONFEDI-UNLaM, San Justo. 2019.

Seguridad en Internet de las Cosas usando soluciones Blockchain

Jorge Eterovic; Marcelo Cipriano; García, Edith; Luis Torres

Instituto de Investigación en Ciencia y Tecnología
Dirección de Investigación Vicerrectorado de Investigación y Desarrollo.
Universidad del Salvador.
Lavalle 1854 – C1051AAB -Ciudad Autónoma de Buenos Aires - Argentina

{jorge.eterovic; cipriano1.618; edithxgarcia}@gmail.com, torreslu@ar.ibm.com

RESUMEN

El mayor reto al que se enfrenta la seguridad en IoT (Internet of Things, Internet de las Cosas) proviene de la propia arquitectura del sistema actual, que se basa por completo en un modelo centralizado conocido como cliente/servidor. Todos los dispositivos se identifican, autentican y conectan a través de servidores en la nube.

Con el desarrollo de hogares, ciudades y autos inteligentes, el Internet de las Cosas se ha convertido en un área de rápido crecimiento que, según una estimación de la consultora internacional Gartner, se calcula que para 2023 podría haber una cantidad de más de 20 veces de dispositivos IoT conectados a la red que de plataformas de TI convencionales [1]. Sin embargo, la realidad nos muestra que la mayoría de estos dispositivos presentan vulnerabilidades factibles de ser atacadas con fines maliciosos.

Por lo general, estos dispositivos de IoT tienen una capacidad de procesamiento, almacenamiento y de conexión a la red limitados, lo que los hacen más vulnerables a los ataques que otros dispositivos con mayor capacidad como por ejemplo los teléfonos inteligentes, tabletas o computadoras.

En este proyecto de investigación se analizan los principales problemas de seguridad en IoT y como Blockchain, que es una de las tecnologías más innovadoras de nuestro tiempo y su uso viene ganando interés desde su aparición gracias a su capacidad para asegurar la integridad de las transacciones y la autenticidad entre cualquier entidad conectada a Internet de manera descentralizada, puede ser una solución para resolver algunos de los problemas de seguridad de IoT.

Palabras Clave:
Seguridad en Internet de las Cosas. Blockchain. Seguridad de las Redes. Seguridad de los Datos.

CONTEXTO

El Vicerrectorado de Investigación y Desarrollo (VRID), perteneciente a la Universidad Nacional del Salvador (USAL), dicta las políticas referidas a la investigación, concibiéndola como un servicio a la comunidad y entendiendo que los nuevos conocimientos son la base de los cambios sociales y productivos. Con el impulso de las propias Unidades Académicas se han venido desarrollando acciones conducentes a concretar proyectos de investigación uni/multidisciplinarios, asociándolos a la docencia de grado y postgrado y vinculando este accionar, para potenciarlo, con otras instituciones académicas del ámbito nacional e internacional.

La Dirección de Investigación, dependiente del VRID, brinda soporte a las distintas Unidades de Investigación y a sus investigadores para el desarrollo de Proyectos y Programas de Investigación, nacionales e internacionales, como así también, apoyo y orientación de recursos para la investigación.

A ella pertenece el Instituto de Investigación en Ciencia y Tecnología (RR 576/12) en el

cual se enmarca este proyecto denominado "Análisis de la seguridad de los datos en Internet de las Cosas usando tecnología Blockchain", con una duración de 2 años (2019-2020).

1. INTRODUCCIÓN

Durante los últimos años, el Internet de las Cosas se ha ido introduciendo gradualmente en nuestras vidas gracias a la alta disponibilidad de sistemas de comunicación inalámbricos [2].

El paradigma de IoT abarca muchos conceptos: dispositivos inteligentes que recopilan datos del entorno, muchas tecnologías diferentes para permitir su conexión, servicios y estándares, y todos los elementos que participan [3,4].

IoT puede traer muchos beneficios a la sociedad de muchas maneras diferentes, pero es muy importante estar atentos para que se encuentre la mejor solución para proteger la privacidad de los datos [5,6]. El gran desafío del IoT es encontrar un entorno de comunicación confiable que garantice la seguridad de los datos transmitidos entre todos los dispositivos conectados [7].

Una de las posibles soluciones podría ser la convergencia entre la tecnología IoT y Blockchain [8].

Blockchain ha sido propuesta por la industria y la comunidad de investigación como una tecnología disruptiva que estaría preparada para desempeñar un papel importante en la gestión, el control y, lo más importante, la seguridad de los dispositivos IoT.

Blockchain puede ser una tecnología clave para proporcionar soluciones de seguridad viables a los desafiantes problemas de seguridad de IoT [9,10]. Con el desarrollo de la cadena de bloques Ethereum, que implementa contratos inteligentes, el espacio de uso potencial de Blockchain se ha vuelto muy importante.

Ethereum se lanzó y se abrió para uso público en julio de 2015. Después, surgieron plataformas de Blockchain de contrato inteligente similares. Entre ellas se incluyen Hyperledger [11], Eris [12], Stellar [13], Ripple [14] y Tendermint [15].

A diferencia de la cadena de bloques de bitcoin, que se usa principalmente para transacciones de moneda digital, la cadena de bloques de Ethereum tiene la capacidad de almacenar registros y, lo que es más importante, ejecutar contratos inteligentes. El término contrato inteligente fue acuñado por Nick Szabo en 1994. Un contrato inteligente es básicamente un protocolo de transacción computarizado que ejecuta los términos del contrato.

Para dar una definición simple, los contratos inteligentes son programas escritos por los usuarios para ser cargados y ejecutados en la cadena de bloques. El lenguaje de programación o scripting para contratos inteligentes se llama Solidity, que es un lenguaje similar a JavaScript.

Ethereum Blockchain usa EVM´s (máquinas virtuales Ethereum), que son básicamente los nodos mineros. Estos nodos son capaces de realizar la ejecución y aplicación de estos programas o contratos inteligentes de manera segura usando criptografía a prueba de manipulaciones de confianza [16].

En el contexto de IoT, se espera que Blockchain basado en contratos inteligentes desempeñe un papel importante en la administración, el control y, lo más importante, la seguridad de los dispositivos [1].

Algunas de las características intrínsecas de Blockchain que pueden ser inmensamente útiles para IoT en general, y para la seguridad de IoT en particular son:

1. Identidad de las cosas.
2. Autenticación de datos e Integridad
3. Autenticación, Autorización y Privacidad.
4. Comunicaciones seguras.

Identidad de las cosas. La IDoT (Identity of Things, identidad de las cosas) para IoT debe abordar una serie de problemas desafiantes de manera eficiente, segura y confiable. El desafío principal se refiere a las relaciones de propiedad e identidad de los dispositivos IoT.

La propiedad de un dispositivo cambia durante la vida útil del mismo, pasando del fabricante al proveedor, distribuidor y usuario [17,18]. La propiedad del usuario de un dispositivo IoT se puede cambiar o revocar, si el dispositivo se revende, se da de baja o ha sido comprometido.

La gestión de los atributos y las relaciones de un dispositivo IoT es otro desafío. Los atributos de un dispositivo pueden incluir fabricante, marca, tipo, número de serie, coordenadas GPS de implementación, ubicación, etc. Además de los atributos, capacidades y características, los dispositivos IoT tienen relaciones. Las relaciones de IoT pueden incluir: dispositivo a humano, dispositivo a dispositivo o dispositivo a servicio. Las relaciones de un dispositivo IoT pueden implementarse, utilizarse, enviarse, venderse, actualizarse, repararse, etc.

Autenticación de datos e integridad. Por seguridad, los datos transmitidos por dispositivos IoT conectados a la red deberían estar firmados criptográficamente por el verdadero remitente, que posee una clave pública y un Globally Unique Identifier (GUID) únicos, para asegurar la autenticación e integridad de los datos transmitidos.

Autenticación, autorización y privacidad. En la actualidad se usan complejos protocolos de autenticación, autorización y administración de dispositivos IoT tales como: Role Based Access Management (RBAC), OAuth 2.0, OpenID, OMA DM y LWM2M.

Comunicaciones seguras. Los protocolos de comunicación de aplicaciones IoT como HTTP, MQTT, CoAP o XMPP, o incluso los protocolos relacionados con el enrutamiento como los de RPL y 6LoWPAN, no son seguros por diseño. Dichos protocolos deben incluirse dentro de otros protocolos de seguridad como DTLS o TLS para aplicaciones de mensajería y para proporcionar una comunicación segura. Por lo tanto, se deben buscar protocolos de seguridad livianos que sean más factibles de usar, considerando los limitados recursos informáticos y de memoria de los dispositivos IoT.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Se analizará cómo Blockchain puede ser una tecnología clave para proporcionar soluciones de seguridad viables a los desafiantes problemas actuales de seguridad de IoT y se propondrán alternativas de seguridad con tecnología Blockchain para los distintos aspectos de seguridad antes analizados: Identidad de las cosas; Autenticación de datos e Integridad; Autenticación, Autorización y Privacidad y Comunicaciones seguras.

Se estudiará si Blockchain tiene la capacidad de resolver estos desafíos de manera fácil, segura y eficiente, basándonos en que la cadena de bloques se ha utilizado ampliamente para proporcionar un registro de identidad confiable y autorizado, para el seguimiento de la propiedad y el monitoreo de productos, bienes y datos.

Se propondrán enfoques como TrustChain [19] para permitir transacciones confiables utilizando Blockchain mientras se mantiene la integridad de las transacciones en un entorno distribuido, ya que Blockchain se puede usar para registrar y dar identidad a dispositivos IoT conectados con un conjunto de atributos y relaciones complejas que se pueden cargar y almacenar en el libro mayor distribuido de Blockchain.

En esa línea de investigación se sabe que Blockchain proporciona una gestión descentralizada y confiable, permitiendo el control y el seguimiento en cada punto de la cadena de suministro y del ciclo de vida de un dispositivo IoT, por lo que se propondrá eliminar la administración y distribución de claves, ya que cada dispositivo IoT tendría un GUID y un par de claves asimétricas únicas cuando se instale y se conecte a la red.

La privacidad de los datos también se puede garantizar mediante el uso de contratos inteligentes que establezcan las reglas de acceso, las condiciones y el tiempo para permitir que ciertos individuos o grupos de usuarios o máquinas posean, controlen o

tengan acceso a los datos en reposo o en tránsito.

Los contratos inteligentes también pueden explicar quién tiene el derecho de actualizar y parchear el software o el hardware de IoT, restablecer el dispositivo IoT, proporcionar nuevos pares de claves, iniciar un servicio o solicitud de reparación y cambiar la propiedad de un dispositivo.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados obtenidos hasta el momento fueron presentados en un artículo aprobado para su publicación en la revista digital ReDDI (Revista Digital del Departamento de Ingeniería), editada por el Departamento de Ingeniería e Investigaciones Tecnológicas (DIIT) de la Universidad Nacional de La Matanza (UNLaM), San Justo, Argentina.

En ese trabajo se hace un análisis de las posibilidades de integración de la tecnología Blockchain con IoT, destacando los desafíos y los beneficios que dicha integración supone. También se presentan las futuras líneas de investigación en la integración de Blockchain con IoT y se deja abierto el interrogante de cómo la combinación de ambas puede proporcionar un entorno más seguro que permita desarrollar nuevos modelos de negocios y aplicaciones distribuidas.

Los objetivos de este proyecto de investigación son:

- Exponer en detalle los conceptos de IoT (Internet de las cosas) y Blockchain y hacer un estudio de su impacto en la sociedad y sus perspectivas de futuro como posibles aplicaciones.
- Hacer un estudio de los artículos de investigación y publicaciones que relacionen estos conceptos, destacando las contribuciones más importantes.
- Analizar la privacidad en el intercambio de datos en las principales tecnologías de IoT y cuáles son las mejores contramedidas para evitar las posibles amenazas.
- Analizar los desafíos y oportunidades de la convergencia entre IoT y la tecnología Blockchain.
- Proponer la convergencia de las tecnologías IoT y Blockchain para aplicaciones seguras.

4. FORMACIÓN DE RECURSOS HUMANOS.

El equipo de investigadores pertenece al cuerpo docente de Tecnologías Aplicadas de la Facultad de Ingeniería, específicamente al área de la Seguridad Informática, de la Universidad del Salvador.

A este proyecto, se incorporaron un docente investigador con amplia experiencia en la industria y 2 alumnos que se encuentran promediando la carrera de Ingeniería en Informática.

Esto redundará en un aumento del activo académico e investigativo representado por su cuerpo de docentes investigadores, como así también sembrará las bases para la investigación a futuro, a través de la participación de alumnos de la Facultad de Ingeniería.

5. BIBLIOGRAFÍA.

[1] Gartner Top 10 Strategic Technology Trends for 2020; 2020. https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2020/.

[2] F. Mattern and C. Floerkemeier; "From the Internet of Computers to the Internet of Things"; ACM Digital Library; From Active Data Management to Event-Based Systems and More; 2010. http://www.vs.inf.ethz.ch/publ/papers/Internet-of-things.pdf.

[3] R. Minerva, A Biru, D. Rotondi; "Towards a definition of the Internet of Things (IoT)"; IEEE Xplore Digital Library; 2015. https://iot.ieee.org/images/files/pdf/IEEE_IoT_Towards_Definition_Internet_of_Things_ Revision1_27MAY15.pdf

[4] Shaddad Abdul-Qawy, P. P. J, E. Magesh, T. Srinivasulu; "The Internet of Things (IoT): An Overview"; Directory of Open Access

Journals; International Journal of Engineering Research and Applications; V.5, N. 12; 2015.

[5] D.Mendez, I. Papapanagiotou, B Yang; "Internet of Things: Survey on Security and Privacy"; Cornell University Library; 2017. https://arxiv.org/abs/1707.01879

[6] J. Granjal, E. Monteiro, J. Sá Silva; "Security for the Internet of Things: A Survey of Existing Protocols and Open Research Issues"; IEEE Xplore Digital Library; IEEE Communications Surveys & Tutorials, 2015, Volume 17, Number 3; http://0-ieeexplore.ieee.org.cataleg.uoc.edu/ document/7005393

[7] A. Bahga, V. K. Madisetti, "Blockchain Platform for Industrial Internet of Things"; Scientific Research; Vol.9, No.10, October 2016. https://www.scirp.org/Journal/Paper Information.aspx?PaperID=71596

[8] E. Gaetani, L. Aniello, R. Baldoni, F. Lombardi, A Margheri, and V. Sassone; "Blockchain-based Database to Ensure Data Integrity in Cloud Computing Environments"; First Italian Conference on Cybersecurity (ITASEC17); Venice, Italy; 2017. http://ceur-ws.org/Vol-1816/paper-15.pdf

[9] G.Zyskind, O. Nathan, A. Sandy Pentland; "Decentralizing Privacy: Using Blockchain to Protect Personal Data"; IEEE Xplore Digital Library; Security and Privacy Workshops (SPW); 2015. http://ieeexplore.ieee.org/ document/ 7163223/

[10] T. Tuan Anh Dinh, R. Liu; "Untangling Blockchain: A Data Processing View of Blockchain Systems"; Cornell University Library; 2017. http://www.comp.nus.edu.sg/ ~ooibc/blockchainsurvey.pdf

[11] Linux-Foundation, Blockchain technologies for business, 2017.
URL https: //www.hyperledger.org/.

[12] C. Kuhlman, What is eris? 2016 edition, 2016. URL https://monax.io/2016/04/03/wtf-is-eris/.

[13] Stellar, Stellar network overview, 2014. URL https://www.stellar.org/developers /guides/get-started/

[14] Ripple, Ripple network, 2013. URL: https://ripple.com/network.

[15] All-In-Bits, Introduction to tendermint, 2017. URL: https://tendermint.com/intro.

[16] K. Christidis, "Blockchains and Smart Contracts for the Internet of Things" IEEE Xplore Digital Library, IEEE Access, Volume 4; 2016. http://0-ieeexplore.ieee.org.cataleg.uoc.edu/document/7467408/

[17] I. Friese, J. Heuer, N. Kong, Challenges from the Identities of Things: Introduction of the Identities of Things discussion group within Kantara initiative, in: 2014 IEEE World Forum on Internet of Things (WF-IoT), 2014, pp. 1–4. http://dx.doi.org/10.1109/WF-IoT. 2014.6803106.

[18] P.N. Mahalle, B. Anggorojati, N.R. Prasad, R. Prasad, Identity authentication and capability-based access control (iacac) for the internet of things, J. CyberSecur. Mobility 1 (4) (2013) 309–348.

[19] P. Otte, M. de Vos, J. Pouwelse, TrustChain: A Sybil-resistant scalable blockchain, Future Gener. Comput. Syst. 2017. http://dx.doi.org/10.1016/j.future.2017.08.048

Avances en Reconocimiento de Patrones de Tecleo para la Identificación de Personas en Ambientes Web

Jorge Ierache, Hernán Merlino, German Concilio, Enrique Calot, Nahuel González,.

Laboratorio de Sistemas de Información Avanzados, Departamento Computación
Facultad de Ingeniería, Universidad de Buenos Aires.
Av. Paseo Colón 850 - C1063ACV - Buenos Aires -Argentina
Tel +54 (11) 4343-0893 / 4343-0092.

{jierache,hmerlino,ecalot,gconcilio}@lsia.fi.uba.ar

Resumen

El Laboratorio de Sistemas de Información Avanzados (LSIA) de la Facultad de Ingeniería de la Universidad de Buenos Aires (FIUBA) cuenta con la línea de investigación *Dinámica de Tecleo*. En el marco de un Proyecto de desarrollo Estratégico (PDE) se han hecho aportes relevantes al área, donde se destaca la integración de los resultados teóricos obtenidos por el LSIA —en un proyecto UBACYT previo— con la industria. Se demostró el potencial de la aplicación práctica de las técnicas desarrolladas en el contexto de un *e-commerce*, adaptando los algoritmos de autenticación e identificación de personas por su ritmo de tecleo para su utilización en un entorno web. En un entorno web no se tiene un control sobre la población de usuarios que utilizan la plataforma y hay una gran variedad de tipos de dispositivos de entrada que afectan la efectividad final del método.

Palabras clave: *LSIA, Dinámica de tecleo, Keystroke Dynamics, ambientes web, e-commerce*

Contexto

El LSIA fue creado en el 2011 y cuenta con diversas publicaciones [LSIA, 2020]. Dentro de sus líneas de investigación, se le presta especial atención a la línea de *Dinámica de Tecleo*. En 2019 se inició un Proyecto de desarrollo Estratégico (PDE) vinculado a la transferencia a la industria reconocido por la UBA (PDE-44-2019). Su objetivo es estudiar el reconocimiento de patrones de dinámica de tecleo en ambientes web, en particular en el contexto e-commerce.

El proyecto PDE utiliza como base de los algoritmos desarrollados en un proyecto UBACYT 20020130200140BA finalizado. EL proyecto estaba centrado en los métodos de educción de cadencia de tecleo centrado en el contexto emocional de un individuo aplicando Interfaces Cerebro-Máquina (BMI). A partir del mismo, se generaron bases de conocimiento fundamentales en el área de reconocimiento de patrones de tecleo y se sentaron las bases del algoritmo que luego se adaptó al contexto de aplicación práctica propuesto en el proyecto PDE.

Introducción

Los patrones neurofisiológicos que vuelven única a una firma manuscrita se

pueden observar también en el ritmo de tecleo de un usuario [Joyce & Gupta, 1990]. La técnica que analiza este tipo de patrones se llama Dinámica de Tecleo o Keystroke Dynamics [Calot et al. 2014], [González et al. 2015, 2016]. Desde el LSIA se estudian las técnicas de identificación y autenticación de personas por su ritmo de tecleo.

En este artículo se enuncian los avances en materia de reconocimiento de patrones de tecleo para la identificación de personas en ambientes web, en particular en el contexto e-commerce.

Resultados y Objetivos

Se realizó un trabajo colaborativo con una de las principales empresas de *e-commerce* de América Latina, quien dio acceso a un conjunto de datos de experimentación. Se obtuvo un conjunto de datos de más de 2000 usuarios reales la plataforma elegidos al azar, sin sesgo.

La base del algoritmo utilizado para identificar usuarios a partir de su ritmo de tecleo es producto del del proyecto UBACYT 20020130200140BA [Calot et al., 2013, 2014, 2015], [González et al. 2015, 2016]. Se han hecho adaptaciones pertinentes al algoritmo para mejorar su efectividad en los entornos web.

Para aumentar la calidad de los datos de entrenamiento se realizó un refinamiento del lote de pruebas. Se configuró un mínimo de teclas alfanuméricas presionadas en una sesión y se descartaron las sesiones preparadas offline, es decir, cuando la secuencia capturada es sólo una combinación de las teclas CTRL + V. Este tipo de sesiones no reflejan el normal comportamiento del usuario al teclear y en consecuencia las muestras simplemente se descartaron.

Las pausas del individuo al teclear tampoco reflejan su cadencia característica y son consideradas fuentes de ruido. Para reducir esta fuente de ruido de los datos, se particiona el texto de entrada utilizando dos estrategias: por delimitadores de tecla o separador de palabras, y por intervalos de silencio máximos –parámetro configurable en el algoritmo-. Además se filtran todas aquellas particiones con menos de un mínimo de teclas alfanuméricas presionadas.

Las letras y números son generalmente utilizados para escribir secuencias que se repiten con frecuencia y por lo tanto tienden a ser ejecutadas con una cadencia relativamente estable. Contrariamente, la mayoría de las teclas especiales presentan patrones que distan notablemente de la predictibilidad. Sin embargo, la consideración de las teclas especiales en los contextos de las teclas comunes mejora la clasificación.

Para probar el algoritmo de distancia entre usuarios enrolados se asumió que la persona real detrás de un usuario es siempre es la misma. Para cada usuario, sus sesiones se subdividieron en dos subconjuntos mutuamente exclusivos, etiquetados como A y B. De esta manera, generamos dos patrones con la misma identidad de usuario.

Utilizamos varias muestras de usuarios para evaluar el comportamiento de distintos métodos de cálculo de distancia entre usuarios con distintos aspectos personalizables. Por ejemplo, para la un grupo A inicial de 40 usuarios, se obtuvieron 80 patrones, dos por usuario, uno perteneciente al grupo ~A y otro a ~B. A través del modelado de cadencia de tecleo de cada patrón, intentamos descubrir qué otro patrón es más probable que comparta el mismo identificador de usuario. Contamos con un éxito cuando identificamos el usuario del conjunto ~A con la menor distancia de patrón para el mismo usuario del conjunto B. Además, se estudió cómo influyen la cantidad de texto

ingresado por el usuario y la cantidad de usuarios enrolados en la plataforma en el porcentaje de acierto del método.

Se ha logrado demostrar que se puede distinguir a una persona de un lote de 2000 usuarios con una efectividad superior al 89% utilizando únicamente nuestro algoritmo de identificación por cadencia de tecleo [Concilio et al., 2018]. Además, en un conjunto de 100 usuarios seleccionados al azar se pudo autenticar sesiones con un EER de 7,82 %. El resultado obtenido es comparable con resultados de otros autores del estado del arte, a pesar de que el set de datos presenta dificultades extra, como la no heterogeneidad de los teclados usados para tipiar, sistemas operativos o el nivel educativo de los individuos.

Por último, se propuso un algoritmo novedoso para asociar usuarios con una huella biométrica característica de la persona real. Para probar este algoritmo se generó un modelo de cadencia de tecleo para cada usuario del conjunto de pruebas. Luego se calculó la distancia definida por el algoritmo a cada usuario del conjunto de datos y se tomó un umbral de sensibilidad debajo del cual dos usuarios se dice que la misma huella biométrica. Decimos que un usuario tiene un cruce con otro usuario cuando comparten la misma huella biométrica.

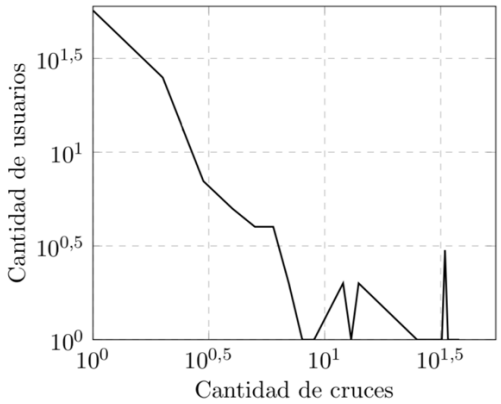


***Figura 1**. Distribución de la cantidad de cruces generados por el algoritmo*

La Figura 1 muestra que aproximadamente un 60% de los usuarios no tienen cruces con ningún usuario –notar que siempre un usuario cruza consigo mismo por definición del experimento–. En un proceso de revisión empírica manual se determinó que este grupo de usuarios en su mayoría no tienen indicios de compartir cuenta con otro usuario de la plataforma.

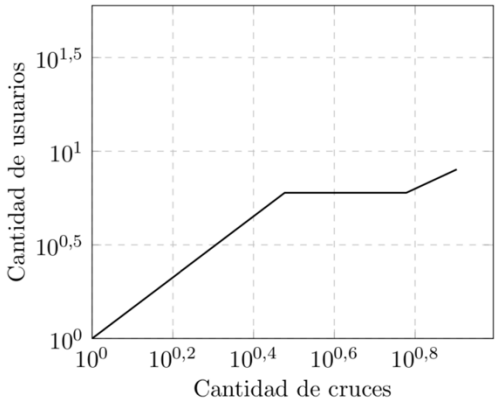


***Figura 2**. Distribución de cantidad de cruces para los usuarios con comportamiento fraudulento previamente detectado.*

Los usuarios con más de 7 cruces representan el 8% de la población de usuarios evaluada. Se determinó en un proceso de revisión empírica que existía realmente una misma persona física detrás de esos usuarios. Dentro de este grupo de usuarios con muchos cruces, se lograron aislar 30 usuarios con comportamiento fraudulento confirmados. Se concluye que los resultados experimentales muestran que es posible detectar usuarios controlados por una misma persona física, con interés fraudulento, a través de patrones de tecleo compartidos por varios usuarios de la plataforma [Concilio et al., 2018].

En otro estudio se ha logrado demostrar que la distancia euclediana es la mejor opción a la hora de querer mitigar el ruido producido por las emociones de las personas al momento de teclear [Calot, et al., 201]. Además, en [Calot et al., 2019 b]

de han evaluado mejoras al algoritmo base original, concluyendo que una distancia específica de Minkowsky reduce el EER de 21,9% a 17,4% con el set de datos de experimentación en condiciones reales.

Formación de Recursos Humanos

El laboratorio actualmente se conforma de dos investigadores formados, un investigador formado invitado, dos estudiantes de doctorado, un alumno investigador. **S**e han radicado en la temática de cadencia de tecleo una tesis de doctorado y una tesis de grado de la Facultad de Ingeniería de la UBA ambas finalizadas y defendidas.

Referencias

CALOT, E. 2015. "Keystroke Dynamics keypress latency dataset". Base de datos para investigación. http://lsia.fi.uba.ar/pub/papers/kd-dataset/

CALOT, E.; PIRRA, F.; RODRIGUEZ, J.M.; PEREIRA, G.; IRIBARREN, J.; IERACHE, J. 2014. "Métodos Adaptativos de Educción de Dinámica de Tecleo Centrado en el Contexto Emocional de un Individuo aplicando Interfaz Cerebro Computadora". XVI Workshop de Investigadores en Ciencias de la Computación, ISBN 978-950-34-1084-4.

CALOT, E.; RODRIGUEZ, J.M.; IERACHE, J. Improving versatility in keystroke dynamic systems. En Proceedings del XIX Congreso Argentino de Ciencias de la Computación, number 5606, 2013. ISBN 978-987-23963-1-2. URL http://hdl.handle.net/10915/32428

CALOT, E.; RODRIGUEZ, J.M.; IERACHE, J., 2014. Improving versatility in keystroke dynamic systems. En Jorge Raúl Finochietto y Patricia Mabel Pesado, editors, Computer Science & Technology Series. XIX Argentine Congress of Computer Science, Selected papers, páginas 289–298. Editorial de la Universidad Nacional de La Plata (EDULP), 2014b. ISBN 978-987-1985-49-4. URL http://lsia.fi.uba.ar/papers/calot14b.pdf

CALOT, E., ROSSI F., GONZÁLEZ N., HASPERUÉ, W., IERACHE, J.. Avances en educción de dinámica de tecleo y el contexto emocional de un individuo aplicando interfaz cerebro computadora. En WICC 2016, Entre Ríos, Argentina), páginas 872–876, jun. 2016. ISBN 978-950-698-377-2. URLhttp://hdl.handle.net/10915/53247

CALOT, E., IERACHE, J.. Multimodal biometric recording architecture for the exploitation of applications in the context of affective computing. En Proceedings del XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017), number 10529, páginas 1030–1039, 2017. ISBN 978-950-34-1539-9. URL http://hdl.handle.net/10915/63866

CALOT, E., IERACHE, J., HASPERUÉ, W.. 2019. Robustness of keystroke dynamics identification algorithms against brain-wave variations associated with emotional variations. En Advances in Intelligent Systems and Computing. Springer, c. En prensa

CALOT, E., IERACHE, J., HASPERUÉ, W.. 2019. Document Typist Identification by Classification Metrics Applying Keystroke Dynamics Under Unidealised Conditions. En International Conference on Document Analysis and Recognition Workshops (ICDARW) IEEE, 2019. p. 19-24.

CONCILIO, G., IERACHE, J., MERLINO, H., CALOT, E.. Application of Keystroke Dynamics Modelling Techniques to Strengthen the User Identification in the Context of E-commerce. En XXIV CACIC 2018. En prensa. URL http://lsia.fi.uba.ar/papers/concilio18.pdf

JOYCE, R.; GUPTA, G. 1990. Identity authentication based on keystroke latencies. Commun. ACM 33, 2 (February 1990), 168-176. http://doi.acm.org/10.1145/75577.75582

GONZÁLEZ, N., CALOT , E. Of Keystroke Dynamics In Free Text. En Biometrics Special Interest Group (Biosig), 2015 International Conference Of The, Páginas 1–5, Sep. 2015. ISBN 978-3-88579-639-8. DOI: 10.1109/Biosig.2015.7314606

GONZÁLEZ, N., CALOT, E. Y IERACHE, J... A Replication Of Two Free Text Keystroke

Dynamics Experiments Under Harsher Conditions. En 2016 International Conference Of The Biometrics Special Interest Group (Biosig), Páginas 1–6, Sep. 2016. DOIi: 10.1109/Biosig.2016.7736905

LSIA: http://lsia.fi.uba.ar vigente marzo 2020

Implementación de un nodo minero institucional en la red Ethereum Blockchain Federal Argentina

Jorge Eterovic; Jonatan Uran Acevedo; Alejandro Rusticcini; Nora Gigante

Programa PROINCE / Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza
Florencio Varela 1903 (B1754JEC), San Justo, (5411) 4480-8900

eterovic@unlam.edu.ar; juran@unlam.edu.ar; arusticcini@unlam.edu.ar; ngigante@unlam.edu.ar

RESUMEN

Blockchain Federal Argentina (BFA) es la primera plataforma multiservicios abierta y participativa de Argentina pensada para integrar servicios y aplicaciones sobre la Blockchain de Ethereum.

Las organizaciones, tanto públicas como privadas, pueden formar parte de Blockchain Federal Argentina, ejerciendo distintos roles de control dentro de la organización. También pueden desplegar aplicaciones sobre la plataforma.

El proyecto de investigación consiste en implementar un nodo Minero (en adelante nodo Sellador) en la Universidad Nacional de La Matanza, dentro de la red Blockchain Federal Argentina, para lo cual se celebrará un contrato de colaboración público-privada entre la UNLaM, representada por el Departamento de Ingeniería e Investigaciones Tecnológicas (DIIT) y el consorcio Blockchain Federal Argentina.

Este contrato hará a la UNLaM parte del mencionado consorcio y permitirá montar un nodo Sellador dentro de la infraestructura Blockchain-BFA. En el marco de este proyecto, se desarrollará e implementará una dApp (Aplicación Distribuida) que quedará disponible para su uso libre para toda la comunidad académica.

La relevancia de este trabajo radica en la importancia que tiene para una institución formar parte de una red con las características de BFA. Hasta el momento, la UNLaM, no forma parte de BFA ni de otras plataformas similares.

Palabras clave:
Ethereum. Blockchain. Nodo Sellador. Contrato Inteligente. BFA.

CONTEXTO

Este proyecto de investigación está siendo presentado como un Programa de Incentivos a Docentes Investigadores de la Secretaría de Políticas Universitarias (PROINCE) en el Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza.

El presente proyecto es del tipo investigación aplicada y consiste en el desarrollo e implementación de un nodo sellador en la UNLaM dentro de la red Blockchain Federal Argentina y de una dApp para uso académico.

1. INTRODUCCIÓN

Blockchain, en español cadena de bloques, es una tecnología que permite administrar un registro de datos en la nube. Tiene como característica la transparencia y es prácticamente incorruptible [1].

Una Blockchain puede ser vista como un gran libro contable. Allí solo pueden ingresarse nuevas entradas y las entradas anteriores no pueden ser modificadas ni eliminadas. Esas entradas se llaman transacciones, las cuales se van agrupando en bloques que se agregan sucesivamente a una cadena [2].

Cada uno de los bloques hace referencia al bloque inmediatamente anterior de modo que, si alguna transacción intenta ser modificada, esa referencia cambia y ese bloque y todos los posteriormente agregados son inválidos.

Por lo anteriormente dicho, si quisiéramos corregir información ya registrada, solo lo podemos hacer mediante el agregado de nueva información. Los datos originales siempre van a permanecer en la cadena y pueden ser inspeccionados en cualquier momento.

Blockchain puede ser vista también como una base de datos pública y distribuida que contiene un histórico irrefutable de información. La Blockchain (esa cadena de nodos compuesta por transacciones) no se encuentra almacenada en un solo servidor centralizado, sino que se encuentra replicada en un gran conjunto de dispositivos conocidos como nodos que conforman lo que se conoce como red de pares.

Cada vez que se agrega una nueva transacción, ésta se integra a un bloque y posteriormente se agrega a la cadena y ésta es actualizada en todas las réplicas de los nodos. Blockchain no solo está protegida por este modelo de red descentralizada, sino que también está validada por métodos criptográficos que garantizan que nada pueda ser borrado o alterado sin que todos los usuarios puedan darse cuenta de ello.

Blockchain permite garantizar la identidad de las partes involucradas, ya que todas las transacciones son firmadas criptográficamente. Se puede certificar la fecha y hora de cada transacción. La información es inmutable e inalterable. Además, toda la información almacenada en la cadena es completamente auditable. Blockchain funciona sin intermediarios, esto es, no hace falta una persona, empresa o institución que legitime la información guardada en la cadena.

Ethereum es una plataforma descentralizada de código abierto (open source), que permite la creación de contratos inteligentes sobre una blockchain. En diciembre de 2013, Vitalik Buterin comenzó el desarrollo de Ethereum, con la primera prueba de concepto (PdC) [3].

Ethereum provee una criptomoneda que se llama “ether”. Se pueden intercambiar ether entre cuentas diferentes (es decir, puede ser utilizado como intercambio de valor). Pero existe una bifurcación de la cadena de bloques de Ethereum a partir de julio de 2016, que dio como resultado dos líneas de Ethereum activas: Ethereum y Ethereum Clásico.

Ethereum funciona de manera descentralizada a través de una máquina virtual llamada Ethereum Virtual Machine (EVM). Esta máquina ejecuta un código intermedio o bytecode el cual es una mezcla de lenguaje de programación LISP, un ensamblador y bitcoin script [4].

Los programas que realizan contratos inteligentes son escritos en lenguajes de programación de alto nivel de tipo Turing completos, como Solidity, que es un lenguaje de alto nivel orientado a contratos. Su sintaxis es similar a la de JavaScript y está enfocado específicamente a la EVM para crear los contratos inteligentes [5].

Un contrato inteligente (en inglés Smart Contract) es un programa informático que ejecuta un flujo de trabajo que generalmente representa acuerdos registrados en una Blockchain, entre dos o más partes (por ejemplo, personas u organizaciones) [6]. Dichos contratos se ejecutarán como resultado de que se cumplan una serie de condiciones especificadas previamente.

Un contrato inteligente es un programa que “vive” en un sistema no controlado por ninguna de las partes, y que ejecuta un contrato automático el cual funciona como una sentencia

if-then (si-entonces) de cualquier otro programa de computadora. Cuando se dispara una condición preprogramada, no sujeta a ningún tipo de valoración humana, el contrato inteligente ejecuta la cláusula contractual correspondiente.

Los Smart Contract tienen como objetivo brindar una seguridad superior a un contrato tradicional y reducir los costos de transacción asociados a la contratación. La transferencia de valor digital mediante un sistema que no requiere confianza (por ejemplo, bitcoins) abre la puerta a nuevas aplicaciones que pueden hacer uso de los contratos inteligentes.

Los contratos inteligentes se componen de una interfaz de usuario y a veces emulan la lógica de las cláusulas contractuales.

Los desarrolladores pueden escribir la lógica de negocio y acuerdos en forma de contratos inteligentes, los cuales se ejecutan automáticamente cuando sus condiciones son satisfechas por ambas partes e informadas a la red. Estos contratos pueden almacenar datos, enviar y recibir transacciones e incluso interactuar con otros contratos, independientemente de cualquier control.

Solidity es un lenguaje de programación orientado a objetos utilizado para escribir contratos inteligentes en la plataforma Ethereum. Fue desarrollado por Gavin Wood y otros programadores [4]. Es un lenguaje de scripting tipado estáticamente. Esto quiere decir que las variables deben ser declaradas junto con su tipo antes de ser utilizadas. Se hace el proceso de verificar y hacer cumplir las restricciones en tiempo de compilación, antes de que se ejecute el programa.

Cuenta con un IDE oficial llamado Remix. Un IDE (Integrated Development Environment, entorno de desarrollo integrado), es una aplicación que proporciona servicios para facilitarle al programador el desarrollo de software [7].

Remix es un entorno de desarrollo, compilación y despliegue de contratos inteligentes basado en un navegador web.

Una dApp es una aplicación distribuida sobre la Ethereum Blockchain. Esta tiene múltiples capas y componentes y no depende de un sistema centralizado, sino que depende de la comunidad de usuarios que la utiliza. Puede ser Web o Mobile. Una dApp es una aplicación que tiene su Back-end construido sobre contratos inteligentes, en contraposición con los Back-end tradicionales [8].

Blockchain Federal Argentina es una plataforma multiservicios abierta y participativa pensada para integrar servicios y aplicaciones sobre blockchain [9]. Una iniciativa confiable y completamente auditable que permite optimizar procesos y funciona como herramienta de empoderamiento para toda la comunidad.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

En el presente proyecto de investigación, se estudiarán y analizarán los derechos y obligaciones emanados de la firma del contrato de colaboración público-privada que se debería celebrar con Blockchain Federal Argentina.

Luego de firmado el acuerdo, se procederá a instalar el hardware necesario para montar el nodo sellador. Seguido a esto, se implementará el software para el correcto funcionamiento del nodo.

Asimismo, se desarrollará e implementará una dApp (Aplicación Distribuida) en los servidores de la UNLaM. Esto se hará mediante el desarrollo de un Contrato Inteligente (Smart Contract), siguiendo con el desarrollo de una API y por último el diseño, desarrollo e implementación de una aplicación Front-end.

Se escribirán y presentarán informes de avances que incluyan el progreso del proyecto y las conclusiones de cada una de las actividades que forman parte del mismo.

Se redactará un informe integral final con el contrato y el software implementado y desarrollado acompañado de recomendaciones y buenas prácticas como conclusión del trabajo de investigación realizado.

3. RESULTADOS OBTENIDOS/ESPERADOS

El objetivo principal de este proyecto de investigación es implementar un nodo Sellador dentro de Blockchain Federal Argentina (BFA).

El objetivo secundario es desarrollar e implementar una dApp (Aplicación Distribuida) perteneciente a la UNLaM.

El objetivo principal incluye la celebración de un contrato de colaboración público-privada entre la Universidad Nacional de La Matanza, representada por el Departamento de Ingeniería e Investigaciones Tecnológicas (DIIT) y el consorcio Blockchain Federal Argentina (BFA). Esto otorgará los permisos necesarios por parte de Blockchain Federal Argentina para montar un nodo Sellador sobre su red, perteneciente a la UNLaM. Luego se procederá a su implementación.

El objetivo secundario incluye el desarrollo e implementación de una dApp (Aplicación Distribuida) perteneciente a la UNLaM. Dicha dApp funcionará sobre la Blockchain de BFA.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo está integrado por docentes-investigadores que pertenecen a distintas cátedras de la carrera de Ingeniería en Informática y de la Tecnicatura de Aplicaciones Web de la UNLaM, alguno de los cuales está haciendo sus primeras experiencias en investigación.

Uno de los miembros del equipo de investigación se encuentra desarrollando su trabajo de tesis de posgrado de la Maestría en Ciberdefensa y Ciberseguridad de la Universidad de Buenos Aires y su tutor es el Mg. Jorge Eterovic, integrante del proyecto de investigación.

5. BIBLIOGRAFÍA

[1] Albert Szmigielski; Bitcoin Essentials; ISBN 978-1-78528-197-6; Ed. Packt Publishing Ltd.; Birmingham, UK. 2016

[2] Mohammad Dabbagh, Mehdi Sookhak, Nader Sohrabi Safa; The Evolution of Blockchain: A Bibliometric Study; IEEE Access PP (99):1-1; 2019

[3] Vitalik Buterin; A Next Generation Smart Contract & Decentralized Application Platform; 2020. https://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf

[4] Gavin Wood; Ethereum: A secure decentralized generalized transaction ledger; Ethereum project yellow paper, 2014.

[5] Chris Dannen; Introducing Ethereum and Solidity; ISBN-13 (pbk): 978-1-4842-2534-9; Ed. Springer Science; New York, USA. 2017.

[6] Loi Luu, Duc-Hiep Chu, Hrishi Olickel, Prateek Saxena, Aquinas Hobor; Making Smart Contracts Smarter; CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security; Pages 254–269. 2016.

[7] Susan Elliott Sim, Rosalva E. Gallardo-Valencia; Finding Source Code on the Web for Remix and Reuse; ISBN 978-1-4614-6595-9: Ed. Springer Science; New York, USA. 2013.

[8] Andrea Pinna, Simona Ibba, Gavina Baralla, Roberto Tonelli, Michele Marchesi¸ A Massive Analysis of Ethereum Smart Contracts. Empirical study and code metrics. DOI: 10.1109/ACCESS.2019.2921936. IEEE Access. 2019.

[9] Blockchain Federal Argentina; 2020; htpps://www.bfa.org

Avances en Aspectos de Seguridad Aplicados a Sistemas de Voto Electrónico

Pablo García[1] Silvia Bast[1] Germán Montejano[2] Martín Lobos[1]

[1]Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad Nacional de La Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-425166– Int. 28
[pablogarcia, silviabast]@exactas.unlpam.edu.ar

[2]Departamento de Informática
Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 – (5700) San Luis – San Luis – Argentina
Tel.: +54-2652-424027 – Int. 251
gmonte@unsl.edu.ar – web: http://www.unsl.edu.ar

RESUMEN

El debate sobre las fortalezas y debilidades del voto electrónico se encuentra en un momento de fuerte polémica, no solamente en la sociedad en general, sino también en el ámbito académico. Las argumentaciones a favor y en contra de su aplicación son muy diversas.

Los que opinan a favor, se basan en la velocidad con la que se conocen los resultados y en la (supuesta) exactitud del proceso. Los que se oponen, afirman que resulta imposible asegurar la transparencia de los sistemas de voto electrónico.

La postura asumida por este equipo de investigación, es que se trata de un sistema de seguridad crítica, y que la confianza del electorado es de máxima importancia para lograr su aceptación, tal como afirman McGaley y Gibson en [1]: "Un sistema de votación es tan bueno como el público cree que es". Este grupo de trabajo percibe estos sistemas como objetos de investigación, por lo tanto, está dedicado al análisis y evaluación de las condiciones de seguridad que deben cumplir y también al estudio de las soluciones que diferentes autores han propuesto hasta el momento, para intentar generar un modelo que facilite el desarrollo de un sistema robusto y confiable.

Se siguen, en forma paralela, dos líneas de trabajo que representan esquemas diferentes que pueden aplicarse a los sistemas de voto electrónico:

a. Basado en criptografía homomórfica.
b. Basado en criptografía One Time Pad.

En este trabajo se exponen los avances que se llevaron a cabo para cada una de las mencionadas líneas.

Palabras clave: Sistemas de Voto Electrónico, Anonimato, Transparencia, Criptografía Homomórfica, One Time Pad, Verificabilidad E2E, Prueba Física.

CONTEXTO

El presente trabajo pertenece al ámbito del Proyecto de Investigación: "Aspectos de Seguridad en Proyectos de Software", que se desarrolla en la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa (Resolución N° 488/14 del Consejo Directivo de la Facultad de Ciencias Exactas y Naturales) y es dirigido por el Doctor Germán Antonio Montejano (Universidad Nacional de San Luis) y codirigido por el Magister Pablo Marcelo García (FCEyN - UNLPam) e incluye a la Magister Silvia

Gabriela Bast, al Magister Daniel Vidoret, al Analista Programador Adrián García y al Programador Superior Claudio Ponzio como investigadores.

Surge desde la línea de Investigación "Ingeniería de Software y Defensa Cibernética", presentada en [2], que a su vez se enmarca en el Proyecto "Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la Profesión de Ingeniero de Software" de la Facultad de Ciencias Físico - Matemáticas y Naturales de la Universidad Nacional de San Luis (UNSL) (http://www.sel.unsl.edu.ar/pro/proyec/2012/index.html) y que incluye acciones de cooperación con la Universidad Federal de Minas Geráis (UFMG, Brasil).

1. INTRODUCCIÓN

a. Modelo Basado en Criptografía Homomórfica

La criptografía homomórfica presenta una característica muy conveniente a los efectos de implementar sistemas de voto electrónico: permite realizar operaciones sobre los datos cifrados directamente, sin necesidad de descifrarlos ni de conocer ninguna clave. Ese atributo le otorga un gran atractivo a su aplicación práctica.

Los primeros modelos de este estilo se originan a fines de los ´70. Ellos se denominan parcialmente homomórficos, porque sólo permiten un tipo de operación sobre los datos cifrados (suma o multiplicación). Son ejemplos de este tipo El Gamal [3], Benaloh [4] y Paillier [5].

A partir de [6] se produce un avance decisivo en la materia, porque se presenta el primer esquema homomórfico completo, es decir que soporta operaciones de adición y producto sobre datos cifrados.

El modelo a implementar, además de seleccionar el esquema criptográfico apropiado, deberá cumplir con una serie de requisitos que se exigen actualmente a los sistemas de votación electrónica:

- Evidencia física que garantice la transparencia del proceso [7].
- Utilización de métodos criptográficos cuya seguridad incluya formalidad matemática, tanto en lo referente al anonimato del votante como a la transparencia de los resultados de los comicios.
- Aplicación del concepto de independencia del software [8].
- Definición de un modelo concreto para la aplicación de verificabilidad "End to End" (E2E) [9].
- Aplicación de técnicas que eviten que un votante pueda demostrar cuál fue su elección.
- Selección de una interface apropiada, con un fuerte análisis de alternativas y una fundamentación sobre la elección. Esto supone un desafío relacionado con características tales como la usabilidad y accesibilidad, pero sobre todo debe tener en cuenta que del diseño dependen determinados comportamientos electorales. En efecto, el formato final de la interfaz no es trivial. Según [10], "Es lógico esperar que la tecnología que se utiliza y las características específicas de la misma tengan un efecto potencial sobre el comportamiento de los actores que interactúan con ella".

b. Modelo basado en One Time Pad (OTP-Vote)

El modelo OTP- Vote se describe en [11] y focaliza especialmente en la confidencialidad e integridad de los datos de un sistema de voto electrónico. Se basa en la siguiente premisa fundamental: en los sistemas de voto electrónico es necesario proteger:

- Indefinidamente, la privacidad del votante, aún después de finalizada la elección.
- Mientras dure el proceso electoral, la seguridad de los datos de los votos, luego la información se hace pública.

El modelo hace uso de:

Claves One Time Pad, que cumplen con las hipótesis y condiciones del "Secreto Perfecto" de Shannon [12]: son aleatorias y tan largas como el mensaje.

Archivos de Datos que Almacenan Bits, que son elementos básicos en el modelo propuesto y se modifican en el transcurso del proceso, ellos son:

• Archivo Binario de Votos (ABV) cuya generación está basada en el modelo de almacenamiento Múltiples Canales Dato Único (MCDU) propuesto por García en [13], que se analiza en [14] y [15] y surge como una propuesta de resolución a las limitaciones de Birthday Paradox [16].

• Clave de Descifrado (CD): surge a partir de operaciones XOR (⊕) [17] de claves OTP.

Tablas Relacionales: que almacenan los datos de la configuración de la elección, cargos, candidatos e identificadores de votos y de los votos planos resultantes del proceso eleccionario.

El modelo propone:

- Anonimato incondicional.
- Seguridad computacional que puede llevarse a cualquier nivel exigible durante el proceso eleccionario.

El proceso incluye las etapas de:

- Configuración de la elección.
- Desarrollo de la elección.
- Cierre de la elección y recuento de votos.

El modelo teórico presentado supone, para cada una de las etapas mencionadas, el cumplimiento de algunas condiciones que resultan imprescindibles para alcanzar el normal funcionamiento del sistema. Estas condiciones se relacionan con aspectos tales como:

- La inalterabilidad de algunos elementos de datos durante el desarrollo del proceso electoral.
- El control de modificación de algunos elementos por determinados procesos.
- El desarrollo del proceso de auditoría en las diferentes etapas.
- El aseguramiento de la comunicación entre usuario y sistema en los momentos en que la misma se produce.
- El aseguramiento de la comunicación que realiza la transmisión de datos entre estaciones y servidor.

El modelo requiere entonces de la especificación de los puntos mencionados, para demostrar que el sistema resultante de la investigación es confiable y seguro.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El grupo de trabajo investiga, básicamente sobre dos líneas paralelas para generar modelos que pudieran aplicarse a los sistemas de voto electrónico:

- Basados en criptografía homomórfica, que es la línea a cargo del Magíster Pablo Marcelo García.
- Basados en criptografía One Time Pad, que es la línea que lleva adelante la Magíster Silvia Gabriela Bast.

3. RESULTADOS Y OBJETIVOS

Los avances del grupo de trabajo que han surgido durante 2019 fueron:

a. En el ámbito de la criptografía homomórfica, este grupo de trabajo ha realizado las siguientes acciones:

   - Análisis de los resultados obtenidos en una encuesta online destinada a personas de todo nivel (desde expertos informáticos hasta votantes comunes) para obtener opiniones sobre la forma concreta que debería tener la interface de un sistema de voto electrónico.
   - Desarrollo de una serie de entrevistas a expertos informáticos para complementar los resultados obtenidos en la encuesta online. Esto se está realizando en la actualidad.
   - Incorporación al proyecto de dos especialistas específicos en comunicaciones de datos para proporcionar metodologías de transmisión de datos que garanticen los niveles de seguridad

exigibles. Los mismos se encuentran realizando el análisis de alternativas para implementar un esquema de basado en Virtual Nets. El objetivo será proporcionar un modelo de comunicación que cumpla con los requisitos.
- Se continuó trabajando en el análisis de métodos homomórficos existentes, avanzando en la selección final del esquema definitivo.

A futuro, se pretende llevar a cabo las siguientes acciones:

- Elegir la interface exacta del modelo en base a los datos obtenidos en encuestas y entrevistas.
- Aplicar la interface seleccionada y publicar los fundamentos de la selección.
- Definir un modelo de transmisión de datos e implementarlo.
- Realizar un relevamiento de aplicaciones orientadas al voto electrónico, que permita detectar falencias y proponer mejoras en el nuevo modelo.
- Selección de un método basado en criptografía homomórfica para aplicar en una futura implementación.

b. En cuanto a la línea de OTP Vote se ha avanzado sobre:
   - Aspectos de auditoría del modelo en cada una de las etapas del proceso.
   - Propuesta de Verificabilidad End to End.
   - Mejoras de la integridad de datos en el modelo, mediante la inclusión de bits de control que permiten verificar la integridad de los datos y detectar posibles intentos de intrusión. Modelo OTP-Vote
   - Se profundizó el trabajo sobre el refinamiento de los procesos de recuperación y generación de votos planos.

Se continuará avanzando en esta línea de investigación en aspectos tales como:

- Aseguramiento de la comunicación entre usuario-sistema y sistema-servidor de datos
- Refinamiento de la propuesta de modelo de auditoría de terceros.
- Automatización del proceso de generación de las tablas relacionales.
- Refinamiento de la propuesta de Verificabilidad End to End.

4. FORMACIÓN DE RECURSOS HUMANOS

En el marco del presente proyecto se presentan los siguientes puntos relacionados con la formación de recursos humanos:

- Pablo García y Silvia Bast completaron el cursado de la totalidad de los créditos exigidos en el Doctorado en Ingeniería Informática en la Facultad de Ciencias Físico Matemáticas y Naturales de la Universidad Nacional de San Luis (UNSL).
- Pablo García y Silvia Bast presentaron las modificaciones oportunamente exigidas a su Plan de Tesis Doctoral, en el marco del Doctorado en Ingeniería Informática en la Facultad de Ciencias Físico Matemáticas y Naturales de la Universidad Nacional de San Luis (UNSL). Las mismas se encuentran en proceso de evaluación.

5. BIBLIOGRAFÍA

[1] **McGaley M., Gibson J.:** "A critical analysis of the council of Europe recommendations on e-voting". EVT'06: Proceedings of the USENIX/Accurate Electronic Voting Technology Workshop

2006 on Electronic Voting Technology Workshop. 2006.
[2] Uzal R., van de Graaf J., Montejano G., Riesco D., García P.: "Inicio de la Línea de Investigación: Ingeniería de Software y Defensa Cibernética". Memorias del XV WICC. Ps 769-773. ISBN: 9789872817961. 2013.http://sedici.unlp.edu.ar/handle/10915/27537.
[3] **El Gamal T**. "A public key cryptosystem and a signature scheme based on discrete logarithms". In Proceedings of CRYPTO 84 on Advances in cryptology, pages 10–18. Springer-Verlag New York, Inc. 1985.
[4] **Benaloh, J.:** "Dense Probabilistic Encryption". Workshop on Selected Areas of Cryptography. pp. 120–128. 1994.
[5] **O'Keeffe M.**: "The Paillier Cryptosystem: A Look Into The Cryptosystem And Its Potential Application". The College of New Jersey Mathematics Department. 2008.
[6] **Gentry G.:** "Fully Homomorphic Encryption Using Ideal Lattices". In the 41st ACM Symposium on Theory of Computing (STOC), 2009.
[7] **Hao, F, Ryan P.:** "Real -World Electronic Voting. Design, Analysis And Deployment". Cr Press. ISBN-13: 978- 1498714693. ISBN-10: 1498714692. 2017.
[8] **Rivest R**.: "On the notion of 'software independence' in voting systems". Philosophical Transactions of The Royal Society A, 366(1881):3759–3767. 2008.
[9] **Kelsey J., Regenscheid A., Moran T., Chaum D.**: "Attacking Paper-Based E2E Voting Systems". In: Chaum D. et al. (eds).
[10] **Ruiz Nicolini, J.:** "La(s) boleta(s) única(s)".https://www.elestadista.com.ar/?p=7104. 2015.
[11] **Bast S:**. "Confidencialidad e Integridad de Datos en Sistemas de E-Voting – Un Modelo para la Implementación Segura de un sistema de Voto Presencial" - Editorial Académica Española. https://www.eae-publishing.com-ISBN 978-3-639-53793-2. 2017.
[12] **Shannon, C.:** "Communication Theory of Secrecy Systems" - Bell System Technical Journal - 1949.
[13] **García, P.**: "Una Optimización para el Protocolo Non Interactive Dining Cryptographers" - Editorial Académica Española (https://www.eae-publishing.com/ - ISBN-13: 978-3-639-85270-7. ISBN-10: 3639852702. EAN: 9783639852707 – 2017.
[14] **van de Graaf J., Montejano G., García P.**: "Manejo de Colisiones en un Protocolo Non Interactive Dining Cryptographers". Anales de las 42° Jornadas Argentinas de Informática e Investigación Operativa (JAIIO, ISSN: 1850-2776). Workshop de Seguridad Informática (WSegI 2013, ISSN: 2313-9110). Páginas 29 a 43. Disponible en: http://42jaiio.sadio.org.ar/proceedings/simposios/Trabajos/WSegI/03.pdf. 2013.
[15] **García P., Montejano G., Bast S., Fritz E.:** "Codificación de Sufragios con Detección de Colisiones en NIDC con Canales Paralelos de Slots" Congreso Nacional de Ingeniería en Informática / Sistemas de Información. CoNaIISI 2016.
[16] **García, P., van de Graaf J., Hevia A., Viola A.**: "Beating the Birthday Paradox in Dining Cryptographers Networks". En "Progress in Cryptology – Latincrypt 2014". Springer International Publishing. ISSN: 0302-9743. ISSN (electrónico): 1611-3349. ISBN: 978-3-319-16294-2. ISBN (eBook): 978-3-319-16295-9. Ps. 179 – 198. Octubre, 2014.
[17] **Murdocca M., Heuring V**. "Principles of Computer Architecture. Appendix A: Digital Logic". Editor: Addison Wesley; Edición: US ed (29 de noviembre de 1999) Idioma: Inglés - ISBN-10: 0201436647 - ISBN-13: 978-0201436648

Participación y despliegue de CTFs como herramienta para fortalecer la formación en ciberseguridad

Javier Díaz, Paula Venosa, Nicolás Macia, Einar Lanfranco, Alejandro Sabolansky, Mateo Durante, Damián Rubio, Jeremías Pretto
Laboratorio de Investigación de Nuevas Tecnologías Informáticas (LINTI).
Facultad de Informática. Universidad Nacional de La Plata
50 y 120 La Plata
{jdiaz, pvenosa, nmacia, einar, asabolansky, mdurante, drubio, jpretto}@linti.unlp.edu.ar

RESUMEN

A través de diferentes proyectos y con un grupo de investigación que se encuentra en continuo crecimiento, el Laboratorio en Nuevas Tecnologías Informáticas (LINTI) desarrolla de manera ininterrumpida desde el año 2000 una línea de investigación en ciberseguridad [1][2]. Entre los hitos más destacados se encuentran la puesta en marcha de la autoridad de certificación UNLP PKIGRID [3] en el año 2007 la cual fue acreditada por TAGPMA [4] y que actualmente forma parte del repositorio de autoridades de certificación académicas confiables TACAR [5][6], así como la creación del primer CSIRT académico de la Argentina CERTUNLP [7]. Años más tarde, los docentes e investigadores del área comenzaron a participar más activamente en competencias de tipo CTF[1].
En el presente trabajo se describen los CTFs, la experiencia y logros del equipo, la implicancia de su aplicación en lo que hace a la metodología de enseñanza, su aplicación en materias de grado relacionadas a la temática así como en lo que hace a formación de recursos humanos en seguridad, y en particular a la colaboración en la formación de equipos de respuesta de incidentes, comenzando por la comunidad académica local y generando sinergias a partir de la coordinación de actividades en grupos de trabajo regionales e internacionales.

Palabras clave: Ciberseguridad, seguridad inteligente, CTFs, Seguridad en aplicaciones, IoT

CONTEXTO

La línea de investigación en ciberseguridad que incluye la “Participación y despliegue de CTFs como herramienta para fortalecer la formación en ciberseguridad” presentada en este trabajo, se inserta en el proyecto de investigación "Internet del Futuro: Ciudades Digitales Inclusivas, Innovadoras y Sustentables, IoT, Ciberseguridad, Espacios de Aprendizaje del Futuro" [8] del Programa Nacional de Incentivos a docentes investigadores, que se desarrolla en el LINTI de la Facultad de Informática de la Universidad Nacional de La Plata (UNLP). Este proyecto está acreditado por la UNLP y financiado por partidas del presupuesto nacional.

1. INTRODUCCIÓN

Los CTFs (Captura de bandera o Capture The Flag por su sigla en inglés) son competencias de

[1] CTF: Capture the flag

seguridad informática que posibilitan el aprendizaje de distintas cuestiones vinculadas con la ciberseguridad de manera lúdica. El objetivo de las competencias es descubrir "flags". Las llamadas "flags" (banderas) son piezas de información que se ocultan, ya sea en servidores solo accesibles a través de algún protocolo o en una aplicación vulnerable, cifrándolas en un archivo o colocándolas no disponibles a simple vista. Durante el tiempo que dura la competencia, se liberan distintos desafíos en donde los equipos participantes aplican diferentes técnicas entre las que se encuentran la ingeniería inversa, hacking, criptoanálisis, ingeniería social para hacerse con las flags.

Este tipo de competencias suelen ser de dos modalidades: (i) los CTFs del tipo *jeopardy* cuentan con desafíos que al resolverlos entregan un texto secreto que se denomina bandera o *flag*. Este *flag* no es más que un texto bajo un formato específico, y el descubrirlo implica la resolución del desafío y (ii) los CTFs de ataque-defensa, donde cada equipo debe defender un servidor o una red con servicios vulnerables del resto de los participantes. Cada equipo tiene tiempo para arreglar los problemas en sus servicios evitando ser atacados y desarrollar tools para aprovechar vulnerabilidades presentes en la red del resto de los equipos. Los puntos se obtienen al proteger los servicios propios (puntos de defensa) y al atacar exitosamente a los servicios de los otros equipos (puntos de ataque) [9][10].

Muchos de los desafíos o retos están directamente relacionados con problemas de seguridad actuales o contemporáneos al momento en que se desarrolla la competencia y los patrones de resolución de problemas son similares a los que se aplican en la vida real. El hecho de resolverlos actúa, en muchos casos, como disparador para encontrar nuevos problemas en el ámbito laboral donde los miembros del equipo se desempeñan.

Tras la culminación de las competencias, es habitual que quienes resuelven retos, publiquen las resoluciones de los mismos [11]. Esta práctica sirve como fuente de aprendizaje para quien consulte dichas resoluciones. Para los que hayan llegado a resolver el ejercicio, sirve para aprender sobre posibles alternativas de resolución. Para aquellos que no hayan logrado resolver el ejercicio, sirve para aprender algo nuevo o para medir qué tan cerca se estuvo de alcanzar la solución con su análisis e intentos previos.

Para que un CSIRT preste sus servicios y crezca, resulta fundamental el desarrollo de habilidades específicas relacionadas con la ciberseguridad como ser el testeo de la seguridad de un sistema o la corrección del sistema para que no sea vulnerable. Investigar sobre nuevas vulnerabilidades, conocer técnicas y herramientas para detectar dichas vulnerabilidades, como así también estándares adecuados para llevar a cabo el proceso de testeo de seguridad, son tareas que forman parte del desafío de contar con sistemas cada vez más seguros en las organizaciones.

Tanto la participación en eventos del tipo CTF como la lectura de las resoluciones (o writeups) favorecen el desarrollo de habilidades relacionadas con la seguridad de sistemas, redes, comunicaciones y servicios. Esto representa un método alternativo de capacitación y actualización contínua para los distintos recursos humanos que forman parte del grupo de investigación.

La participación en este tipo de competencias suele ser en equipo. Esto fomenta el espíritu de compartir conocimientos y experiencias, como

así también el de relacionarse con personas interesadas en las mismas temáticas.

Los CTFs pueden utilizarse como herramienta para entender la ciberseguridad a partir de una actividad lúdica. La aplicación de la gamificación como una metodología de enseñanza alternativa se aplica en varios países del mundo. Las competencias de tipo CTF se consideran una forma de llevar a cabo esta metodología [12][13].

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

En la actualidad, como se ha mencionado, la línea de trabajo de formación y actualización a través del desarrollo de competencias mediante la continua participación en concursos de tipo CTF, aquí presentada, constituye uno de los ejes fundamentales del grupo de investigación en ciberseguridad.

Además de ello, el grupo continúa investigando y trabajando en las siguientes actividades:

- Gestión de incidentes de seguridad.
- Monitoreo de seguridad inteligente.
- Detección y análisis de vulnerabilidades en distintos tipos de dispositivos, protocolos y tecnologías.
- Gestión de seguridad informática en infraestructuras de red y servicios.
- Forensia digital.
- Infraestructura de clave pública PKI.
- Desarrollo seguro de software.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Como principales objetivos se plantean algunos generales como:

- Consolidar la línea de investigación en ciberseguridad y su aplicación en la docencia y la extensión, trabajando sobre los temas emergentes asociados a las metodologías y paradigmas que surgen día a día.
- Transmitir la experiencia adquirida en distintos proyectos y actividades a los alumnos de las cátedras de grado y postgrado con contenidos afines de nuestra Facultad.

Y otros más específicos como:

- Consolidar el equipo de CTF de la UNLP (llamado CERTUNLP o SYPER)
- Mejorar los resultados de años anteriores, tanto en conferencias como en el ranking mundial mantenido por CTFtime.
- Organizar nuevos eventos a nivel nacional y regional.

Entre los resultados que se han obtenido en este último tiempo:

- En marzo de 2020 se organizó una competencia de ciberseguridad utilizando una metodología basada en CTFs en el marco del evento de Metared denominado “Jornada de Ciberseguridad para Universidades 2020”.
- En el año 2018 y en el año 2019, el grupo obtuvo el primer puesto en el CTF organizado por Ekoparty, la conferencia de Seguridad más importante de la región. Desde el año 2017, el equipo ocupa el 1er puesto de Argentina en el ranking mantenido por CTFtime [14] [15].
- Mejora contínua de los servicios prestados por CERTUNLP a la comunidad

académica, aplicando conceptos aprendidos a partir de la participación en CTFs.

- Aplicación de la metodología de enseñanza utilizando competencias de tipo CTF en las siguientes materias de grado: Introducción a la Ciberseguridad, Desarrollo Seguro de Aplicaciones, Introducción a la Forensia Digital y Seguridad y Privacidad en Redes.
- Formación de un equipo interclaustro, compuesto por docentes y alumnos, que participan en competencias de tipo CTF y que se reúnen periódicamente para intercambiar conocimientos adquiridos en distintas temáticas de interés. El equipo se constituyó a partir de la convocatoria realizada por este grupo en el año 2017, invitando a todos los alumnos y docentes interesados a participar.
- Definición de un circuito para el desarrollo de nuevos retos, así como también la infraestructura a utilizar para su instalación y mantenimiento. Esto fue desarrollado en de manera conjunta entre distintos docentes parte del grupo de investigación, de modo que las distintas cátedras involucradas en temáticas de ciberseguridad, utilicen las mismas soluciones para la implementación de sus CTFs.
- Tanto en las cátedras de grado como en las "Jornada de Ciberseguridad para Universidades 2020" se utilizó la plataforma Open Source CTFd [16]. La misma, ha sido desplegada y configurada en servidores del LINTI.

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de investigación trabaja un grupo de docentes/investigadores del LINTI (Laboratorio de Investigación en Nuevas Tecnologías Informáticas) de la Facultad de Informática de la UNLP (Universidad Nacional de La Plata). Este equipo de trabajo también forma parte de CERTUNLP, el CSIRT Académico de la Universidad Nacional de La Plata, ámbito en el que se aplican las distintas temáticas incorporadas utilizando las competencias de tipo CTF.

En el último año, se realizaron dos tesinas de grado, dirigidas por los profesores Einar Lanfranco y Paula Venosa, autores de este trabajo.

La primera de las tesinas, denominada "Automatizando la resolución de problemas en competencias de seguridad informática" de los alumnos Jeremías Pretto y Facundo Basso, profundizó la investigación y el desarrollo de herramientas que pueden ser utilizadas en competencias tipo CTF. Como resultado de este trabajo se realizaron dos herramientas para resolver retos de los que se presentan habitualmente. La primera es para ayudar a los participantes novatos a acercarse a la solución de problemas y la segunda para enfrentar los retos que involucran distintos tipos de ataques a RSA.

La segunda tesina titulada "'Capture the flag' aplicada a la enseñanza de ciberseguridad en escuelas secundarias" fue realizada por los alumnos Patricio Bolino y Gabriela Suárez y tuvo como objetivo organizar CTFs para alumnos de escuelas secundarias, adecuando contenidos y retos para sortear barreras que dificultan la participación de los adolescentes sin ningún tipo de conocimientos previos.

5. REFERENCIAS

[1] - Díaz, Francisco Javier, Molinari, Lía Hebe, Venosa, Paula, Macia, Nicolás, Lanfranco, Einar Felipe, Sabolansky, Alejandro Javier (2018). "Investigación en ciberseguridad: un enfoque integrado para la formación de recursos de alto grado de especialización". WICC 2018 (Workshop de Investigadores en Ciencias de la Computación). UNNE, Corrientes, Argentina. Abril de 2018. Libro de Actas XX Workshop de Investigadores en Ciencias de la Computación. ISBN 978-987-3619-27-4.

[2] Díaz, Francisco Javier, Molinari, Lía Hebe, Venosa, Paula, Macia, Nicolás, Lanfranco, Einar Felipe, Sabolansky, Alejandro Javier (2019). "Investigación en ciberseguridad: nuevos desafíos para adaptarse a nuevos paradigmas". WICC 2019 (WICC 2019, Universidad Nacional de San Juan). Libro de Actas XXI Workshop de Investigadores en Ciencias de la Computación.pp 905-910. ISBN 78-987-3984-85-3

[3] PKIGrid: http://www.pkigrid.unlp.edu.ar/

[4] TAGPMA: The Americas Grid Policy Management Authority http://www.tagpma.org/

[5] TACAR: Trusted Academic Certification Authority Repository https://www.tacar.org/

[6] Clave de la UNLP PKIGrid en TACAR: https://www.tacar.org/cert/install/7

[7] CERTUNLP: CSIRT académico de UNLP https://cespi.unlp.edu.ar/cert

[8] Proyecto: F020 - Internet del Futuro. Ciudades Digitales, Inclusivas, Innovadoras y Sustentables, IoT, Ciberseguridad, Espacios de aprendizaje del Futuro. https://cyt.proyectos.unlp.edu.ar/projects/11-f020

[9] "CTF: Learn Hacking by Playing" (14 de agosto de 2019). https://www.sothis.tech/en/ctf-learn-hacking-by-playing/

[10] "CTF: Entrenamiento en seguridad informática" (26 de febrero de 2014). https://www.incibe-cert.es/blog/ctf-entrenamiento-seguridad-informatica

[11] Writeups de CTFtime. https://ctftime.org/writeups

[12] Mc Daniel Lucas, Talvi Erik, Hay Brian, "Capture the Flag as Cyber Security Introduction" en IEEE Computer Society Washington, "Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS)", University of Alaska Fairbanks, 2016.

[13] Cheung Ronald S., Cohen Joseph P., Lo Henry Z., Elia Fabio, "Challenge Based Learning in Cybersecurity Education", Department of Computer Science, University of Massachusetts, Boston, MA, USA, 2011.

[14] SYPER https://ctftime.org/team/2003

[15] Clasificación de SYPER en Argentina: https://ctftime.org/stats/2017/AR

[16] CTFd https://github.com/CTFd/CTFd

Criptología Simétrica Encriptado-Autenticado

Castro Lechtaler, Antonio[1, 2]; Cipriano, Marcelo[1,3]; García, Edith[1],
Liporace, Julio[1]; Maiorano, Ariel[1]; Malvacio, Eduardo[1], Pazo Robles María Eugenia[1].

[1]Laboratorio de Investigación en Técnicas Criptográficas y Seguridad Teleinformática.
Facultad de Ingeniería del Ejército -FIE. Universidad de la Defensa Nacional - UNDEF
[2] CISTIC/FCE - Universidad de Buenos Aires.
[3] Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes UNQ.

{acastro@, marcelocipriano}@fie.undef.edu.ar,
{edithxgarcia; jcliporace; maiorano; edumalvacio; eugepazorobles}@gmail.com

RESUMEN

Con este proyecto se propone investigar dentro de la criptografía simétrica, aquellos algoritmos de clave secreta que encriptan y autentican datos al mismo tiempo, logrando así confidencialidad y autenticidad (además de integridad) de los mensajes que se transmiten..

Nos orientaremos en el estudio de algunos modelos de algoritmos presentados en el concurso Cifrado Autenticado: Seguridad, Adaptabilidad y Robustez (CAESAR) y en la competencia NIST Lightweight Cryptography (Criptología liviana, que incluye encriptado autenticado e integridad) .

Analizaremos en cada caso, las características de diseño de los algoritmos presentados en los mencionados concursos y estableceremos una serie de criterios que es necesario tener en cuenta a la hora de garantizar en un mismo esquema de cifrado de clave simétrica, seguridad, simplicidad y velocidad (Security, Simplicity, Speed S^3).

Este conjunto de técnicas nos permitirá establecer un marco de referencia para el diseño y para la evaluación (a través del criptoanális) del nuevo paradigma para el desarrollo de algoritmos criptográficos simétricos.

Palabras Clave

Criptología Simétrica, Cifrado Autenticado,Cifrado Autenticado con Datos Asociados.

.

CONTEXTO

En el marco de la carrera de grado de Ingeniería en Informática, el posgrado en Criptografía y Seguridad Teleinformática, y de la Maestría de Ciberdefensa que se dictan en la Facultad de Ingeniería del Ejército (FIE) "Gral. Div. Manuel N. Savio" (EST), Universidad de la Defensa Nacional (UNDEF) se llevan adelante tareas de I+D+i por parte del Grupo de Investigación en Criptología y Seguridad Informática (GICSI).

GICSI depende del Laboratorio de Investigación en Técnicas Criptográficas y Seguridad Teleinformática (CriptoLab) perteneciente al Laboratorio Informática (InforLab). Y está conformado por docentes investigadores, profesionales técnicos y alumnos de dicha área.

1. INTRODUCCIÓN

Quizás las dos propiedades fundamentales que debe garantizar la criptografíaía de clave simétrica son la *confidencialidad* (privacidad) y la *autenticidad* (e *integridad*) de los datos que se transmiten a través de canales inseguros. Estas dos características han sido tradicionalmente estudiadas, formalizadas e implementadas por separado[1-2], a través de primitivas criptográficas diferentes (por ejemplo: un algoritmo simétrico en modo CBC para confidencialidad y un CBC MAC para autenticación). Sin embargo tal separación raras veces aparece en la práctica. Por el contrario, en

la gran mayoría de las aplicaciones la autenticidad resulta acompañar a la confidencialidad. Por lo tanto, el problema de alcanzar privacidad y autenticidad simultáneamente con un mismo esquema de cifrado de clave secreta requiere un análisis profundo, exhaustivo y sistemático.

Para tal fin, Bellare y Rogaway (e independientemente Katz y Yung) propusieron en el año 2000 un enfoque formal para solucionar este problema [3-4]. Al mismo tiempo Bellare y Namprempe investigaron la seguridad combinando un esquema de cifrado convencional y una MAC para alcanzar un encriptado autenticado AE [5]. Poco después, aparecieron modelos orientados AE de clave secreta: OCB (2001) [6], CCM (2002)[7-8], y GCM (2004)[9]. La relevancia de eficientes AE en aplicaciones en el mundo real, implicó la estandarización de varios esquemas: el modo CCM en IEEE 802.11i, IPsec ESP e IKEv2; el modo GCM aparece en NIST SP 800-38D; el modo EAX está especificado en ANSI C12.22; e ISO/IEC 19772:2009 tiene seis esquemas AE (cinco diseños orientados AE y un método de composición genérica). En particular el AE se utiliza para evitar lo que se denomina chosen ciphertext attack, donde un atacante puede solicitar a un oráculo que descifre un paquete cifrado y así poner en riesgo todo el criptosistema.

Luego de esta nueva tendencia de diseño de esquemas de AE, surgieron varias cuestiones relativas a la seguridad: errores en las pruebas para medir el nivel de fortaleza, claves débiles, TAGs de tamaño corto y también problemas en la performance e implementación de los algoritmos en diferentes aplicaciones. Una serie creciente de problemas ocasionaron lo que se calificó como *"Desastres"* de la criptografía simétrica [10]. Por otro lado si bien se pueden implementar dos primitivas como las que ya existen para el cifrado y otra para generar una MAC, existe la necesidad de lograr eficientemente realizar ambas operaciones en un mismo esquema de cifrado, implementando un algoritmo y algún modo de operación que realice ambas operaciones y se ajuste a las necesidades de por ejemplo: la transmisión de datos, como ser redes de datos con Pc´s y gran ancho de banda y por otro lado pequeños dispositivos de hardware como los que se usan en IoT(*Internet of Things*) .

Todo esto marcó la necesidad de investigar aún más en el campo del Cifrado Autenticado, y motivó en 2013 hacer la convocatoria de AE CAESAR[11] y en 2015 NIST Lightweight Cryptography[12]. Así se abrió un nuevo campo de investigación que incluye resultados en técnicas de criptoanálisis, diseño y construcción de primitivas criptográficas, análisis de seguridad como así también nuevos modelos formales de seguridad criptológica.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Para llevar adelante este proyecto se siguen las siguientes líneas de investigación y desarrollo:

- Estudio de material actualizado, asistencia a Cursos, Congresos y Workshops específicos, profundización en el estado del arte del encriptado autenticado y en los nuevos ataques que se hayan desarrollado para tales modelos.
- Estudio, análisis y selección de esquemas de AEAD de clave simétrica.
- Relevamiento de los métodos de diseño que se aplican a los algoritmos AEAD.
- Estudio de técnicas criptográficas para el diseño de esquemas de cifrado autenticado y cifrado autenticado con datos asociados.

- Implementación en diferentes soportes de los esquemas AE.
- Análisis y conclusiones de los resultados obtenidos.

3. RESULTADOS OBTENIDOS / ESPERADOS

Al realizar el estudio, análisis de las técnicas y herramientas para el diseño de AE y AEAD, se persigue establecer un marco de referencia para el diseño de algoritmos de encriptado autenticado:

- Establecer protocolos y métodos de diseño para tales esquemas de cifrado autenticado.
- Evaluar a través de diferentes técnica criptográficas el nivel de seguridad de las primitivas de AE.
- Desarrollar nuevas y buenas prácticas en el diseño en general de algoritmos de clave simétrica.
- Implementar de manera eficiente los AE en diferentes plataformas.

4. FORMACIÓN DE RECURSOS HUMANOS

Los docentes investigadores del proyecto dictan las asignaturas Criptografía y Seguridad Teleinformática, Matemática Discreta y Paradigmas de Programación I, II y son docentes en las materias Criptología y Criptología Avanzada de la Especialización en Criptografía y Seguridad Teleinformática y de la Maestría en Ciberdefensa Desde esas cátedras se invita a los alumnos a participar. Es por ello que 3 de ellos han demostrado su interés y se han sumado en calidad de colaboradores. En particular, el alumno Leiras, Facundo ha recibido la beca "Estímulo a las Vocaciones Científicas" (EVC) otorgadas por el Consejo Interuniversitario Nacional (CIN) [13].

Se desea destacar que el incremento del Know-How que tendrá el grupo de investigadores a lo largo de la vida del proyecto será una importante y económica Formación de Recursos Humanos en beneficio de sus integrantes y de la institución en la cual desarrollan sus actividades científico-docentes.

Por último y atendiendo a la responsabilidad ética y social que compete a la actividad científica y tecnológica, el Grupo Integrante de este Proyecto de Investigación, ya sea durante su ejecución o por la aplicación de los resultados obtenidos, desea expresar su compromiso a no realizar cualquier actividad personal o colectiva que pudiera afectar los derechos humanos, o ser causa de un eventual daño al medio ambiente, a los animales y/o a las generaciones futuras.

5. BIBLIOGRAFÍA

[1] BELLARE, M.—DESAI, A.—JOKIPII, E.—ROGAWAY, P.: A concrete security treatment of symmetric encryption, in: 54th Annual Symp. on Found. of Comput. Sci.– –FOCS '97, Miami Beach, FL, 1997, IEEE Comput. Soc., 1997, pp. 394–403.

[2] BELLARE, M.—KILIAN, J.—ROGAWAY, P.: The security of the cipher block chaining message authentication code, J. Comput. Syst. Sci. 61 (2000), pp. 362–399.

[3] BELLARE, M.—ROGAWAY, P.: Encode-then-encipher encryption: How to exploit nonces or redundancy in plaintexts for efficient cryptography, in: Advances in Cryptology—ASIACRYPT '00 (T. Okamoto, ed.), 6th Internat. Conf. on the Theory and Appl. of Cryptology and Inform. Security, Kyoto, Japan, Lecture Notes in Comput. Sci., Vol. 1976, Springer, Berlin, 2000, pp. 317–330.

[4] KATZ, J.—YUNG, M.: Unforgeable encryption and chosen ciphertext secure modes of operation, in: Fast Software Encryption—FSE '00 (Schneier, B. ed.), 7th Internat. Workshop—FSE '00, New York, NY, USA, 2000, Lecture Notes in Comput. Sci., Vol. 1978, Springer, Berlin, 2001, pp. 284–299.

[5] BELLARE, M.—NAMPREMPRE, C.: Authenticated encryption: relations

among notions and analysis of the generic composition paradigm, in: Advances in Cryptology– –ASIACRYPT ’00 (T. Okamoto, ed.), 6th Internat. Conf. on the Theory and Appl. of Cryptology and Inform. Security, Kyoto, Japan, Lecture Notes in Comput. Sci., Vol. 1976, Springer, Berlin, 2000, pp. 531–545.
[6] ROGAWAY, P.—BELLARE, M.—BLACK, J.—KROVETZ, T.: OCB: A block-cipher mode of operation for efficient authenticated encryption, in: Proc. of the 8th ACM Conf. on Computer and Comm. Security ACM—CCS ’01, ACM New York, NY, USA, 2001, pp. 196–205.
[7] DWORKIN, M.: Recommendation for Block Cipher Modes of Operation: The CCM Mode for Authentication and Confidentiality. NIST Special Publication 800-38C, Gaithersburg, 2004.
[8] WHITING, D.—HOUSLEY, R.—FERGUSON, N.: Counter with CBC-MAC (CCM). IETF RFC 3610 (Inform). Sep.2003.
http://www.ietf.org/rfc/rfc3610.txt
[9] MCGREW, D. A.—VIEGA, J.: The security and performance of the galois/counter mode (GCM) of operation, in: Progress in Cryptology—INDOCRYPT ’04 (A. Canteaut et al., eds.), 5th Internat. Conf. on Cryptology in India, Chennai, India, 2004, Lecture Notes in Comput. Sci., Vol. 3348, Springer, Berlin, 2004, pp. 343–355.
[10] BERNSTEIN, D. J.: Cryptographic competitions: Disasters, https://web. archive.org/web/20130418063008 /http://cr.yp.to/talks/2013.03.12/slides.pdfhttps://competitions.cr.yp.to/disasters.html (consultada 1-3-20).
[11]https://competitions.cr.yp.to/caesar.html (consultada 1-3-20).
[12]https://csrc.nist.gov/projects/lightweight-Cryptography (consultada 1-3-20).
[13]http://evc.cin.edu.ar/informacion (consultada el 2/3/20).

Análisis de Frameworks de Nube: Microsoft Azure y Amazon Web Services, mediante versiones privadas de prueba en entornos educativos

Silvia Edith Arias[1], Laura Mónica Vargas[2,3], Alejandra Di Gionantonio[1], Diego Serrano[1], Adriana Cucchi[1], Paula Sosa[1], Ezequiel Ambrogio[1], Daniel Arch[1]

s_autn@hotmail.com, {laura.monica.vargas, ing.alejandradg, diegojserrano, adriana.beat, sosa.pau, ezequielambrogio} @gmail.com, daniel.arch@pjn.gov.ar

[1]Laboratorio de Investigación de Software, Departamento de Ingeniería en Sistemas de Información, Facultad Regional Córdoba, Universidad Tecnológica Nacional

[2] Laboratorio de Redes y Comunicaciones de Datos, Departamento de Computación, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba

[3] Laboratorio de Procesamiento de Señales, Departamento de Matemática, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba

I. RESUMEN

En la presente investigación nos dedicamos a indagar sobre las peculiaridades de dos plataformas de Cloud Computing. Es el caso de Microsoft Azure y Amazon Web Services, haciendo uso de versiones privadas de prueba en entornos educativos.

Teniendo en cuenta sus características principales, se procede a realizar una comparación de las plataformas mencionadas, con la finalidad de elaborar una Tabla comparativa de los frameworks mencionados para presentar una serie de recomendaciones acerca de las características de las plataformas más seguras y adecuadas a la hora de decidir por una u otra. Con el fin de alcanzar el objetivo de este trabajo que resulta ser el de indagar y adquirir conocimientos teóricos sobre marcos de trabajo de plataformas de Cloud Computing que permitan alojar imágenes médicas con inserción de Marcas de Agua en los EHRs. Posteriormente elaborar una comparación cuantitativa y cualitativa de sus características principales, y recomendar cuáles son las plataformas más seguras y adecuadas. Para garantizar de algún modo que las imágenes médicas se puedan almacenar y distribuir en forma segura preservándolas de cualquier intento de distorsión.

Palabras clave: seguridad informática, cloud computing.

II. CONTEXTO

El presente trabajo se realiza en el Laboratorio de Investigación de Software, Departamento de Ingeniería en Sistemas de Información de la Facultad Regional Córdoba de la Universidad Tecnológica Nacional. En el marco del Proyecto "Análisis comparativo entre Plataformas de Cloud Computing, para el caso de almacenamiento de imágenes médicas con marcas de agua" acreditado y financiado por la Secretaría de Ciencia y Técnica de Código: CCUTNCO0004961. El cual se lleva a cabo en el Laboratorio de Investigación de Software de la Facultad Regional Córdoba de la Universidad Tecnológica Nacional (Argentina).
La temática de watermarking ha sido presentada por los cuatro primeros autores en los Proyectos homologados por SeCyT-UTN PID, "Marcas de Agua múltiples en imágenes digitales fijas para autenticación y detección de adulteraciones". Código SeCyT – UTN1166, 2010-2011, Resolución 26/10, 2010 SeCyT del Rectorado de UTN y "Marcas de Agua Seguras en Imágenes para identificación del propietario". Proyecto ID promocional. Código SeCyT- UTN EIPRCO753, 2008-2009, Resolución 75/08 SeCyT del Rectorado UTN.

La segunda y la tercera autora han publicado en

la Revista de la FCEFyN de la UNC un artículo de difusión de marcas de agua. "Marcas de Agua: una Contribución a la Seguridad de Archivos Digitales". Revista de la Facultad de Ciencias Exactas, Físicas y Naturales de la UNC. ISSN 2362-2539 (Versión electrónica). Año 3 – Nº 1 (2016).

III. INTRODUCCIÓN

El presente proyecto de investigación se inscribe dentro de los lineamientos de Seguridad Informática, ya que se trata de información sensible que circula a través de la red y demanda de un canal y de un almacenamiento con infraestructura segura frente a sustracciones, ataques, incidentes y pérdidas de información.
Siendo el objetivo del mismo analizar las plataformas de Cloud Computing y las particularidades que ofrecen para el almacenamiento seguro de imágenes médicas donde han sido embebidos metadatos con información del paciente y del profesional médico, a modo de marca de agua que asegura la detección de una adulteración en la imagen.
Cloud Computing es un mecanismo que creció en los últimos años, basado en la Web que permite escalar y virtualizar recursos de TI que son proporcionados como servicios a través de la red. Características inherentes y esenciales que deben ser provistas por las aplicaciones de cloud computing son: servicio bajo demanda, acceso ubicuo, escalabilidad, elasticidad, independiza al usuario del mantenimiento y pago por uso, siendo la seguridad todavía un desafío [1].
Se trabajó en las características de las dos principales arquitecturas de Cloud Computing privadas en la actualidad como son Amazon y Azure seleccionadas para este estudio.
Consideramos las mencionadas plataformas para alojar y consultar las imágenes médicas, que una vez generadas en las instituciones médicas, fueron tratadas posteriormente con marcas de agua como componente de seguridad informática.
Es necesario asegurar que estas imágenes no sean alteradas o manipuladas durante el proceso de transmisión especialmente si se utiliza una red pública, como así también proporcionar privacidad en las cadenas de datos de los Registros Electrónicos de los Pacientes o Electronic Health Records (EHR), y una plataforma de Cloud Computing para su almacenamiento o consulta. En investigaciones anteriores vinculados con la seguridad en el procesamiento de imágenes, hemos visto las ventajas que ofrece la tecnología de las marcas de agua para tal fin.
La Ley N° 10590 aprobada en Diciembre del año 2018 en la provincia de Córdoba crea el Sistema Provincial de Historia Clínica Electrónica (HCEU) que tiene como finalidad “el registro indeleble de los datos de salud y enfermedad de cada persona, desde su nacimiento hasta su fallecimiento”. [2]
La misma ley en su artículo quinto y refiriéndose a la Historia Clínica Electrónica establece que “el almacenamiento, actualización y uso se efectúa en estrictas condiciones de seguridad, integridad, autenticidad, confiabilidad, exactitud, inteligibilidad, conservación, disponibilidad y acceso”. Se puede ver en el extracto la honda preocupación de los legisladores respecto de los objetivos que persigue este trabajo, que claramente podría contribuir a una mejora en los servicios de salud prestados tanto por instituciones públicas como privadas.
A partir de los datos obtenidos en esta investigación, elaboramos un cuadro comparativo de las plataformas citadas y establecemos una serie de recomendaciones acerca de las características de las plataformas más seguras y adecuadas para alojar imágenes médicas.

IV. OBJETIVOS Y LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto se inscribe dentro de los lineamientos de investigación en Seguridad Informática.

El objetivo de este proyecto de investigación es analizar el estado del arte de cloud computing para servicio de almacenamiento de imágenes médicas con marcas de agua y difundir los resultados obtenidos para realimentar el proceso de desarrollo de los algoritmos de watermarking.

V. MATERIALES Y MÉTODOS

Cloud Computing es una plataforma computacional de trabajo que ofrece recursos, tales como infraestructura, aplicaciones, procesamiento para ser consumidos bajo demanda como un servicio más en Internet.

Cloud computing es un modelo de computación que brinda grandes beneficios a sus usuarios finales como así también a organizaciones públicas y privadas.

Para el desarrollo de este trabajo se aplica el Método Empírico-Analítico, que se basa en la experimentación, lo cual asegura que el conocimiento procede de la experiencia y de tomar decisiones basándose en lo que se conoce. [3]

En el desarrollo de esta investigación se creó un marco de trabajo ágil SCRUM, la que permitió tener una visibilidad de las tareas y poder identificar impedimentos.

Se trabajó con bloques de tiempos definidos llamados Sprint de dos semanas donde nos reuníamos como equipo y planificamos las tareas. Esta forma de trabajo permitió mejorar la división de tareas, logrando un mayor dinamismo.

Se realizaron planificaciones dentro de sprint, dailies diarias y semanales, y al finalizar cada sprint el equipo se reunió para ver los avances y poder plantear mejoras dentro de la forma de trabajo realizada.

Dentro del equipo de trabajo se identificaron diferentes roles que aplica el marco de trabajo SCRUM, contamos con un Scrum Master que ayuda a cumplir el objetivo del sprint, eliminando impedimentos que surjan dentro del equipo de trabajo y un Product Owner que prioriza las tareas.

Atentos a las primeras Review y Retrospective, para analizar la efectividad que pueda haber tenido la aplicación de esta nueva forma de trabajo en el avance del proyecto. Ver figura 1.

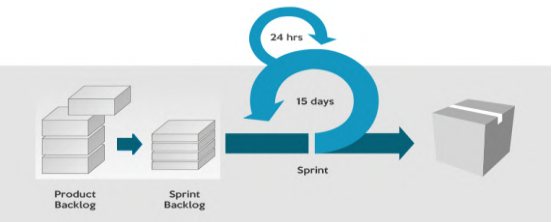


Figura 1. Ejemplo de metodología ágil (fuente: https://luis-goncalves.com/es/que-es-la-metodologia-scrum/)

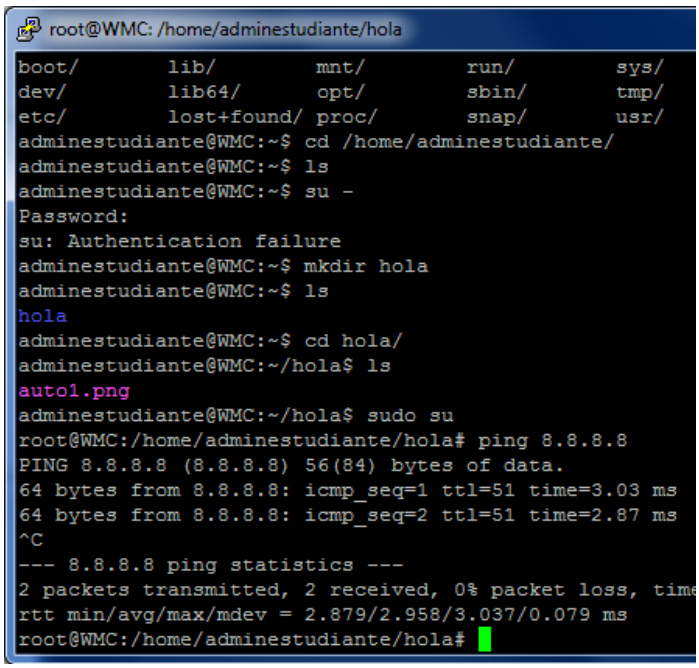
Comparativa entre Microsoft Azure y Amazon Web Services (AWS)

El eje sobre el que gira nuestro trabajo está centrado en la evaluación de frameworks de Cloud Computing, en este caso Azure y Amazon para el almacenamiento de imágenes médicas. Dicha evaluación está basada en el análisis de los principales servicios que prestan los frameworks de Cloud computing seleccionados. Resaltando diferencias entre ambos frameworks.

Diferencias entre Microsoft Azure y Amazon Web Services (AWS)

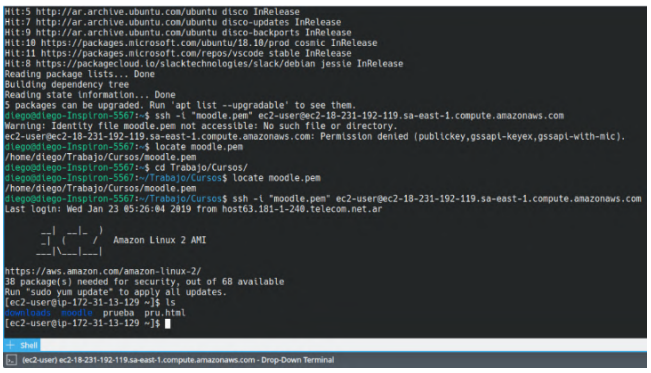
- AWS está enfocado más en modelos serverless proporcionando herramientas completas para dejar aún lado el hardware, licenciamiento y administración costosa.
- Azure está enfocado en modelos de nubes híbridas con la ventaja de que los modelos on-premise conviven mejor por tener ya licenciamiento Microsoft. Permite una fácil integración con otras herramientas Microsoft y garantía seguridad con back up multi-nube. [4]
- AWS cuenta con escalabilidad y flexibilidad natural sin mínimos de consumo.
- Azure otorga paquetes de almacenamiento predeterminados.
- AWS cuenta ya con productos para soluciones de machine learning altamente automatizados sugiriendo el comportamiento de los algoritmos.
- Azure cuenta con también con productos para ML solo que menos automatizados, es necesario más desarrollo y con ciertas limitantes para la operación y adaptabilidad. [5]
- AWS proporciona varias capacidades y servicios de seguridad para mejorar la privacidad y controlar el acceso de redes. Entre ellos se incluyen:
  - o Los firewalls de red integrados en Amazon VPC y las capacidades de firewall para aplicaciones web existentes en AWS WAF permiten crear redes privadas y controlar el acceso a las instancias y aplicaciones
  - o Cifrado en tránsito con TLS en todos los servicios
  - o Opciones de conectividad que permiten conexiones privadas o dedicadas desde la oficina o entorno on-premise [6] [7]
  - o Azure ofrece una amplia gama de opciones de seguridad configurables, así como la capacidad de controlarlas, por lo que puede personalizar la

seguridad para satisfacer los requisitos exclusivos de las implementaciones de su organización. [8]
Se tomaron capturas de pantalla de las pruebas realizadas tanto en la plataforma de Azure como en Amazon.
Azure:
A través de una consola, desde Linux o desde alguna herramienta como Putty, nos conectarnos remotamente utilizando el protocolo SSH, con el usuario "adminestudiante" a la máquina virtual (VM) de Linux que fue creada previamente llamada "WMC", en la misma se puede ver que tenemos el árbol básico de directorios y archivos de Linux, donde pudimos crear carpetas dentro del mismo, pasar de ser un usuario común a ser usuario con privilegio de root (administrador), teniendo conectividad a la red (internet), entre otros. Ver Figura 2.



Amazon:
Las mismas pruebas se realizaron en Amazon
Con el objetivo de reflejar las propiedades a analizar existentes en las plataformas de cloud, es necesario seleccionar, clasificar, comparar, analizar y hacer abstracción de las principales características, generalizarlas y explicarlas. Ver Figura 3.

Tabla 1. Comparación entre Azure y Amazon



VI. RESULTADOS Y AVANCES

Con el fin de inducir sugerencias e ideas de una manera sistemática y ordenada, se confeccionó la tabla 1 cuyo contenido muestra la información esencial para el proceso de toma de decisión por parte de la institución médica a la hora de elegir una Plataforma de Nube Privada.
Cabe aclarar que el análisis y las pruebas en laboratorio no están concluidas. Motivo por el cual esta Tabla 1 será completada paralelamente con el avance del proyecto de investigación.

VII. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está conformado por docentes-investigadores pertenecientes a la carrera de grado de Ingeniería en Sistemas de Información.
El grupo está compuesto por una Directora, cuatro profesores investigadores de apoyo, tres ingenieros aspirantes a incorporarse a la carrera de investigador.
Este proyecto contribuirá a la formación y crecimiento de la carrera de investigador de los integrantes del mismo.
Además existe la colaboración de una docente investigadora de la FCEFyN-UNC.
Se dirigirán trabajos finales sobre la temática abiertos a estudiantes de Ingeniería en Sistemas de Información.

El Proyecto comenzó el 1 de enero de 2018 por un período de tres años, hasta el 31 de diciembre de 2020.

CARACTERISITICA	AMAZON EC2	MICROSOFT AZURE
Servicio de cómputo	Elastic Compute Cloud (EC2)	Virtual Machines (VMs)
Escalabilidad automática (auto scaling)	Amazon Cloud Watch	Autoscaling application block y Azure Fabric controller
Blueprints (imágenes para acelerar el aprovisionamiento)	(AMI) Imagen de máquina Amazon	Imágenes provistas en una galería y también imágenes propias guardadas.
App Hosting	Amazon Elastic Beanstalk	Cloud Services Azure Batch Azure Scheduler Logic Apps
Soporta Sistema Operativo Windows	Windows Server 2003 R2. - Windows Server 2008. - Windows Server 2008 R2. - Windows Server 2012.	Windows Server 2012 Data Center. -Windows Server 2008 R2 SP1.
Soporta Sistema Operativo Linux	SUSE Linux Enterprise Server. - Red Hat Enterprise Linux.	OpenSUSE 12.3. -SUSE Linux Enterprise Server 11 SP2. - Ubuntu Server 12.04 LTS. - Ubuntu Server 12.10. -Ubuntu Server 13.04. -OpenLogic CentOS 6.3. -Ubuntu Server 12.10 DAILY.
Soporte para almacenamiento de datos	Amazon S3. -Amazon Relational DB Service. -Amazon SimpleDB. -SQL Server Express. -SQL Web. -SQL Server STD. -Amazon Redshift	SLQ Relacional. -Almacenes de tablas NoSQL. -Blob no estructurado. -Amazon Dynamo DB
Servidor Web	Apache. -IIS. -Otros	IIS v7.5
Alternativas de Hipervisores	XEN y LXC (Linux Containers)	XEN y LXC (Linux Containers)

REFERENCIAS

[1] Briand, L. C., Daly, J., and Wüst, J. (1999)., "A unified framework for coupling measurement in objectoriented systems", *IEEE Transactions on Software Engineering*.

[2] Título: LEY N° 10590 – Salud. Provincia de Córdoba. 2018. Sistema Provincial de Historia Clínica Electrónica Única (HCEU). Creación.

[3] González Carlos D. - Evaluación de calidad web: Métodos, técnicas y uso de métricas de usabilidad (2016) http://www.usabilidadweb.com.ar/metodos_eval_calidad_web.php

[4] Azure comparativo general. (2018). Recuperado de: https://www.inbest.cloud/comunidad/aw s-vs.-azure-comparativo-general

[5] Bravent. (2017). Recuperado de: https://www.bravent.net/migrar-a-la-nube-azure-o-aws

[6] Amazon. (2019). Recuperado de Web: http://aws.amazon.com/

[7] Amazon- web-services. (2019). Recuperado de: https://www.ticportal.es/temas/cloud-computing/amazon-webservices https://searchdatacenter.techtarget.com/ es/definicion/Microsoft-Azure-Windows- Azure

[8] Amazon Security. (2019). Recuperado de: https://aws.amazon.com/es/security/

MÉTODOS Y HERRAMIENTAS PARA EL ANÁLISIS FORENSE DE DISPOSITIVOS MÓVILES

Graciela Viaña, Liliana Figueroa, Cecilia Lara, Analía Méndez, Norma Lesca, Daniel Ghunter

Instituto de Investigación en Informática y Sistemas de Información, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero
gv857@hotmail.com; lmvfigueroa@yahoo.com.ar; laraceciliacristina@gmail.com; anmendez725@yahoo.com; norma.lesca@gmail.com; dgunther@unse.edu.ar

RESUMEN

La evidencia digital genera grandes desafíos al utilizarse en el sistema procesal penal como prueba en la investigación de delitos, por lo que la justicia necesita contar con una regulación adecuada que permita obtener evidencias digitales legalmente aceptables, que ayuden a resolver conflictos según métodos científicos de recolección, análisis y validación.

En este sentido, Santiago del Estero requiere de guías de buenas prácticas y/o protocolos de actuación que orienten el trabajo pericial informático, así como contar con estrategias que permitan almacenar y mantener las evidencias digitales obtenidas de dispositivos móviles.

En esta etapa de la investigación, y con la intención de validar el conjunto de lineamientos obtenido como resultado del proyecto "Computación Móvil: desarrollo de aplicaciones y análisis forense", se propone establecer un instrumento de validación teniendo en cuenta el cumplimiento de estándares internacionales de calidad.

Por otro lado, se encuentran en etapa de evaluación los aspectos a considerar en el diseño de un repositorio para el almacenamiento de la evidencia digital obtenida de dispositivos móviles, así como la exploración sistemática y análisis del sistema de clasificación de herramientas forenses de dispositivos móviles [1].

Palabras clave:

Computación móvil, análisis forense, lineamientos para el tratamiento de evidencia digital, repositorio digital.

CONTEXTO

La presente línea de investigación se encuentra inserta en el proyecto "Métodos y herramientas para el análisis forense de dispositivos móviles" (23/C156), iniciado en el año 2019, financiado por el Consejo de Ciencia y Técnica de la Universidad Nacional de Santiago del Estero [5]. Este proyecto es una continuación de trabajos realizados desde el año 2017 en el proyecto "Computación Móvil: desarrollo de aplicaciones y análisis forense" [4].

Como producto de estas investigaciones se lograron resultados referidos al análisis de la obtención legal de la evidencia digital, estudio de antecedentes jurisprudenciales sobre tratamiento de evidencia digital en dispositivos móviles, investigación y análisis de protocolos vigentes en otras jurisdicciones y una propuesta de protocolo de actuación para la obtención de evidencia digital en móviles en el ámbito del Ministerio Público Fiscal de Santiago del Estero [17].

Actualmente, el equipo de investigación tiene como propósito fijar estrategias de validación de la propuesta de protocolo para que sea factible su implementación tomando como principal referencia la familia de normas ISO/IEC 27000, realizar el estudio de repositorios que permitan la construcción de un modelo de datos para la gestión de las evidencias digitales y analizar la aplicabilidad de las herramientas de los niveles superiores de la pirámide móvil forense.

Con este fin, se establecieron convenios con el Ministerio Público Fiscal y el Poder Judicial de Santiago del Estero, instituciones

donde se pretende llevar a cabo la validación en campo de los resultados alcanzados.

1. INTRODUCCIÓN

La justicia se encuentra actualmente ante el desafío de la utilización de la evidencia digital como prueba fundamental en la investigación de diferentes delitos [10]. El empleo de dispositivos móviles se incrementó notablemente y por consiguiente su uso en actividades delictivas [9].

En este contexto, es necesaria una regulación adecuada para la obtención de evidencia digital desde dispositivos móviles, apoyada en métodos científicos, que considere los aspectos de fragilidad, anonimidad y volatilidad que diferencian a la evidencia digital de la física, y que garantice su posterior validez en juicio.

La Informática Forense involucra la aplicación de la metodología y la ciencia para identificar, preservar, recuperar, extraer, documentar e interpretar [18] evidencias procedentes de fuentes digitales, para facilitar la reconstrucción de los hechos en la escena del crimen [12] y utilizarlas posteriormente como material probatorio en un proceso judicial [2,3].

A diferencia de la Informática Forense clásica, la Informática Forense sobre móviles es un campo relativamente nuevo, donde las normas y procedimiento se encuentran aún en desarrollo [16]. La admisibilidad de un análisis forense sobre móvil está determinada por la formalidad con que se realice el proceso de recolección, control, análisis y presentación de la evidencia, donde los protocolos desempeñan un rol fundamental.

Aunque en algunas provincias de nuestro país existen guías de buenas prácticas y protocolos de actuación relacionados con la evidencia digital, Santiago del Estero no cuenta con normativa al respecto. En este sentido, las principales contribuciones desarrolladas en el país son:

- *"Guía de obtención, preservación y tratamiento de la evidencia digital"* [15] de la Unidad Fiscal Especializada en Ciberdelitos de la Procuración General de la Nación. Aborda cómo se debe obtener, observar y tratar la evidencia digital para mejorar la eficiencia en el ámbito penal.
- *"Protocolo de actuación para pericias informáticas del Poder Judicial de Neuquén"* [11]. Aporta un procedimiento para pericias informáticas sobre telefonía celular y el uso del Universal Forensic Extraction Device (UFED).
- *"Guía integral de empleo de la informática forense en el proceso penal"* [6], del Laboratorio de Investigación y Desarrollo de Tecnología en Informática Forense (InfoLab). Presenta los aspectos básicos a considerar en la búsqueda, obtención, preservación, examen pericial y presentación de evidencias digitales.
- *"Guía de procedimientos para pericias de dispositivos móviles"* [14] del Poder Judicial de Río Negro. Propone una metodología del ciclo de vida de la evidencia digital.

Como respuesta a la necesidad planteada en la provincia de Santiago del Estero, se propuso un conjunto de lineamientos a los que recurrir para la obtención de evidencia digital de dispositivos móviles [17], que se encuentra en proceso de validación para su utilización como protocolo por el Ministerio Publico Fiscal de Santiago del Estero. La mencionada propuesta, organizada en fases, abarca el proceso completo de tratamiento de la evidencia digital, con especial énfasis en las actividades y técnicas relacionadas con dispositivos móviles, constituyendo una herramienta que permita a los peritos y operadores judiciales realizar una adecuada planificación y control de las actividades periciales durante la investigación penal preparatoria.

Para diseñar las actividades de validación de la propuesta del protocolo, tendientes a garantizar la calidad de los procesos aplicados y sus resultados, se tomaron como

principal referencia los siguientes estándares internacionales:

- La norma ISO/IEC 27042:2015 "*Guidelines for the analysis and interpretation of digital evidence*" [8], que propone una serie de definiciones relacionadas a la evidencia digital.
- La norma ISO/IEC 27037:2012 "*Information technology - Security techniques - Guidelines for identification, collection, acquisition and preservation of digital evidence*" [7], que establece los principios fundamentales que definen la formalidad de una investigación y son condiciones necesarias y suficientes para que se recaben, aseguren y preserven elementos probatorios sobre medios digitales.
- Las directrices RFC 3227 – "*Guidelines for Evidence Collection and Archiving*" [13], que ofrecen una guía de alto nivel para recolectar y archivar evidencia, estableciendo principios a respectar durante dichos procesos, así como consideraciones legales.

Durante el desarrollo de la investigación, se logró determinar que los principales problemas con los que se enfrentan los peritos informáticos de la provincia son: la ausencia de herramientas que permitan la extracción de datos desde todos los dispositivos móviles del mercado, los requerimientos recibidos no se encuentran claramente definidos desde el inicio, y la imposibilidad de compra de paquetes de software especializado debido a su alto costo.

Para abordar dichos problemas y realizar los ajustes pertinentes en la propuesta de protocolo, se tomó como referencia "El Sistema de Clasificación de Herramientas Forenses de Dispositivos Móviles" [1], que presenta la Pirámide Móvil Forense, poniendo especial atención en los niveles superiores de la misma. Según esta propuesta el nivel de extracción y el análisis requerido dependen de la solicitud y los detalles de la investigación, y cada nivel tiene su propio conjunto de requerimientos de aplicación. Los niveles más altos requieren un examen más completo y de habilidades adicionales de los peritos, que pueden no ser aplicables en todos los dispositivos y situaciones.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El presente proyecto de investigación se refiere a:

Informática Forense: métodos y herramientas para el análisis forense de dispositivos móviles.

Se proponen dos líneas de investigación derivadas, consideradas desde el ámbito de la justicia penal de Santiago del Estero:

- Protocolo de actuación para la extracción de evidencias digitales de dispositivos móviles.
- Repositorio digital para la gestión de evidencias digitales extraídas de dispositivos móviles.

3. RESULTADOS OBTENIDOS Y ESPERADOS

El objetivo general de la investigación propuesta es:

- *Contribuir al mejoramiento de la calidad del proceso de obtención de evidencias digitales obtenidas de dispositivos móviles en el ámbito judicial de Santiago del Estero.*

Los objetivos específicos que permitirán alcanzar el objetivo general son:

- *Validar el Protocolo para la obtención de evidencias digitales de dispositivos móviles en el ámbito judicial de Santiago del Estero.*
- *Ampliar el Protocolo de obtención de evidencias digitales de dispositivos móviles considerando el sistema de clasificación de herramientas forenses de dispositivos móviles.*
- *Analizar alternativas de construcción de repositorio digital para la gestión de*

evidencias digitales extraídas de dispositivos móviles.

En esta propuesta se pretende llevar adelante una investigación aplicada para validar el Protocolo de Actuación propuesto, para el cumplimiento de buenas prácticas que garanticen tanto la calidad de los procesos aplicados como de los resultados obtenidos, tomando como referencia estándares internacionales mencionados anteriormente.

La hipótesis planteada es la siguiente:

El uso de un protocolo preestablecido de Informática Forense para móviles y de un repositorio especializado, optimiza la gestión de evidencias digitales extraídas de los dispositivos móviles.

Como puede observarse, la variable a estudiar es la "optimización de la gestión de evidencias criminales obtenidas de dispositivos móviles", que podrá ser evaluada a través de indicadores cuantitativos que se pueden aplicar a casos de prueba especialmente diseñados.

Para realizar la validación de la aplicabilidad del protocolo propuesto se planificaron reuniones de trabajo y experiencias de puesta en marcha en el ámbito del Ministerio Público Fiscal de Santiago del Estero, con el fin de lograr un consenso técnico en la materia y permitiendo que se generen nuevas versiones del mismo a partir de las sugerencias y recomendaciones derivadas de los informes elaborados como resultado de dichas experiencias.

Se considera que la validación de la propuesta de protocolo, y su posterior aceptación por parte del Ministerio Público Fiscal y el Poder Judicial de Santiago del Estero, traerán un beneficio significativo para la justicia santiagueña, dado que actualmente no existe un procedimiento claro y definido que brinde un marco de trabajo a las actividades periciales. Esto permitiría mejorar la calidad de las evidencias digitales y ayudará en la labor de los fiscales y peritos de la provincia.

4. FORMACIÓN DE RECURSOS HUMANOS

La Directora y Codirectora del proyecto pertenecen al Departamento de Informática de la Universidad Nacional de Santiago del Estero. El asesor es experto en Informática Forense y jefe del área de Informática Forense del Poder Judicial de Río Negro.

Los investigadores constituyen un equipo interdisciplinario conformado por cuatro docentes de la UNSE y un investigador externo, con profesión en Informática, Electromecánica y Derecho. Estos poseen distintas categorías de investigación y algunos desempeñan sus actividades profesionales en el Gabinete de Ciencias Forenses del Ministerio Público Fiscal de Santiago del Estero y en el Juzgado Federal de Santiago del Estero.

El equipo de investigación se encuentra asistiendo y asesorando a alumnos de grado y posgrado de UNSE que realizan sus trabajos finales de carrera en temáticas relacionadas con esta línea de investigación, los que se encuentran actualmente en etapa de finalización.

5. REFERENCIAS

1. AYERS, R.; BROTHERS, S.; JANSEN, W. (2014). Guidelines on Mobile Device Forensics. NIST Special Publication 800-101. Revision 1. http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-101r1.pdf
2. CASTILLO, C.; ROMERO, A.; CANO, J. (2008). Análisis Forense Orientado a Incidentes en Teléfonos Celulares GSM: Una Guía Metodológica. XXXIV Conferencia Latinoamericana de Informática, Centro Latinoamericano de Estudios en Informática (CLEI). http://www.clei2008.org.ar/
3. DEL PINO, S. (2007). Introducción a la informática forense. Pontificia Universidad Católica del Ecuador. http://www.criptored.upm.es/guiateoria/gt_m592b.htm

4. FENNEMA, M.; FIGUEORA, L.; VIAÑA, G.; LESCA GOMEZ, N.; LARA, C. (2017). Tratamiento de evidencias digitales forenses en dispositivos móviles. XIX Workshop de Investigadores en Ciencias de la Computación. ISBN 978-987-42-5143-5.
5. HERRERA S.; FIGUEROA L.; GHUNTER D.; LARA C.; VIAÑA G.; MÉNDEZ A.; LESCA N. (2019)"Métodos y herramientas para el análisis forense de dispositivos móviles". XXI Workshop de Investigadores en Ciencias de la Computación. http://redunci.info.unlp.edu.ar/docs/Libro_WICC_2019-con_paginas.pdf
6. INFO-LAB. (2015). Guía integral de empleo de la informática forense en el proceso penal. http://info-lab.org.ar/images/pdf/14.pdf
7. ISO/IEC 27037:2012 (en) Information technology — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence. https://www.iso.org/obp/ui/#iso:std:iso-iec:27037:ed-1:v1:en
8. ISO/IEC 27042:2015 (en) Information technology — Security techniques — Guidelines for the analysis and interpretation of digital evidence. https://www.iso.org/obp/ui/#iso:std:iso-iec:27042:ed-1:v1:en
9. MELLAR, B. (2004). Forensic Examination of Mobile Phones. Digital Investigation – Elsevier. United Kingdom. http://faculty.colostate-pueblo.edu/dawn.spencer/Cis462/Homework/Ch4/Forensic%20examination%20of%20mobile%20phones.pdf
10. ORTA MARTINEZ, R. (2013). Informática Forense como Medio de Pruebas. http://www.dragonjar.org//informatica-forense-como-medio-de-prueba.xhtml
11. PODER JUDICIAL DE NEUQUÉN. (2013). Pericias informáticas sobre telefonía celular. http://200.70.33.130/images2/Biblioteca/ProtocoloPericiasTelefoniaCelular.pdf
12. REITH, M.; CLINT, C.; GUNSCH G. (2002). An Examination of Digital Forensic Models. International Journal of Digital Evidence, Air Force Institute of Technology, Volume 1 Issue 3. www.utica.edu/academic/institutes/ecii/publications/articles/A04A40DC-A6F6-F2C1-98F94F16AF57232D.pdf.
13. RFC3227 - Guidelines for Evidence Collection and Archiving. https://www.ietf.org/rfc/rfc3227.txt
14. SUP. TRIBUNAL DE JUSTICIA DEL PODER JUDICIAL DE RÍO NEGRO. (2015). Guía de procedimientos para pericias de dispositivos móviles. http://digesto.jusrionegro.gov.ar/bitstream/handle/123456789/455/Ac011-15.pdf?sequence=1&isAllowed=y
15. UNIDAD FISCAL ESPECIALIZADA EN CIBERDELITOS. (2016). Guía de obtención, preservación y tratamiento de evidencia digital. http://www.fiscales.gob.ar/procuracion-general/wp-content/uploads/sites/9/2016/04/PGN-0756-2016-001.pdf
16. VARSALONE, J., KUBASIAK, R. (2009). Mac Os X, iPod and iPhone Forensic Analysis DVD Toolkit. Syngress Publishing, Inc, pp. 355-475.
17. VIAÑA, G.; FIGUEORA, L.; LARA, C.; LESCA GOMEZ, N. (2018). Protocolo de actuación para recolección y preservación de la evidencia digital móvil en el Sistema Procesal Penal de Santiago del Estero. VI Congreso Nacional de Ingeniería en Informática y Sistemas de Información.
18. ZDZIARSKI, J. (2008). iPhone Forensics, Recovering Evidence, Personal Data & Corporate Assets. O"Reilly Media, Inc.

Tecnología Informática Aplicada en Educación

Incorporación de la Tecnología Móvil en el Proceso Educativo

Mg. Roberto Bertone[1], Mg. José Luis Filippi[2], Lic. Guillermo Lafuente[3], Mg. Carlos Ballesteros[4], Lic. Gustavo Lafuente[5], I.S. Daniel Perez[6], I.S. Sofía Aguirre[7], A.S. Alejandra Mansilla[8]
LIAU[9] - Facultad de Ingeniería – UNLPam.
pbertone@ada.info.unlp.edu.ar[1]
{filippij[2], lafuente[3], balleste[4], gustavo[5], perezd[6], aguirres[7], mansilla[8]}@ing.unlpam.edu.ar
[9]Laboratorio de Investigación de Ambientes Ubicuos

Resumen

La integración de las TICs en las instituciones educativas es un proceso que se desarrolla desde hace varias décadas en la República Argentina. Si bien existe un fuerte consenso en la universalización del acceso a las diferentes tecnologías de la información y las comunicaciones, la discusión vigente reside en la incorporación de los dispositivos móviles (teléfonos de última generación) como instrumento mediador del Proceso de Enseñanza-Aprendizaje.

En la actualidad la educación no se encuentra limitada a los entornos formales conformado por las instituciones educativas en sus diferentes niveles, se integra con procesos formativos que tienen lugar a través de los nuevos dispositivos tecnológicos de última generación, caracterizados por su miniaturización, movilidad y conectividad permanente; haciendo posible el acceso a múltiples contextos de aprendizaje virtuales. Resulta entonces imprescindible estudiar las tendencias del aprendizaje móvil en el mundo desarrollado, realizar las adaptaciones necesarias e implementarlas en el ámbito educativo.

El propósito del proyecto, de naturaleza teórico-práctico, es el de indagar las posibilidades que ofrecen los dispositivos móviles (teléfonos inteligentes, tabletas digitales) como instrumentos aplicados a diferentes situaciones de enseñanza y aprendizaje.

Bajo esa premisa se intenta llevar a cabo un trabajo experimental en la Facultad de Ingeniería - UNLPam, desarrollando un escenario de aplicación real con la utilización de éste tipo de dispositivos.

Palabras claves: Aprendizaje Móvil. Dispositivo Móvil. Aprendizaje Colaborativo.

Contexto

Tipo de Investigación: Aplicada
Campo de Aplicación Principal: 7 1802 Computación, 7 1803 Comunicaciones.
Campos de Aplicación posibles: 13 1040 Ciencia y Tecnología, 7 4399 Otras – Educación – Tecnología Aplicada a la Educación
Institución que Coordina el Proyecto: Facultad de Ingeniería, Universidad Nacional de La Pampa.

Introducción

Con más de 5900 millones de usuarios de telefonía móvil en el mundo, los dispositivos de última generación han transformado nuestra manera de vivir. Aunque en todos los sectores de la sociedad se hace uso intensivo de esta tecnología, los educadores y los responsables de formular las políticas educativas no han aprovechado su potencial para mejorar el proceso educativo. **[1]**

Múltiples necesidades se presentan en el ámbito educativo en sus diferentes niveles entre los cuales se pueden mencionar: desarrollo de contenidos bibliográfico digital de producción local y regional, incentivar el uso de aplicaciones móviles que posibiliten el intercambio de información entre los actores

que conforman la comunidad educativa (videoconferencias entre docentes, estudiantes, directivos, y comunidad en general), incorporar el uso de plataformas de formación virtual abiertas, que posibiliten incorporar recursos y/o materiales didácticos en diferentes formatos, mantener un canal de comunicación fluido entre la institución educativa y el contexto local, regional, y nacional, e implantar estrategias metodológicas para alcanzar un aprendizaje cooperativo y colaborativo.

Mark Weiser **[2]** menciona que: "*vamos camino a ambientes ubicuos, ambientes poblados de numerosos sensores que gracias a la miniaturización de los dispositivos son invisibles al usuario y están en permanente rastreo de la actividad humana*". Aquí radica el objetivo primario del proyecto: indagar diferentes aplicaciones tecnológicas móviles a través de las cuales se puedan ofrecer servicios que satisfagan las necesidades de los usuarios que transitan en el ámbito educativo, profundizando en el aprendizaje móvil.

Situación Actual del Problema

La aparición de dispositivos tecnológicos caracterizados por su miniaturización, gran capacidad de almacenamiento y gran velocidad de procesamiento, dieron origen a nuevos entornos en la comunicación de las personas, con una hipercomunicación caracterizada por la multimedialidad. El envío de mensajes textuales seguidos de imágenes, sonidos y videos es parte de la actividad cotidiana.

La educación influenciada por ésta realidad, acepta la necesidad de reconfigurar el proceso educativo, partiendo de algunas premisas:

a. Redefinir nuevas acciones para docentes y estudiantes.

El creciente número de dispositivos móviles y la conectividad a internet de forma permanente, permite a los estudiantes adquirir nuevos contenidos en línea y configurar sus conocimientos a partir de intereses propios. El estudiante es artífice de su futuro, asume un papel activo, decide que aprender y cómo.

Frente a la realidad del mundo digital que rodea a los estudiantes de hoy en día, es justo que los docentes replanteen su rol en la construcción del conocimiento sobre ellos. De allí que la competencia exigida a un profesor del Siglo XXI es preciso que se enfoquen "en las necesidades de los alumnos, supervisando su búsqueda de información e intentando facilitar la búsqueda de información individual de los alumnos ya que el papel de suministrador de conocimiento ha sido superado por las TIC" [3].

Mucho se habla de cuáles serían las competencias digitales para un docente, un acercamiento a ello lo propone Núñez-Torrón Stock [4], quien plantea cinco competencias básicas para el docente en la educación, y entre ellas se pueden destacar dos:

1. Usar herramientas de trabajo en línea.
2. Utilizar dispositivos móviles en el aula.

En la mayoría de las universidades de todo el mundo se observa la incorporación de los sistemas de Gestión de Aprendizaje (LMS), para implementar entornos virtuales que permiten la gestión del aprendizaje. **[5]** Sin embargo algunas investigaciones indican que las motivaciones de los estudiantes son vistas como una obligación al participar en entornos de aprendizaje institucionales cerrados, mientras que la motivación es mayor cuando el proceso formativo es abierto a diferentes tecnologías en contextos informales. Estos espacios son conocidos como PLE (Personal Learning Environments) donde los estudiantes configuran su propio entorno. **[6]**

Algunos autores ya incorporan un nuevo concepto, el de mPLE (Mobile Personal Learning Environments) como una nueva estructura de aprendizaje para la generación que hace uso de los dispositivos móviles. **[7]**

b. Continuidad del aprendizaje.

El aprendizaje móvil permitirá a los estudiantes dar continuidad a su formación a lo largo de toda su vida, a partir de plataformas que posibilitan el aprendizaje móvil caracterizado por:

- Poseer gran cantidad de información.
- Ser omnipresentes.
- Disponer de gran diversidad de recursos materiales.
- Fomentar la participación colaborativa.

Actualmente los estudiantes y futuros profesionales acceden a la información a través del aprendizaje informal, han dejado de ser únicamente consumidores para ser productores de información.

En el caso particular de los estudiantes universitarios no solo deben dominar las disciplinas correspondientes a su carrera, además deben incorporar habilidades transversales como el pensamiento crítico, resolución de problemas, persistencia y trabajo colaborativo. Sin embargo, en muchos países no se están desarrollando estas habilidades. **[8]**

c. Grandes volúmenes de datos. Big data.

Disponer y aprovechar ésta tecnología es una obligación para todas aquellas instituciones que manipulan grandes volúmenes de datos, en nuestro caso particular la universidad.

La gran cantidad de información accesible a partir de variadas herramientas conlleva una serie de consideraciones éticas relacionadas a la propiedad de los datos y la privacidad.

Los investigadores que estudian el aprendizaje en línea, los sistemas de tutoría inteligente, los laboratorios virtuales, las simulaciones y los sistemas de gestión del aprendizaje están explorando maneras de entender y utilizar mejor la analítica del aprendizaje, a fin de mejorar la actividad del docente y seguir avanzando en la educación para todos. **[9]**

En éste contexto el punto de partida debe ser siempre pedagógico. No se trata de innovar por innovar. Hay que explotar el potencial que ofrecen las nuevas herramientas tecnológicas que hacen posible el aprendizaje móvil, a partir de una planificación docente previa dentro de un marco teórico formalizado.

La herramienta en sí misma no ofrece por si sola resultados, pero un buen docente sabrá aprovechar las ventajas que brinda para un aprendizaje móvil de calidad.

Por ello el aprendizaje móvil (mLearning) puede incorporarse como refuerzo del aprendizaje formal, ampliando la oferta educativa y la modalidad. Lo importante es innovar en el proceso educativo en su conjunto y no solo con el dispositivo tecnológico.

El entorno educativo cambia **[10]**, la educación se presenta como la formación de los educandos en competencias, destrezas, habilidades para desempeñarse en un nuevo espacio social, el digital **[11]**. Nuevos escenarios educativos, los mismos actores con un nuevo rol, y la implantación de las tecnologías móviles es el desafío actual **[12].**

Línea de Investigación y Desarrollo

El plan de actividades corresponde al proceso de investigación aplicada, con objetivos de corto, mediano y largo plazo, y una duración prevista de cuatro años.

Primer año.

- Identificar el porcentaje de inserción de dispositivos móviles en la institución (teléfonos y/o tablets).
- Analizar la inserción del aprendizaje móvil en las universidades de todo el mundo.
- Estudiar como el aprendizaje móvil puede cerrar la brecha entre el aprendizaje formal y el informal.
- Identificar los escenarios educativos que muestren aspectos móviles en las

actividades formativas del ámbito académico propio.
- Examinar herramientas tecnológicas móviles orientadas al proceso educativo de acceso libre y gratuito.

Segundo y Tercer año.

- Definir el enfoque pedagógico que posibilite un aprendizaje móvil de calidad.
- Estudiar las tecnologías disponibles para implementar acciones de aprendizaje móvil en el contexto educativo universitario.
- Definir prioridades según necesidades de inmediatez en la virtualización de las disciplinas que se han de impartir.
- Gestionar el uso de plataformas de formación virtual que posibilite una doble modalidad, presencial y mediada por las nuevas tecnologías móviles.
- Incorporar el uso de redes sociales que favorezcan la práctica educativa.
- Confeccionar objetos de aprendizaje acorde a las herramientas disponibles en la nube y a las características de los dispositivos móviles que van a operar el producto final.
- Desarrollar aplicaciones móviles a partir de las necesidades que se presenten durante el transcurso del proceso educativo virtual.
- Capacitar a la comunidad educativa en general en el desarrollo de objetos de aprendizaje a partir de los requerimientos de carácter institucional.
- Difundir los avances a toda la comunidad universitaria los progresos a medida que van trascurriendo, a través de jornadas, congresos y/o revistas científicas.
- Propiciar el intercambio de información permanente con grupos de investigación que den valor agregado a nuestra actividad profesional.

Cuarto año.

- Instituir las aplicaciones desarrolladas para dispositivos móviles que hayan alcanzado buen nivel de aceptación.
- Conformar un repositorio de objetos de aprendizajes de acceso libre.
- Registrar y difundir los resultados alcanzados con la finalidad de que se puedan utilizar, ampliar y mejorar a través de trabajos futuros.
- Disponer de un catálogo de trabajos realizados en el sitio web del grupo de investigación GIAU (Grupo de Investigación de Ambientes Ubicuos) a partir del cual los interesados podrán acceder al material requerido.

Presentar en jornadas, congresos y/o revistas de todo el mundo los resultados alcanzados.

Resultados Obtenidos/Esperados

El proyecto da inicio a su actividad de investigación y desarrollo durante el año 2019, por lo que sus integrantes se encuentran efectuando las tareas enmarcadas en el segundo año.

A partir de la tarea realizada por los integrantes del proyecto de I+D durante el primer año se concretó:

a. Virtualizar las materias Introducción a la Informática, Programación Procedural, Programación Web y Autómatas y Lenguajes. El proceso consistió en el diseño de materiales formativos de diferente índole: videos instructivos, exposiciones multimedia, tareas de programación, bibliografía digital, autoevaluación, espacios que facilitan el intercambio de información y trabajo colaborativo. El producto logrado se presento en las Jornadas que se realizan anualmente entre las Facultades de Ingeniería y Cs. Veterinarias de la UNLPam, en Octubre de 2019.
b. Se elaboró un procedimiento para la programación de dispositivos inteligentes, su aplicación en el aula se documentó para ser expuesta la experiencia en CACIC 2019.
c. Se construyó un modelo teórico de una plataforma de educación a distancia para personas con discapacidad visual. El resultado del trabajo se presentó en JATIC 2019.

Al finalizar el proyecto se espera contribuir en la incorporación de aplicaciones tecnológicas

móviles a través de las cuales se puedan ofrecer servicios que satisfagan las necesidades de los usuarios que transitan en el ámbito educativo de la UNLPam y la UNLP, profundizando en el aprendizaje móvil.

Formación de Recursos Humanos

Director de Proyecto
Co-Director de Proyecto
6 Investigadores

Referencias

[1] Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2013). El futuro del aprendizaje móvil. Implicaciones para la Planificación y la Formulación de Políticas. http://unesdoc.unesco.org/images/0021/002176/217638s.pdf

[2] Weiser M. (1991), The Computer for the Twenty-First Century.

[3] webdelmaestrocmf.com, "Estas son las capacidades digitales de un docente moderno." [Online]. Available: http://webdelmaestrocmf.com/portal/5-competencias-digitales-que-deben-tener-los-profesores-actuales/?utm_source=blogsterapp&utm_medium=Facebook. [Accessed: 27-Oct-2017].

[4] A. Núñez-Torrón Stock, "5 competencias digitales que deben tener los profesores actuales." [Online]. Available: http://www.ticbeat.com/educacion/5-competencias-digitales-profesores/. [Accessed: 27-Oct-2017].

[5] M. P. Prendes, "Plataformas de campus virtual con herramientas de software libre: Análisis comparativo de la situación actual en las universidades españolas," Murcia, Project Report EA-2008-0257, Jun. 2009. [Online]. Available: http://www.um.es/campusvirtuales/informe_final_CVSL_SF.pdf

[6] F. J. García-Peñalvo, "Docencia," in Libro blanco univ. digital 2010, J. Laviña Orueta and L. Mengual Pavón, Eds. Barcelona, Spain: Ariel, 2010, pp. 29-61.

[7] G. Attwell, J. Cook, and A. Ravenscroft, "Appropriating Technologies for Contextual Knowledge: Mobile Personal Learning Environments," in Best Practices for the Knowledge Society. Knowledge, Learning, Development and Technology for All, M. D. Lytras, P. O. de Pablos, E. Damiani, D. Avison, A. Naeve, and D. G. Horner, Eds. Springer Berlin Heidelberg, 2009, pp. 15-25.

[8] (World Economic Forum, 2015).

[9] Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief.US Department of Education, Office of Educational Technology, 1-57.

[10] Weiser, M. (1998) *The future of Ubiquitous Computing on Campus*. Communications of ACM, 41-1, January 1998, 41-42.

[11] Malani R,. Griswold W, Simon B, (2009) *Public Digital Note-Taking in Lectures*. Ubicomp 2009.

[12] Richards M, Woodthorpe J, (2009), *Introducing TU100 "My Digital Life": Ubiquitous computing in a distance learning environment*. Ubicomp 2009.

Puesta en valor de un simulador de entrenamiento mediante la incorporación de experiencia inmersiva y analíticas de aprendizaje

Franco Lanzillotta[1], Stella Maris Massa[2], Adolfo Tomás Spinelli[3]
Grupo de Investigación en Tecnologías Interactivas
Facultad de Ingeniería, Universidad Nacional de Mar del Plata
[1]flanzi@fi.mdp.edu.ar, [2]smassa@fi.mdp.edu.ar, [3]spinelliadolfo@fi.mdp.edu.ar

RESUMEN

Los simuladores de entrenamiento son utilizados para generar entornos simulados que imitan la realidad en mayor o menor medida, a través de los cuales los usuarios pueden entrenar distintas habilidades y adquirir nuevos conocimientos. De esta manera, posibilita capacitar sin incurrir en costos operativos y sin los riesgos inherentes del escenario real. Pero la evaluación del aprendizaje y de las habilidades adquiridas durante la simulación está lejos de ser sencilla y pide métodos y modelos adicionales que produzcan evaluaciones válidas y evidencias de aprendizaje, lo que requiere datos adicionales de los aprendices.

En este marco, se ha propuesto llevar a cabo la puesta en valor de un simulador de entrenamiento para el uso de artillería antiaérea, incorporando mejoras en la simulación para aumentar la experiencia inmersiva, conjuntamente con la medición de datos y presentación de informes. Este sistema deberá inspeccionar cómo interactúa cada aprendiz con el simulador, almacenando información detallada y generando informes sobre las interacciones y los cambios en el estado interno de la simulación para un análisis posterior. De esta manera, permitirá al instructor evaluar objetivamente el desempeño del aprendiz, además de poder medir su progreso a lo largo de las sesiones de entrenamiento.

Palabras clave: Serious Games, Simuladores, Interacción Persona-Computadora, Analíticas de Aprendizaje.

CONTEXTO

Lo reportado en este artículo forma parte del Trabajo Final de Grado de Franco Lanzillotta, estudiante avanzado de la carrera Ingeniería en Informática de la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata (UNMDP). Este trabajo se enmarca en el proyecto "Analíticas de Aprendizaje en Simuladores de Entrenamiento", el cual forma parte de la beca de investigación del estudiante. Su directora, la Dra. Stella Maris Massa, es también la directora del Grupo de Investigación en Tecnologías Interactivas, en el cual se desarrolla este proyecto.

Este plan articula con el Proyecto de Investigación 15/G521 (2018-2019): "Modelos y herramientas para el proceso de desarrollo de serious games", Facultad de Ingeniería, UNMDP. El objetivo general de este proyecto es: "Proponer un Proceso de Elicitación para Serious Games en donde se combine el contexto y el perfil de usuario para mejorar tanto la experiencia del juego como la educativa". En particular, uno de los objetivos específicos es: "Construir una solución de software que proporcione información sobre el aprendizaje de los estudiantes en un Serious Game para realizar Analíticas de Aprendizaje.".

1. INTRODUCCIÓN

Los simuladores son definidos por Vidal-Gomel y Fauquet-Alekhine [1] como "artefactos que simulan (parcial o completamente) la operación o el comportamiento de un sistema técnico, un equipamiento o un fenómeno natural". Estos sistemas son utilizados para generar entornos simulados que imitan la realidad en mayor o menor medida, a través de los cuales los usuarios pueden entrenar distintas habilidades y adquirir nuevos conocimientos [2].

Los Simuladores de Entrenamiento (TS) se utilizan especialmente cuando es necesario que los usuarios puedan controlar o manipular sistemas complejos [3]. Su principal ventaja es el hecho de que su uso elimina los costos que conlleva capacitar en un escenario real y los riesgos inherentes a la actividad que se desea entrenar [4].

El uso de TS puede verse ampliamente difundido en muchas áreas en la actualidad, tales como el adiestramiento militar, el entrenamiento civil y las aplicaciones comerciales [5]. Otro ejemplo es el ámbito médico, en el cual se ha reducido el riesgo asociado con errores humanos y se ha incrementado la seguridad de los pacientes a través de la capacitación con TS [6].

Coincidiendo con Zyda [7], los TS pueden ser analizados como un caso particular de Serious Games (SG), debido a que la interacción del usuario con el entorno simulado genera una alta inmersión sensorial. Por otro lado, de acuerdo a la definición de SG que brinda el autor, lo que los separa de los videojuegos tradicionales es el hecho de que hacen uso del entretenimiento como medio para formar a sus usuarios en diversos ámbitos. Por lo tanto, los estudios realizados en esta temática pueden ser aprovechados para el desarrollo de TS.

Una etapa muy importante del entrenamiento basado en simulación es la de retroalimentación, momento en el cual tanto los aprendices como sus instructores pueden analizar los puntos fuertes y los aspectos a mejorar de las actividades realizadas [2]. Para llevar a cabo esta tarea, hacen falta herramientas que brinden información a partir de la cual se pueda evaluar el desempeño del aprendiz.

Las LA han sido definidas en el marco de la International Conference on Learning Analytics and Knowledge (LAK '11) como "la medida, colección, análisis y presentación de datos sobre los aprendices y su contexto, con el propósito de comprender y optimizar el aprendizaje y los entornos en los que se produce". Esta definición ha sido ampliada por Ferguson [8], quien ha añadido que las LA hacen uso de datos preexistentes que pueden ser leídos por una computadora y que las técnicas empleadas permiten el manejo de grandes conjuntos de datos, los cuales no podrían ser tratados de forma manual.

Tal como señalan Westera, Nadolski y Hummel [9], el análisis de los datos del desempeño de los aprendices en los SG posee dos diferentes etapas. En la primera, se analiza la información recopilada de cada usuario individualmente, con el fin de mejorar y personalizar su interacción con el sistema. En la segunda etapa, se trabaja con los datos de una población de jugadores para así poder evaluar y mejorar el videojuego en sí mismo.

De acuerdo con Baalsrud Hauge et al. [10], los tipos de datos a emplear en estas dos etapas son muy similares. Debido a esto, resulta imprescindible tener en cuenta la integración de las LA desde la etapa de diseño de los SG. La recomendación de estos autores es definir e incluir en el sistema una capa que se se encargue de traducir las acciones de los usuarios en

información relevante para medir el comportamiento del jugador dentro del SG, para poder así analizar el cumplimiento de los objetivos de aprendizaje de acuerdo a los criterios de evaluación.

El Modelo de Proceso de Desarrollo para Serious Games (MPDSG) [11] es una metodología para la construcción de Serious Games desarrollada por el Grupo de Investigación en Tecnologías Interactivas (GTI) de la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata. El MPDSG resulta de la combinación del Modelo de Proceso para el desarrollo de Objetos de Aprendizaje (MPOBA) [12], el modelo de Líneas de Producción de Software (LPS) [13] y el Diseño Centrado en el Usuario (DCU) [14].

El MPDSG propone un proceso de desarrollo iterativo en el cual se involucran a los distintos expertos que se encuentran relacionados con la producción del SG. En cada ciclo de este proceso se realiza una etapa de especificación, destinada a establecer los requerimientos y las características del producto. En esta etapa pueden identificarse tres fases [15]:

- **Elicitación:** se obtienen los requerimientos del videojuego, los cuales han de ser redactados de forma compresible y no ambigua.
- **Especificación:** se elaboran, en base a los requerimientos, los documentos que describen el videojuego que se ha de desarrollar.
- **Validación:** los expertos involucrados validan los requerimientos obtenidos y los documentos redactados.

Los escenarios de Leite et al. [16] son una herramienta utilizada para realizar la descripción de los requerimientos del videojuego de forma clara y concisa en lenguaje natural. Estos escenarios cuentan las acciones que pueden ocurrir dentro del videojuego, cuáles son los sujetos y objetos involucrados, qué cambios de estado pueden sufrir estos y cómo ocurren.

De acuerdo con lo desarrollado por Spinelli et al. [17], cuando se trata de un SG en el cual se van a implementar LA, resulta esencial definir el conjunto de habilidades a ser adquiridas por los aprendices a través de la extracción del conocimiento de los expertos. Luego, se deben relacionar con los escenarios construidos y, finalmente, determinar las variables a partir de las cuales se permite medir el alcance de estas habilidades. Todos estos elementos y relaciones debe ser validados por el juicio de los expertos y ajustarse a lo largo de las iteraciones del proceso de desarrollo.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Las principales líneas de investigación que se abordan en este proyecto son:

- Serious Games
- Simuladores de entrenamiento
- Analíticas de aprendizaje

3. RESULTADOS OBTENIDOS/ESPERADOS

El objetivo general del proyecto es llevar a cabo la puesta en valor de un simulador de entrenamiento para el uso de artillería antiaérea, incorporando mejoras en la simulación para aumentar la experiencia inmersiva, conjuntamente con la medición de datos y presentación de informes que permitan realizar una evaluación objetiva del desempeño del aprendiz.

En ese marco se establecieron los siguientes objetivos específicos:

- Analizar herramientas disponibles para el desarrollo de simuladores de entrenamiento.
- Identificar información relevante del proceso de aprendizaje de los aprendices sobre la base de sus datos de interacción en el simulador de entrenamiento.
- Construir una solución que proporcione información sobre el aprendizaje de los aprendices en el simulador de entrenamiento para poder realizar analíticas de aprendizaje.

Se ha llevado a cabo la elicitación del simulador propuesto a través de entrevistas a los expertos, los cuales fueron clasificados de acuerdo a su conocimiento del dominio. A partir de las entrevistas realizadas, se confeccionaron los escenarios de Leite para describir los requerimientos del simulador. También, se determinaron las habilidades a ser adquiridas por los aprendices a través de las sesiones de simulación. A cada una de estas habilidades se le asociaron distintas variables que permitan cuantificar el desempeño de los aprendices, para que pueda ser evaluado por sus instructores. A lo largo de las iteraciones del proceso de desarrollo, todos estos elementos fueron validados con los expertos con el fin de garantizar que el producto resultante cumpla con sus expectativas y necesidades. Actualmente, se está finalizando la fase de producción con la puesta en marcha del sistema, el cual continuará siendo probado mientras es utilizado, para así poder corregir sus errores, ajustarlo y mejorarlo.

4. FORMACIÓN DE RECURSOS HUMANOS

El estudiante Franco Lanzillotta se encuentra realizando su Trabajo Final de Grado de la carrera Ingeniería en Informática de la Facultad de Ingeniería de la UNMDP, paralelamente con la Beca de Investigación de Estudiante Avanzado de la misma universidad, siendo la Dra. Stella Maris Massa directora de ambos proyectos, mientras que el Ing. Adolfo Tomás Spinelli es co-director del Trabajo Final de Grado.

5. BIBLIOGRAFÍA

[1] Vidal-Gomel, C. y Fauquet-Alekhine, P. (2016). Reflections and theoretical contributions regarding trainers' practice and simulation. Simulation Training: Fundamentals and Applications – Improving professional practice through simulation training, Springer, Cham, pp. 1–29.

[2] Vázquez-Mata, G., Guillamer Lloveras, A. (2009). El entrenamiento basado en la simulación como innovación imprescindible en la formación médica. Educación Médica, vol. 12, no. 3, pp. 149–145.

[3] Drews, F. A. y Bakdash, J. Z. (2013). Simulation Training in Health Care. Reviews of Human Factors and Ergonomics, vol. 8, no. 1, pp. 191–234.

[4] Lechtaler, A. C., Blanc, C. F., Carden, M. L., Köhler, A., Polak, A. G. y Señoriño, J. M. (2015). Simulación Inmersiva con Realidad Aumentada. XVII Workshop de Investigadores en Ciencias de la Computación (WICC), Salta, Salta, Argentina.

[5] Camarasa, D. A. y Bianchi, O. M. (2012). Desarrollo de Software de Simulación Inmersiva para Fracciones Heterogéneas. XIV Workshop de Investigadores en Ciencias de la Computación (WICC), Posadas, Misiones, Argentina.

[6] Luciano, C., Banerjee, P. y DeFanti, T. (2009). Haptics-based virtual reality periodontal training simulator. Virtual Reality, vol. 13, no. 2, pp. 69–85.

[7] Zyda, M. (2005). From visual simulation to virtual reality to games. Computer, vol. 38, no. 9, pp. 25–32.

[8] Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, vol. 4 no. 5/6, pp. 304–317.

[9] Westera, W., Nadolski, R., Hummel, H. (2014). Serious Gaming Analytics: What Students Log Files Tell Us about Gaming and Learning. International Journal of Serious Games, vol. 1 no. 2, pp. 35–50.

[10] Baalsrud Hauge, J., Berta, R., Fiucci, G., Fernández Manjón, B. (2014). Implications of learning analytics for serious game design. Proceedings of the 14th International Conference on Advanced Learning Technologies (ICALT), Athens, Greece, pp. 230–232.

[11] Evans, F., Spinelli, A., Zapirain, E., Massa, S. M., Soriano, F. (2016). Proceso de desarrollo de Serious Games. Diseño centrado en el usuario, jugabilidad e inmersión. III Congreso Argentino de Ingeniería (CADI) – IX Congreso Argentino de Enseñanza de la Ingeniería (CAEDI), Resistencia, Chaco, Argentina.

[12] Massa, S. M. (2013). Objetos de Aprendizaje: Metodología de Desarrollo y Evaluación de Calidad. Tesis Doctoral, Facultad de Informática, UNLP, La Plata.

[13] Clements, P., Northrop, L. M. (2002). Software Product Lines: Practices and Patterns. Addison-Wesley.

[14] Hassan Montero, Y., Ortega Santamaría, S. (2009). Informe APEI sobre usabilidad. Informe APEI 3.

[15] Spinelli, A. T., Massa, S. M. (2018). Elicitación en Serious Game. Congreso Bienal IEEE ARGENCON 2018, San Miguel de Tucumán, Tucumán, Argentina.

[16] Leite, J. C. S. P., Haddad, G. D. S., Doorn, J. H. y Kaplan, G. N. (2000). A Scenario Construction Process. Requirement Engineering, vol. 5, no. 1, pp. 38-61.

[17] Spinelli, A. T., Massa, S. M., Rico, C., Kühn, F. (2018) Diseño de Serious Games. Requerimientos del Juego – Competencias y Habilidades. XX Encuentro Internacional VirtualEduca, Buenos Aires, Argentina.

METODOLOGÍA DE IMPLEMENTACIÓN DE UN CHATBOT COMO TUTOR VIRTUAL EN EL ÁMBITO EDUCATIVO

Orozco-González, Martha; Panizza, Luciano; Vegega, Cinthia; Pytel, Pablo & Pollo-Cattaneo, Ma. Florencia

Programa de Maestría en Ingeniería en Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina.

Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS). Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina.
{mmorozcog, ppytel, flo.pollo}@gmail.com

Resumen

Los ChatBots son instrumentos de usos múltiples que se utilizan hoy día en diversos ámbitos. Una de estas posibles aplicaciones es como Tutores Virtuales, apoyando el proceso de enseñanza-aprendizaje dado que ayudan a los estudiantes a avanzar en su desarrollo académico dentro de una asignatura universitaria. Éstos son sistemas software basado en Inteligencia Artificial y Procesamiento del Lenguaje Natural capaces de percibir su entorno, procesar lo que perciben y brindar una respuesta de manera racional, generando una conversación coherente. Aunque existen disponibles muchas herramientas y frameworks que facilitan la implementación de ChatBots, la construcción de un Tutor Virtual todavía no es una tarea sencilla. Por lo tanto, se identifica la necesidad de proponer una guía para el desarrollo de ChatBots, teniendo en cuenta las características del ámbito educativo dentro de las Universidades Latinoamericanas.

Palabras clave:

Inteligencia Artificial, Procesamiento del Lenguaje Natural, ChatBot, Tutor Virtual, Proceso Enseñanza-Aprendizaje.

Contexto

En el marco de las actividades del Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS) se comienza una nueva línea de trabajo que se articula dentro de los objetivos de GEMIS en el campo de la Informática, vinculando la Inteligencia Artificial (que está asociada al Procesamiento del Lenguaje Natural y específicamente al desarrollo de ChatBots) y de la Educación (buscando la automatización de tutoría de alumnos fuera del aula). De esta manera, se permite la generación de nuevos conocimientos en el área de la Ingeniería de Software con la aplicación de tecnologías no convencionales provenientes del Aprendizaje Automático (o Machine Learning en inglés), por lo que sus actividades se desarrollan dentro del ámbito del PID con incentivos UTN UTI5103TC, y una Tesis de Maestría.

Líneas de Investigación, Desarrollo e Innovación

Todavía es muy lejano pensar que la Inteligencia Artificial implemente en la realidad escenas de una película de ciencia ficción. De todas maneras, gracias

a la evolución reciente en términos tecnológicos, se encuentra presente en nuestras vidas. Tal es el caso de los Agentes Conversacionales, ChatBots y/o Asistentes Virtuales que pueden ser desarrollados a través del Procesamiento del Lenguaje Natural [1], una rama importante de la Inteligencia Artificial. Se puede definir a un Agente Conversacional o ChatBot, como un programa computacional diseñado para simular una conversación inteligente con una o más personas o, con otros sistemas software, a través de métodos auditivos o textuales [2]. Esta tecnología se inició en la década de 1960, siendo su objetivo engañar a los usuarios para hacerles creer que eran seres humanos reales. Sin embargo, estos sistemas no sólo se construyen para imitar la conversación humana o para entretener a los usuarios [3], sino que también tienen muchas aplicaciones prácticas. Grandes compañías como Microsoft, Google, Facebook, IBM y Amazon están invirtiendo cada vez más en los ChatBots [4] e incluso han desarrollado su propios frameworks para el desarrollo de los mismos. En la actualidad son diversos los usos de esta tecnología [5], y su aplicación en Latinoamérica ha ido aumentando a pasos agigantados. Según el estudio, realizado en Chile y Argentina por ChatBot Chocolate [6], ha sido posible establecer los cinco sectores en los cuales esta tecnología ha impactado con mayor fuerza en el año 2019, los cuales son el sector Bancario, Recursos Humanos, Turismo, E-commerce y Salud.

Por otra parte, en los últimos años, se identifican proyectos basados en Inteligencia Artificial que prometen revolucionar tanto la enseñanza como el proceso de aprendizaje [7], identificando tecnologías para hacer más eficientes sus procesos y las gestiones realizadas por los miembros de la comunidad universitaria. Como es sabido, los servicios de mensajería instantánea y las redes sociales son muy populares en todas partes del mundo, especialmente entre los adolescentes [4]. Es por ello que frecuentemente son utilizados dentro del ámbito educativo para compartir información relacionada sobre un tema específico. Dicha práctica está dando buenos resultados, por lo cual es posible mejorarla con la implementación de una herramienta más estructurada y con los conocimientos de un experto, que pueda llenar el hueco generado por los inconvenientes de comunicación en el aula. De esta manera se logra la implementación de Tutores Inteligentes que tengan la capacidad de:

- interactuar con los estudiantes para brindarles información,
- automatizar cuestionarios y sistemas de preguntas frecuentes,
- agilizar los procesos rutinarios dentro y fuera del aula, y
- apoyar el trabajo de los docentes permitiéndoles hacer seguimiento a la evolución de los estudiantes.

Para lograr estos objetivos existen comunidades y/o equipos de personas impulsadas por Universidades que se encuentran trabajando en la construcción de Tutores Inteligentes mediante la aplicación de Chatbots. Es posible encontrar diversas investigaciones relacionadas con el desarrollo e

implementación de ChatBots utilizadas como trabajos de grado y posgrados [8-10], además de varios ChatBots que han sido aplicados con éxito en el ámbito educativo [11-14].

No obstante, se han identificado algunas problemáticas por las cuales un Tutor Virtual o ChatBot no puede emular al 100% una conversación fluida y coherente como lo haría un interlocutor humano. Asimismo, es importante resaltar que los principales referentes de estos desarrollos tienen su base de conocimiento en idioma Inglés, por ser un idioma más estandarizado y fácil de parametrizar, mientras que el Español presenta una problemática mayor debido a la variación lingüística del lenguaje [15]. El principal problema para realizar Procesamiento de Lenguaje Natural en idioma Español es la variación lingüística del lenguaje, el cual consiste en que un mismo concepto se puede expresar de formas diferentes modificando la expresión, como el empleo de sinónimos, alteraciones en la estructura sintáctica, entre otros. Es un hecho conocido que una misma oración puede tener intenciones o interpretaciones diferentes, según el contexto en el que se enuncia. Esto genera, que uno de los problemas más importantes corresponde a la interpretación pragmática de la conversación [16].

Teniendo en cuenta estas dificultades, no es suficiente la utilización de herramientas o frameworks para el desarrollo de ChatBots, sino que también se debe contemplar la utilización de metodologías para la implementación de Chatbots en idioma Español aplicados al ámbito educativo. A pesar de la poca documentación disponible relacionada con dichas metodologías, se han identificado los siguientes casos [5; 17-20], donde se ha aplicado un proceso ingenieril. A partir de un estudio comparativo de estos casos [21] ha sido posible establecer que aún quedan varios aspectos que se deben profundizar y mejorar. Se puede afirmar que muchas metodologías dejan de lado aspectos importantes del desarrollo del ChatBot, con limitaciones que deben ser solucionadas.

En este contexto, se identifica la necesidad de proponer una nueva guía para el desarrollo de ChatBots, teniendo en cuenta las características del ámbito educativo dentro de las Universidades Latinoamericanas. Para llegar al planteamiento de la solución propuesta, ya se ha comenzado un análisis exhaustivo de la información recopilada en el transcurso de la investigación por lo que es posible detectar los objetivos que tendrán sus fases:

- En la Fase I, se debe buscar una clara definición de los requisitos, establecidos por los futuros usuarios (es decir docentes y estudiantes), para la implementación de un ChatBot aplicado a una asignatura. La correcta definición de los requisitos permitirá desarrollar un producto de calidad y buen rendimiento, logrando satisfacer las necesidades de dichos usuarios.
- La Fase II, se debe ocupar del diseño del ChatBot, así como en el de cualquier otra clase de software necesaria para su funcionamiento. Para ello se debe tomar en cuenta los requisitos funcionales y no

funcionales, definidos a través de casos de uso, escenarios, definición de interfaces, modelo de datos, modelos de implementación, y, por supuesto, el modelo de despliegue de la conversación, el cual hace referencia a las capacidades para el manejo del lenguaje natural.

- La Fase III, debe ocuparse de la selección de la herramienta o framework a utilizar para construir el ChatBot, teniendo en cuenta las necesidades de la asignatura identificadas en las fases anteriores.
- La Fase IV, lleva a cabo la construcción de los componentes del ChatBot. Para ello se utilizan los entregables obtenidos en las fases anteriores, así como también la herramienta seleccionada.
- En la fase V, se debe medir el nivel de calidad del ChatBot, así como también, el nivel de satisfacción de los usuarios. De esta manera es posible confirmar su correcto funcionamiento.
- Por último, la Fase VI se ocupa de la puesta en Operación del ChatBot. Para ello no basta con verificar las configuraciones básicas de conexión al ChatBot, sino que también se debe evaluar la integración de los componentes entre sí y con otros sistemas software.

Resultados y Objetivos

La presente línea de trabajo busca proponer una metodología que determine todas las fases, actividades y tareas de la construcción de un ChatBot de manera que cuente con mejor calidad de dialogo y respuesta en idioma Español, y que pueda ser aplicada exitosamente en el ámbito educativo universitario. De esta manera se espera asistir a un equipo de desarrollo de software en las actividades de análisis, diseño e implementación de un ChatBot para ser aplicado en una asignatura de grado de una carrera universitaria de una Universidad Latinoamericana logrando satisfacer las necesidades de estudiantes y docentes.

Asociados a este objetivo general se definen los siguientes objetivos específicos:

- Especificar las acciones necesarias para identificar los objetivos del ChatBot junto con las características de la asignatura de grado y el alcance del proyecto.
- Establecer las acciones necesarias que permitan generar los modelos de diseño del ChatBot, teniendo en cuenta los requisitos recolectados.
- Definir un conjunto de pasos para seleccionar la herramienta software más adecuada para el desarrollo del ChatBot, teniendo en cuenta las características del proyecto.
- Identificar los recursos técnicos con los que se cuenta para realizar el desarrollo del ChatBot, así como establecer los pasos necesarios para llevar a cabo la implementación de cada uno de sus componentes.
- Determinar las acciones necesarias para asegurar la calidad del ChatBot construido.
- Estipular los pasos requeridos para realizar apropiadamente la puesta en marchar del ChatBot.

- Efectuar una prueba de concepto para demostrar la validez de la aplicación de la nueva metodología.

Formación de Recursos Humanos

El equipo de trabajo se encuentra conformado por tres investigadores formados, una tesista de maestría, y un alumno en la carrera de Ingeniería en Sistemas de Información. Asimismo, se prevé incorporar más alumnos avanzados en la carrera de Ingeniería en Sistemas de Información con posibilidades de articular sus Trabajos Finales de Carrera de Grado, así como también tesistas de posgrado que desarrollarán sus Trabajos Finales Integrador de Especialidad. De esta manera se espera generar un verdadero espacio integrado de investigación en carreras de grado y posgrado.

Referencias

1. Van Grove, J. (2016) *Do people want to talk to bots?*. The San Diego Union-Tribune.
2. Jiménez, P. & Sánchez, J. (2015) *De Eliza a SIRI: la evolución*. Revista de Ciencia, Tecnología y Medio ambiente, Universidad Alfonso X el Sabio, Madrid.
3. Abushawar, B. & Atwell, E. (2007) *Chatbots: are they really useful?* LDV-Forum: Zeitschrift für Computerlinguistik und Sprachtechnologie, 22(1) pp. 29–49. doi: 10.1.1.106.1099.
4. Franceschin, T. (2016) *Chas ChatBots: Una Tecnología que puede revolucionar el Sistema Educativo*. Publicado en Edu4.me. Disponible en: http://edu4.me/los-chatbots-una-tecnologia-que-puede-revolucionar-el-sistema-educativo/
5. Gómez-Róspide, C. & Puente, C. (2012) *Agente virtual inteligente de ayuda al aprendizaje*, p. 178.
6. IA LATAM. (2019) *El uso de chatbots en Latinoamérica viene creciendo a pasos agigantados*. Publicado en http://ia-latam.com/2019/04/01/el-uso-de-chatbots-en-latinoamerica-viene-creciendo-a-pasos-agigantados/
7. Universia. (2018) *6 usos de los chatbots en la Universidad*. Publicado en https://noticias.universia.es/ciencia-tecnologia/noticia/2018/11/12/1162419/6-usos-chatbots-universidad.html
8. Toledo-Cambizaca, A. & Cordero-Zambrano, J. (2018) *Desarrollo de un chatbot que ayude a responder a preguntas frecuentes referentes a becas en la Universidad Técnica Particular de Loja*. Trabajo de titulación previo a la obtención del título de Ingeniero en Sistemas Informáticos y de Computación. Universidad Técnica Particular de Loja, Ecuador.
9. Garcia, L. & Menéndez, R. (2018) *Asistente Virtual tipo ChatBot*. Trabajo de titulación previo a la obtención del título de Ingeniero de Sistemas. Universidad Católica de Colombia.
10. Condori, W. (2017) *Desarrollo de un asistente virtual utilizando facebook messenger para la mejora del servicio de atención al cliente en la universidad privada de Tacna en el 2017*. Trabajo de titulación previo a la obtención del título de Ingeniero de Sistemas. Universidad Privada de Tacna.
11. Chatbots.org (2019) *Soy Diego*. Publicado en: https://www.chatbots.org/chatbot/diego/
12. BotGenes Restless Evolution. (2014) *Asistente Virtual de Clase*. Disponible en: http://botgenes.com/avc
13. A.L.I.C.E. A.I Foundation (2019) *Dave*. Disponible en: https://www.chatbots.org/chat_bot/dave/
14. Georgia Profesional Tech Education, (2019) Meet Jill Watson: Georgia Tech's first AI teaching assistant. Disponible en: https://pe.gatech.edu/blog/meet-jill-watson-georgia-techs-first-ai-teaching-assistant
15. Ferro, J. V. (2006) *Aplicaciones del Procesamiento de Lenguaje Natural en la Recuperación de Información en Español*, 36, pp. 57–58.
16. Rodríguez, J., Merlino, H., & Fernández, E. (2014) *Comportamiento Adaptable de ChatBots Dependiente del Contexto*. Tesis de grado en Ingeniería en Informática, Universidad de Buenos Aires.
17. Kuz, A. (2015) S*itio de la herramienta para Agente Inteligente*. Publicado en www.agenteinteligente.net
18. Ruiz-Tadeo, A., García, M. & Farias, N. (2009) *Sistema Inteligente Conversacional para la Orientación Vocacional*. Tesis para obtener grado de maestra en computación. Colima, México – 2009.
19. Limón, M. (2016) *Construir un prototipo de ChatBot unido a un sistema de consultas, que utilizando procesamiento de lenguaje natural permita a quien lo use extraer la información contenida en la base de datos y apoyar el proceso de toma de decisiones*. Tesis de grado en Licenciatura en Ciencias Informáticas. Instituto Politécnico Nacional, Ciudad de México.
20. Meza, M. & Álvarez, F. (2010) *Inteligencia Artificial Aplicada a Objetos de Aprendizaje a Través de la Tecnología del Chatbot Experto en Temas Específicos*. Tesis para Obtener el Título de Doctor en Ciencias de la Computación, Universidad Autónoma de Aguas Calientes, México.
21. Orozco-González, M., Pytel, P. & Pollo-Cattaneo, M.F. (2018). *Estudio de Herramientas y Metodologías Disponibles para la Implementación de ChatBots como Tutores Virtuales*. Memorias del 6to Congreso Nacional de Ingeniería Informática y Sistemas de Información (CONAIISI 2018). Workshop de Educación en Ingeniería. Artículo Nro 82-1322-1. ISSN 2347-0372.

El proceso de desarrollo de serious games
Modelos, herramientas y analíticas de aprendizaje

Stella Maris Massa[1], Lucrecia Ethel Moro, Gustavo Bacino, Adriana Pirro, Felipe Evans, Hernán Hinojal, Adolfo Spinelli, Esteban Zapirain, Carlos Rico, Franco David Kühn y Franco Lanzillota

[1]Facultad de Ingeniería/ Universidad Nacional de Mar del Plata/Argentina
(7600) Av. Juan B. Justo 4302, +54-223-481660
smassa4@gmail.com, lucreciamoro@gmail.com, gustavo@gmail.com, adriana.pirro@gmail.com, evansfelipe@gmail.com, hhinojal@fi.mdp.edu.ar, estebanzapirain@gmail.com, carlos@fi.mdp.edu.ar, fdkuhn@mail.com, franco.lanzi96@gmail.com

RESUMEN

Los serious games (SG) son aplicaciones interactivas que se concentran en el uso de los principios de diseño de juegos para otros fines no meramente lúdicos.

En este trabajo se describe el Proyecto de investigación "El proceso de desarrollo de SG. Modelos, herramientas y analíticas de aprendizaje" que será desarrollado durante los años 2020/2021 en la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata (UNMDP), donde se proseguirá abordando el desarrollo de SG desde las problemáticas relativas a su aplicabilidad en diferentes dominios como defensa, salud, educación, gestión de emergencias, planificación urbana e ingeniería.

Como educadores y desarrolladores de videojuegos, la validez de los SG con respecto a sus objetivos educativos debe ser tanto mensurable como medida. La naturaleza interactiva de los videojuegos hace que la aplicación de analíticas de aprendizaje se constituya en una herramienta útil para capturar los datos de interacción de los estudiantes o aprendices con el propósito de interpretar el proceso de aprendizaje. Sin embargo, existen escasas investigaciones y normas para comunicar información entre videojuegos y sus módulos de seguimiento.

Este proyecto tiene como objetivo consolidar el proceso de análisis de la interacción de los aprendices en SG a través de la extracción de información relevante del aprendizaje y la generación de reportes que visualicen los resultados. En este proyecto se combinan dos tendencias principales en la investigación del aprendizaje con tecnologías: SG y analíticas de aprendizaje.

***Palabras clave*:** Modelos, procesos de desarrollo, serious games, analíticas de aprendizaje.

CONTEXTO

Desde el año 2018 los integrantes del Grupo de Investigación en Tecnologías Interactivas (GTI) de la Facultad de Ingeniería de la UNMDP vienen desarrollando tareas de investigación relacionadas con los SG.

En el marco del Proyecto 15/G521: "Modelos y herramientas para el proceso de desarrollo de SG" se focalizó la investigación en los modelos y herramientas para el desarrollo de los SG, que son un tipo de videojuegos donde el objetivo trasciende el mero entretenimiento y busca incorporar un nuevo conocimiento en el jugador. Y en particular, se precisaron marcos conceptuales amplios, desarrollando y consolidando metodologías y herramientas para su desarrollo [1 – 4]. Además, se presentaron publicaciones internacionales que ponen en valor buenas experiencias disruptivas en contextos de Escuelas y Facultades de Ingeniería y Humanidades [5 – 7].

1. INTRODUCCIÓN

Los SG son aplicaciones interactivas creadas con una intencionalidad educativa, que proponen la explotación de la jugabilidad como experiencia del jugador. Presentan a los jugadores retos y misiones que implican toma de decisiones, resolución de problemas, búsqueda de información selectiva, cálculos, exploración, experimentación, competencia, cooperación etc., logrando el efecto inmersivo en el juego, como una prolongación de la experiencia vital del usuario[8 – 9].

La efectividad de los SG como herramientas de simulación y entrenamiento radica en su nivel de inmersión visual, sonora y emocional con el usuario y en su característica de interactividad constante. La experiencia de juego permite a un usuario entrar en estado flow (de inmersión y concentración total) en un tiempo mucho menor que por otros medios [10].

Corresponde destacar que las dificultades para medir los resultados del aprendizaje alcanzado a través del uso de los SG han constituido una barrera en el éxito de su adopción en educación [11 – 12]. Este conflicto entre aprendizaje y desempeño será mayor a medida que los SG ofrezcan más opciones abiertas y libertad de movimiento a los estudiantes, ya que la evaluación del aprendizaje requiere de métodos y modelos adicionales que produzcan evaluaciones válidas y evidencias de aprendizaje basado en juegos, lo que demanda datos adicionales de los jugadores.

Tradicionalmente, la validación de los SG se ha hecho a través de experimentos formales, mediante cuestionarios que los estudiantes completan antes y/o después de interactuar con el juego [13]. No se descartan estas metodologías y han sido trabajadas desde el proyecto 15/G520: "Ambientes de aprendizaje enriquecidos con tecnología" [14 – 15], donde han sido exploradas las características de los ambientes enriquecidos con tecnología con el objetivo de consolidar los estudios propios y su aplicación, en particular al caso de los SG en el aula [16 – 17].

La integración de Analíticas de Aprendizaje en el diseño de SG ofrece nuevas oportunidades para rastrear y analizar datos del comportamiento de los estudiantes sobre la base de su interacción individual o grupal, personalizar e interpretar el proceso de aprendizaje y realizar recomendaciones [18 – 20].

En la actualidad, la riqueza de datos reunidos a través de las web-log, motores de seguimiento, eyetrackers, localización y detectores de movimiento, en combinación con las emergentes Analíticas de Aprendizaje, representan una excelente oportunidad para mejorar la supervisión y evaluación del aprendizaje basado en juegos.

Para aplicar esta tecnología emergente en los SG, los datos pueden ser capturados mediante la adición de un "rastreador" que envía datos de la interacción del jugador (también denominados rastreos) a un servidor. El análisis de las huellas puede producir información relevante sobre las interacciones de los estudiantes con el juego, haciendo que el conjunto de acciones, errores y aciertos del jugador sea significativo [21].

Sin las Analíticas de Aprendizaje, los SG en la educación resultan similares a las cajas negras: simplemente proporcionan un estado final que demuestra los resultados del juego, normalmente en forma de métricas relativas a la puntuación final del jugador, pero que no suministran información con respecto al proceso de aprendizaje [22].

En el proyecto 15/G521 se realizó un estudio exploratorio mediante la revisión sistemática de la literatura (años 2008 a 2017) respecto a los enfoques y herramientas propuestas para una correcta implementación de Analíticas de Aprendizaje en SG. En el proceso de análisis de los artículos presentados, se comprobó que el desarrollo de los SG y su aplicación en el ámbito del aprendizaje real es prometedor, pero bastante pobre en cuanto a los recursos que se utilizan para su implementación. La

mayoría de las publicaciones están enfocadas en la implementación de estas tecnologías en el ámbito educativo, pero las soluciones propuestas son variadas, sin encontrarse lineamientos generales más allá de la definición de las propias analíticas de aprendizaje [23].

La manera particular en que se obtengan los datos determinará profundamente las posibilidades tanto en el análisis como en la visualización, ya que implicará la cantidad y la riqueza de la información recibida [24]. Las visualizaciones de resultados también deberán ajustarse a las necesidades de los diferentes agentes involucrados: estudiantes y profesores, como parte del uso educativo del juego, así como para desarrolladores, que estarán más interesados en el correcto funcionamiento del juego, pero también desean conocer si el juego está cumpliendo con los objetivos de aprendizaje previstos.

La principal diferencia de diseño entre los SG y los videojuegos, radica en la especificación de sus requerimientos. En ellos deben incluirse los contenidos y objetivos educativos junto a la evaluación del aprendizaje y la calidad educativa. En artículos previos a este proyecto se hace referencia a las consideraciones teóricas y prácticas adoptadas para obtener los requerimientos educativos [25] y a la refinación del proceso, presentando un mecanismo para relacionar los eventos del serious game y las habilidades que debe alcanzar el jugador a través del juego en relación a los objetivos educativos planteados [26]. En una de las tesis de Grado en Ingeniería Informática de la Facultad, se presentó un caso de estudio, con el correspondiente desarrollo del prototipo de software [27].

Es importante señalar que, se considera que las Analíticas de Aprendizaje en SG no sólo deben centrarse en el análisis o la visualización de datos, sino también deberían proveer a los profesores de las herramientas que les permitan seguir el aprendizaje de sus estudiantes [28].

2. LÍNEAS DE INVESTIGACIÓN

Las principales líneas de investigación que se abordan en este proyecto son:

- Interacción persona-ordenador (IPO).
- Serious games y ludificación.
- Analíticas de aprendizaje y SG.

3. RESULTADOS ESPERADOS Y OBJETIVOS

El objetivo general del proyecto de investigación que se presenta en este artículo es el de consolidar el proceso de análisis de la interacción de los estudiantes en SG para obtener información relevante del aprendizaje.

En ese marco, se establecen los siguientes objetivos específicos:

- Determinar las tareas y secuencias que permitan conformar el proceso de incorporación de Analíticas de Aprendizaje en los SG.
- Seleccionar las tecnologías adecuadas para la visualización de las interacciones de los estudiantes en los SG.
- Generar una estructura de visualización de las interacciones de los estudiantes en los SG.
- Producir reportes que visualicen la información relevante del proceso de aprendizaje de los estudiantes sobre la base de sus datos de interacción en SG.
- Desarrollar un trabajo de campo en el que se analizará la viabilidad y posibilidades del Proceso de incorporación de Analíticas de Aprendizaje para SG.

A partir del cumplimiento de los objetivos planteados, se espera consolidar el proceso de análisis de la interacción de los estudiantes en un serious game, a través de la extracción de información relevante del aprendizaje y la

generación de reportes que visualicen los resultados. Para ello se combinarán dos tendencias principales en la investigación del aprendizaje con tecnologías: SG y Analíticas de Aprendizaje.

Sistematizar los procesos de producción de SG, particularmente la selección de estrategias para extraer información relevante del proceso de aprendizaje del jugador/estudiante, contribuirá con la difusión de buenas prácticas en un sector en expansión y permitirá la apertura de líneas de investigación mediante la elaboración de un marco de referencia sobre metodologías y tecnologías emergentes.

4. FORMACIÓN DE RECURSOS HUMANOS

La formación de recursos humanos se considera un aspecto imprescindible e insoslayable. Integrantes del proyecto se encuentran desarrollando y dirigiendo cinco tesis de postgrado, en el marco del proyecto de investigación, correspondientes al Doctorado en Humanidades y Artes- Mención Ciencias de la Educación (UNR); al Doctorado en Modelado y Simulación Computacional, a la Especialización en Docencia Universitaria (UNMdP) y a la Maestría en Ingeniería de Software (UNLP). También se encuentra en ejecución, una Beca de Investigación de estudiante avanzado de la UNMdP.

5. REFERENCIAS BIBLIOGRÁFICAS

[1] Evans, F., Bacino, G., Rico, C., (2017). Modelo de Proceso de un Serious Game., En S. M. Massa y G. Bacino (Eds.), *Videojuegos en serio: creando serious games para aprender jugando* (pp. 41-59). Mar del Plata: Universidad Nacional de Mar del Plata.

[2] Massa, S. M. y Bacino G. (2017). *Videojuegos en serio: creando serious gamespara aprender jugando.* Mar del Plata: Universidad Nacional de Mar del Plata.

[3] Spinelli, A. T. y Massa S. M. (2018a). Elicitación en SG. En *Actas del IV Congreso Bianual de la IEEE Argentina (Argencon 2018)*, Universidad Nacional de Tucumán, Universidad Tecnológica de Tucumán, San Miguel de Tucumán, Argentina.

[4] Zapirain, E., Massa, S. M., Cardoso, J. P. (2019). Quantum Penny Flip: an open source serious game for quantum computing assessment. En *Actas del XIV Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2019)*, Universidad Nacional de San Luis. San Luis, Argentina.

[5] Jiménez Alcázar, J. F., Rodríguez, G. y Massa, S. M. (Coords.) (2018). *El videojuego en el aula de ciencias y humanidades.* Murcia: Universidad de Murcia.

[6] Pirro, L., Fernández, M., Hinojal, H. (2018). Bitácora de Mr. Jones. Una experiencia Virtual,enJ. F. Jiménez Alcázar, G. Rodríguez y S. M. Massa (Coords.)*, El videojuego en el aula de ciencias y humanidades,* (pp. 121-134). Murcia: Universidad de Murcia.

[7] Massa, S.M. (2018). Educación con videojuegos: nuevos desafíos, enJ. F. Jiménez Alcázar, G. Rodríguez y S. M. Massa (Coords.)*, El videojuego en el aula de ciencias y humanidades,* (pp. 69-88). Murcia: Universidad de Murcia.

[8] del Moral Pérez, M. E. (2013). Advergames & edutainment: fórmulas creativas para aprender jugando. *Ponencia inaugural del Congreso Internacional de Videojuegos y Educación (CIVE, 2013)*. Facultad de Formación del Profesorado de Cáceres (Universidad de Extremadura), Cáceres, España.

[9] Westera, W., Nadolski, R., Hummel, H. & Wopereis, I. (2008). SG for Higher Education: a Framework for Reducing Design Complexity. *Journal of Computer-Assisted Learning (Wiley), 24*(5), 420-432.

[10] Hamari, J. &Koivisto, J. (2014). Measuring flow in gamification: Dispositional Flow Scale – 2. *Computers in Human Behavior, 40*, 133-143.

[11] Alvarez, J. & Michaud, L. (2008). *SG – Advergaming, edugaming, training and more*. IDATE Consulting and Research.
[12] de Freitas, S. &Liarokapis, F. (2011). Serious games: A New Paradigm for Education? En M. Ma, et al. (Eds.) *SG and Edutainment Applications*, (pp. 9-23). UK: Springer.
[13] Calderón, A., & Ruiz, M. (2015). A systematic literature review on *serious games*evaluation. An application to software project management. *Computers & Education, 87,* 396–422.
[14] Hinojal, H. (2018). El uso de simuladores en el aula universitaria y los *serious games*.*Revista e-tramas, 1,* 34-50.
[15] Hinojal, H. y Massa, S. M. (2018). SG en el aprendizaje de Redes de Computadoras. Un análisis desde la experiencia óptima.En *Actas del IV Congreso Bianual de la IEEE Argentina (Argencon 2018)*, Universidad Nacional de Tucumán, Universidad Tecnológica de Tucumán, San Miguel de Tucumán, Argentina.
[16] Moro L. y Massa S. M. (2018). Características de un ambiente de aprendizaje enriquecido con TIC. Un estudio de caso. En A*ctas del XX Encuentro Internacional Virtual Educa 2018,* Ciudad Autónoma de Buenos Aires, Argentina.
[17] Moro, L. E., Farías, Y. B. y Morcela, O. A. (2018). Aprender Ciencias jugando. Educación con videojuegos: nuevos desafíos, EnJ. F. Jiménez Alcázar, G. Rodríguez y S. M. Massa (Coords.)*, El videojuego en el aula de ciencias y humanidades,* (pp. 105-120). Murcia: Universidad de Murcia.
[18] Johnson, L., Adams Becker, S. & Cummins, M. (2012). *NMC Horizon Report: 2012 K-12 Edition*. Austin, Texas: The New Media Consortium.
[19] Serrano-Laguna, A., Torrente, J., Moreno-Ger, P. & Fernández-Manjón, B. (2014). Application of Learning Analytics in Educational Videogames. *Entertainment Computing, 5*(4), 313-322.
[20] Siemens, G. (2010). What are learning analytics. *Retrieved March, 10*, 2011.
[21] Loh, C. S., Sheng, Y., &Ifenthaler, D. (2015). SG analytics: Theoretical framework. In *SG analytics* (pp. 3-29). Springer, Cham.
[22] Alonso-Fernandez, C., Calvo, A., Freire, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. (April, 2017). Systematizing game learning analytics for serious games. In *2017 IEEE Global Engineering Education Conference (EDUCON),* (pp. 1111-1118). IEEE.
[23] Massa, S. M. y Kühn, F. D. (2018). Analíticas de Aprendizaje en serious games: una revisión sistemática de la literatura, En *Actas del IV Congreso Bianual de la IEEE Argentina (Argencon 2018)*, Universidad Nacional de Tucumán, Universidad Tecnológica de Tucumán, San Miguel de Tucumán, Argentina.
[24] Alonso- Fernández, C. (2016). Gaming learning analytics for SG.Trabajo fin de Grado. *Facultad de Informática. Universidad Complutense de Madrid*.
[25] Spinelli, A. T. y Massa S.M., (2018b). Elicitación de Requerimientos Educativos en un SG. En *Actas del XIII Congreso Nacional de Tecnología en Educación y Educación en Tecnología (TE&ET 2018)*. RedUnci, Universidad Nacional de Misiones, Posadas, Misiones, Argentina.
[26] Spinelli, A. T. y Massa, S. M. (2018c). Diseño de SG, requerimientos del juego – Competencias y habilidades. En *Actas del XX Encuentro Internacional Virtual Educa.* Ciudad Autónoma de Buenos Aires, Argentina.
[27] Kühn, F. (2019a). *Juegos Serios y Analíticas de Aprendizaje: Implementación en el entorno educativo*. Trabajo Final de Carrera de Grado en Ingeniería en Informática. Facultad de Ingeniería. Universidad Nacional de Mar del Plata. 2019
[28] Kühn, F. D. (2019b). Proceso de implementación de analíticas de aprendizaje en juegos serios dentro de espacios educativos. *Revista e-tramas, 3,* 20-39.

Aplicación de rúbrica C.O.d.A para evaluación de calidad objetos de aprendizajes basados en realidad aumentada

Lucas Kucuk[1], Jorge Ierache[2]
[1] Departamento de Informática, Facultad de Ciencias Exactas, Químicas y Naturales, Universidad Nacional de Misiones, [1]kucuk@hotmail.es
[2]Grupo de Realidad Aumentada, Departamento de Ingeniería e Investigaciones Tecnologías, Universidad Nacional de la Matanza, jierache@unlam.edu.ar

RESUMEN

En el proceso de enseñanza en el ámbito educativo, los objetos de aprendizajes mediados por tecnologías emergentes han tomado un rol significativo. En este trabajo se aborda en particular la implementación de la rúbrica C.O.d.A (Calidad de Objetos de Aprendizajes) para evaluar la calidad de un objeto de aprendizaje desarrollado con realidad aumentada utilizando un marco de trabajo. Se muestran los resultados de su despliegue en un caso de estudio con docentes y alumnos de la enseñanza de nivel medio de la educación técnica. Los resultados obtenidos destacan la motivación e interés de los alumnos en el uso de esta tecnología.

CONTEXTO

El trabajo se desarrolló en el marco de un trabajo de final de maestría de la Maestría en Tecnologías de la Información de la Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad Nacional de Misiones en conjunto con la Universidad Nacional del Nordeste.

1. INTRODUCCIÓN

La evaluación de los productos de software que se utilizan en educación es un área que amerita un análisis en profundidad, ya que existe un vasto contenido académico que se centra en el desarrollo e implementación y no tanto en la evaluación de la calidad de los mismos.

Entre las definiciones más aproximadas, la realidad aumenta (RA) se puede definir en base al Reality-Virtuality Continuum presentado por Milgram y Kishino [1], donde se la define como la integración de elementos reales y virtuales, pero considerándola más cercana al mundo real. Es una tecnología que complementa la percepción e interacción con el mundo real y permite al usuario estar en un entorno real aumentado, con información adicional generada por la del entorno real y la Virtualidad Aumentada (más cerca del entorno virtual).

Un Objeto de Aprendizaje (OA) es un recurso educativo que permite facilitar el proceso de enseñanza aprendizaje la definición de objetos de aprendizaje, y que en su estado digital se convierte en recurso educativo abierto (REA) [2]. Las características de los OA (desde ahora objetos de aprendizaje) son la flexibilidad, modularidad, reutilización, granularidad, Interoperabilidad, accesibilidad y portabilidad, conceptos que serán analizados con más detalle en el trabajo [3].

Un objeto de aprendizaje se compone de contenidos vinculados al aprendizaje objetivo, actividades que integran los contenidos de forma interactiva y que además generan una retroalimentación; el OA también debe ser identificable a través de elementos de identificación [4].

La evaluación de la calidad de los objetos de aprendizaje: (1) ayuda a los autores a crear mejores materiales didácticos en formato digital si se aplica durante la creación de los mismos de forma que vaya guiando la toma de decisiones y acciones para la obtención de las mejores puntuaciones en la evaluación de calidad; (2), sirve para valorar la dedicación que requiere la producción de OA de calidad, y (3) las evaluaciones obtenidas con C.O.d.A.

(Herramienta de Evaluación de la Calidad Objetos de Aprendizaje) [5], ayudan a los profesores y estudiantes a localizar en los repositorios -o contenedores- de objetos de aprendizaje los más adecuados a sus necesidades didácticas y técnicas.

Esta herramienta C.O.d.A. está dirigida principalmente a los profesores, investigadores y estudiantes que son creadores y usuarios de los OA, expertos en sus disciplinas, pero no necesariamente expertos en informática. Permite mejorar sus OA mediante el cumplimiento del mayor número de criterios posible. Además, puede ser utilizada por los revisores externos cuando los autores deseen o necesiten un reconocimiento de la calidad de sus colecciones de OA. C.O.d.A. se puede utilizar para guiar la creación de los OA, antes de su utilización real o para valorar su efectividad tecnológica y didáctica potencial del OA posteriormente a su utilización. Consiste en un formulario con diez criterios de calidad puntuables de 1 (mínimo) a 5 (máximo) y una guía de buenas prácticas para orientar la puntuación de los OA. Los cinco primeros criterios son de carácter didáctico, mientras que los otros cinco son tecnológicos de manera que ambos aspectos tienen el mismo peso [5].

1. Objetivos y coherencia didáctica
2. Calidad de los contenidos
3. Capacidad de generar reflexión, crítica e innovación
4. Interactividad y adaptabilidad
5. Motivación
6. Formato y Diseño
7. Usabilidad
8. Accesibilidad
9. Reusabilidad
10. Interoperabilidad (excluido)

Para cada uno de los criterios propuestos se evaluará la calidad con una puntuación de 1 a 5, siendo 1 el mínimo y 5 el máximo. Si se considera que alguno de los criterios no es relevante para el OA evaluado, o si el revisor no se considera cualificado para juzgar ese criterio, siempre se puede evitar seleccionando la opción "No Aplicable" (NA). Este modelo puede ser utilizado en la revisión individual, autoevaluación, en grupo y en la evaluación por pares [5].

Para desarrollar C.O.d.A., se han tenido en cuenta los modelos de evaluación de la calidad de contenidos educativos desarrollados en Universidades Españolas, entre las que destacan la UNED (UNED), el Campus Virtual de la UCM, la Universidad de Murcia [5]. Asimismo, sintetiza los criterios comúnmente compartidos por la mayoría de los modelos de evaluación de OA publicados como el de Paulsson y Naeve [6] y Kurilovas y Dagiene [7]. La presentación de C.O.d.A. está inspirada en la herramienta LORI (Learning Object Review Instrument) [8].

2. LINEAS DE INVESTIGACIÓN Y DESARRROLLO

El desarrollo del objeto de aprendizaje fue abordado teniendo en cuenta el marco de trabajo utilizado en [9]. La implementación de C.O.d.A es parte de uno de los procesos contemplados en el marco de trabajo para el desarrollo y evaluación de objetos de aprendizajes (OA) basados en RA que se puede ver en [9]. Para el caso de estudio se planteó la construcción de un OA utilizando un tablero de papel cartulina, una tarjeta (en papel) por cada video a mostrar, y un mazo de cartas (en papel) que contenga afirmaciones correctas e incorrectas y en el dorso contenga un target (marcador de realidad aumentada); también se desarrolló una aplicación que visualiza RA hecho con Aumentaty Creator. Cada tarjeta muestra un video educativo en RA, luego de observar las mismas, el usuario procederá a mirar todas las afirmaciones y debe colocar las que les parezca correctas en una ubicación específica del tablero mostrando el dorso de los naipes. Finalmente, el usuario utiliza el celular y la aplicación enfocando a los naipes, se visualiza en RA si las respuestas son correctas o incorrectas. El tablero dispone de un instructivo disponible en RA. Cable aclararse que a diferencia del trabajo hecho en [9], es que aquí se desarrolló otro OA, este basado en contenidos de Química.

La implementación del objeto de aprendizaje en total duró 180 minutos dividida en dos momentos de 90 minutos. El total de usuarios que utilizaron el framework es de 50 alumnos y 1 docente, hubo 5 ausentes (Figura 1).



Figura 1. Alumnos utilizando el OA

Una vez finalizado el proceso de prueba, el docente y los alumnos procedieron a realizar la evaluación de usabilidad. Se imprimieron 42 copias de la herramienta C.O.d.A. en formato de cuestionario. Cada criterio de la herramienta realiza preguntas al respecto de cada característica a evaluar. Los alumnos procedieron a calificar del 1 al 5 cada una de las características analizadas.

También se incorporaron procedimientos para realizar un análisis cuantitativo y cualitativo de los datos obtenidos de los cuestionarios. Consiste en codificar los datos dispuestos en las tablas de los cuestionarios (C=criterio, P=pregunta, A=5 hasta E=1) y resolver las siguientes fórmulas:

Total de puntos por Pregunta-Característica *(ptcr)*: es el resultado de sumar los productos del total de elecciones (valoraciones) por su correspondiente puntaje.

$$ptcr = puntaje\ total\ de\ la\ Pregunta - Característica$$
$$r_((n)) = respuesta$$
$$ptcr = r_((a)) * 5 + r_((b)) * 4 + r_((c)) * 3 + r_((d)) * 2 + r_((e)) * 1$$

Formula 1 - Total de puntos por Pregunta-Característica

Valoración máxima por característica *(vtcar)*: es la valoración máxima que se puede obtener por cada característica, para obtenerlo se debe multiplicar 5 (máxima valoración) por la cantidad de encuestados (*CE*).

$$vtcar = 5 * CE$$

Formula 2 - Valoración máxima por característica

Porcentaje de valoraciones por Pregunta-Característica (*PV):* se obtiene de dividir el total de valoración obtenido por cada Pregunta-Característica *(pctr)* y *vtcar*. Dividido el total de la valoración máxima obtenible por característica.

$$PV = pctr/vtcar$$

Formula 3 - Porcentaje de valoración por Dimensión-Criterio

3. RESULTADOS OBTENIDOS

Luego de realizar los cálculos se han completado una serie de tablas, que se muestran a continuación, que contienen la información de resultados obtenidos por cada criterio analizado. Se procedió a calcular la mediana aritmética del porcentaje de puntos obtenidos en cada Pregunta-Característica (*Me*). La mediana se utiliza para devolver la tendencia central en el caso de distribuciones numéricas sesgadas.

Tabla 1 - Tabla de Valores obtenidos en "Objetivos y coherencia didáctica"

C1								
Código Pregunta-Características	A	B	C	D	E	Ptcr	% de puntos	*Me*
P1	40	3	2	0	0	203	97%	92%
P2	35	5	5	0	0	199	92%	
P3	32	10	3	0	0	193	92%	
P4	34	10	1	0	0	200	94%	

Tabla 2 - Tabla de Valores obtenidos en "Calidad de los contenidos"

C2								
Código Pregunta-Características	A	B	C	D	E	ptcr	% de puntos	*Me*
P5	39	1	4	0	0	201	96%	92%
P6	35	10	0	0	0	191	91%	
P7	32	8	5	0	0	194	92%	

P8	34	6	5	0	0	198	92%	
P9	40	3	2	0	0	202	97%	
P10	35	8	2	0	0	199	92%	
P11	36	6	3	0	0	201	94%	
P12	38	2	5	0	0	201	94%	
P13	36	6	3	0	0	194	92%	

Tabla 3 - Tabla de Valores obtenidos en "Capacidad de generar reflexión, crítica e innovación"

C3								
Código Pregunta-Características	A	B	C	D	E	ptcr	% de puntos	*Me*
P14	35	3	7	2	0	188	95%	94%
P15	27	10	8	3	0	186	94%	
P16	38	4	3	0	0	190	96%	

Tabla 4 - Tabla de Valores obtenidos en "Interactividad y adaptabilidad del OA"

C4								
Código Pregunta-Características	A	B	C	D	E	ptcr	% de puntos	*Me*
P18	35	9	0	1	0	194	93%	93%
P19	35	9	0	1	0	194	93%	
P20	32	4	4	5	0	183	87%	
P21	41	3	1	0	0	203	97%	
P22	42	3	0	0	0	206	98%	
P23	35	7	2	1	0	194	93%	

Tabla 5 - Tabla de Valores obtenidos en "Motivación"

C5								
Código Pregunta-Características	A	B	C	D	E	ptcr	% de puntos	*Me*
P24	45	0	0	0	0	210	100%	100%
P25	45	0	0	0	0	210	100%	
P26	45	0	0	0	0	210	100%	

Tabla 6 - Tabla de Valores obtenidos en "Formato y Diseño"

C6								
Código Pregunta-Características	A	B	C	D	E	ptcr	% de puntos	*Me*
P27	32	8	3	0	0	198	94%	96%
P28	35	10	0	0	0	198	96%	
P29	43	2	0	0	0	208	99%	
P30	40	0	3	2	0	205	96%	
P31	38	2	2	2	1	194	96%	

Tabla 7- Tabla de Valores obtenidos en "Usabilidad"

C7								
Código Pregunta-Características	A	B	C	D	E	ptcr	% de puntos	*Me*
P32	38	7	0	0	0	200	93%	93%
P33	41	4	0	0	0	204	95%	
P34	35	5	5	0	0	193	92%	

Tabla 8 - Tabla de Valores obtenidos en "Reusabilidad"

C8								
Código Pregunta-Características	A	B	C	D	E	ptcr	% de puntos	*Me*
P35	40	2	3	0	0	193	96%	97%
P36	42	3	0	0	0	205	97%	
P37	44	1	0	0	0	209	99%	

Tabla 9 - Tabla de Valores obtenidos en "Interoperabilidad"

C9								
Código Pregunta-Características	A	B	C	D	E	ptcr	% de puntos	*Me*
P38	39	2	4	0	0	193	96%	97%
P39	43	2	0	0	0	205	97%	
P40	44	1	0	0	0	209	99%	

Finalmente se procedió a realizar un gráfico considerando las medianas aritméticas obtenidas por cada Dimensión-Criterio evaluada (Figura 2).

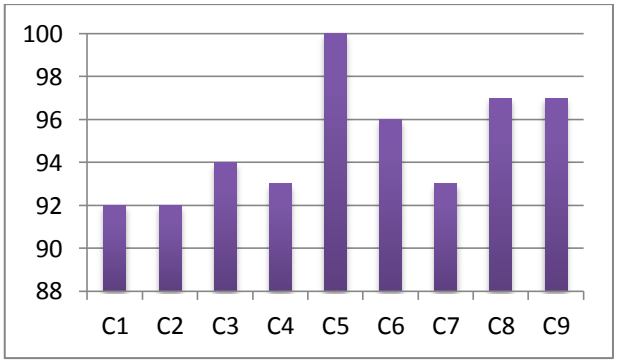


Figura 2 - Porcentajes de valoraciones por Dimensión-Criterio

Los resultados indican que existe una mayor valoración en la dimensión de "Motivación", con una mediana aritmética de 100% sobre los porcentajes de puntos obtenidos por características evaluadas. En segundo lugar, se hallan las dimensiones "Reusabilidad" y "Interoperabilidad" con 97%. En tercer lugar, se sitúa la dimensión "Formato y Diseño" con 96%, en cuarto lugar, se sitúa la dimensión "Capacidad de generar reflexión, crítica e innovación" con 94%, en quinto lugar, valoradas se hallan las dimensiones "Interactividad y adaptabilidad del OA" y "Usabilidad" con 93% y en sexta posición

quedaron las dimensiones “Objetivos y coherencia didáctica” y “Calidad de los contenidos” con 92%.

La tendencia central que describen las medianas aritméticas halladas muestran que los valores obtenidos en las características evaluadas por cada dimensión-criterio se hallan cercanas a los valores de 4 y 5, es decir, a las mejores puntuaciones. Destacando el resultado de “motivación”

La herramienta C.O.d.A demostró ser aplicable por usuario de bajo nivel de formación en evaluación de calidad de OA basados en RA.

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de trabajo participan los directores, docentes investigadores de la Universidad Nacional de La Matanza, con antecedentes de investigación en los temas que se abordan. El desarrollo del proyecto fue realizado por un estudiante de la Maestría en Tecnologías de la Información de la Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad Nacional de Misiones.

5. BIBLIOGRAFÍA

[1] P. Milgram, H. Takemura, A. Utsumi, F. Kishino. “Augmented Reality: A class of displays on the reality-virtuality continuum”.. Telemanipulator and Telepresence Technologies, 2012.

[2] L. Kucuk, J. S. Ierache y G. Dapozo. “Marco de trabajo para la usabilidad en objetos de aprendizajes basados en realidad aumentada”, In XX Workshop de Investigadores en Ciencias de la Computación, ISBN: 978-987-3619-27-4, Abril 2018.

[3] M. Gértrudix, et al., "Acciones de diseño y desarrollo de objetos educativos digitales: programas institucionales." International Journal of Educational Technology in Higher Education (ETHE)4.1, 2007.

[4] F. García Jiménez, et al. "Una experiencia didáctica en el diseño e implementación de objetos de aprendizaje para la enseñanza de la física.", Revista Educación en Ingeniería 11.22, p.p. 13-20, 2016.

[5] P. Fernández, R. Domínguez, R. Armas, “Herramienta COdA de Evaluación de la Calidad de Objetos de Aprendizaje, desarrollada en el marco de los Proyectos de Innovación y Mejora de la Calidad de la Docencia”, PIMCD 268/2010-2011 y PIMCD 236/2011-2012 financiados por el Vicerrectorado de Desarrollo y Calidad de la Docencia de la Universidad Complutense de Madrid, 2012.

[6] F. Paulsson y N. Ambjörn. "Virtual workspace environment (VWE): A taxonomy and service oriented architecture framework for modularized virtual learning environments-Applying the learning object concept to the VLE.", International Journal on E-learning 5.1, 45-57, 2006.

[7] E. Kurilovas,y V. Dagiene. "Multiple Criteria Evaluation of Quality and Optimisation of e-Learning System Components.", Electronic Journal of e-Learning 8.2, pp. 141-151, 2010.

[8] S. A. Rodríguez, “Instrumento para evaluar Recursos Educativos Digitales”, LORI – AD,2015

[9] L. Kucuk, J. S. Ierache y G. Dapozo, “Marco de trabajo para desarrollo e implementación de objetos de aprendizajes basados en realidad aumentada para la evaluación de usabilidad”, XIV Congreso Nacional de Tecnología en Educación y Educación en Tecnología ISBN: 978-987-733-196-7, Julio 2019.

Desarrollo de Herramientas IoT para la Enseñanza de Sistemas de Control

Marco Miretti[†], Joaquín Bicego[*], Paula B. Olmedo[‡], Emanuel Bernardi[#]
Applied Control & Embedded System - Research Group (AC&ES-RG)
Universidad Tecnológica Nacional
Facultad Regional San Francisco (UTN-FRSFCO)
San Francisco, Córdoba, Argentina
[†]marco.miretti@gmail.com, [*]joaquinbic@gmail.com,
[‡]paulabeatrizolmedo@gmail.com, [#]ebernardi@sanfrancisco.utn.edu.ar

RESUMEN

La línea de investigación formulada en este proyecto está orientada a la enseñanza de Sistemas de Control en ámbitos académicos. El mismo expone un conjunto de herramientas del internet de las cosas (IoT), donde sus dispositivos de hardware constituyen sistemas dinámicos, típicos del área, que posibilitan el diseño de controladores y su posterior validación. En tanto el paquete de software involucrado, diseñado en *Python*, se compone de una interfaz intuitiva y amigable que facilita la interacción con el hardware, permitiendo su utilización por parte de usuarios ajenos al ámbito de los sistemas embebidos.

***Palabras clave**:* Innovación Educativa, Sistemas de Control, Internet de las Cosas, Python.

CONTEXTO

Esta línea de investigación se enmarca en el área de Control Automático de Sistemas, del Departamento de Ingeniería Electrónica de la Universidad Tecnológica Nacional, Facultad Regional San Francisco (UTN-FRSFCO). En tanto, las actividades involucradas se llevarán a cabo por los integrantes del grupo de investigación sobre control aplicado y sistemas embebidos *Applied Control & Embedded Systems - Research Group* (AC&ES-RG), integrado por becarios alumnos, docentes investigadores y becarios doctorales. Además, es importante destacar que parte de este trabajo se constituirá como proyecto final de carrera de dos estudiantes, ambos participantes del mencionado grupo.

1. INTRODUCCIÓN

La experimentación como herramienta de aprendizaje, es en sí un método sumamente enriquecedor, que al ser complementada con conceptos teóricos posibilita la generación de resultados sobresalientes. Esto es particularmente relevante en áreas de formación ingenieril, donde realizar ensayos prácticos constituye una parte íntegra del estudio [1]. Entonces, a la hora de percibir un fenómeno, u observar el comportamiento de un sistema, la práctica resulta mandatoria.

En numerosas oportunidades, investigadores de distintos centros han desarrollado herramientas para complementar la enseñanza en un área particular [2, 3], con excelentes resultados. Por otro lado, se observa una creciente incorporación de sistemas IoT aplicados a la educación, por ejemplo herramientas de pruebas remotas [4, 5, 6] cuyas bondades son altamente ventajosas en el contexto de nuestra región, por tratarse de una institución a la que concurren alumnos de múltiples municipios circundantes.

1-A. Propuesta

En base a los fundamentos mencionados, a través de este proyecto se plantea el desarrollo de un conjunto de herramientas de acceso remoto para la experimentación en sistemas de control. Específicamente, se busca construir una serie de sistemas dinámicos que permitan a los

estudiantes el ensayo y validación de las técnicas de control instruidas.

Así, el objetivo de estas herramientas es centrar el aprendizaje en las múltiples técnicas de control disponibles actualmente [7, 8, 9], y no en las tediosas tareas involucradas en la programación de bajo nivel, i.e. programación de sistemas embebidos, configuración de las comunicaciones, etc. Es por ello, que se diseñará una librería intuitiva y simple de utilizar, implementada en lenguaje *Python*, de modo que para su utilización se necesiten conocimientos mínimos de programación.

Por otro lado, está planteada la necesidad de compatibilizar la librería desarrolla con el software *GNU Octave*, debido a que este posee un extenso paquete dedicado al control de procesos y resulta ampliamente popular entre los estudiantes e investigadores de estas áreas. Además, en la actualidad existen proyectos para llevar la accesibilidad de Octave un paso adelante, como por ejemplo con el desarrollo de una Interfaz Gráfica de Usuario (GUI)[1] para la enseñanza de control de procesos con *GNU Octave* [10].

Como consecuencia, se considera una característica fundamental el acceso continuo a estas herramientas. Por lo que, se prevé la configuración de un servidor encargado de gestionar el conjunto de sistemas dinámicos, posibilitando así, la disponibilidad remota de las herramientas. El diagrama del sistema propuesto, se muestra en la Fig. 1.

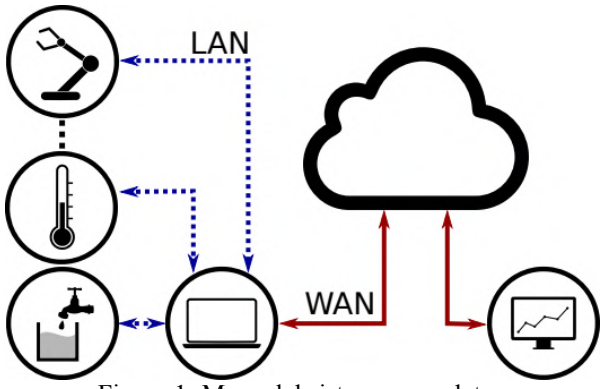


Figura 1: Mapa del sistema completo.

1-B. Sistemas dinámicos

En una primera etapa se proponen tres sistemas dinámicos a controlar, sin embargo el proyecto está pensado para que sea escalable, de modo que puedan añadirse cuantos sistemas sean requeridos según el ámbito y nivel académico.

La red de interconexión de los sistemas para la experimentación remota, está dividida en una sub-red de área amplia (WAN)[2] y otra de área local (LAN)[3]. Por lo tanto, los estudiantes se conectarán de forma remota a un servidor contenido en la red WAN, que a su vez está conectado a la red LAN. De este modo, dicho servidor accionará los actuadores y accederá a las mediciones de los sensores a través de peticiones *HTTP*. Es importante remarcar que la conexión de área local es una red inalámbrica (WLAN)[4] implementada en el módulo ESP8266, programado en lenguaje *C*.

I-B1. Péndulo Aeropropulsado: Uno de los sistemas a controlar, es el péndulo aeropropulsado. Un bosquejo de sus componentes se presenta en la Fig. 2.

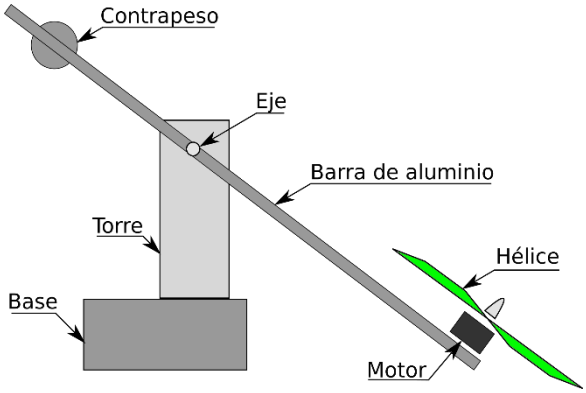


Figura 2: Partes del péndulo autopropulsado.

Su principio de operación se basa en establecer un *set-point* (ángulo respecto al vector de gravedad), para que en base a la manipulación de la fuerza de propulsión de la hélice, se compense la desviación entre el ángulo actual y el de *set-point*. Es de destacar, que este es un sistema no

[1] del inglés, *Graphical User Interface.*

[2] del inglés, *Wide Area Network.*

[3] del inglés, *Local Area Network.*

[4] del inglés, *Wireless Local Area Network.*

lineal SISO[5].
Entre sus posibilidades, este sistema permite realizar pruebas con filtros de Kalman aplicando fusión de sensores, ya que posee un acelerómetro/giróscopo, junto a un *encoder* incremental, acoplados en su eje de rotación. Esto, se ha realizado previamente en otro trabajo [11] y su aplicación en este contexto resulta relevante.

I-B2. Sistema de Cuatro Tanques: Este dispositivo se compone de cuatro tanques interconectados, donde a través de la manipulación del caudal de entrada a los mismos es posible realizar el control de nivel sobre los tanques. Las interacciones entre los depósitos, junto a las no linealidades inherentes (hidráulicas, electromecánicas, etc) transforman a este sistema en un desafío complejo.
En consecuencia, este sistema es ampliamente utilizado en laboratorios de sistemas de control, ya que el mismo además es MIMO[6], haciéndolo propicio para el diseño y análisis de las más variadas técnicas de control en tiempo real [12].

I-B2. Intercambiador de calor: Los intercambiadores de calor constituyen un equipamiento indispensable en la industria de procesos. Por lo que, a través del dispositivo propuesto se pretende dominar los fenómenos físicos involucrados en el funcionamiento de los mismos. En particular la inclusión de tiempos de retardo, las no linealidades y el ingreso de perturbaciones convierten a este sistema MIMO en un adecuado desafío para la enseñanza de sistemas de control [13].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El presente proyecto consta de los siguientes ejes de investigación y desarrollo:

- Estudio de las técnicas de control de interés en la enseñanza de grado y postgrado.
- Análisis de las necesidades de hardware y software requeridas para construir el sistema.
- Desarrollo del firmware necesario para el control de los sistemas dinámicos.
- Diseño e implementación del software educativo para el acceso remoto a los sistemas.
- Generación de documentación adecuada para su correcta utilización.
- Divulgación de los resultados parciales y finales de la investigación.

3. OBJETIVOS

Objetivo General:
Desarrollar un conjunto de sistemas dinámicos con interfaces de usuario remotas, que permitan a los estudiantes implementar controladores de forma práctica, obteniendo resultados tangibles.
Objetivos específicos:

- Analizar las necesidades del alumnado y las herramientas de programación requeridas para la utilización de la librería a desarrollar.
- Definir las funciones necesarias para que un estudiante logre controlar cada uno de los sistemas propuestos, mediante múltiples técnicas de control.
- Construir los sistemas dinámicos a ser controlados.
- Considerar los aspectos a tener en cuenta para el uso del sistema en cursos de sistemas de control.
- Divulgar el proyecto y sus capacidades, particularmente en docentes del área de control.
- Organizar clases y trabajos prácticos que involucren estos sistemas, para ser utilizados en ámbitos académicos.
- Recopilar resultados y cuantificar el impacto de estas herramientas en las cátedras.

4. RESULTADOS ESPERADOS

Se espera que la aplicación de esta herramienta tenga un impacto positivo en las cátedras relacionadas a los Sistemas de Control. En tanto,

[5] del inglés, *Simple Input Simple Output.*

[6] del inglés, Multiple Input Multiple Output.

debido a que el proyecto se encuentra en una fase intermedia de desarrollo, se buscan los siguientes resultados:

- Crear sistemas dinámicos que permitan visualizar de forma intuitiva y directa las técnicas de control.
- Desarrollar una librería sencilla de utilizar, pero a la vez versátil.
- Desplegar una plataforma para el uso remoto de dichos dispositivos.
- Planificar su utilización en las cátedras.

5. FORMACIÓN DE RECURSOS HUMANOS

El grupo de investigación en Control Aplicado y Sistemas Embebidos *(AC&ES-RG)* de la UTN Facultad Regional San Francisco está conformado por dos becarios doctorales, tres ingenieros electrónicos y cinco estudiantes de las carreras ingeniería electrónica, industrial y en sistemas de información. Además, el presente proyecto forma parte de la tesina de grado de dos estudiantes del grupo.

6. BIBLIOGRAFÍA

[1] Phillip C Wankat y Frank S Oreovicz. «Teaching engineering». Purdue University Press, 2015.

[2] Walid Balid, Mahmoud Abdulwahed e Imad Alrouh. «Development of an educationally oriented open-source embedded systems laboratory kit: a hybrid hands-on and virtual experimentation approach». En: *International Journal of Electrical Engineering Education* 51.4 (2014), págs. 340-353.

[3] Ricardo J Costa, Paulo Portela y Gustavo R Alves. «An educational kit to teach and learn Operational Amplifiers». En: *2017 4th Experiment@ International Conference (exp. at'17)*. IEEE. 2017, págs. 137-138.

[4] Luis Gomes y col. «Remote experimentation for introductory digital logic course». En: *2009 3rd IEEE International Conference on E-Learning in Industrial Electronics (ICELIE)*. IEEE. 2009, págs. 98-103.

[5] Mohamed Shaheen, Kenneth A Loparo y Marcus Rbuchner. «Remote laboratory experimentation». En: *Proceedings of the 1998 American Control Conference*. ACC Vol. 2. IEEE. 1998, págs. 1326-1329.

[6] SH Chen y col. «Development of remote laboratory experimentation through Internet». En: *Proceedings of the 1999 IEEE Hong Kong symposium on robotics and control*. Vol. 2. Hong Kong. 1999, págs. 756-760.

[7] Katsuhiko Ogata. «Modern Control Engineering». 5.ª ed. Pearson Education, 2010. ISBN: 978-013-615-673-4.

[8] Farid Golnaraghi y Benjamin C. Kuo. «Automatic Control Systems». 9.ª ed. Wiley, 2009, pág. 944. ISBN: 978-047-004-896-2.

[9] Eduardo J. Adam. «Instrumentación y Control de Procesos. Notas de Clase». 3.ª ed. Santa Fe: Ediciones UNL, 2018. ISBN: 978-987-749-122-7.

[10] E Sergio Burgo y Eduardo J Adam. «Desarrollo de Interfaces Gráficas para Enseñanza de Control de Procesos».

[11] Marco Miretti y col. «Estimación híbrida de posición angular en dispositivos de captura de imágenes aéreas».En: VIII Jornadas de Ciencia y Tecnología para Alumnos (CyTAL) (ago. De 2018).

[12] Wael A Altabey. «Model optimal control of the four tank system». En: International Journal of Systems Science and Applied Mathematics 1.4 (2016), págs. 30-41.

[13] Choon Khon Ng y Mohd Fauzi Zanil. «Development of IoT based Heat Exchanger Control Trainer for Undergraduate Process Control Programme». En: *Research Communication in Engineering Science & Technology* 2 (2019), págs. 6-16.

Uso de las TICs para la Construcción de Espacios Institucionales - Generación Automática de Sitios Web Mediante Frameworks

Julieta Gatica, Camila Olguín, José Montejano, Ayrton Marini, Mariano Luzza, Mario Berón, Analía Zaldua, Germán Montejano
Universidad Nacional de San Luis
Facultad de Ciencias Físico Matemáticas y Naturales
Área de Programación y Metodologías de Desarrollo de Software
Correo Electrónico: {mluzza,mberon,gmonte}@unsl.edu.ar
{jag81295,camimt25,jose.p.montejano,anazaldua}@gmail.com {ayrtonmarini10}@hotmail.com

Resumen

La educación es uno de los principales pilares sobre los que se cimienta el crecimiento social. En muchas situaciones los distintos contextos sociales y geográficos limitan a los miembros de una comunidad de recibir una educación de calidad. Es por esto que la elaboración de estrategias que permitan romper barreras de cualquier índole, que dificulten la educación, es una actividad de valor para todo tipo de sociedad.

Una de las formas de paliar los inconvenientes mencionados en el párrafo precedente consiste en el uso de las Tecnologías de la Información y de la Comunicación para brindar contenidos a distancia. En la actualidad, la Internet y la posibilidad de adquirir aparatos tecnológicos que se conectan a ella permite disminuir abruptamente las barreras existentes entre las sociedades situadas en distintas regiones. Esta peculiaridad es muy importante para diseñar soluciones que propendan a una inclusión educativa de calidad.

En este artículo se presenta una línea de investigación que se enfoca en el diseño y construcción de espacios institucionales para fortalecer la vinculación educativa entre instituciones secundarias y la universidad, proporcionando un medio que permita, a los estudiantes de distintos niveles, interactuar, establecer una interacción de conocimiento mutuo, fortalecer sus conocimientos, etc., y así evitar que la internet y tecnologías asociadas estén subutilizadas por la falta de diseño de un software que permita la integración de las actividades educativas.

Palabras clave

Vinculación educativa - Biblioteca Digital - Gestor de sitios web institucionales.

Contexto

La presente línea de investigación se enmarca en el Proyecto Bilateral (PO/16/93) de "Fortalecimiento de la Seguridad de los Sistemas de Software mediante el uso de Métodos, Técnicas y Herramientas de Ingeniería Reversa". Realizado en conjunto con la Universidade do Minho Braga, Portugal. Aprobado por el Ministerio de Ciencia Tecnología e Innovación Productiva (Mincyt). Y por el Proyecto de Investigación (P-031516): "Ingeniería de Software: Conceptos, Prácticas y Herramientas para el Desarrollo de Software de Calidad" de la Facultad de Ciencias Físico Matemáticas y Naturales, Universidad Nacional de San Luis. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el Programa de Incentivos.

1. Introducción

Hoy en día el uso de las tecnologías de la información y de la comunicación (TICs) en las instituciones educativas es una realidad insoslayable. Muchas campañas gubernamentales están orientadas a proporcionar, a los distintos miembros de la comunidad educativa, computadoras y tablets con el propósito de fomentar el uso de las TICs en educación y de incrementar las posibilidades educativas de la sociedad. Concomitantemente con este esfuerzo, el acceso a la tecnología por parte de los ciudadanos es más factible (en la mayoría de los hogares se puede encontrar al menos una computadora, smartphone, tablet, etc.). Esta característica hace que las estrategias educativas puedan salir del ámbito escolar o institucional y trasladarse a cualquier parte del planeta.

Lo mencionado en el párrafo precedente es una ventaja que los avances tecnológicos proveen. Los mismos, en la mayoría de los casos, no son explotados como se debe. Es común asistir a las instituciones escolares y observar la existencia de computadoras, netbooks, tablets, conexiones a internet, etc. las cuales en el mejor de los casos se utilizan para buscar información para resolver prácticos de las materias que cursan. Sin lugar a dudas, lo previamente mencionado es una ventaja respecto de épocas pasadas donde tal tipo de actividades eran impensadas. Sin embargo, muchas de las posibilidades que proporcionan las TICs no son utilizadas. A modo de ejemplo, se puede mencionar la poca vinculación que existe entre las escuelas y las instituciones de nivel superior e incluso con aquellas del mismo nivel. Esto se debe a que es necesario construir aplicaciones que posibiliten dicha interacción. Para alcanzar el objetivo antes mencionado, se debe contar con docentes o profesionales capacitados en el desarrollo de software. Si bien las instituciones educativas cuentan con profesionales con ese perfil, los mismos están dedicados a tareas docentes y generalmente no disponen del tiempo y recursos suficientes para poder llevar adelante un proyecto de software que provea una solución al problema de integración educativa. Es importante mencionar que lo descrito en los párrafos anteriores no es un mero enunciado teórico sino más bien una realidad concreta observada y vivida por los docentes en su quehacer educativo.

2. Líneas de investigación y desarrollo

Las TICs se definen como "un conjunto diverso de herramientas y recursos tecnológicos usados para comunicar, crear, diseminar, almacenar y administrar información"[1]. Cuando se usan apropiadamente, ayudan a expandir el acceso a la educación, a elevar la calidad educativa, a fortalecer su importancia en el lugar de trabajo, etc., posibilitando que tanto la enseñanza como el aprendizaje se conviertan en un proceso activo conectado a la vida real. Sin embargo, la integración de las TICs es un proceso complejo, ya que no solo requiere de las tecnologías, sino también de pedagogía, preparación institucional, competencia docente, entre otras cosas.

En la actualidad, existen numerosas investigaciones acerca de la mejor forma para aplicar las TICs en contextos educativos. A través de una encuesta realizada a docentes de ingeniería eléctrica en la Universidad de Bechar, Algeria, se obtuvo que los docentes que hacen uso de las TICs en su proceso de enseñanza notaron que éstas apoyan la participación del alumno, crean un marco de justificación, facilitan el trabajo colaborativo y aumentan los factores de autonomía y creatividad en el alumno [2].

Otros estudios afirman que el avance de las TICs ha permitido a muchos países del mundo avanzar y crear una ventaja competitiva, y

consideran que su implementación en la educación se considera el papel más crítico en la creación de una sociedad competitiva [3].
Teniendo en cuenta lo antes mencionado, se puede pensar que el problema principal en cuanto al uso de las TICs en la educación es la falta de un software que facilite la interacción hombre-máquina, y que ayude tanto a los profesores como a los estudiantes en el proceso de aprendizaje/enseñanza.
La evolución de las TICs, trajo consigo una gran variedad de herramientas para facilitar la vida a los desarrolladores. Entre estas, se encuentran los denominados "Frameworks". Meiert [4], los compara con una biblioteca de patrones de diseño que contiene, generalizadamente, todas las funcionalidades necesarias para un desarrollador de software. Por lo tanto, se puede definir un framework como una abstracción que otorga una funcionalidad genérica y se puede cambiar selectivamente mediante un código adicional escrito por el desarrollador, proporcionando así un software específico de aplicación.
Lo descrito anteriormente lleva a pensar en la rapidez y facilidad con la que se podría crear un sitio web para facilitar la vinculación entre las distintas instituciones con estas tecnologías. Suleman et al. [5] hacen referencia a varias instituciones que usan tecnologías de componentes para resolver problemas de forma rápida y correcta. Entre estos problemas existe la creación de varias Bibliotecas Digitales, las cuales han alcanzado un nivel significativo de diversificación y desarrollo, y ahora se han convertido en un importante componente en la educación superior debido a los procesos de creación, publicación y distribución de contenido, aumentando su potencial e impacto. Un ejemplo de ello es el proyecto "Health Education Assets Library (HEAL)" [6], pensado para proporcionar contenido digital gratuito de la más alta calidad que satisfaga las necesidades de los educadores y estudiantes de ciencias de la salud. Otro ejemplo es el Sistema de bibliotecas de la UDLA [7], la cual es una red de bibliotecas que ofrece recursos electrónicos y brinda servicios de información acerca de la biblioteca.
Las razones por las que se desarrollan tecnologías relacionadas a las bibliotecas digitales es que hacen que la investigación científica y otra información académica producida, puedan ser íntegras y estén centralizados a largo plazo. Además evita la existencia de información académica dispersa y garantiza el uso efectivo del conocimiento. Al mismo tiempo, esto puede mejorar el status y el valor académico de la universidad, ya que permite un acceso abierto global mejorando los canales de distribución así como también, el medio ambiente [8].
Otra razón para la creación de una biblioteca es la posible reutilización de los recursos educativos. Esto se ha descrito como Economía de Objetos de Aprendizaje [9], donde el material se comparte, luego se reutiliza y finalmente se mejora.
Por otro lado, existe una gran variedad de publicaciones que dan pautas de diseño y estructura de sitios web escolares, que indican los aspectos que debería contener un sitio web escolar ideal. Por ejemplo, Tubin y Klein [10], presentan la idea de que las escuelas no le sacan el máximo provecho a los sitios web o que incluso mantienen el sitio desactualizado, de cara a esto, dan una aproximación al contenido, la estructura y sensibilidad que debiera tener el sitio web ideal para una institución educativa. Investigadores del Cogent Computer Solutions [11], afirman que los sitios web escolares deberían tener como su principal función ser, para los padres de los alumnos, una puerta abierta a la educación de sus hijos. Ferrer [12], plantea cómo debe ser el diseño de la página web y también las herramientas necesarias para crear la misma. Sin embargo, a pesar de todo este material existente sobre diseño y estructura, no se encuentran antecedentes del desarrollo de un generador de plantillas de sitios web exclusivamente escolares para su libre uso.
Finalmente, existen numerosas investigaciones referidas a la vinculación entre instituciones educativas y universidades, como el programa realizado por la Facultad de Ciencias Económicas de la UNICEN [13] cuyo objetivo es realizar un acercamiento entre la Universidad y la comunidad educativa de la

Región Centro de la Provincia de Buenos Aires.

Por otro lado, mediante un proyecto, realizado por docentes y estudiantes de la Universidad De Las Américas (UDLA), se desarrolló un sitio web que tiene como objetivo vincular a la comunidad universitaria con el contexto ecuatoriano mediante programas y proyectos que generen impacto y empoderamiento en grupos vulnerables[14]. Estas investigaciones, sin embargo, no tratan temas de cómo hacer uso de las TICs para minimizar la brecha existente entre escuelas y universidades y la sociedad en general.

3. Resultados obtenidos/esperados

A través de los trabajos realizados por los integrantes de esta línea de investigación se han podido obtener diferentes resultados, los cuales se mencionan a continuación: (i) Vinculación 3.0: es una herramienta que tiene como objetivo facilitar la interacción entre las diferentes instituciones educativas y sus comunidades asociadas. Para tal fin, la aplicación proporciona diferentes funcionalidades que son muy útiles para las instituciones y sus comunidades educativas, como lo son: asignación de espacio en disco, creación de manera automática de sus sitios web, una biblioteca digital de recursos educativos abiertos, etc. Es un sistema que responde a diversos retos como la escasez de recursos financieros, cambios tecnológicos, transformaciones de recursos educativos, etc. (ii) Una biblioteca digital: es una herramienta que puede convertirse en un instrumento eficaz de promoción de la universidad, al establecer una interacción de conocimiento mutuo entre el nivel medio y el nivel superior, con el fin de que la percepción de su imagen sea acorde con la realidad vivida por los estudiantes y por los distintos usuarios de la comunidad educativa de dicho repositorio. (iii) Retroalimentación de las instituciones: La experiencia de utilizar Vinculación 3.0 en un colegio fue muy satisfactoria dado que las funcionalidades del sistema resultaron cubrir varios requerimientos de las instituciones educativas según la opinión del equipo de conducción de la institución (usuarios reales de Vinculación 3.0).

Obviamente, Vinculación 3.0 puede seguir creciendo a través del perfeccionamiento de las funcionalidades que posee y la incorporación de otras nuevas que irán surgiendo a medida que sea utilizado por más instituciones. Como trabajo futuro se pretende: (i) Incorporar mayor contenido a la biblioteca, y obtener una mayor comunidad que comparta, reutilice, y mejore los recursos educativos abiertos. (ii) Actualizar las GUIs del sitio de acuerdo a las recomendaciones y el feedback obtenido por parte de los usuarios, para mejorar la amigabilidad y la gestión por parte de los administradores.

4. Formación de recursos humanos

Los progresos obtenidos en esta línea de investigación sirven como base para el desarrollo de tesis de posgrado, ya sea de doctorado o maestría en Ingeniería de Software y desarrollo de trabajos finales de las carreras Licenciatura en Ciencias de la Computación, Ingeniería en Informática e Ingeniería en Computación de la Universidad Nacional de San Luis, en el marco de los Proyectos de Investigación mencionados en el contexto del presente documento.

5. Bibliografía

[1] Victoria L. Tinio (2003). ICT in Education (First Edition).

[2] M. Belagra, C. Benachaiba, B. Guemid. “Using ICT in Higher Education: Teachers of Electrical Engineering Department at the University of Bechar: Case Study”. IEEE

Global Engineering Education Conference (EDUCON), 17, 20 April 2012.
[3] H. D. Hermawan, D. N. Yunita, N. Deswila. “Implementation of ICT in Education in Indonesia during 2004-2017”. International Symposium on Educational Technology (ISET). July - Aug, 2018.
[4] J. O. Meiert. “The Little Book of HTML/CSS Frameworks”. (First Edition).
[5] H. Suleman, E. A. Fox. “A Framework for Building Open Digital Libraries”. Revista D-Lib, Vol. 7 Nro. 12, 2001.
[6] S. A. McIntyre, S. H. J. Uijtdehaage, S. E. Dennis, C. S. Candler. “A Digital Library for Health Sciences Educators: The Health Education Assets Library (HEAL)”. ACM/IEEE Conference on Digital Libraries, June, 2004.
[7]“Home–Vinculación”. http://vinculacion.udla.edu.ec. Visto el 31/08/2018.
[8] L. Naiwen, Z. Xin. “The Study On Constructing Institutional Repository Of University”. Fourth International Conference on Multimedia Information Networking and Security. Nov, 2012.
[9] C. Duncan, R. House. “Learning object economies: Barriers and drivers” ELearnInternational, vol. 18, p. 19, 2004.
[10] D. Tubin, S. Klein. “Designing a School Website: Contents, Structure, and Responsiveness”. Planning and Changing Vol. 38, No. 3&4, 2007, pp. 191–207.
[11] Cogent Computer Solutions. “Structure and function of school websites”. Summary report. Junio 2015.
[12] R. Ferrer. “Diseño de páginas web en educación”. Universidad-o-Autónoma de Madrid, Tendencias Pedagógicas 10, 2005.
[13] “Programa de Vinculación Universidad - Escuela Secudaria”. https://www.econ.unicen.edu.ar/escuelas-secundarias.html. Visto el 07/09/2018.
[14] “Nuevo modelo de vinculación con la comunidad se aplica en la UDLA”. https://www.udla.edu.ec/2016/07/29/nuevo-modelo-de-vinculacion-con-la-comunidad-se-aplica-en-la-udla. Visto el 07/09/2018.

Simulación y equipo real Mikrotik en la enseñanza de redes de computadoras en el nivel universitario (resultados parciales)

Daniel Arias Figueroa, Ernesto Sánchez, Loraine Gimson, Álvaro Gamarra, Gustavo Gil, Rodolfo Baspineiro, Romina Chacón, Agustín Colque

Facultad de Ingeniería de la Universidad Católica de Salta
C.I.D.I.A. (Centro de Investigación y Desarrollo de Informática Aplicada)
Facultad de Ciencias Exactas – Universidad Nacional de Salta
Av. Bolivia Nº 5150 – Campo Castañares – Salta – Tel. 0387-4255476
daaf@cidia.unsa.edu.ar, esanchez@cidia.unsa.edu.ar

Resumen

Nuestra línea de investigación, pretende evaluar la influencia de la utilización de software de simulación, en la enseñanza de contenidos de redes de computadoras en carreras de grado y cursos de postgrado a fin de contrastar con la enseñanza con laboratorios con equipo real Mikrotik. Para ello nos planteamos un estudio aplicado, longitudinal y experimental que considera aspectos cuantitativos y cualitativos, que permitan determinar si los estudiantes que realizan prácticas en un entorno de simulación aprenden significativamente más que aquellos estudiantes que se capacitan con laboratorios con equipos reales de red.

Palabras clave: protocolo TCP-IP, emulación, simulación, enseñanza, redes, Mikrotik, GNS3.

Contexto

La línea de investigación se encuentra apoyada por el C.I.D.I.A. (Centro de Investigación y Desarrollo de Informática Aplicada) que depende de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta y por la Facultad de Ingeniería de la Universidad Católica de Salta, por lo tanto se cuenta con toda la infraestructura disponible para esta investigación. El proyecto cuenta con el financiamiento del CIUNSa – Consejo de Investigación de la Universidad Nacional de Salta y el Consejo de Investigación de la Universidad Católica de Salta.

Introducción

Los altos costos de equipos específicos necesarios para montar un laboratorio de red, sumados a los escasos recursos con los que cuentan las mayorías de las universidades en materia de infraestructura de red para la enseñanza, hacen considerar a las herramientas de simulación como una posible solución para que las prácticas puedan ser mejor aprovechadas por los estudiantes, posibilitando además la utilización de estas herramientas fuera de los horarios de clases. De acuerdo a lo dicho anteriormente, se podría resumir la problemática de trabajar con laboratorio de red con equipo real de la siguiente manera:

- El docente debe plantear los trabajos prácticos de laboratorio adecuándose a las características del equipamiento disponible, generalmente escaso.

- Si bien la cantidad de estudiantes habitualmente no es muy elevada (aproximadamente 15-20 estudiantes cada año), contrasta con la cantidad de equipos que se disponen.

- Los equipos de hardware (enrutadores, conmutadores, concentradores, cableado de red, conectores, etc.) son costosos, y su actualización y mantenimiento también significa costos elevados, por lo que usualmente se puede contar con, a lo sumo, uno o dos dispositivos por comisión o grupo de estudiantes. Esto hace impracticable los laboratorios con equipo real.

- La curva de aprendizaje para la administración de los dispositivos es alta, lo mismo ocurre con la conectorización física para definir una determinada topología, ya que se disponen diferentes tipos de interfaces de red tal como Ethernet, FastEthernet y puertos Seriales. Esto impide realizar demasiados grupos que accedan al hardware de red.

La mayoría de las universidades ya cuentan con toda una infraestructura de red, laboratorios con pcs y conectividad a Internet al servicio de las diferentes asignaturas. Todos estos elementos sirven a nuestro propósito de mejorar las actividades prácticas utilizando métodos de simulación con el equipamiento ya disponible, minimizando de esta manera las inversiones en compra y mantenimiento de equipo específico de red.

Uno de los objetivos claves en la enseñanza de las redes de computadoras en las carreras de sistemas, es transmitir conceptos básicos y fundamentos a los estudiantes. Sin embargo, desde hace un tiempo se viene investigando cómo facilitar la relación entre la realidad y las teorías y modelos, es decir, entre lo concreto y lo abstracto. Así, las computadoras personales, con la variedad de software que se ha desarrollado, tienen en sí mismas un gran potencial para mejorar el proceso de enseñanza y aprendizaje haciendo que:

- el aprendizaje sea más interesante,

- sea un aprendizaje activo y no pasivo en las aulas,

- sea al ritmo del estudiante en forma personalizada,

- los estudiantes estén más motivados y la educación sea permanente.

Líneas de Investigación, Desarrollo e Innovación

Esta investigación pretende dar respuesta a la pregunta, ¿cuál es la influencia del uso de herramientas de simulación en el aprendizaje de conceptos y fundamentos sobre redes de computadoras en estudiantes del nivel universitario? De esta pregunta principal se derivan las siguientes preguntas:

- ¿Existe diferencia estadísticamente significativa en el nivel de comprensión de los conceptos que se van a evaluar sobre redes IP, entre estudiantes que reciben instrucción mediada por herramientas de simulación y estudiantes que reciben instrucción utilizando laboratorios con equipos reales de red?

- ¿Existe relación entre la utilización de herramientas de simulación y la actitud de los estudiantes hacia el aprendizaje de las

asignaturas en cuestión y de otras asignaturas del plan de estudio de la carrera Licenciatura en Análisis de Sistemas?

Es importante destacar que, en el marco del proyecto de investigación, se realizarán experiencias con temáticas tales como: enrutamiento IP estático, enrutamiento IP dinámico con el protocolo RIP, protocolo DNS, protocolo ARP, Direccionamiento IP y otros. Todas estas temáticas corresponden a las asignaturas de redes de computadoras del plan de estudios de la carrera Licenciatura en Análisis de Sistemas de la UNSa y de la carrera Ingeniería en Telecomunicaciones de UCASAL. Para ello se llevará a cabo un estudio aplicado, longitudinal y experimental que, asumiendo la complejidad del fenómeno educativo, considerará aspectos cuantitativos y cualitativos, contrastando la enseñanza con simulación con la enseñanza con equipo real de redes Mikrotik.

Asimismo, la investigación persigue en el marco del paradigma cualitativo los siguientes objetivos:

- Realizar entrevistas y estudios de casos con los alumnos durante los períodos previstos.
- Diseñar actividades prácticas y de investigación adecuadas para ser realizadas con los simuladores y con laboratorio con equipo real.
- Fomentar el aprendizaje cooperativo entre los estudiantes.
- Introducir a los estudiantes de licenciatura e ingeniería en la cultura científica actual, que concibe la simulación por computadora como una herramienta fundamental para el estudio, la investigación y la experimentación.
- Contratar el resultado producto de la investigación con estudios similares en otras universidades.

Objetivos

El objetivo general del estudio aquí planteado consiste en determinar el impacto del uso de herramientas de simulación en el aprendizaje de conceptos y fundamentos de redes de computadoras en estudiantes del nivel universitario. A partir de este objetivo se plantearon los siguientes objetivos específicos:

- Establecer la relación entre la utilización de la simulación y la actitud de los estudiantes hacia el aprendizaje de las redes de computadoras y de temáticas de otras asignaturas del plan de estudios;
- Determinar si existe diferencia estadísticamente significativa en el nivel de comprensión de los principios de redes de computadoras, entre estudiantes que reciben instrucción mediada por herramientas de simulación y estudiantes que reciben instrucción con laboratorios con equipo real.

En esta primera etapa del proyecto se realizó el relevamiento, análisis y selección de las herramientas de simulación que se utilizarán en la experimentación. Se realizó la capacitación de los tutores en el SO. Mikrotik y en la plataforma de virtualización GNS3. También se están desarrollando las guías prácticas para los temas de la capa de red, para utilizarse con la herramienta GNS3. Se pudo realizar una experiencia piloto que nos permitió una primera validación del material elaborado.

Esta planificado realizar durante el año 2020 las primeras experiencias en el aula en el marco de las asignaturas Redes de Computadoras I y Redes de Computadoras II de la carrera Licenciatura en Análisis de Sistemas y en las asignaturas Redes I y Redes II de la carrera Ingeniería en Telecomunicaciones de la Universidad Católica de Salta.

Formación de Recursos Humanos

El grupo de investigación conformado se caracteriza por una constitución heterogénea de profesionales vinculados a la informática. El Director es Doctor en Ciencias Informáticas y el Codirector Master en Ingeniería de Software. Dentro de los investigadores se cuenta con un Magister en Redes de Datos, una Magister en Ingeniería de Software, un Ingeniero en Sistemas y un Técnico Universitario en Programación. También integran el grupo dos alumnos avanzados de la carrera Licenciatura en Análisis de Sistemas de la UNSa. En el transcurso del proyecto se tiene como objetivo consolidar la formación en investigación de los integrantes de menos antecedentes y también está contemplado que uno de los integrantes complete el cursado de la Maestría en Redes de Datos y la certificación Mikrotik.

Referencias

SAKAR, N. I. (2006) Teaching TCP/IP Networking Using Practical Laboratory Exercises, *International Journal of Information and Communication Technology Education*, Vol. 2, No. 4, pp. 39-50.

GOLDSTEIN, G., M LEISTEN, S, STARK, K.,& TICKLE, A. (2005) Using a Network Simulation Tool to Engage Students in Active Learning Enhances Their Understanding of Complex Data Communications Concepts, Proceedings of the *7th Australasian conference on Computing Education*, pp. 223-228.

JAVIDI, G. & SHEYBANI, E. (2008) Content-Based Computer Simulation of a Networking Course: An Assessment, *Journal of Computers*, Vol. 3, No. 3, pp. 64-72.

CAMERON, B. (2003): Effectiveness of simulation in a hybrid online networking course. *Quarterly Review of Distance Education*, 4(1), 51.

GATTO, D. (1993): The use of interactive computer simulations in training. *Australian Journal of Educational Technology*, 9(2), 144-156.

YAVERBAUM, G., & NADARAJAN, U. (1996): Learning basic concepts of telecommunications: an experiment in multimedia and learning. *Computers & Education*, 26(4), 215-224.

ZHU, S. Y. (2011).Teaching Computer Networks through Network Simulation Programs. Faculty of Business, Computing and Law – School of Computing. University of Derby. *Learning Teaching & Assessment Conference*.

KUROSE, J.F. & ROSS, K.W. (2015). Computer Networking: A Top-Down Approach. 6th Edition. *Pearson Education*. ISBN: 9780132856201.

AVILA BLAS, Orlando José (2003). Probabilidad y estadística inferencial: teoría y aplicaciones. ISBN: 978-987-9381-23-6.*Editorial: Univ. de Salta.*

GLIEM, J & GLIEM, R. (2003). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. Conference in Adult, Continuing, and Community Education. Midwest Research to Practice.

CÁMARA, ALZUGARAY (2011). Trabajos Prácticos, Métodos de Simulación y aprendizaje significativo. La Tecnología Educativa al servicio de la Educación Tecnológica.

ARIAS FIGUEROA, D. SÁNCHEZ, E. Colaboración de BASPINEIRO (2018). Redes de Computadoras II con Packet Tracer. Editorial de la Universidad Nacional de Salta – EUNSa – ISBN 978-987-633-527-0; 1a ed. - Salta - E-Book - CDD 004.678.

ARIAS FIGUEROA, D. (2015). Redes de Computadoras I con Packet Tracer. Editorial de la Universidad Nacional de Salta – EUNSa – ISBN 978-987-633-132-6; 1a ed. - Salta - E-Book - CDD 004.68.

Analítica de aprendizaje aplicada al contexto de la enseñanza superior mediante la definición de métricas para la valoración de rendimiento académico.

Corso Cynthia, Constable Leticia, Colacioppo Nicolás, Chávez Marcelo.

Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información
Departamento Ingeniería en Sistemas de Información
Facultad Regional Córdoba. Universidad Tecnológica Nacional
Maestro M. López esq. Cruz Roja-Ciudad Universitaria-Córdoba
cynthia@bbs.frc.utn.edu.ar/leticiaconstable@gmail.con/nicolas_colacioppo@hotmail.com/jcmarcelochavez@gmail.com

RESUMEN

Este trabajo expone una experiencia analítica mediante el establecimiento de una serie de métricas que permitirá monitorear aspectos relacionados con el proceso aprendizaje de los estudiantes de la cátedra Paradigmas de Programación de la Universidad Tecnológica Nacional Facultad Regional Córdoba. Esta definición de variables y métricas facilitará su medición y análisis mediante la construcción de un sistema de soporte de decisión que favorecerá el proceso de toma de decisiones en el ámbito de la enseñanza superior. La propuesta de variables y métricas presentada se organiza en tres categorías: Pedagogía, Tecnología y Contexto.

.

Palabras claves: *analítica de aprendizaje, definición de métricas, educación superior, sistema de soporte de decisión.*

CONTEXTO

Este trabajo hace referencia al proyecto "Integración de recursos del Paradigma Analítico y de la Inteligencia de Negocios como estrategia para el fortalecimiento en el proceso de toma de decisiones" PID-SIUTNCO0005101, que ha sido homologado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional. El contexto de desarrollo de la presente investigación es el Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información *(GIDTSI)* radicado en la U.T.N Facultad Regional Córdoba.

1. INTRODUCCIÓN

La evaluación de la calidad en las instituciones educativas es un tema de vital de importancia, más aún cuando lo que se mide es la calidad del proceso de enseñanza y aprendizaje. Actualmente el uso de nuevas tecnologías de información y comunicación se ha extendido en las prácticas docentes. Cada vez es más frecuente la implementación de aulas virtuales en el contexto educativo como un recurso de apoyo al dictado presencial de ciertas asignaturas. La interacción de los estudiantes y profesores con las aulas virtuales propicia la generación de una cantidad significativa de datos que pueden ser aprovechados.

Una alternativa para aprovechar los datos y sobre todo tener la posibilidad de cuantificarlos es la *analítica de aprendizaje* que permite la medición,

recopilación y análisis de lo que ocurre en el interior de los procesos educativos mediados por la tecnología, pero precisa de una interpretación adecuada que permita la mejora de los procesos educativos y la posibilidad de ajustar de mejor manera la intervención educativa con la finalidad de mejorar el desempeño académico de los estudiantes [1][2] [3][4][5]. Por ello la importancia de contar con un buen diseño previo que permita identificar qué aspectos didácticos pueden verse mejorados para favorecer un mejor desempeño académico de los estudiantes.

Este trabajo expone el diseño de una experiencia analítica, que complementa la definición y categorización variables y métricas en un contexto educativo. El foco del estudio es la cátedra de Paradigmas de Programación perteneciente al segundo nivel de la carrera de Ingeniería en Sistemas de Información de la UTN Facultad Regional Córdoba. Desde la cátedra se promueve la incorporación de herramientas tecnológicas que permitan contribuir a una educación de calidad a través del asesoramiento y acompañamiento de las actividades académicas desarrolladas. La herramienta utilizada en la cátedra es Moodle, considerada no solo para la publicación de materiales teóricos y prácticos sino también para el desarrollo de actividades complementarias y de refuerzo del aprendizaje [6].

El esquema de métricas planteadas será considerado como base para la construcción de un *sistema de soporte de decisión* que agilice la información relacionada con el rendimiento académico de los estudiantes y de aspectos vinculados con el aprendizaje mediante el uso de la plataforma virtual Moodle. La importancia de transferir los resultados al ámbito de la educación superior radica en la posibilidad de disponer nuevos recursos que faciliten el mejoramiento en materia de calidad educativa basado en la toma de decisiones inteligentes.

2. LINEAS DE INVESTIGACIÓN y DESARROLLO

De este proyecto se desprenden al menos cuatro líneas de investigación y desarrollo. A continuación se presenta una descripción sintética de las mismas:

✓ Inteligencia de Negocios: se pretende el estudio y análisis de herramientas de BI haciendo foco en los sistemas de soporte de decisión, su arquitectura y la revisión metodologías para su diseño e implementación.

✓ Paradigma Analítico: en este proyecto se considerará el estudio de las analíticas de aprendizaje cuya finalidad es el seguimiento del uso de los entornos virtuales.

✓ Entornos virtuales de enseñanza y aprendizaje: se investigarán y analizarán herramientas de seguimiento para la recolección de los datos que surgen de las interacciones de los alumnos a las actividades y con los profesores.

3. RESULTADOS OBTENIDOS/ESPERADOS

Objetivo General

Desarrollar una metodología para el diseño e implementación de un sistema de información que agilice el proceso de toma de decisiones, considerando la analítica digital y métricas de interés para el dominio bajo estudio.

Objetivos Particulares

- Relevar la situación actual del dominio bajo estudio e identificar los aspectos de los que se pretende realizar un seguimiento para su control y optimización.
- Definir objetivos y métricas analíticas alineadas con los requerimientos detectados.
- Elaborar un esquema que permita implementar un almacén con el fin de consolidar los datos de interés que provienen de diferentes fuentes.
- Diseñar y desarrollar una interfaz que permita la integración de los datos almacenados para la detección de desviaciones de los objetivos propuestos en el dominio de análisis.

Resultados parciales obtenidos

Para la construcción del esquema de métricas de interés se realizó una revisión de la literatura sobre educación e-learning y evaluación [7] [8] [9] [10] [11] [12]. Con la finalidad de tener claridad sobre los elementos bajo estudio y a la vez determinar sus alcances y limitaciones. De los trabajos consultados se aprecia que las métricas específicas consideradas para el proceso de aprendizaje son escasas, y la gran mayoría de estas refieren a infraestructura y acceso a las TICs. Como así también se resaltan ciertas cuestiones que influyen en la evaluación del e-learning como lo es el diseño instruccional del curso de manera que permita al estudiante y al docente conocer los objetivos del aprendizaje, actividades a realizar, forma de evaluación; otros criterios importantes son la usabilidad, la navegabilidad, el acceso del sistema implementado, de igual forma la comunicación es un proceso indispensable para la adecuada interiorización del conocimiento adquirido.

Propuesta de variables y métricas: se realiza una propuesta sobre las métricas que se deben
tener en cuenta para evaluar del e-learning en la cátedra bajo estudio.

La propuesta de métricas se basa en el análisis de los sistemas de métricas de e-learning de la sociedad de la información y de trabajos de evaluación de e-learning, además se proponen otras métricas que son necesarias para evaluar el proceso de aprendizaje del estudiante y que son de interés para la cátedra bajo análisis. A continuación se detalla una propuesta de métricas para medir el proceso de aprendizaje en la cátedra bajo estudio, que para un mejor análisis se ha estructurado en diferentes categorías:

- Pedagogía: la intención es medir el proceso de aprendizaje del estudiante, con base en los materiales de estudio ofrecidos, el diseño de aula virtual de los cursos, las retroalimentaciones a las actividades presentadas.
- Tecnología: en el cual se evalúa la navegabilidad y usabilidad de la herramienta virtual utilizada como soporte en el proceso educativo y las habilidades en el manejo de TICs por parte de los estudiantes.
- Contexto: todo proceso de evaluación depende de las características específicas del contexto de referencia, es por ello que se evalúa el contexto social, económico y cultural.

En la Tabla 1. se presentan las principales variables y métricas por categoría que serán consideradas para la evaluación del proceso de enseñanza/aprendizaje de los estudiantes de la cátedra bajo estudio.

Categoría	Variable	Métrica/s
Pedagogía	Característica del curso	Cantidad de estudiantes inscriptos.
Pedagogía	Rendimiento	% de estudiantes con aprobación directa. % de estudiantes con promoción práctica. % de estudiantes regulares. % de estudiantes libres.
Pedagogía	Retención	✓ % de estudiantes que aprueban la asignatura=Nro. de estudiantes que aprueban la asignatura/Nro. de estudiantes inscriptos en el curso * 100. ✓ % de estudiantes que aprueban la asignatura dentro de los que finalizan el cursado= Nro. de estudiantes que aprueban la asignatura/Nro. de estudiantes que finalizan el cursado * 100. ✓ % de estudiantes que finalizan el cursado=Nro. de estudiantes que no finalizan el cursado/Nro. de estudiantes inscriptos en el curso * 100.
Pedagogía	Características de los estudiantes	✓ % de asistencia del curso. ✓ % de inasistencia del curso.
Pedagogía	Modelo de pedagogía	✓ Nro. de horas de estudios a la semana. ✓ Tasa de actividades prácticas realizadas= Nro. de prácticos realizados/Nro. de prácticos programados * 100.
Tecnología	Herramienta Virtual	✓ Frecuencia de acceso en el aula virtual. ✓ Tiempo total de acceso en el aula virtual. ✓ Frecuencia de participación en los foros.
Contexto	Económico	✓ % de estudian que trabajan. ✓ Promedio de horas trabajadas.
Contexto	Social	✓ % de estudiantes que realizan otras actividades académica y o culturales. ✓ Frecuencia de desarrollo de actividades culturales.

Tabla1. Variables y métricas de interés para la evaluación del proceso de aprendizaje de los estudiantes de la cátedra bajo estudio.

4. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto está conformado por docentes-investigadores pertenecientes a la carrera de grado de Ingeniería en Sistemas de Información. Todos los integrantes docentes del PID han participado del proceso de categorizaciones en investigación dentro del Programa de Incentivos del MECyT; así como en la categorización interna que posee la U.T.N. Además se prevé la participación de alumnos avanzados en la

carrera que realizan su práctica supervisada como requisito para el otorgamiento del título de grado de Ingeniero. En este proyecto participa un becario alumno con el objetivo de complementar su formación académica con un acercamiento al ámbito de la investigación científica.

5. REFERENCIAS

[1] Reich, J. (2015) "Rebooting MOOC Research. Improve assessment, data sharing, and experimental design", *Science*, 347 (6217), pp. 34-35. http://science.sciencemag.org/content/sci/347/6217/34.full.pdf?sid=d136a464-3083-4b2c-bfa8-17b76aa3a9dd

[2] Khalil, M.; Kastl, C.; Ebner, M. (2016) "Portraying MOOCs Learners: a Clustering Experience Using Learning Analytics". En Khalil, M.; Ebner, M.; Kopp, M.; Lorenz, A.; Kalz. M. (Eds.), *Proceedings of the European Stakeholder Summit on experiences and best practices in and around MOOCs* (EMOOCs 2016) BookOnDemand: Norderstedt, pp. 265-278.

[3] Zapata-Ros, M. (2013) "Analítica de aprendizaje y personalización", *Campus Virtuales*, 2 (2), pp. 88-118. http://uajournals.com/ojs/index.php/campusvirtuales/article/view/41/40

[4] Martos-Garcia, D.; Usabiaga, O.; Valencia-Peris, A. (2017) "Students' Perception on Formative and Shared Assessment: Connecting two Universities through the Blogosphere". *Journal of New Approaches in Educational Research*, 6 (1), pp. 64-70. DOI: http://dx.doi.org/10.7821/naer.2017.1.194

[5] Urrutia, M. L., Cano, E. V., & López-Meneses, E. (2017). Analítica de aprendizaje en MOOC mediante métricas dinámicas en tiempo real. *@ tic. revista d'innovació educativa*, (18), 38.

[6] Nóbile Cecilia Inés., Luna Álvaro Enrique. Los entornos virtuales de Enseñanza y Aprendizaje en la Universidad Nacional de La Plata. Una aproximación a los usos y opiniones de los estudiantes. INNOEDUCA, Editorial: Universidad de Málaga,Vol. 1, pp. 3-9, 2015.

[7] Paladea, C. G. (2019). Detección automática del nivel de conocimiento de estudiantes en entornos de e-learning.

[8] Mora-Vicarioli, F. (2019). Estado del arte de la evaluación de los aprendizajes en la modalidad del e-learning desde la perspectiva de evaluar para aprender: precisiones conceptuales. *Revista Electrónica Calidad en la Educación Superior*, 10(1), 58-95.

[9] Mercado Borja, W. E., Guarnieri, G., & Luján Rodríguez, G. (2019). Análisis y evaluación de procesos de interactividad en entornos virtuales de aprendizaje (Analysis and Evaluation of Interactivity in Virtual Learning Environments). *Trilogía Ciencia Tecnología Sociedad, 11*(20).

[10] Roig-Vila, R., Martínez, J. M. A., & Lledó, A. (2018). Memorias del Programa de Redes-I3CE de calidad, innovación e investigación en docencia universitaria. Convocatoria 2017-18.

[11] Vargas-Cubero, A. L., & Villalobos-Torres, G. (2018). El uso de plataformas virtuales y su impacto en el proceso de aprendizaje en las asignaturas de las carreras de Criminología y Ciencias Policiales, de la Universidad Estatal a Distancia de Costa Rica. *Revista Electrónica Educare*, *22*(1), 20-39.

[12] Ibáñez, J. C., Sáiz, M. S. I., & Gómez, G. R. (2018). Propuesta metodológica de evaluación para evaluar competencias a través de tareas complejas en entornos virtuales de aprendizaje. *Revista de Investigación Educativa*, *36*(1), 159-184.

ANÁLISIS DESDE LA PERSPECTIVA INTERACCIÓN HOMBRE- MÁQUINA DE LOS PROCESOS DE INCLUSIÓN EN EL CONTEXTO UNIVERSITARIO.

Alejandra Noemi Marquesin[1], Silvina Daniela Traverso, Alejandro Javier Hadad.

Laboratorio de Análisis, Procesamiento, Almacenamiento y Control de Datos- LAPACDa.

Facultad de Ciencia y Tecnología- Universidad Autónoma de Entre Ríos (UADER).

Código Postal 3.100- Oro Verde- Entre Ríos- Argentina.

alemarquesin@gmail.com, daniellatraverso@gmail.com, alejandrojavierhadad@gmail.com.

RESUMEN

El presente proyecto se plantea identificar factores que influyen en la trayectoria de los estudiantes con discapacidad desde dos cuestiones esenciales: acceso físico al edificio y materiales digitales accesibles.

Actualmente las Tecnologias de Información y Comunicación (TIC´S) están presentes en muchos ámbitos, y su uso esta cada vez más popularizado dentro de las organizaciones educativas, por donde transitan personas con discapacidad. Sin embargo, su utilización presenta ciertas dificultades para las personas con discapacidad, atento a que los diseños están dirigidos a usuarios sin inconvenientes y no se tienen en cuenta sus necesidades especiales. En particular la presencia de estudiantes con discapacidad en la Universidad es una realidad.

En el año académico 2019, se solicitó información para realizar un relevamiento de la población estudiantil con discapacidad, al Área de Sistemas y Departamento de Alumnado de dos facultades pertenecientes a la Universidad Autónoma de Entre Ríos (de aquí en adelante UADER): Facultad de Ciencia y Tecnología (de aquí en adelante FCyT) y la Facultad de Ciencias de la Vida y la Salud (de aquí en adelante FCVyS).

Los resultados fueron los siguientes:

- FCyT: 61 estudiantes revelaron tener una discapacidad, de los cuales 18 con discapacidad visual, 2 con discapacidad auditiva, 5 con discapacidad motriz y 36 con otro tipo de discapacidad[1].
- FCVyS: 177 estudiantes revelaron tener una discapacidad, de los cuales 38 con discapacidad visual, 10 con discapacidad auditiva, 11 con discapacidad motriz y 118 con otro tipo de discapacidad.

Es a partir de este grupo de estudiantes, que se genera la necesidad de explorar un camino didáctico innovador, con la finalidad de guiar y acompañar en el objetivo máximo del proyecto, que es el que estos educandos logren acceder a la institución y a sus materiales de estudio, sin inconveniente, considerando ajustes razonables, ambientes propicios, disminución de las barreras, entre otros; con la finalidad de obtener un diseño flexible.

PALABRAS CLAVES: inclusión, usabilidad, estudiante con discapacidad y universidad.

CONTEXTO

En la Argentina, actualmente, puede observarse mayor presencia de estudiantes con discapacidad dentro de las instituciones de educación superior.

Existen abundantes leyes de jurisdicción nacional, que remiten a los derechos de las personas con discapacidad. A

[1] Cómo por ejemplo: Trastornos del lenguaje, Neurológica, entre otros.

partir de ellas, las organizaciones educativas se ven impulsadas a transitar procesos tendientes a eliminar barreras físicas, culturales, comunicacionales y sociales, en pos de materializar el objetivo propuesto por esta tendencia educativa de carácter igualitario, que no es otra que el ejercicio del derecho a la educación, de todos los ciudadanos que la requieran.

A partir de lo anterior, es que las universidades tienen la necesidad de transformarse en ámbitos inclusivos. Sin embargo, puede observarse que, a pesar del esfuerzo institucional realizado por la UADER, existen escasos progresos en cuanto a la accesibilidad.

En concordancia con esa premisa y para atender a las demandas sociales y propias de la institución, se encuentran actualmente trabajando en conjunto dos facultades pertenecientes a UADER:

- FCyT, Sede Oro Verde y Diamante
- FCVyS, Sede Paraná y sede Ramirez.

Se elaboró, y actualmente está en ejecución, un Proyecto de Investigación y Desarrollo de Inserción[2] (PIDIN), denominado: *"Análisis desde la perspectiva interacción hombre- máquina de los procesos de inclusión en el contexto universitario"*. El objetivo del proyecto es proporcionar información sobre la población estudiantil con discapacidad, que sirva como insumo para modificar prácticas didácticas y pedagógicas al interior de las cátedras, y también las dinámicas de gestión institucional sobre un estudiante con discapacidad.

Además, entendemos que, es la primera de muchas acciones sistematizadas que las facultades involucradas deben desarrollar, para trabajar en constituirse de forma concreta, en facultades inclusivas.

[2] Proyectos centrados en la formación de noveles investigadores en funciones de dirección y ejecución, con una duración de un año para su ejecución y hasta seis meses adicionales para la presentación del informe.

INTRODUCCIÓN

Para el desarrollo del presente trabajo, se comenzó por realizar un relevamiento de la población estudiantil con discapacidad de FCyT y FCVyS. Como se mencionó anteriormente, la información la brindo el Área de Sistemas y Departamento de Alumnado de cada facultad, quienes tomaron como dato primario, la información del Sistema SIU Guaraní[3].

En simultáneo, se realizaron acciones de difusión en el ámbito de influencia regional de ambas facultades, para convocar a personas con discapacidad relacionadas al ámbito académico, que estuvieran interesadas en colaborar con el proyecto. Esta acción produjo respuestas positivas por lo que se sumaron más interesados.

Con la información obtenida del relevamiento más aquellas personas con discapacidad que desearon participar en el proyecto de manera colaborativa, se seleccionó una muestra de personas según el siguiente criterio en cuanto a la discapacidad: visual y motriz.

La discapacidad visual, en particular, es aquella que se muestra por una disminución total o parcial de la vista. Barraga (1985) distingue entre:

- Ceguera total: ceguera total o sólo percepción de luz que el individuo no puede utilizar para la adquisición de ningún conocimiento o información.

- Ceguera parcial: percepción de bultos.

- Baja visión: el déficit visual incapacita al individuo para algunas actividades usuales, precisando de adaptaciones o métodos específicos, como puede ser la lectoescritura braille, para llevar a cabo algunas de ellas. Puede ver objetos a pocos centímetros.

[3] **SIU-GUARANÍ** es un sistema de gestión académica que registra y administra todas las actividades académicas de la Universidad y sus Facultades, desde que los alumnos ingresan como aspirantes hasta que obtienen el diploma.

- Visión límite: el déficit visual no incapacita al individuo para las actividades habituales, pero precisa de adaptaciones sencillas para poder llevar a cabo algunas de ellas. Puede leer en tinta con ayudas ópticas o con ampliaciones.

Por otro lado, respecto de la discapacidad motriz, Casado (2013) dice: "abarca desde las personas que no son capaces de mover ninguna extremidad (tetraplejia) hasta las personas que pueden andar, pero necesitan ayuda (como bastones y muletas), pasando por quienes no pueden mover las extremidades inferiores (paraplejia)".

El autor añade: "es importante tener en cuenta que hay ocasiones en las que una persona puede presentar variaciones simultáneas en varias de sus funciones. Un ejemplo de esto son las personas sordociegas, que combinan pérdida de visión con pérdida de oído. Estas personas dependen casi exclusivamente de la modalidad táctil para poder comunicarse".

Para obtener información, respecto a la situación actual de las facultades en estudio, se trabajaron diferentes técnicas de indagación, destinadas a, docentes, de manera de conocer cuáles son las ideas propias y aspectos formativos en relación con la inclusión de personas con discapacidad.

En segundo término, desde la palabra del estudiante con discapacidad, con el objetivo de conocer las experiencias que atravesaron en base a estrategias de inclusión llevadas a cabo por la institución, para el ingreso, permanencia y egreso. Este trabajo, no abordo el análisis, desde la perspectiva de la inclusión, de procedimientos administrativos mediante los cuales la organización gestiona las acciones de los estudiantes de estas características, ejemplo de ello, los trámites de inscripción a carrera, inscripción a cátedras, tramitación de un título.

Otro tópico que se incluyó en la interacción con las personas con discapacidad, fue indagar sobre cuán importante son los entornos tecnológicos accesibles para acceder al material de estudio. Esto último, a partir de la idea de que lo que se pretende es obtener los requerimientos de los estudiantes con discapacidad para una trayectoria amena en la universidad.

El autor Loucopoulos (1995) define requerimiento como: "una sucesión sistemática de desarrollo de los requisitos a través de un proceso cooperativo iterativo de analizar el problema, la documentación de la observación que resulta en variedad de formatos de representación y la comprobación de la exactitud de la comprensión."

Hasta el momento las técnicas de indagación trabajadas fueron:

- Entrevistas para estudiantes, definida por Lores (2001) como: "entrevistar a los usuarios respecto de su experiencia en un sistema interactivo resulta una manera directa y estructurada de recoger información. Además las cuestiones se pueden variar con tal de adaptarlas al contexto"
- Cuestionarios para estudiantes y docentes, definida por Lores (2001) como: "el cuestionario es menos flexible que la entrevista, pero puede llegar a un grupo más numeroso y se puede analizar con más rigor."
- Observación de campo para estudiantes, para analizar como utiliza las tecnologías para acceder a los documentos digitales, Lores (2001) explica" objetivo principal consistirá en observarlos para entender cómo realizan sus tareas."

Cuando los usuarios y las computadoras interactúan lo hacen a través de una interfaz.

Según (Lores, 2001) una interfaz es: "una superficie de contacto que refleja las propiedades físicas de los que interactúan, las funciones a realizar y el balance de poder y control. En el caso de la Interacción Persona Ordenador, la interfaz es el punto en el que seres humanos y ordenadores se ponen en

contacto, transmitiéndose mutuamente tanto información, órdenes y datos como sensaciones, intuiciones y nuevas formas de ver las cosas."

Sin embargo, una interfaz puede ser un límite en la comunicación. Si no logra expresarse de forma adecuada genera que la interfaz se convierta en una barrera al acceso del conocimiento, debido a problemas de diseño por la escasa atención respecto a: contexto de uso, las tareas a realizar y las características del usuario.

Para una interacción persona- ordenador, debemos tener en cuenta un término fundamental: *usabilidad*. Conforme a la norma ISO 9241-11 la define como: "medida en que un producto puede ser usado por usuarios específicos para alcanzar ciertas metas con efectividad, eficiencia y satisfacción en un determinado contexto de uso". Para que un sistema interactivo cumpla sus objetivos tiene que ser usable y accesible a la mayor parte de la población.

En el proyecto, se seleccionó un material académico digital al azar, y con colaboración de un usuario con discapacidad, se realizó la prueba de test *pensando en voz alta*. Lores (2001) lo explica como: "en este método de evaluación se les pide a los usuarios que expresen en voz alta sus pensamientos, sentimientos y opiniones mientras que interaccionan con el sistema".

Con toda la información colectada, por medio de diferentes métodos, más una lectura bibliográfica, se proponen criterios de producción de materiales accesible, destinada a los actores institucionales y público en general. Además, realizar recomendaciones respecto a software o hardware que se adapten a las necesidades del estudiante universitario con discapacidad.

Vale la pena aclarar, que para este informe, se trabaja todo aquello vinculado a las Tecnologías de Información y Comunicación, dejando para otra oportunidad cuestiones relacionadas al acceso físico al edificio.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Siguiendo la línea de investigación correspondiente a lo expuesto en este trabajo, se llevarán a cabo actividades relacionadas con las siguientes áreas temáticas:

- Interacción Hombre Máquina.
- Tecnologías inclusivas.
- Educación inclusiva.

RESULTADOS OBTENIDOS O ESPERADOS

Hasta el momento se han obtenido:

- Relevamiento de estudiantes con discapacidad que concurren a la FCyT Y FCVyS. Cantidad de abandono, permanencia y egreso.
- Requerimiento de los estudiantes de dichas facultades respecto a su trayectoria por las instituciones.
- Requerimientos de los usuarios con discapacidad en cuanto a materiales académicos accesibles y tecnología de apoyo existentes.
- El conocimiento que tiene el docente universitario respecto a la inclusión.
- Pautas de uso de Software y hardware que se adaptan a diversas discapacidades.
- Criterios de producción de materiales accesibles en procesadores de texto.

Se plantea como trabajo futuro, ampliar los casos de estudio incorporando a los estudiantes sordos.

FORMACIÓN DE RECURSOS HUMANOS

Este es el primer proyecto de investigación en esta temática del equipo. Actualmente forman parte de un nuevo Laboratorio de Investigación, creado por la FCyT e inaugurado en marzo del 2020, denominado: Laboratorio de Análisis, Procesamiento, Almacenamiento y Control de Datos- LAPACDa.

Forman parte del equipo una Directora (FCyT), un Codirector (FCVyS), dos participantes pertenecientes a FCyT, incluido una estudiante, y dos participantes pertenecientes a FCVyS.

Se prevé la capacitación y formación relacionadas a la temática de recursos humanos, a través de cursos de actualización y posgrado en el área de estudio. Además la transferencia de conocimiento y resultados; brindar charlas informativas para difundir los resultados en el ámbito de UADER y la región.

BIBLIOGRAFIA

- Barraga, n. (1985). "Disminuidos visuales y aprendizaje". Madrid, once.
- Casado Martínez c (2011). "Interacción persona- ordenador". Barcelona.
- ISO 9241-11:1998. Requisitos ergonómicos para trabajos de oficina con pantallas de visualización de datos (pvd) - parte 11: orientación en la usabilidad. 2014.
- Lorés j. (2001). AIPO- Asociación introducción a la interacción persona-operador. España.
- Loucopoulos p. (1995). "Requisitos del sistema en ingeniería".

Dispositivos móviles para la educación en colegios de nivel secundario

Pizarro Rubén, Testa Oscar, Camiletti Pablo, Ascheri María Eva
Departamento de Matemática / Facultad de Ciencias Exactas y Naturales/ Universidad Nacional de La Pampa
Uruguay 151, 00-54-02954-245220 int 7124
rubenpizarro71@gmail.com ; otesta@gmail.com

RESUMEN

Los dispositivos móviles han modificado el entorno de estudiantes y profesores, provocando profundos cambios en las formas de hacer y aprender de todos. Es necesario un rediseño metodológico y pedagógico para el abordaje de la educación de nivel secundario. Para estudiar este problema surge nuestro proyecto de investigación aprobado y financiado por la Facultad de Ciencias Exactas y Naturales de la UNLPam.

En una primera etapa del proyecto se ha realizado una revisión bibliográfica sobre el tema de investigación, las aplicaciones existentes destinadas a la enseñanza en el nivel de la educación secundaria y las experiencias de inclusión de dispositivos móviles en este nivel. Posteriormente definimos las instituciones, el diseño y elaboración de herramientas para la recopilación de información para avanzar con la experiencia directa en el ámbito educativo y la recopilación de la información propiamente dicha. Además hemos desarrollado capacitaciones a docentes de nivel medio, vinculadas con la inclusión de dispositivos móviles en educación secundaria. Nos resta organizar, analizar y elaborar conclusiones.

Nos proponemos analizar la compleja influencia del acceso a las redes y la utilización de dispositivos móviles y cómo aprovechar estas características para experimentar, analizar y proponer mejoras en las formas de utilizar estas tecnologías en los colegios de nivel secundario de la Ciudad de Santa Rosa

Palabras clave: *dispositivos móviles, aplicaciones, enseñanza aprendizaje.*

CONTEXTO

El proyecto de investigación que se presenta está radicado en el Departamento de Matemática de la Facultad de Ciencias Exactas y Naturales de la UNLPam y financiado por dicha Facultad.

Tiene una duración de cuatro años, siendo este el cuarto año de su desarrollo. Está vinculado con otros proyectos del Departamento ya ejecutados y en ejecución, relacionados con la enseñanza en nivel secundario y la inclusión de tecnologías en educación.

1. INTRODUCCIÓN

Desde que en 1989 Salomon, Perkins y Roy Pea entre otros, compartieran unas jornadas sobre computadoras y aprendizaje a partir de las cuales apareciera años más tarde el título "Distributed cognitions" comenzamos a considerar con distintas palabras y distintos matices, que el conocimiento está presente

entre otras cosas en los artefactos y sus modos de usos, que conforman instrumentos y medios que nos transforman. Uno de ellos, quizás el artefacto simbólico omnipresente en la mayor parte de nuestras mediaciones hoy en día es el celular o dispositivo móvil, que evidentemente han modificado el entorno de estudiantes y profesores, provocando profundos cambios en las formas de hacer y aprender de todos. Precisamente porque observamos que los modos de conocer están íntimamente vinculados con las personas, los artefactos disponibles y las relaciones establecidas entre personas a través del uso de artefactos, es que proponemos encontrarnos con docentes en las escuelas para indagar las formas en las que ellos conviven con estos aparatos y las redes de relaciones que con ellos construyen, los contenidos que comparten y producen. Estos dispositivos que surgieron como herramientas de comunicación y entretenimiento han llegado a desarrollar un papel fundamental en el mundo de la economía y el conjunto de la sociedad (Shuler et al, 2013).

Sin dudas los dispositivos móviles están presentes en la vida de la mayoría de docentes y estudiantes para realizar diversas actividades. Ante esta realidad los Ministerios de Educación en diferentes países han iniciado políticas tendientes a que estos dispositivos cobren mayor importancia también en educación. En esta línea podemos mencionar los modelos 1 a 1 implementados en mucho países que consisten en la provisión por parte del estado de un dispositivo móvil (netbook o tableta) a cada alumno y docente. También existen iniciativas del tipo "Trae tu propia tecnología" denominada (BYOT, por sus siglas en inglés 'Bring Your Own Technology'), consistentes en que cada estudiante lleve su propio dispositivo móvil para utilizarlo en la escuela. Esta metodología se ha implementado en sectores sociales más desarrollados en los cuales todos pueden acceder a estos dispositivos.

Los proyectos mencionados anteriormente, entre otros de similares características, permiten sin dudas un mayor acceso a la tecnología y aumenta las posibilidades de acceso a la información de los docentes y estudiantes. Permitirían así acceder a las principales características tecnológicas del aprendizaje móvil que según indica Cantillo Valero, et. al. (2012), son las siguientes:

- Portabilidad, debido al pequeño tamaño de los dispositivos.
- Inmediatez y conectividad mediante redes inalámbricas.
- Ubicuidad, ya que se libera el aprendizaje de barreras espaciales o temporales.
- Adaptabilidad de servicios, aplicaciones e interfaces a las necesidades del usuario. También existe la posibilidad de incluir accesorios como teclados o lápices para facilitar su uso.

Vivimos indudablemente momentos en los cuales las tecnologías avanzan vertiginosamente pero existe una amplia brecha entre el desarrollo y la implementación (Montoya, 2009); es así que continuando con el análisis de este autor, nos indica que para lograr la inclusión de los dispositivos móviles, en especial los teléfonos inteligentes, se puede temporizar tres pasos:

1. Apropiación del objeto
2. Apropiación de la funcionalidad
3. Apropiación de la nueva forma de aprendizaje

Las dos primeras etapas como señala Díaz Bilbao (2014), serían alcanzadas rápidamente

por los estudiantes, quedando la tercera etapa a cargo de las instituciones educativas y los docentes.

Sin duda la mayoría de los autores coinciden en indicar que la inclusión de dispositivos móviles en el aula, produce grandes beneficios: motivan a los estudiantes, aumenta las habilidades sociales mejorando la cooperación y la colaboración mejorando la creatividad y la capacidad cognitiva.

Sin embargo como ha sucedido con otras tecnologías, percibimos que los dispositivos móviles actualmente no son tenidos en cuenta ni aprovechados en las formas ni en la magnitud que observamos sería conveniente. Por el contrario en algunos ámbitos han sido prohibidos por causar distracción y otros problemas entre los estudiantes y la comunidad educativa.

Como lo expresan Castells, Palamidessi y muchos otros, los cambios que se han producido en las formas de comunicarse, acceder a información y conocer el mundo, mucho tienen que ver el acceso a las redes y los nuevos dispositivos y esto pareciera ser de alguna manera todavía resistido en el ámbito educativo. Al mismo tiempo se reconoce una crisis identificada en las dificultades que se presentan con la falta de atención, falta de aprendizaje o productividad, de motivación, de malestar de estudiantes y profesores entre otros problemas.

¿Tendremos que plantearnos abordar el trabajo en las instituciones educativas desde nuevos paradigmas? Pensar alternativas para las instituciones educativas implica salir de lugares comunes, observar que el mundo y la sociedad han cambiado mucho, en comparación las instituciones educativas, emblemáticas a la hora de hablar del conocimiento, parecen no haber dado signos claros de entender la metamorfosis del conocimiento a la que hoy asistimos.

Es evidente que este contexto no puede ser abordado con propuestas simplistas, pensar que la sola inclusión de dispositivos móviles puede generar las condiciones necesarias para mejorar las posibilidades de enseñanza y aprendizaje en la educación no alcanza.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación que se abordan con el presente proyecto son las siguientes:

A. Nivel de penetración de los dispositivos móviles en las actividades de enseñanza aprendizaje en aulas de nivel medio.
B. Búsqueda, experimentación y análisis de aplicaciones educativas para la enseñanza en el nivel medio
C. Características metodológicas de las clases de nivel medio en las que se incluye dispositivos móviles.

3. RESULTADOS OBTENIDOS / ESPERADOS

Dado que este proyecto nos lleva a insertarnos en la escuela secundaria de la ciudad de Santa Rosa provincia de La Pampa, no dudamos en hacer un relevamiento propio de los docentes de alguna de las instituciones en las que estamos en contacto. Este relevamiento lo hicimos mediante un cuestionario cuyas dimensiones de análisis fueron:

- Características de los teléfonos celulares de los docentes

- Frecuencia y tipo de utilización de su teléfono celular
- Teléfonos celulares en la institución educativa

Respecto a una de las dimensiones indicadas, "Características de los teléfonos celulares de los docentes", relacionado al acceso a dispositivos móviles y conectividad, los docentes encuestados respondieron de la siguiente forma:

- el 100% tiene teléfono móvil,
- el 84% indica tener sistema operativo Android,
- teléfonos móviles con una antigüedad de entre uno y dos años,
- el 100% tiene internet en su hogar,
- el 90% accede a internet desde su teléfono móvil utilizando paquete de datos
- el 68% accede por wifi

Claramente la utilización de dispositivos móviles no es una actividad ajena a los docentes.

Analizando la dimensión "Frecuencia y tipo de utilización de su teléfono celular", vemos que los docentes utilizan los teléfonos móviles principalmente para comunicarse con familiares y para chequear redes sociales. Siendo relativamente bajo el porcentaje de docente que busca información en internet utilizando el teléfono móvil y muy pocos los que juega utilizando estos dispositivos Relacionada a la "Frecuencia y tipo de utilización de su teléfono celular" y a los "Teléfonos celulares en la institución educativa", se consultó entre otros aspectos, la consideración que tenía del teléfono móvil tanto para su vida cotidiana como para su actividad laboral.

De los datos obtenidos vemos que la utilidad que se le asigna al dispositivo móvil en la cotidianeidad, la mayoría opina que el mismo resulta indispensable. Aunque en el ámbito laboral no se presenta la misma opinión, ya que la mayoría de los encuestados indican su desacuerdo en considerar indispensables a los teléfonos móviles. Además se analizó el rol de los teléfonos celulares en la institución educativa.

Hemos desarrollado además acciones de capacitación de docentes de nivel medio sobre la utilización de estas tecnologías en las aulas. Los resultados de las mismas están en análisis y se espera extender y profundizar las mejoras que los dispositivos móviles y el tipo de actividades asociados a ellos puedan aportar al ámbito educativo.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo es de carácter interdisciplinario, está conformado con especialistas del área de Educación, Matemática y Computación. Participan además estudiantes del Profesorado en Computación.

Se realizan, también, actividades de capacitación y transferencia referidas a la utilización de aplicaciones educativas para dispositivos móviles con sistema operativo Android.

5. BIBLIOGRAFÍA

- Argentina, Ministerio de Educación de Argentina. (2015). *Conectar Igualdad, fundamentos del programa.* Disponible en http://www.conectarigualdad.gob.ar/seccion/sobre-programa/fundamentos-del-programa-17

- Castells, M. (2016). ¿Comunidades virtuales o sociedad en red?. Plaza & Janés Editores. Barcelona, España

- Castells, M., & Schmalenberger, M. (2006). *La sociedad red: una visión global* (No. U10 1087). IICA, Bs As (Argentina).

- Cantillo Valero, C., Roura Redondo, M. & Sánchez Palacín, A. (2012). Tendencias actuales en el uso de dispositivos móviles en educación. Educational Portal of the Americas – Department of Human Development, Education and Culture. OEA-OAS ISSN 0013-1059 La Educ@ción Digital Magazine N 147 – www.educoas.org

- Díaz Bilbao (2014) El smartphone como herramienta educativa. Trabajo final de maestría. Máster en formación del profesorado de educación secundaria (ámbito tecnológico). Universidad Pública de Navarra.

- Montoya, M. (2009). Mobile learning – mlearning- technology resources and their relationship with distance learning environments: applications and research studies. Tecnológico de Monterrey, ITESM (México).

- Miranda Ruiz, E. (2010). *Paradigma Interpretativo en Investigación.* Disponible http://www.monografias.com/trabajos97/paradigma-interpretativo/paradigma-interpretativo.shtm1

- Palamidessi, M. (2006). La escuela en la sociedad de redes. *Una introducción a las tecnologías de la* información y la comunicación.

- RINALDI, M. (2012). *Revolución Mobile Leraning*. Disponble en: http://es.slideshare.net/crossmedialab/revolucin-mobile-learning Consultado 06/08/2013.

- Shuler C, Niall Winters & Mark West. (2013) El futuro del aprendizaje móvil. Implicaciones para la formulación y planificación de políticas et al. UNESCO working paper series on mobile learning. Disponible en http://unesdoc.unesco.org/images/0021/002196/219637s.pdf Consultado 10/08/16.

- UNESCO (2012) Aprendizaje móvil para docentes. Temas globales. Recuperado el 17 de marzo de 2017 de http://unesdoc.unesco.org/images/0021/002164/216452s.pdf

Construcción de una Herramienta de Capacitación a Distancia y Recursos Educativos Abiertos bajo un Enfoque de Diseño Universal orientados a Personas con Discapacidad Visual

Guillermo Javier Lafuente[1], Carlos Ballesteros[2], José Luis Filippi[3], Gustavo Hernán Lafuente[4]

GIAU[5] – Facultad de Ingeniería – UNLPam.
Calle 110 esq. 9 nº 390
[1]gjlafuente@gmail.com, {balleste[2], filippij[3],gustavo[4]}@ing.unlpam.edu.ar
[5]Grupo de Investigación de Ambientes Ubicuos – http://giau.ing.unlpam.edu.ar/

Resumen

Este proyecto de "Inclusión en la web, como diseño Universal para personas con discapacidad visual" liderado por un equipo interdisciplinario de la Facultad de Ingeniería de la Universidad Nacional de La Pampa tiene como finalidad implementar una plataforma de educación a distancia accesible y utilizable que permita brindar cursos de capacitación inclusiva a personas no videntes o con disminución visual y a todas aquellas personas que deseen utilizarla como herramienta de aprendizaje. Las capacitaciones serán dictadas por la Fundación BienEstar[1], especializada en el trabajo con personas con discapacidad.

Para contribuir a la resolución del proyecto, se viene trabajando desde hace dos años en aspectos de accesibilidad web y en la actualidad se ha implementado un EAD (Entorno de Educación a Distancia) que contempla las características de Accesibilidad explicitadas en el (Artíc. 9° de la Convención Internacional de los Derechos de las Personas con Discapacidad Ley (26.378). En forma paralela, se encuentra en proceso de desarrollo un conjunto de Recursos Educativos Abiertos y Accesibles (REAAs), con el objetivo de ser instrumentos de capacitación en el contexto del EAD implementado, en el marco de alguna disciplina de interés para la Fundación BienEstar.

Palabras clave: Inclusión Social, Accesibilidad, Educación a Distancia, Discapacidad Visual.

Contexto

Este proyecto se lleva a cabo en el dentro de las líneas de I+D desarrolladas por el grupo GIAU (Grupo de Investigación de Ambientes Ubicuos) de la Facultad de Ingeniería de la UNLPam.

Introducción

Este proyecto involucra el diseño e implementación de una plataforma de capacitación a distancia utilizable y abierta para personas con discapacidad visual. El concepto de diseño universal utilizado en varios campos y acuñado por Ronald Mace [1], [2], se refiere a la idea de diseñar productos que sean estéticos

[1] Fundación BienEstar, Realicó, La Pampa http://www.fundacionparaelbienestar.org/

y usables para cualquier persona independientemente de su edad, habilidad y estado. Los términos más comunes utilizados para referirse a diseño universal son: simple, intuitivo, flexible, equitativo, perceptible y tolerable al error. El término diseño universal está muy relacionado a otros términos como accesibilidad o usabilidad [3].

Con el crecimiento exponencial de las nuevas tecnologías, el término accesibilidad también se ha extendido al ámbito de la informática. Junto al crecimiento de Internet, crecieron estrategias de accesibilidad, principalmente las de accesibilidad para la Web. Diversos autores [4] [5] han discutido y desarrollado este tópico, describiendo así, tecnologías que asisten a la navegación web como el *reconocimiento de voz*, o la *magnificación de pantalla y lectores de pantalla*. En 1999 la Web Accessibility Initiative (WAI) publicó las guías para contenido web accesible WCAG [6], [7], para mejorar la accesibilidad de la web para personas con discapacidad [8].

Situación actual del problema

Con el surgimiento de los dispositivos móviles, la interacción humano - computadora ha cambiado significativamente, apareciendo nuevos desafíos que conlleva a involucrar nuevas técnicas para la evaluación de usabilidad [9], [10]. Ha ocurrido un cambio radical en el desarrollo de pantallas táctiles basadas en dispositivos móviles. En poco tiempo la interacción basada en gestos se ha convertido en un estándar para el desarrollo de interfaces en los dispositivos móviles. Las pantallas táctiles proveen una gran flexibilidad y acceso directo a controles e información, aunque por otro lado lo hacen menos accesibles para usuarios ciegos o con impedimentos visuales.

Dado que la mayoría de las aplicaciones están diseñadas para usuario con visión, las características de accesibilidad no siempre son adecuadas para obtener resultados aceptables. Otro parámetro a tener en cuenta es la relación con los métodos del proceso de enseñanza y aprendizaje que utilizan los docentes de todos los niveles, que sólo enfatizan la mayoría de los métodos visuales. Quizá algunos profesores desconozcan las técnicas y estrategias para impartir clase a los estudiantes contemplando aspectos de discapacidad con ceguera parcial o total. Una aplicación diseñada para usuarios videntes, con una capa extra que incluya características de accesibilidad no ha mostrado ser una opción utilizable. El usuario con discapacidad visual puede utilizar esta aplicación, pero la interfaz de usuario no ha sido concebida para ciegos: Por consiguiente, una buena experiencia de usuario para el caso de ciegos, no está para nada asegurada. De acuerdo a esta premisa, se necesitan aplicaciones que contemplen a usuarios con baja visión o no videntes, logrando una mejor experiencia de usuario posible, con un Desarrollo Centrado en este tipo de Usuario (DCU).

Fundamentación

Existen diversas barreras a las cuales se enfrentan las personas con discapacidad visual, no sólo físicas, sino también de acceso a la información, quedando muchas veces excluidas del sistema socioeducativo. Según Piñeros [11], la desventaja radica, en el ámbito social, que no logra integrar a la persona discapacitada. Los usuarios con discapacidad visual "*tienen las mismas necesidades de información que el resto de los ciudadanos*". Estas personas deben recibir información accesible, que les permita tomar decisiones y realizar una vida independiente [11].

La Facultad de Ingeniería de la UNLPam, busca dar respuesta a la necesidad que en la actualidad presenta la "Fundación para el BienEstar" brindando tecnologías como ámbitos de acceso, difusión de la información y servicios a la comunidad. Se intenta sentar las bases de inclusión digital en una sociedad, que desafíe las diferencias, que profundice los vínculos y que permita alcanzar mayor igualdad

social y educativa para personas con discapacidad a través de un diseño para todos, con las configuraciones de apoyo tecnológico que deban adoptarse.
Atento a esto, es que se diseña una plataforma de educación a distancia para capacitar a personas con disminución visual, no videntes y demás beneficiarios que quieran hacer uso de la misma, empleando para ello, diferentes mediadores didácticos que orienten a la utilización del mismo.
En la primera conferencia de E-learning hacia la inclusión social se declararon los siguientes términos: "*El e-learning no ha de limitarse a ser cursos en línea para universidades y grandes compañías. No ha de estar centrado exclusivamente en cómo aumentar los beneficios. Cuando se desarrollen módulos de e-learning, se deberá estar seguro de que todos los grupos sociales tienen acceso a las técnicas, y darle a todo el mundo los medios para usar las TIC en su desarrollo profesional y personal, y así poder aprender en la sociedad de la información*" [12]. A través de esta línea de trabajo, los investigadores entienden que deben provocar un impacto social positivo, que tenga mayor conciencia social inclusiva; utilizando buenas prácticas que respeten la diversidad en un marco de igualdad.

Metodología de Trabajo

Desde el punto de vista del desarrollo, el proyecto está siendo abordado mediante la metodología Design Thinking [13], combinadas junto a la metodología Lean Startup [14] para disponer de un producto mínimo viable que responda rápidamente a los requerimientos de usuario.
Desde el punto de vista de la accesibilidad en los artefactos desarrollados, se están implementando dos estrategias para realizar la validación y verificación de la accesibilidad teniendo en cuenta principalmente dos tipos de usuario, el no vidente y el disminuido visual. Para ello, se elaboraron dos estrategias de evaluación automatizada para verificar la accesibilidad como así también una estrategia de evaluación centrada en el usuario.

a) Estrategia de evaluación automatizada de la accesibilidad

Cada artefacto de software que se produce actualmente es evaluado para alcanzar el nivel máximo de accesibilidad. Para ello, se utiliza un conjunto de herramientas automatizadas [15], [16], [17] que asisten a los procesos de desarrollo, para obtener un mayor grado de accesibilidad sobre los productos construidos. Estas herramientas examinan la accesibilidad bajo las directrices de accesibilidad web WCAG, e informan sobre las barreras conocidas y probables o potenciales (en forma de un informe independiente y / o anotaciones del producto evaluado).
También se trabaja con lectores de pantalla para detectar problemas de interpretación de texto en la lectura de documentos de distinta fuente o formato.
Los datos obtenidos de las herramientas [15], [16], [17] son agrupados y tabulados para preparar una lista única de problemas potenciales de accesibilidad para cada sección de los artefactos de software construidos. Se involucra en este proceso cada unidad de información contenida en los REAAs elaborados.

b) *Estrategia de evaluación de accesibilidad centrada en el estudiante/usuario*

En la actualidad se ha seleccionado un grupo definido de estudiantes el cual se encuentra integrado de la siguiente forma: la mitad por estudiantes sin problemas y la otra mitad con estudiantes no videntes y disminuidos visuales. Los expertos de la Fundación para el Bienestar colaboran en esta tarea de selección. Paralelamente, se está avanzando en un laboratorio de pruebas y software para que los participantes elegidos completen algunas actividades de un REAA. El laboratorio consistirá de una estación de trabajo en un aula universitaria de acuerdo con un formato tradicional de laboratorio de pruebas de

usabilidad como los propuestos por Nielsen [18] y Rubin y Chisnell [19]. La estación de trabajo estará constituida por un escritorio con una PC, Internet por cable, mouse y micrófono externo. Una cámara de video en un trípode estará disponible delante de la estación para capturar las expresiones faciales del estudiante y las interacciones con la PC, y el aula será visible desde una ventana vidriada para realizar una observación adyacente. Cada estudiante deberá completar la actividad en una sesión moderada (en presencia de un investigador y una cámara de video). Luego realizará una actividad en una sesión no modificada (trabajando solo en el aula con la cámara de video apagada). Durante las sesiones moderadas, el investigador deberá indicar al estudiante que piense en voz alta si es necesario siguiendo una forma de comunicación del habla del protocolo de pensar en voz alta y lo registrará [20].

Para relevar estas pruebas de laboratorio se utilizarán herramientas para captura de escenas como OpenVULab o Screencast-O-Matic[2], las cuales permiten crear screencasts de las actividades en pantalla realizada por los estudiantes mientras completan las consignas propuestas. Los screencasts serán sincronizados con las verbalizaciones de pensamiento en voz alta realizada por los participantes con el fin de registrar todos los detalles de la interacción con los softwares y contenidos desarrollados. Toda la actividad de laboratorio será documentada por los investigadores participantes del proyecto junto con personal idóneo propuesto por la Fundación para el BienEstar.

A futuro, también se piensa armar un escenario similar, pero para estudiar el comportamiento de lo producido en pantallas táctiles o sobre teléfonos inteligentes.

Líneas de Investigación, Desarrollo e Innovación

[2] OpenVULab (http://openvulab.org) Screencast-O-Matic (https://screencast-o-matic.com/)

A continuación, se describen sintéticamente las líneas de investigación en las cuales se avanza para llevar a cabo el objetivo del proyecto.

- Entornos de Educación a Distancia, REAA.
- Usabilidad, Diseño para usabilidad (UX), Diseño Centrado en el Usuario (DCU).
- Diseño de Interacción (IxD)
- Evaluación de usabilidad para ciegos con entornos existentes.
- Métodos, técnicas y herramientas disponibles para cada etapa del proceso DCU.
- Accesibilidad y estandarización en el ciclo de creación de materiales educativos

Resultados y Objetivos

Desarrollar una plataforma de educación a distancia abierta y accesible, y REAAs a través de un diseño centrado en usuarios con discapacidad visual, que esté diseñada y probada por este tipo de usuarios, haciéndolos partícipes del proceso de desarrollo conforme a las metodologías de trabajo implementadas.

Formación de Recursos Humanos

Actualmente, el proyecto cuenta con la siguiente conformación de grupo de trabajo: un Director de Proyecto, un co-director y tres Investigadores, 2 estudiantes, uno perteneciente a carreras relacionadas a informática en el ámbito de la Facultad de Ingeniería de la UNLPam. y otro estudiante perteneciente la carrera Licenciatura en Ciencias de la Educación de la Facultad de Ciencias Humanas de la UNLpam). La conformación del grupo docentes afectado al proyecto ha sido seleccionado siguiendo un criterio interdisciplinar, teniendo en cuenta a docentes especialistas en Educación Virtual, Tecnología y Educación Especial, los cuales pertenecen a la Facultad de Ingeniería y a la Facultad de

Ciencias Humanas de la UNLPam. Se ha incorporado a un integrante estudiante de la Facultad de Ciencias Humanas de la UNLPam, quien es disminuida visual, siendo uno de los actores de mayor relevancia para alcanzar los objetivos del proyecto.

Por último, es importante resaltar que el avance de esta línea de investigación ha permitido generar la formulación de dos tesis de Maestría una relaciona a la estandarización y accesibilidad de recursos educativos abiertos destinado a estudiantes con ceguera parcial o total, y la otra, relacionada al desarrollo y guía para la elaboración de recursos educativos abiertos contemplando los principios del diseño de Interface de Usuario (IU), de Experiencia de Usuario UX, y de Interacción (IxD).

Referencias

[1] Ronald Mace, "The RL Mace Universal Design Institute." [Online]. Available: http://www.udinstitute.org/. [Accessed: 14-Mar-2018].

[2] R. L. Mace, "What is Universal Design (DO NOT CITE!)," 2005.

[3] M. F. Story, "Maximizing Usability: The Principles of Universal Design," *Assist. Technol.*, 1998.

[4] H. Takagi, C. Asakawa, K. Fukuda, and J. Maeda, "Accessibility designer: visualizing usability for the blind," in *ACM SIGACCESS Accessibility and Computing*, 2004, no. 77–78, pp. 177–184.

[5] J. P. Bigham, A. C. Cavender, J. T. Brudvik, J. O. Wobbrock, and R. E. Lander, "WebinSitu," in *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility - Assets '07*, 2007, p. 51.

[6] S. Leuthold, J. A. Bargas-Avila, and K. Opwisa, "Beyond web content accessibility guidelines: Design of enhanced text user interfaces for blind internet users," *Int. J. Hum. Comput. Stud.*, 2008.

[7] "Web Accessibility Initiative (WAI) - home page | Web Accessibility Initiative (WAI) | W3C." [Online]. Available: https://www.w3.org/WAI/. [Accessed: 06-Oct-2017].

[8] J. Mankoff, H. Fait, and T. Tran, "Is your web page accessible?," in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*, 2005.

[9] J. Kjeldskov and J. Stage, "New techniques for usability evaluation of mobile systems," *Int. J. Human-Computer Stud.*, 2004.

[10] A. Kaikkonen, A. Kekalainen, M. Cankar, T. Kallio, and A. Kankainen, "Usability testing of mobile applications: A comparison between laboratory and field testing," *J. Usability Stud.*, 2005.

[11] I. Piñeros, *El acceso a la información de las personas con discapacidad visual : modelo de servicio para bibliotecas públicas*. Alfagrama Ediciones, 2008.

[12] G. Apostopoulou, L. Baronio, and et all., "E-learning hacia la inclusión social," pp. 1–8, 2004.

[13] B. Tim, "Design Thinking By Tim Brown," *Harv. Bus. Rev.*, 2008.

[14] E. Ries, *The Lean Startup: How Today's Entrepeneurs Use Continuous Innovation to Create Rdically Sucesful Businesses*. 2011.

[15] "WAVE Web Accessibility Tool." [Online]. Available: http://wave.webaim.org/. [Accessed: 16-Mar-2018].

[16] "IDI Web Accessibility Checker : Web Accessibility Checker." [Online]. Available: https://achecker.ca/checker/index.php. [Accessed: 16-Mar-2018].

[17] "Home: AInspector Sidebar." [Online]. Available: http://ainspector.github.io/. [Accessed: 16-Mar-2018].

[18] J. Nielsen, *Usability Engineering*, vol. 44, no. 3. 1993.

[19] D. Rubin, J., & Chisnell, *Handbook of usability testing. How to plan, design, and conduct effective tests (2nd ed.)*. 2008.

[20] M. T. Boren and J. Ramey, "Thinking aloud: Reconciling theory and practice," *IEEE Trans. Prof. Commun.*, vol. 43, no. 3, pp. 261–278, 2000.

MODELO DE SISTEMA DE GESTION DE CALIDAD PARA LA CATEDRA DE PROGRAMACION NUMERICA

Jorge A. Silvera, Angel R. Barberis, Lorena E. del Moral, Valeria A. González
Centro de Investigación y Desarrollo en Informática Aplicada (C.I.D.I.A.)
Facultad de Ciencias Exactas – Universidad Nacional de Salta – Salta - Argentina
jsilvera@unsa.edu.ar, barberis@cidia.unsa.edu.ar, lorena.dms.7@gmail.com,
vgonzalez@cidia.unsa.edu.ar

RESUMEN

En este trabajo se presenta la experiencia obtenida con la aplicación, a un caso real, de una metodología de Ingeniería de Software WebML y el estándar IFML aplicado a la gestión de la calidad en la cátedra de programación numérica de la carrera licenciatura de en análisis de sistemas de la Universidad Nacional de Salta (U.N.S.a), como con una herramienta de apoyo al proceso. La experiencia demostró que la aplicación de la metodología y la herramienta contribuyeron a la mejora de la calidad en la gestión de los procesos de enseñanza-aprendizaje.

Palabras claves:
Normas de Calidad ISO 9.001, Sistema de Gestión de Calidad (SGC), Programación Numerica, Herramientas informáticas de apoyo al SGC. Metodología WebML. IFML.

CONTEXTO

La presente investigación se desarrolla en el marco del proyecto número 2536, titulado "Rediseño Educativo para el Aprendizaje del Cálculo Numérico", aprobado por resolución 430/2018-CI del Consejo de Investigación de la Universidad Nacional de Salta, siendo ésta última, la entidad financiante.

El proyecto tiene por objeto primario el desarrollo de una metodología de enseñanza y aprendizaje que propicie un ambiente de entrenamiento en la programación y desarrollo ágil de software, que promueva en el estudiante una interacción social colaborativa y cooperativa; al mismo tiempo, que genere un proceso que provoque en los alumnos motivación, integración, participación e interés por el estudio, persiguiendo así, un mejoramiento en el rendimiento académico y disminución de la tasa de deserción.

La metodología experimental se viene aplicando desde el año 2015, y desarrolla un conjunto de actividades de gestión que abarcan etapas de planificación, desarrollo, evaluación (diagnósticas, de proceso y sumativa), reorganización-adaptación y análisis de riesgo que dependen en mayor o menor medida de la formación previa que tengan los estudiantes en curso. Sobre estas actividades de gestión es que se intenta aproximar un Modelo de Sistema de Gestión de Calidad bajo las Normas ISO 9001, circunscripta al desarrollo de una cátedra, la de Programación y Cálculo Numérico, como una iniciativa atómica, que de resultar positivo la aplicación del estándar ISO, se extenderá al resto de las cátedras que conforman la carrera de la Licenciatura en Análisis de Sistemas, y así extendiéndose a los niveles superiores de la Facultad y la Universidad toda.

1. INTRODUCCION

La formación y creación de profesionales altamente capacitados y competentes es el mayor objetivo que busca alcanzar una institución universitaria, ya que así contribuye al mejoramiento de la calidad de uno de los elementos de primera necesidad de la Sociedad como es la Educación. Bajo la óptica de una organización que presta servicios, la sociedad le exige a las universidades una contraprestación de excelencia, sustentada sobre la base de una educación de calidad facilitada por docentes y adquiridas por los estudiantes.

Ante un mundo globalizado y en constante desarrollo de la información y las comunicaciones, las nuevas sociedades se exponen a nuevos y exigentes requerimientos haciendo que las necesidades educativas evolucionen y cambien constantemente. Ante esto, las universidades deben instrumentar una permanente actualización y adaptación de sus ofertas académicas para satisfacer las necesidades sociales. Es así como la educación universitaria necesita valerse de ciertos instrumentos que le permita validar que sus prestaciones de servicios responden a los máximos estándares de calidad reconocidos por la sociedad para su aceptación. Estos instrumentos son las Normas ISO 9001, que en su versión 2015 [3], establecen los requisitos que deben reunir y cumplir las organizaciones en general para la obtener el reconocimiento de la calidad en su gestión, con la asistencia de un Modelo de Sistema de Gestión de la Calidad (SGC).

La presente investigación propone entre otras actividades el análisis, diseño e implementación de un software de apoyo al SGC para una cátedra, utilizando como metodología Interaction Flow Modeling Language (IFML) [7], de la mano de la herramienta CASE WebRatio [6], en su versión WebRatio Platform University Edition para estudiantes, en el marco del WebRatio University Program, el cual es un programa gratuito creado para las instituciones de enseñanza de nivel superior interesadas en introducir el nuevo estándar internacional de la OMG, el Interaction Flow Modeling Language (IFML) y el paradigma visual Model-Driven para el desarrollo de aplicaciones Web y Móviles dentro de su curriculum escolar utilizando WebRatio.
Se espera que el presente trabajo sirva de referencia para cualquier implementación de gestión de la calidad en el ámbito Universitario.

La adopción de un sistema de gestión de la calidad es una decisión estratégica para una organización que le ayuda a mejorar su desempeño global y proporcionar una base sólida para las iniciativas de desarrollo sostenible. Los beneficios potenciales para una organización que implementar un SGC basado en la Norma ISO 9001 son:

a) la capacidad para proporcionar regularmente productos y servicios que satisfagan los requisitos del cliente y los legales y reglamentarios aplicables.
b) facilitar oportunidades de aumentar la satisfacción del cliente.
c) abordar los riesgos y oportunidades asociadas con su contexto y objetivos.
d) la capacidad de demostrar la conformidad con requisitos del sistema de gestión de la calidad.

Los requisitos del SGC especificados en esta Norma son complementarios a los exigidos para los productos y servicios que brinda la organización. El cumplimiento permanente de los requisitos y la consideración constante de las necesidades y expectativas futuras, representa un desafío para las organizaciones en un entorno cada vez más dinámico y complejo. Para lograr estos objetivos, la organización debe considerar necesario adoptar diversas formas de mejora además de la corrección y la mejora continua, tales como el cambio abrupto, la innovación y la reorganización.

Promover la calidad en el proceso de enseñanza-aprendizaje en una cátedra es tratar de mejorar de forma continua sus prácticas, de tal manera que permitan [1]:

- Garantizar los resultados en cuanto a rendimiento académico y deserción de alumnos.
- Asegurar la trazabilidad de los procesos que se desarrollan.
- Posibilitar la mejora continua de la satisfacción de los diferentes actores intervinientes en la Educación.

La gestión de la calidad en la Educación debe ser sobre todo flexible y adaptada a las necesidades específicas de las cátedras, sustentándose en tres elementos:

1) Definición de los objetivos que se pretenden alcanzar en función de los criterios de satisfacción de las diferentes partes interesadas.
2) Proceso, que incluye la planificación y desarrollo de actividades tendientes a cumplir los objetivos.
3) Los indicadores que medirán el grado en el que se consiguen los objetivos.

La representación de los tres elementos básicos junto con los principios de gestión de calidad aplicables se muestra en la figura 1.

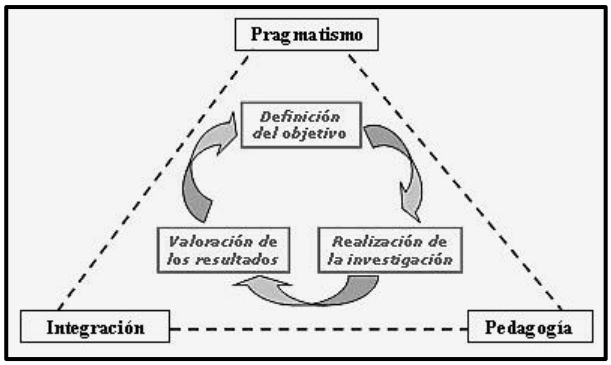


Figura 1- Fases principales de la de la gestión pedagógica de una cátedra.

Se puede definir una metodología para implementar el SGC en el ámbito de una cátedra, utilizado el esquema anterior de manera sistemática, dando lugar a un ciclo de mejora continua, tal y como se representa en la figura 2.

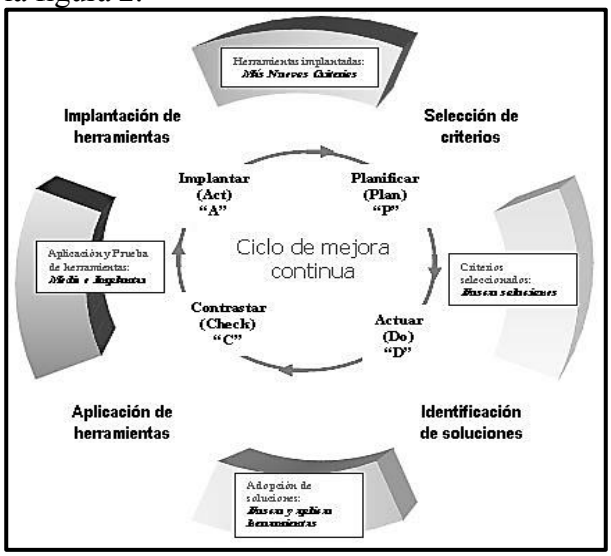


Figura 2- Ciclo de mejora continua.

Un Sistema de Gestión de Calidad se basa en la administración de muchos documentos, de manera estricta y minuciosa [2]. Por esta razón es fundamental el apoyo de la tecnología informática, es decir que no alcanza con un procesador de textos y un espacio de almacenamiento compartido en un servidor de archivos.

Todo documento relacionado con el sistema de gestión de la calidad debe ser desarrollado a través de un proceso perfectamente documentado. Además, una vez aprobado, debe estar disponible para todos quienes participan en el SGC. A su vez las sucesivas revisiones deben quedar claramente identificadas así como los cambios realizados. También, si los documentos obsoletos se mantienen en el SGC para poder ser consultados deben quedar claramente identificados como obsoletos para impedir que sean utilizados como actuales, por error. Por ejemplo, debe impedirse su modificación incluso a personal autorizado para generar documentos.

Esto es claramente el manejo de una base documental asociada a un proceso de decisión y de elaboración conjunta de los que típicamente se realizan con herramientas de trabajo en grupos.

En función a lo expresado, tomando como referencia las especificaciones y directrices investigadas para la aplicación de la familia de norma ISO 9.001:2.008 [4] se realizó el análisis y diseño de un SGC, utilizando la metodología WebML [5], logrando expresar los modelos en el standard IFML (*Interaction Flow Modeling Language*) [7] y se implementó un prototipo funcional de la herramienta para acompañar una implementación de calidad bajo los requisitos de la norma ISO 9.001:2.015 en la cátedra de programación y Cálculo Numérico. Dicho Sistema de Gestión está en permanente actualización para aproximarse tanto como se pueda a los estándares de calidad en el cumplimiento de los objetivos de la cátedra.

Para la especificación de requerimientos de calidad se utilizó el estándar ISO/IEC 25030:2007 [9] que es parte del standard internacional SQuaRE, que permite expresar los requisitos de usuarios como requisitos funcionales, lo que facilita la especificación del sistema de software.

El enfoque de la metodología WebML, y su evolución IFML [8], combina componentes tradicionales bien conocidos por los desarrolladores, como el diseño conceptual de datos usando el modelo Entidad Relación y la especificación de Casos de Usos usando UML, con nuevos conceptos y métodos para el diseño de hipertextos, que son fundamentales para el desarrollo web. No obstante, el valor del enfoque de la propuesta no está en los componentes individuales, sino en la definición de un marco sistemático para que las actividades de desarrollo de aplicaciones web puedan ser organizado de acuerdo a los principios fundamentales de la ingeniería de software, de modo que todas las tareas encuentren un soporte adecuado a partir de los conceptos, notaciones y técnicas propuestas por las metodologías. La característica distintiva de este marco de desarrollo es el énfasis en el modelo conceptual, el cual ha sido probado con éxito en muchos campos de la ingeniería del software, y en diseño de base de datos, donde el modelo Entidad Relación ofrece una notación de alto nivel e intuitiva para la comunicación de los requisitos de información entre los diseñadores y no técnicos, y es la base para la creación de esquemas de bases de datos de alta calidad.

En esencia, WebML e IFML consiste en gráficos simples y conceptuales para expresar el hipertexto de un conjunto de páginas y operaciones. Una página web es vista como una estructura compuesta por unidades de contenido y links.

IFML fue introducido en la versión 7.2 de WebRatio, donde se cambia los nombres de los elementos y la representación visual dentro de Web Project para cumplir con la notación IFML el cual es un estándar de la OMG, donde se propone que un sitio web conste de tres grandes partes conceptuales, la estructura, el hipertexto y la presentación.

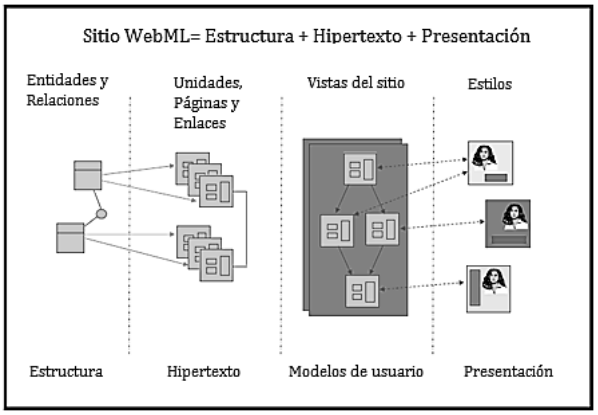


Figura 3- Conceptos principales de WebML e IFML.

La metodología está compuesta por la creación de los Modelos de Datos o Dominio, Hipertexto o iteración de flujo, Presentación y Personalización, expresando los modelos en el standard IFML. En este trabajo se desarrollaron todos los modelos de la metodología y para la implementación se utilizó la herramienta WebRatio Platform University Edition.

En síntesis, los resultados obtenidos a la fecha son los siguientes:

- El Análisis y diseño de un Sistema de Gestión de Calidad para la cátedra con modelos expresados en el estándar IFML, junto a la especificación de requerimientos, aplicados a los procesos específicos relacionados a la enseñanza y aprendizaje en Programación y Cálculo Numérico.
- Un prototipo funcional de un Sistema de Gestión de Calidad, en ambiente Web con HTML y JSP.

2. LINEAS DE INVESTIGACION y DESARROLLO

Los principales ejes temáticos que se están investigando son los siguientes:

- ✓ Tecnología Informática aplicada en Educación.

- ✓ Gestión de Calidad aplicada a la educación superior.
- ✓ Herramientas informáticas para la implementación de un SGC ISO 9.001.
- ✓ Aplicación de la metodología WebML con IFML para el diseño de una herramienta que apoye a la implementación de SGC ISO 9.001.

3. RESULTADOS OBTENIDOS/ESPERADOS

En el presente trabajo se ha abordado la gestión de los sistemas de calidad desde la perspectiva de las actividades de gestión de la enseñanza que realiza una cátedra, de una carrera informática universitaria.

Consideramos que, además de todas las herramientas disponibles, la aplicación de la familia de normas ISO 9001:2015 junto con el apoyo de un modelo adecuado de software, con la especificación de requerimientos adecuado, se constituye en una estrategia importante para alcanzar la satisfacción de los estudiantes, y de cualquier interesado de la sociedad.

Actualmente se está considerando la posibilidad de alcanzar la certificación del proceso metodológico llevado adelante por la cátedra, por algunas de las entidades certificantes, tales como Bureau Veritas Quality International e IRAM, las cuales son las más conocidas en nuestro país.

También se puso a disposición el prototipo del SGC a otras cátedras de la Universidad que estén interesadas en la calidad y la Familia de Normas ISO 9001.

4. FORMACION DE RECURSOS HUMANOS

La estructura del equipo de investigación es de 5 (cinco) miembros incluidos el Director del proyecto.

Un miembro obtuvo la Especialidad en Ingeniería de Software de la Universidad Nacional de La Plata, con el trabajo “Sistema de Gestión de Calidad bajo Normas ISO”. Otro alcanzo el título de Licenciado en Análisis de Sistemas, otorgado por la Universidad Nacional de Salta, con la tesis “Análisis y Diseño de un Sistema de Gestión de Calidad basado en la Norma ISO 9.001”.

Además, se cuenta con la participación de un miembro que realiza el trabajo final de la Especialidad en Psicopedagogía Institucional en la Universidad Nacional de Salta, de igual forma, se cuenta con un integrante realizando el trabajo final de la Especialidad en Tecnología Multimedia para Desarrollos Educativos, en la Universidad Nacional de Córdoba.

Por último, se continúa con la dirección de tesinas de grado de alumnos de la carrera Licenciatura en Análisis de Sistemas de la Universidad Nacional de Salta.

5. BIBLIOGRAFIA

[1] Castellón Murcia, Roberto Aristides, "Aplicación de la norma ISO 9001 en el proceso enseñanza aprendizaje en la educación superior", UFG Editores, Agosto 2009.

[2] Arias Figueroa, Daniel y otros, " Normas ISO y su Aplicación en Centros Educativos”, XIII Workshop de Investigadores en Ciencias de la Computación, 2011.

[3] Norma ISO 9001 (2.015) Elaborada por el Comité Técnico ISO/TC176 de ISO.

[4] Guía de Interpretación de la IRAM-ISO 9.001:2008 para la educación. IRAM 30.000:2001.

[5] WML, sitio oficial de la Metodología http://www.webml.org.

[6] WebRatio, sitio oficial de la herramienta http://www.webratio.com.

[7] IFML sitio oficial del lenguaje http://www.ifml.org

[8] Marco Brambilla; Piero Fraternali, “Migrating from WebML to IFML”. The OMG Press, 2013.

[9] Norma ISO/IEC 25030 (2007). Software Engineering–Software Product Quality Requirements and Evaluation (SQuaRE), 2007.

Evaluación cuantitativa y sistemática de Software Educativo Libre

Estela Fritz[1]; María Eva Ascheri[1]; Alejandra Zangara[2]

[1]Departamento de Matemática
Universidad Nacional de La Pampa
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina
Tel.: +54-2954-245220 – Int. 7125
[fritzem, mavacheri]@exactas.unlpam.edu.ar

[2] Facultad de Informática
Universidad Nacional de La Plata
50 y 120 – (1900) La Plata – Argentina
Tel.: +54-221-4277270 / 4277271
alejandra.zangara@gmail.com

Resumen

El presente trabajo se enmarca en el Proyecto de Investigación: *"Propuesta de Clasificación de software libre utilizado en la enseñanza de la programación"*, que se desarrolla en el ámbito de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa.

El principal objetivo de este trabajo es construir un modelo para evaluar cuantitativamente software educativo y poder incluir en ese modelo las características deseables de un software libre. Por otra parte, se pretende una evaluación desde el punto de vista pedagógico más que desde el punto de vista técnico, o desde la Ingeniería del software. El modelo presentado en este trabajo está basado en el método LSP.

Palabras clave: Método LSP, evaluación de software, software educativo, software libre, métodos cuantitativos.

Contexto

Por Resolución N° 160/18 del Consejo Directivo de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa se acredita el Proyecto de Investigación: *"Propuesta de Clasificación de software libre utilizado en la enseñanza de la programación"*. La Directora es la Mag. María Eva Ascheri de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa (UNLPam) y la Co-Directora es la Dra. Alejandra Zangara de la Facultad de Informática de la Universidad Nacional de La Plata (UNLP). El mencionado proyecto incluye a la Profesora Estela Marisa Fritz como investigadora. Constituye su Trabajo Final para alcanzar el grado de Especialista en Tecnología Informática Aplicada en Educación. Además, será la base para desarrollar su futura tesis de maestría a llevarse a cabo en la UNLP.

1. Introducción

La rápida evolución de las Tecnologías de la Información y la Comunicación (TICs) ha creado un escenario para la enseñanza y el aprendizaje, donde ya es impensado trabajar sin los recursos que brinda el uso del software.

Los recursos de software son abundantes y variados por lo que su selección se vuelve un proceso complejo, si no se tienen en claro, primeramente, los objetivos de dicho proceso.

Por otro lado, el creciente número de repositorios digitales ha posibilitado que los docentes puedan compartir sus materiales educativos, sin límite de fronteras. La mayor parte de estos materiales se comparte bajo la filosofía del software libre.

Por todo lo expuesto, frente a la gran cantidad de material disponible, es necesario que el docente pueda seleccionarlo, según los objetivos que se plantea en el proceso de enseñanza- aprendizaje, los destinatarios y sus características particulares, el contexto en el cual enseña, las actividades que planifica, la evaluación y la metodología subyacente en el software, entre otros aspectos relevantes.

Los criterios para la selección del software deben ser establecidos previamente. Por ello este trabajo propone la construcción de un modelo cuantitativo, elaborado a través de un método sistemático: *el método LSP*.

En este modelo, el docente puede ponderar cada una de las características deseables para el software, de acuerdo a la importancia o relevancia que posea dicha característica en el sistema a seleccionar.

Existen en la bibliografía numerosos modelos para evaluar software educativo. Algunos con énfasis en el proceso de diseño y desarrollo tal como se propone en [1]. Otros enfoques se rigen por los principios propios de la Ingeniería del Software [2]. Se busca en otros casos la calidad del software según los estándares internacionales [3] y [4]. Muchas publicaciones presentan un modelo basado en listas de criterios que se formulan en forma de preguntas. Son ampliamente difundidas en este sentido las listas de control (*check-lists*), aunque contrariamente a lo que propone el método del presente trabajo, la técnica de las listas de control no permite ponderar por relevancia los diferentes ítems. Squires y Mc Dougall [5], proponen un modelo basado en interacciones entre los actores intervinientes desde la etapa de diseño del software hasta su utilización: interacciones entre diseñadores, estudiantes y profesores tomados de dos en dos; algunas de esas interacciones tienen sentido bidireccional. Este modelo propone un enfoque cualitativo y aporta un modo de pensar acerca de la selección del software.

En cuanto a los criterios a tener en cuenta para la evaluación y posterior selección del software, Insa, D. y Morata, R. [6] proponen tres categorías: 1) criterios técnicos que hacen referencia al software en lo concerniente a recursos audiovisuales, a la interacción y la documentación, 2) criterios didácticos, y 3) criterios de integración curricular.

Los enfoques cuantitativos buscan, fundamentalmente, objetividad y se basan en criterios preestablecidos.

El objetivo principal de esta línea de investigación y del presente trabajo, puede resumirse del siguiente modo:

- Proponer un modelo para evaluar cuantitativamente software educativo libre.

Por ello, se hará una referencia breve a los diferentes conceptos que constituyen los ejes de este objetivo: software libre- modelo cuantitativo.

1.1 Software Libre

Cualquier trabajo relacionado con el uso del software libre plantea inevitablemente la cuestión de establecer decisiones relacionadas con dicha utilización.

En este trabajo se plantea un modelo para evaluar software educativo que contempla varios criterios e incluye entre los mismos el de pertenecer a la categoría de software libre. Para precisar el concepto de software libre se exponen algunas consideraciones realizadas por el creador del proyecto GNU [7].

Un software puede considerarse libre si el usuario posee con relación a aquél las siguientes libertades.

Libertad cero: libertad para ejecutar el software con cualquier propósito. *Libertad uno*: libertad que posee el usuario de modificar el software para ajustarlo a sus necesidades (esto podría incluir arreglar “bugs”) o para añadirle nuevas características o para migrarlo a un sistema informático diferente. *Libertad*

dos: libertad de un usuario de distribuir copias del programa. Es una forma de compartir conocimiento. Pueden compartirse las modificaciones realizadas al software. *Libertad tres*: es la libertad o posibilidad que tiene un usuario de publicar una versión mejorada de un software determinado. En la medida en que muchos usuarios realizan contribuciones, se crea una comunidad de software. Esta libertad refuerza la idea de trabajar juntos por el avance del conocimiento. Así las libertades *uno, dos y tres* son las que diferencian el software libre del software típico. Podríamos resumirlo del siguiente modo: *libertad uno* es ayudarse a uno mismo, *libertad dos* es ayudar a otros (distribuir copias) y *libertad tres* es la libertad de construir la comunidad de usuarios. Es necesario aclarar, por último, que software libre no significa que no sea comercial. La libertad tres, fundamentalmente, es lo que justifica plenamente el uso del software libre en la enseñanza ya que se relaciona con la construcción del conocimiento.

1.2 Modelo Cuantitativo

Como se expresó anteriormente, los modelos cuantitativos buscan objetividad en la evaluación. Para ello utilizan un conjunto de criterios preestablecidos relevantes y coherentes en contraste con evaluaciones subjetivas basadas en puntos de vista particulares (el del docente, el del estudiante, el del experto). Estos criterios preestablecidos valoran diferentes componentes del software, características deseables relacionadas con aspectos pedagógicos y didácticos: adecuación a los objetivos de aprendizaje, enfoque pedagógico (constructivista o cognitivista), orientación hacia los alumnos, grado de intervención docente, inclusión de módulos de evaluación y seguimiento, entre otros.

También es posible incluir en el modelo cuantitativo, variables asociadas con las cuatro libertades de un software para ser considerado software libre. Asimismo, sería posible incluir variables (o criterios) relacionados con aspectos técnicos de diseño de software: accesibilidad, adaptabilidad, interoperabilidad, reusabilidad, por nombrar algunos.

Este modelo supone una evaluación previa del software y los criterios seleccionados no están relacionados con la utilización de éste por parte de los estudiantes, ya que el proceso de evaluación tiene como finalidad la selección del software más adecuado y que reúna las características buscadas o se aproxime a ello, lo más posible. Sin embargo, el modelo es muy flexible y podrían incluirse variables (criterios) relacionadas con el uso del software una vez que el estudiante interactúa con el mismo.

1.2.1 Construcción del modelo

Se describe a continuación, en forma muy general, el método en el que se basa el modelo cuantitativo propuesto en este trabajo: El método *LSP (Logic Scoring Preferences)* [8].

LSP es un método cuantitativo basado en el empleo de la lógica continua que posibilita la creación de funciones complejas que intervienen en la evaluación, comparación y posterior selección de sistemas de software de propósito general [9].

Una primera etapa en la construcción de esta función compleja consiste en establecer las características deseables para el software a evaluar. Esas características se transforman según LSP en *variables de performance* (X_1, …, X_n). En el caso del software educativo, algunas variables de performance que se pueden describir son las siguientes: en qué medida es tenido en cuenta el aprendizaje individual, cuantificar fases de aprendizaje inmediato y de largo alcance, gradualidad del software para el aprendizaje (¿Es posible identificar esa gradualidad?), en qué medida el software está basado en el conocimiento de cómo aprende el ser humano, en qué grado contribuye el software a los principios de contigüidad, repetición y refuerzo, qué contenidos de la memoria o capacidades aprendidas el software contribuye a recuperar, qué categorías (en cuanto a resultados del aprendizaje) el software tiende a desarrollar, cuál es el rol del docente en el proceso, en qué

medida el sistema guía al sujeto hacia el autoaprendizaje. Esto por mencionar algunas.

Para cada variable de performance es necesario definir un rango aceptable de valores. Los valores de dichas variables, se transforman, por la aplicación de funciones llamadas *criterios elementales* G_i , $1 \leq i \leq n$, en *preferencias elementales* E_i , $1 \leq i \leq n$. Esos criterios elementales transforman el valor de una variable de performance en un valor perteneciente al intervalo [0,100] o [0,1], es decir, $0 \leq E_i \leq 1$ ó $0 \leq E_i \leq 100\%$ y expresan el grado de cumplimiento con un requisito del sistema que está siendo evaluado. Estas preferencias elementales van siendo agregadas mediante operadores de la lógica continua en estructuras de agregación más complejas que permiten obtener un único valor final denominado *Preferencia global* ***E.*** Existe la posibilidad de ponderar las preferencias elementales dentro de la estructura de agregación, mediante pesos W_i , que son coeficientes que multiplican a las preferencias elementales y les otorgan mayor o menor relevancia en la ***E***. La preferencia global se interpreta como el grado global de satisfacción de todos los requerimientos especificados.

2. Líneas de Investigación y Desarrollo

El proyecto de Investigación antes citado, aborda los siguientes aspectos relacionados a:

- Relevamiento de software libre utilizado en la enseñanza de la programación.
- Descripción de un método cuantitativo para la evaluación y selección del software libre basándose en criterios, fundamentalmente pedagógicos.
- Evaluación del software relevado con el fin de clasificarlo según los criterios preestablecidos.

El presente trabajo corresponde al segundo de los aspectos mencionados.

3. Resultados Obtenidos/Esperados

Después de considerar los tópicos explicados anteriormente, como resultado, puede decirse que el método *LSP* propuesto constituye un muy buen método para construir a partir de él, un modelo cuantitativo de evaluación que incrementa la eficiencia de ésta y la posterior selección de sistemas de software en general y software educativo en particular con filosofía GNU. Representa un método más ventajoso cuando se lo compara con otros métodos cuantitativos [10]. *LSP* tiene principalmente las siguientes ventajas relativas a los métodos cuantitativos para evaluación de sistemas:
(1) La especificación de requerimientos es sistemática, flexible y completa.
(2) El proceso de evaluación es sistemático y racional, el cual refleja explícita y cuantitativamente el nivel global de satisfacción de los requerimientos del usuario.
(3) Es un modelo que permite objetividad en la evaluación.

El método *LSP* puede ser usado para la evaluación y comparación de una amplia variedad de sistemas complejos [11]. El principal poder de la estrategia de *LSP* es la capacidad de construir un modelo versátil de un conjunto de preferencias. Mediante la combinación apropiada de operadores (lógicos) de agregación con el conjunto de preferencias, es posible derivar criterios complejos que tienen poder expresivo y flexibilidad.

El software educativo debe satisfacer una variedad de requerimientos. El nivel global de satisfacción de esos requerimientos (**E**) es usado para su evaluación y comparación.

Por otro lado, se espera que esta investigación brinde a los docentes herramientas cuantitativas para la evaluación y selección de software, en función de los objetivos que desean alcanzar.

4. Formación de Recursos Humanos

En el marco del presente proyecto, como ya se mencionó anteriormente, se espera que la Prof. Fritz alcance el grado de Especialista en Tecnología Informática Aplicada en Educación y que, en un futuro, pueda desarrollar su tesis de maestría en la UNLP.

Actualmente, la Prof. Fritz ha completado la etapa de búsqueda y análisis de métodos de evaluación del software, principalmente métodos cuantitativos, tema en el que no era experta. Durante 2020, realizará todos los aportes relacionados con la definición de criterios para evaluar software educativo libre para la enseñanza de la programación, principalmente desde el punto de vista pedagógico, y tomando como base alguna teoría del aprendizaje en particular. A partir de ello, se espera construir el modelo de Preferencia Global para la posterior clasificación y selección de software. Esto constituye un insumo necesario para los nuevos avances en el proyecto.

5. Bibliografía

[1] **Zaragoza, J., Casado, A.** (1992) "Aspectos técnicos y pedagógicos del ordenador en la escuela". Editorial Bruno, Madrid, España.

[2] **Galvis, A**. (1995). "Fundamentos de tecnología educativa". Editorial de la Universidad Estatal a Distancia. San José, Costa Rica.

[3] **International Standard. ISO/IEC 9126** (1991) Information Technology – Software Product Evaluation – Quality characteristics and guidelines for their use.

[4] **International Standard ISO/IEC 9001** (1991) Quality Systems – Model for quality assurance in design/development production, installation and servicing.

[5] **Squires, D., McDougall, A.** (1997) "Cómo elegir y utilizar software educativo". Traducción Pablo Manzano. Ediciones Morata y Fundación Pandeia, Madrid, 1° edición.

[6] **Insa, D., Morata, R.** (1998) Multimedia e Internet: las nuevas tecnologías aplicadas en la educación.

[7] **Stallman, R.** (2004) "Software libre para una Sociedad libre". Traficantes de sueños, Madrid.

[8] **Dujmovic, J.J. and Elnicki, R.** (1982) "A DMS Cost/Benefit Decision Model: Mathematical Models for Data Management System Evaluation, Comparison, and Selection". National Bureau of Standards, Washington. DC., No NBS-GCR, NTIS No PB82- 170150 (155 pages).

[9] **Dujmovic, J.J.** (1996) "A Method for Evaluation and Selection of Complex Hardware and SoftwareSystems". Proceedings of the 22nd International Conference for the Resource Management and Performance Evaluation of Enterprise CS, CMG 96, 1, 368-378.

[10] **Daso, A**. **et al** (2013) "Desarrollo de Modelos de Evaluación Usando Operadores de una Lógica Continua". XV WICC (Workshop de Investigadores en Ciencias de la Computación), págs. 420-424.

[11] **Dujmović, J. J., Hajime Nagashima**. (2006) "LSP Method and Its Use for Evaluation of Java IDEs." International Journal of Approximate Reasoning 41.1: 3–22. Web.

Diseño de un sistema de medición de desempeño para Moodle en Educación Superior

Salinas Sergio Ariel[1], Tagarelli Sandra[1], Bianchini Germán[2], Caymes-Scutari Paola[2,3], Ontiveros Patricia[4], Rotella Carina[4], Chirino Pamela[2], Galdamez Mariela[2]

[1]Laboratorio de Analítica de Datos, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza
[2]Laboratorio de Investigación en Cómputo Paralelo/Distribuido
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579
[3]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
[4]Laboratorio de Gobierno Electrónico, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza

s4salinas@gmail.com, stagarelli@gmail.com, gbianchini@frm.utn.edu.ar, pcaymesscutari@frm.utn.edu.ar, pontiveros@frm.utn.edu.ar, carinarotella@gmail.com, pamelaachirino@gmail.com, mariela.galdamez.16@gmail.com

RESUMEN

Las universidades locales requieren de un sistema de medición de desempeño para el Sistema de Gestión de Aprendizaje (SGA) denominado Moodle.

Para implementar Moodle las universidades invierten recursos físicos, tecnológicos y tiempo. Este proceso de implementación requiere de un plan estratégico que defina un conjunto de acciones a ejecutar con sus respectivos objetivos a alcanzar. Para evaluar la eficiencia de la ejecución de un plan estratégico se necesita de un sistema de medición de desempeño.

El principal componente de este sistema es un conjunto limitado de indicadores que sintetizan los efectos de las acciones del plan estratégico. El número de indicadores es limitado para facilitar su uso y reducir la complejidad de análisis por parte de los directivos de la institución.

Existen distintas metodologías para diseñar un sistema de medición de desempeño. Sin embargo, estas metodologías sólo proveen lineamientos generales a tener en cuenta a la hora del diseño.

El problema que aborda este trabajo es la identificación de los indicadores apropiados para medir el desempeño de Moodle en una universidad local.

Palabras clave: Moodle, Sistema de medición de desempeño, Indicadores de desempeño, Sistemas de gestión de aprendizaje.

CONTEXTO

La iniciativa de investigación de esta temática surge en el marco de una Maestría en Inteligencia de Negocios y Analítica. Esta maestría es dictada en forma conjunta entre la Universidad HNU de Alemania y la UTN-FRM.

Ambas universidades han invertido diversos recursos en la implementación de una infraestructura para el funcionamiento de Moodle. A partir de esta inversión surge la pregunta: ¿Cuál es el efecto que tiene en la comunidad educativa tal inversión? En este sentido los directivos de ambas universidades carecen de herramientas para medir los resultados de tal inversión.

Actualmente, se trabaja con referentes de ambas universidades en el desarrollo de esta línea de investigación.

En enero de 2020 esta iniciativa de investigación se integró al proyecto denominado "Formación de docentes y alumnos de grado como Investigadores Científicos Iniciales en las áreas de Informática y Ciencias de la Computación".

El proyecto ha sido acreditado, homologado y se encuentra financiado por la UTN bajo el código PID TEUTIME0007658TC.

El mismo se ejecuta en el marco del LICPaD y cuenta con la colaboración de docentes de los grupos ADA-Lab (Laboratorio de Analítica de Datos) y GE-Lab (Laboratorio de Gobierno Electrónico). Ambos grupos pertenecen al Departamento de Ingeniería en Sistemas de Información de la UTN-FRM.

1. INTRODUCCIÓN

En las últimas décadas los avances tecnológicos han cambiado dramáticamente gran parte de las actividades humanas y la educación no ha sido una excepción.

Las instituciones de educación superior utilizan un ecosistema tecnológico compuesto por diferentes plataformas interoperativas.

Un componente clave en el ecosistema tecnológico educativo es el Sistema de Getión de Aprendizaje (SGA). Este sistema es una aplicación para crear, evaluar, gestionar, almacenar, distribuir actividades, entre otras funciones que forman parte del proceso de aprendizaje de los estudiantes [1].

La implementación y gestión de un SGA en una institución requiere de recursos adicionales y del compromiso de la comunidad educativa.

Usualmente, los recursos de una institución son limitados y están destinados a procesos predefinidos donde un SGA no está contemplado. Esto significa que es necesario planificar, adquirir y organizar los recursos necesarios [2].

Además de obtener recursos adicionales a partir del compromiso de directivos, es necesario diseñar estrategias para adoptar la plataforma en la institución. Para ello se deben definir nuevos roles, responsabilidades, objetivos y tareas a cumplir. También, es necesario desarrollar programas de capacitación para fortalecer la colaboración y cooperación entre usuarios [3].

La gestión de recursos necesarios para implementar el sistema y para una exitosa adopción del sistema requieren de un plan estratégico.

Los directivos de la institución son responsables de la elaboración del plan estratégico basado en acuerdos y negociaciones entre distintos actores involucrados. El plan define las fases a seguir, los responsables de cada fase, objetivos, resultados esperados y su correspondiente planificación [4].

Para evaluar los efectos de las acciones del plan los directivos requieren un sistema para medir los efectos de cada acción.
La medición de desempeño es el proceso de cuantificación de una acción en términos de efectividad y eficiencia. Efectividad se refiere al grado en que un requerimiento se cumple mientras que eficiencia es el *ratio* que expresa como los recursos son utilizados económicamente cuando se alcanza un determinado grado de efectividad [5].
Existen distintas metodologías para diseñar un sistema de medición de desempeño [6] a saber, la pirámide *Du Pont* y *ROI* [7], *Balanced Scorecards* [8], *los estándares ISO* [9] y *dashboards* [10].
Estas metodologías son útiles y valiosas, sin embargo, su simplicidad implica un desafío importante al momento de utilizarlas. Esto se debe a que las mismas establecen recomendaciones generales respecto a qué áreas deberían ser medidas. Además, proveen lineamientos limitados respecto a la identificación de las medidas apropiadas a tener en cuenta [11].
Las limitaciones de las metodologías están determinadas por diferentes factores. En primer lugar, las organizaciones son heterogéneas y requieren medidas acordes a la naturaleza de sus actividades. En segundo lugar, el número de procesos y la complejidad de los mismos pueden dar lugar a un número significativamente elevado de medidas.
En este contexto se define el problema que aborda la línea de investigación considerando dos aspectos. En primer lugar, los directivos requieren de un número limitado de indicadores los cuales deben ser seleccionados cuidadosamente. Una elección errónea puede conducir a errores en la percepción de los efectos de las acciones que se requieren medir.
En segundo lugar, es necesario definir de qué manera identificar estos indicadores claves para los directivos.
La principal pregunta que se busca responder en la línea de investigación es ¿cuáles son los principales indicadores de desempeño que deben considerarse para Moodle en un sistema de medición de desempeño?
La motivación para abordar este problema está definida por la necesidad de los directivos de las universidades de HNU y UTN-FRM de disponer de un sistema de medición de desempeño para Moodle.
En este sentido, la hipótesis de trabajo sostiene que es posible mejorar las tareas de gestión de Moodle por parte de los directivos mediante un sistema de medición de desempeño.
Considerando lo expuesto hasta el momento se define el siguiente objetivo principal de la línea de investigación: "identificar los indicadores claves para el sistema de medición de desempeño para Moodle".
Actualmente, la línea de investigación se encuentra en desarrollo y cuenta con la participación e interés de los directivos de las universidades HNU y UTN-FRM.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

En esta iniciativa existe una línea de investigación principal sustentada por tres ejes complementarios.
La principal línea de investigación tiene como objetivo diseñar un sistema de medición de desempeño para Moodle. El sistema contribuirá a analizar los resultados de las acciones estratégicas definidas por los directivos de la institución.
Para diseñar el sistema es necesario conocer sobre los Sistemas de Gestión de Aprendizaje, las metodologías para

diseñar un sistema de desempeño y los requerimientos de información propios de una institución de educación superior.

El primer eje complementario de investigación tiene como objetivo conocer el funcionamiento y servicios provistos por los Sistemas de Gestión de Aprendizaje, entre ellos Blackboard, Canvas y Sakai por mencionar algunos [12]. Esta información permite identificar limitaciones en las soluciones que pueden proponerse a una institución respecto de los requerimientos identificados por la misma.

El segundo eje tiene como objetivo identificar las metodologías existentes para el diseño de un sistema de desempeño. En el diseño del sistema se espera combinar los aspectos más distintivos de cada metodología mediante la complementación de las estrategias propuesta por cada metodología.

El tercer eje tiene como objetivo analizar los requerimientos de la UTN-FRM. En este sentido, conocer las potencialidades de los SGA y las sugerencias de las metodologías de desarrollo de sistemas de medición desempeñan un papel importante.

El primer y segundo eje complementario de investigación contribuirán a disponer de herramientas apropiadas para el análisis de requerimientos de la UTN-FRM. Finalmente, se diseñará el sistema de medición considerando los ejes mencionados.

3. RESULTADOS ESPERADOS

El principal resultado esperado es el diseño de un sistema de medición de Moodle para la UTN-FRM y de aplicación en HNU. Se espera generar resultados indirectos consecuencia de las etapas previas al diseño del sistema. Estos resultados están relacionados con los sistemas SGA, las metodologías de diseño de sistemas de medición y aspectos humanos e institucionales.

El análisis de Moodle permitirá comprender las funcionalidades disponibles en la plataforma accesibles en ambas universidades.

Además de conocer los SGAs, comprender las metodologías de diseño de sistemas de medición de desempeño es clave para el diseño del sistema. Existe una amplia bibliografía al respecto y se espera integrar los aspectos más relevantes de cada metodología.

Considerando que la iniciativa es de interés para ambas universidades, se espera generar lazos de cooperación entre ambas instituciones.

Debido a que la iniciativa surge de docentes pertenecientes al grupo de reciente conformación ADA-Lab, se busca afianzar la presencia del grupo en la UTN-FRM y atraer la atención de alumnos para participar en distintas actividades de investigación.

Para concluir esta sección, cabe mencionar que es importante, a través de esta iniciativa de investigación, publicar todo resultado de interés en congresos pertinentes.

4. FORMACIÓN DE RECURSOS HUMANOS

Durante el proceso de investigación se espera fortalecer las habilidades de investigación de docentes interesados en la investigación.

Además, se espera estimular la incorporación de estudiantes y graduados para que accedan a una primera experiencia en las tareas de investigación.

Es importante generar lazos entre los miembros de la iniciativa y las instituciones involucradas para futuras iniciativas.

5. BIBLIOGRAFÍA

[1] Gorshenin, A. (Nov. 2018). "Toward Modern Educational IT-ecosystems: from Learning Management Systems to Digital Platforms". In: 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), pp. 1–5. DOI:10.1109/ICUMT.2018.

[2] Hustad, Eli and Aurélie Aurilla Bechina Arntzen (2013). "Facilitating teaching and learning capabilities in social learning management systems: Challenges, issues, and implications for design". In: Journal of Integrated Design and Process Science 17.1, pp. 17–35.

[3] Sife, Alfred, Edda Lwoga and Camilius Sanga (Jan. 2007). "New technologies for teaching and learning: Challenges for higher learning institutions in developing countries". In: The International Journal of Education and Development using Information and Communication Technology 3.

[4] Zuvic-Butorac, Marta et al. (2011). "Establishing an Institutional Framework for an E-learning Implementation: Experiences from the University of Rijeka, Croatia". In: Journal of Information Technology Education : Innovations in Practice 10, pp. 43–56.

[5] Neely, Andy (2005). "The evolution of performance measurement research: developments in the last decade and a research agenda for the next". In: International Journal of Operations & Production Management 25.12, pp. 1264–1277.

[6] Kennerley, Mike and Andy Neely (2002). "Performance measurement frameworks: a review". In: Business performance measurement: Theory and practice, pp. 145–155.

[7] Soliman, Mark T. (2008). "The Use of Dupont Analysis by Market Participants". In: The Accounting Review 83.3, pp. 823–853. URL : https://academic.microsoft.com/paper/2121142168.

[8] Kaplan, Robert S and David P Norton (1998). "Putting the Balanced Scorecard to Work". In: Performance measurement, management, and appraisal sourcebook 66.

[9] Van den Berghe, Wouter (1997). "Application of ISO 9000 Standards to Education and Training. Interpretation and Guidelines in a European Perspective.

[10] Vazquez-Ingelmo, A., F. J. Garcia-Penalvo and R. Theron (2019). "Information Dashboards and Tailoring Capabilities - A Systematic Literature Review". In: IEEE Access 7, pp. 109673–109688. ISSN : 2169-3536.

[11] Neely, Andy, John Mills et al. (2000). "Performance measurement system design: developing and testing a process-based approach". In: International journal of operations & production management 20.10, pp. 1119–1145.

[12] Itmazi, Jamil and Miguel Gea (June 2005). "Survey: Comparison and evaluation studies of Learning Content Management Systems".

LOS ESTILOS DE APRENDIZAJE EN LA CONSTRUCCION DE TUTORES INTELIGENTES. UNA REVISION BIBLIOGRAFICA

Jaquelina E. Escalante[1,2,4], Sonia I. Mariño[3], Susana T. Marchisio[4]

[1]Fac. de Humanidades. Las Heras 727. Rcia. Chaco
Universidad Nacional del Nordeste.
[2]Universidad de la Cuenca del Plata. Lavalle 50. Corrientes
[3]Dpto. de Informática. Fac. de Cs. Exactas y Naturales y Agrimensura.
9 de Julio 1449. CP: 3400. Corrientes. Argentina.
[4]Fac. de Cs. Exactas, Ingeniería y Agrimensura. Av. C. Pellegrini 250. Rosario. Universidad Nacional de Rosario.

escalantejaquelina_cen@ucp.edu.ar; simarinio@yahoo.com; smarch@fceia.unr.edu.ar

RESUMEN

En este trabajo se introducen conceptos de los campos en los que incursiona esta investigación, como son: los estilos de aprendizaje en estudiantes y los Tutores Inteligentes (TI). Se presenta una revisión bibliográfica de proyectos o estudios que consideran estilos de aprendizaje en la enseñanza para la construcción de Tutores Inteligentes. Así mismo se detallan las características consideradas esenciales en un prototipo de tutor inteligente. Este sistema se constituye en una herramienta complementaria del proceso de enseñanza aprendizaje en estudiantes de Ingeniería para mejorar la calidad del aprendizaje en el área.

Palabras clave: tutores inteligentes, aprendizaje, estilos de aprendizaje, TIC, Ingeniería

CONTEXTO

El proyecto "*Sistema Tutorial Inteligente orientado a la enseñanza de la formulación de proyectos de investigación en posgrados de Ingeniería*" [1] se desarrolla en el marco del Doctorado en Ingeniería de la Facultad de Ciencias Exactas, Ingeniería y Agrimensura de la Universidad Nacional de Rosario.

1. INTRODUCCIÓN

La Educación, es una de las actividades humanas que más transformaciones ha tenido a lo largo de la historia y los cambios tecnológicos, han contribuido en gran parte a esto. La sociedad del conocimiento requiere de medios tecnológicos para acceder a la información. Desde ese punto de vista y más aún desde la práctica docente se considera a los Estilos de Aprendizaje como una estrategia de aprendizaje. La especialización del recurso humano es un elemento de vital importancia, requiriéndose de las Universidades dar respuesta a demandas de educación. [1]

Se observa en la bibliografía una valoración muy positiva, en una relación docente tutor – alumno a los fines de garantizar la realización de la tutoría personalizada que el aprendizaje de este tipo de saberes requiere [2].

En la enseñanza en Ingeniería, se demanda necesidades y exigencias para lograr que el proceso de formación responda a los requerimientos del contexto. Lo expuesto implica una organización del proceso docente educativo centrado en el estudiante, desarrollado de manera interactiva y colaborativa y que le permita adquirir un aprendizaje para toda la vida. [3]

Para comprender el sentido de esa práctica, es necesario analizar el estilo de aprendizaje del estudiante de Ingeniería, es decir, cómo recibe y procesa la información. Esto constituye una

indagación muy útil para el trabajo cotidiano de los docentes [1], [4], [5].
El profesor debe actuar como facilitador de los procesos para aprender a aprender y aprender a emprender [3], [6], [7], [8], [9].
Por ello, los estilos de aprendizaje se relacionan directamente con la manera en que los estudiantes aprenden, los docentes enseñan y la forma en la que ambos interactúan en la relación enseñanza aprendizaje [6], [10], [11].
Atendiendo al hecho de que se hace indispensable en la época actual que las TIC se apliquen dentro del proceso de enseñanza – aprendizaje, se plantea diseñar herramientas que aporten soluciones pedagógicas a las necesidades expuestas. Aunando en investigación y desarrollo en el área de los Sistemas Tutoriales Inteligentes (en adelante STI) resulta factible su utilización como herramienta pedagógica de enseñanza en Ingeniería.
En la literatura se localizan una diversidad de definiciones en torno al termino Sistema Tutor Inteligente (STI). Una conceptualización puede exponerse considerando que "es un sistema de software que utiliza técnicas de IA para representar el conocimiento e interactúa con los estudiantes para enseñárselo" [12], [13], [14], [15].
Estos STI han evolucionado desde sus inicios y son un ejemplo de sistemas educativos adaptables.
Básicamente un STI consta de los siguientes componentes: un módulo del dominio del conocimiento: la base de conocimiento del tutor, esto es, todos los conceptos y relaciones necesarias en el proceso de enseñanza-aprendizaje, un módulo del estudiante (que es capaz de definir el conocimiento del estudiante en cada punto durante la sesión de trabajo, un módulo del tutor: que genera las interacciones de aprendizaje basadas en las discrepancias entre el especialista y el estudiante y finalmente la interface con el usuario: que permite la interacción del estudiante con un STI de una manera eficiente (conocimiento sobre cómo presentar los contenidos).

Esta propuesta de I+D considera en el diseño e implementación de un STI, más precisamente en el diseño del módulo estudiante, los estilos de aprendizaje del alumno usuario, para posteriormente informar al módulo tutor a cerca del mismo, con el objetivo de que el módulo defina las estrategias más adecuadas a cada perfil. Es decir, se centra en el estudiante**,** en el análisis de cómo adquiere, procesa e incorpora el conocimiento que imparten los docentes.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Con el objetivo de *caracterizar* los *estilos de aprendizaje en estudiantes de Ingeniería*, los aspectos importantes en los que se enfoca esta línea de la tesis es relevar proyectos o estudios que consideren los estilos de aprendizaje del estudiante, con la finalidad de:

- Identificar herramientas de detección de estilos de aprendizaje
- Determinar métodos activos para lograr aprendizajes significativos
- Identificar y seleccionar Tecnología Informática aplicada en Educación.
- Diseñar y modelar un STI para integrar los aspectos teóricos y metodológicos contemplados en el proyecto

3. RESULTADOS Y OBJETIVOS

Los STI permiten modelar y simular el comportamiento de un docente tutor para determinar qué enseñar, cómo enseñar y a quién enseñar a través de un módulo del dominio.
En este artículo, se muestran algunos de los resultados del proyecto de tesis, centrado en la línea identificada como *Relevar proyectos o estudios de investigación, donde se consideren estilos de aprendizaje del estudiante* para modelar el STI propuesto. Por ello,

- Se relevó la existencia de una diversidad de STI, adaptables a las

características o evolución de ciertos tipos de aprendizaje en el alumno.

- Se clasificaron y analizaron los trabajos seleccionados, y se identificaron que estilos de aprendizaje de los estudiantes son mencionados en distintas disciplinas
- Se identificaron las tendencias actuales en torno a metodologías activas de enseñanza en Ingeniería determinadas por el CONFEDI [16] y que privilegian los estilos de aprendizaje.
- Se indagaron en los diferentes Test de detección de estilos de aprendizaje propuestos en la literatura. Este análisis fundamentó aquel más adecuado para el contexto en que se desarrolla esta tesis.
- Se realizó una exhaustiva identificación de trabajos de investigación relevantes, que integran a los estilos de aprendizaje en la construcción de un STI.
- Se relevaron, clasificaron y analizaron diversos STI identificados en la web o en publicaciones. Además, se identificaron aquellos STI que incorporan como uno de sus objetivos el apoyo al logro de aprendizajes significativos
- Se sistematizaron las publicaciones seleccionadas, y se generó un listado de métodos y herramientas software utilizadas para construir el modelo inteligente a introducir en el STI propuesto.

Por lo expuesto, y lo relevado en la revisión bibliográfica de proyectos previamente implementados en la misma área, el STI que surge de esta propuesta contextualizada en la FCEIA, podrá facilitar y apoyar al estudiante de Ingeniería en su proceso de enseñanza aprendizaje en el contexto definido y se podrá adaptar a otros contextos.

Se destaca la determinación y énfasis detectado en torno a cómo los métodos activos se relacionan con los estilos de aprendizajes de los alumnos, con miras a lograr aprendizajes significativos y como las TIC aportan o colaboran en este logro.

4. FORMACIÓN DE RECURSOS HUMANOS

En el marco de este proyecto se desarrolla una tesis comprendida en el Doctorado en Ingeniería, en la FCEIA – Universidad Nacional de Rosario.

El conocimiento adquirido y consolidado a través de la propuesta se aplica en la construcción de un prototipo de Sistema Tutor Inteligente que contemple en el módulo del estudiante sus estilos de aprendizaje. Se pretende que esta estrategia apoye al módulo docente en la selección de estrategias para impartir conocimiento.

Esta experiencia se transferirá al ámbito académico de la FCEIA y como soporte a los procesos de seguimiento. Además, se espera formar y consolidar recursos humanos en los temas tratados para su validación constante y posible adaptación a otros contextos.

BIBLIOGRAFIA

[1] Escalante, J. E., Mariño, S. I., & Marchisio, S. Hacia el Modelado del tutor Experto para la realización de un Sistema Tutor Inteligente para apoyo en Ingeniería X Congreso Argentino de Enseñanza de la Ingeniería. Córdoba. (2018)

[2] Ullauri Moreno, M. I. Influencia de los estilos de aprendizaje en el rendimiento académico del idioma inglés de los estudiantes de la Facultad de Educación, Humanas y Tecnologías de la Universidad Nacional de Chimborazo. Tesis Doctoral, Lima. Perú. (2017).

[3] Capote León, G. E., Rizo Rabelo, N., & Bravo López, G. La formación de ingenieros en la actualidad. Una explicación necesaria. (R. C.-I. 2218-3620, Ed.) Revista Universidad y Sociedad, 8((1)), 21-28. (2015). Recuperado de https://rus.ucf.edu.cu/index.php/rus/article/view/300/297

[4] Ventura, A. C., Palou, I., & Széliga, C. Estilos de enseñanza y de aprendizaje en Ingeniería:

Una propuesta de Educación adaptativa para primer año. Educación en Ingeniería, 9((18)), 178-189. (2014). Doi: https://doi.org/10.26507/rei.v9n18.461

[5] J.I. Pérez Rave. "El artículo de revisión sistemática como vehículo de escritura, investigación y publicación en ingeniería". Investigación y Ciencia de la Universidad Autónoma de Aguascalientes. Número 64: 70-77, (2015).

[6] Coto Jimenez, M. Descubrimiento del estilo de aprendizaje dominante en estudiantes de Matemática Superior. Revista Educación. 44((1)), 1-13. 2020. doi:http://doi.org/10.15517/REVEDU.V44I1.38571

[7] Arenas Landinez, A. L., & Ramírez Prada, D. C. Visión prospectiva de la Formación en Ingeniería. 8° LACCEI Latin American and Caribbean Conference for Engineering and Technology. "Innovation and Development for the Americas"., (págs. 1-10). Arequipa, Perú. (2010). Recuperado de http://www.laccei.org/LACCEI2010-Peru/RP023.html

[8] Paoloni, P. V., Chiecher, A. C., & Elisondo, R. C. Graduados de ingeniería y competencias genéricas. Cinco estudios de la última década que recuperan sus valoraciones y experiencias. Educación en Ingeniería., 14((28)), 54-64 (2019). DOI: http://doi.org/10.26507/rei.v14n28.986

[9] Vega Gonzalez, L. R. La educación en ingeniería en el contexto global: propuesta para la formación de ingenieros en el primer cuarto del Siglo XXI. Ingeniería Investigación y Tecnología., XIV (2), 177-190. (2013) Recuperado de http://www.revistaingenieria.unam.mx/numeros/v14n2-04.php

[10] Alonso, C. M., Gallego, D., & Honey, P. Los estilos de Aprendizaje. Procedimientos de diagnóstico y mejora. (7° edición). Bilbao, España: Ediciones Mensajero. (1995). ISBN: 978-84-271-1914-7.

[11] Felder, R. M., & Silverman, L. K. Learning and Teaching Styles. Journal of Engineering Education, 78((7)), 674-681. (1988). Recuperado de https://www.researchgate.net/publication/257431200_Learning_and_Teaching_Styles_in_Engineering_Education

[12] J. R. Anderson, M. Matessa, C. Lebiere, "ACT-R: A theory of higher level cognition and its relation to visual attention". Human-Computer Interaction, 12(4), 439-462, (1997).

[13] L. M. M. Giraffa, M. A. Nunes, and R.M. Viccari, "Multi-Ecological: a Learning Environment using Multi-Agent architecture". Proc. of MASTA'97 (EPIA'97 workshop on Multi-agent Systems: Theory and Applications), Coimbra, Brasil, (1997)

[14] A. Gordillo Guillen, H. Andrade Gómez, R. Rivera López. "Modelo de un sistema tutor inteligente para el desarrollo del pensamiento computacional" *Revista Investigación Educativa. Journal CIM* Vol. 5, Núm. 2, pp. 1479 – 1485. (2017).

[15] K. VanLehn, *Student Modelling*, en M.C. Polson, and J.J. Richardson, (Eds.) Foundations of Intelligent Tutoring systems. Taylor & Francis, (2013).

[16] CONFEDI. Consejo Federal de Decanos de Ingeniería. (2018) Recuperado de https://confedi.org.ar/download/documentos_confedi/LIBRO-ROJO-DE-CONFEDI-Estandares-de-Segunda-Generacion-para-Ingenieria-2018-VFPublicada.pdf

Enseñanza de TICs, mediante el desarrollo de videojuegos utilizando metodologías STEAM

Romina Stickar, Rodrigo René Cura

Departamento de Informática, Facultad de Ingeniería,
Universidad Nacional de la Patagonia San Juan Bosco.
Puerto Madryn, Chubut, Argentina. +54 280-4472885 – Int. 116
romistickar@gmail.com, rodrigo.renecura@gmail.com

RESUMEN

La enseñanza de TICs, y en particular la de programación, es un desafío que se basa en dos pilares fundamentales, la experiencia práctica y la motivación. Las metodologías STEAM cumplen con el primero de los pilares por medio del aprendizaje activo, y los videojuegos, en general, son una motivación para los estudiantes. El desarrollo de videojuegos es el escenario perfecto para la combinación de estos ejes, y además, provee un espacio en donde experimentar con diferentes disciplinas y lograr una adquisición más profunda de competencias variadas.

La enseñanza de TICs es un requerimiento cada vez más fundamental en todos los niveles educativos, ya que estas tecnologías están presentes de forma transversal en todos los ámbitos, tanto profesionales como cotidianos. La adopción de las mismas se ha convertido en una necesidad tan básica como cualquiera de las otras disciplinas, con la particularidad de que se encuentra de forma homogénea entre estas y no puede estudiarse aisladamente como las asignaturas más tradicionales.

El uso de las TICs implica una nueva concepción del proceso de enseñanza y aprendizaje, en donde el estudiante se convierte en un participante activo y constructor de su propio aprendizaje y el docente se transforma en un guía y facilitador de estos procesos.

Por tal motivo, este proyecto busca hacer un aporte a la aplicación y adaptación de estas metodologías, en los distintos niveles educativos, usando el desarrollo de videojuegos como marco integrador de los distintos ejes disciplinarios y que les permitan a los estudiantes la apropiación activa de las capacidades relativas a las TICs.

Palabras clave: STEAM, VIDEOJUEGOS, PROGRAMACIÓN, PEDAGOGÍA.

CONTEXTO

Este grupo de investigación y desarrollo se lleva a cabo dentro de la sede Puerto Madryn de la Facultad de Ingeniería de la Universidad Nacional de la Patagonia San Juan Bosco (FI-UNPSJB).

El mismo surgió porque es necesario que los adolescentes y jóvenes puedan acercarse cada vez más temprano a la resolución de problemas a través de la algorítmica y programación. Los videojuegos son una forma atractiva de introducirlos a este paradigma.

Existen algunas herramientas de programación de videojuegos que sirven para aplicar metodologías STEAM, algunas de la cuales hemos utilizado en distintos ámbitos educativos, como Scratch [3], Pilas Engine [4] y Alice [5].

Desde el año 2013, se vienen realizando distintas experiencias con las herramientas arriba mencionadas con la participación de alumnos y docentes de la FI. Entre ellas

podemos mencionar: “Programación de videojuegos con la herramienta Pilas Engine en el nivel medio” [1]. “Curso de Alice”, en colaboración con la fundación Sadosky. Taller de programación de videojuegos en el Club de Ciencias y Tecnologías de Puerto Madryn.

Este proyecto se lleva a cabo en el marco de los proyectos presentados y avalados por Ciencia y Técnica de la Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB).

El financiamiento del proyecto es realizado por la UNPSJB.

1. INTRODUCCIÓN

En los últimos años, a nivel mundial, hay una tendencia creciente a incluir programación como una asignatura más de los trayectos formativos obligatorios en los niveles primario y secundario. Dado que las personas están expuestas a las tecnologías de la información y la comunicación (TICs) desde una temprana edad, se pretende que reciban una introducción anticipada en la algorítmica y programación con el fin de obtener capacidades que les permitan no ser simples usuarios sino que puedan aprovechar al máximo el potencial de las herramientas digitales.

La naturaleza transversal de las TICs no solo requiere una enseñanza directa de las mismas sino que se puede adquirir al aplicarlas en otras áreas de conocimiento, así también como fomentar su aprendizaje mediante otras disciplinas. Esto último se puede lograr siguiendo metodologías STEAM.

STEAM es un conjunto de metodologías que pretenden desarrollar proyectos interdisciplinares y busca que la enseñanza-aprendizaje de la Ciencia, Tecnología, Ingeniería, Arte y Matemática se realice de manera integrada [2,6,7].

En una Educación STEAM, los estudiantes trabajan en equipo y aprenden a resolver problemas reales sobre los que deben tomar decisiones y reflexionar; aumentan su capacidad para resolver problemas de forma creativa así como el pensamiento crítico individual, su autoestima e impulsan sus capacidades comunicativas. La experimentación en primera persona les permite mejorar la retención de los conceptos aprendidos a largo plazo. Además, el uso de tecnologías emergentes minimiza la sensación "intimidatoria" que estos producen. A través de la explicación de hipótesis e ideas, hacen conexiones entre los objetivos de la resolución de problemas y los procesos realizados. [8, 9,10]

La aproximación a solucionar problemas mediante procesos algorítmicos, a través de la construcción de un videojuego donde se deben respetar ciertas reglas para alcanzar las metas deseadas, hacen de este un ambiente ideal para el abordaje del aprendizaje activo, pues el estudiante cuenta con una herramienta de simulación evolutiva. Este ambiente evolutivo es el que le permite al estudiante adquirir capacidades innovadoras respecto de cómo solucionar ciertos problemas. El desarrollo de videojuegos, provee un escenario enriquecedor para descubrir y comprender conceptos de TICs aplicando de forma natural metodologías STEAM.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este proyecto busca investigar y adaptar las metodologías STEAM mediante el desarrollo de videojuegos para la enseñanza de TICs en todos los niveles educativos.

Para cumplir con el objetivo se llevarán a cabo las siguientes líneas de acción específicas:

- Selección de material bibliográfico e investigación y estudio del estado del arte de las distintas metodologías STEAM.
- Generación de materiales didácticos para los niveles educativos primarios y secundarios, los cuales puedan ser utilizados por los docentes de esos niveles sin necesidad de recurrir a la asistencia brindada por la universidad.

- Búsqueda de estrategias de articulación entre asignaturas y la programación de videojuegos. Se construirá un escenario ficticio como modelo de resolución de problemas, permitiendo a los estudiantes contar con una mirada y aproximación diferente y que puedan recrear sus propias reglas para el funcionamiento de un ambiente determinado, comprobando en la práctica sus resultados.
- Elaboración de recursos y actividades en forma de proyectos prácticos de desarrollo de videojuegos enfocados en contenidos puntuales de las áreas de ciencia, tecnología, ingeniería, arte y matemática.
- Continuación de las actividades realizadas por el Laboratorio de Investigación en Informática (LINVI) perteneciente al Departamento de Informática de la UNPSJB, así como el aumento de proyectos acreditados y de trabajos publicados por este.

3. RESULTADOS OBTENIDOS/ESPERADOS

Actualmente el proyecto se encuentra en una etapa inicial que consta de jornadas de trabajo con los integrantes del mismo.

Los miembros del equipo ya nos encontramos trabajando la selección de material bibliográfico y base conocimiento sobre las metodologías STEAM, herramientas de programación de videojuegos, experiencias de aplicación de metodologías STEAM y de programación de videojuegos en los distintos ámbitos educativos.

Sobre la finalización del proyecto se espera obtener:

- Un conjunto de proyectos prácticos de desarrollo de videojuegos.
- Planificaciones completas de las actividades para ser replicadas en futuros talleres.
- Realización de un evento público en la universidad para que los participantes expongan los resultados obtenidos a lo largo del año.
- Realización de varias actividades de transferencia, entre ellas actividades de capacitación.
- Catálogo de actividades ludificadas relacionadas a las asignaturas en las que se participó durante la ejecución del proyecto. Detalladas de tal forma que puedan ser recreadas y adaptadas más allá del contexto del proyecto aquí presentado.
- Firma de nuevos acuerdos particulares con otras escuelas primarias o secundarias de la ciudad con el fin de transferir la experiencia.
- Inclusión de alumnos de la carrera de Licenciatura en Informática en el proyecto.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo está integrado por 2 docentes del Departamento de Informática de la FI-UNPSJB sede Puerto Madryn.

Los 2 docentes se encuentran realizando actualmente la Maestría en Enseñanza en Escenarios Digitales.

Es intención del proyecto incorporar alumnos para que puedan participar de la experiencia de ser capacitadores de los talleres de videojuegos.

Se diseñará como parte del proyecto un taller de capacitación destinado a proveer los conocimientos mínimos a los nuevos integrantes.

5. BIBLIOGRAFÍA

[1] Programación de videojuegos en el nivel medio. Stickar, Defosse, René Cura, Barry. TE&ET 2015.

[2] Teaching and Learning STEM: A Practical Guide. Felder & Brent (2016)

[3] Scratch https://scratch.mit.edu/ (2020)

[4] Pilas Engine https://pilas-engine.com.ar/ (2020)

[5] Alice https://www.alice.org/ (2020)

[6] How People Learn: Bridging Research and Practice. Donovan S. M., Bransford J. D., Pellegrino J. W. (1999)

[7] "STEM, STEM Education, STEMmania". The Technology Teacher. International Technology Education Association. Sanders, M (2009)

[8] Analysis on the Effects of the Augmented Reality-Based STEAM Program on Education. Ahn & Choi (2015)

[9]Second-Graders Beautify for Butterflies. Journal of STEM Arts, Crafts, and Constructions. Anderson & Meier, 2016

[10]Using Creativity from Art and Engineering to Engage Students in Science. Journal of STEM Arts, Crafts, and Constructions. Kuhn, Greenhalgh, & Mcdermott, 2016

Evento de Transferencia DeCoDev, Desarrollo Colaborativo de un Videojuego

Lic. Angela Belcastro [1], APU Pablo Dibez [2], APU Macarena Quiroga [3], Mg. Rodolfo Bertone [4]
[1]Departamento de Informática. Facultad de Ingeniería (FI). UNPSJB. Comodoro Rivadavia.
[2, 3]Estudiante de grado. Departamento de Informática.FI. UNPSJB. Comodoro Rivadavia.
[3]Auxiliar de segunda. Departamento de Matemáticas.FI. UNPSJB. Comodoro Rivadavia.
[4] III LIDI – Facultad de Informática – UNLP. (1900) La Plata, Buenos Aires, Argentina
[1]angelab@ing.unp.edu.ar,[2]msinf.quiroga@gmail.com, [3]pdibez@gmail.com
[4]pbertone@lidi.info.unlp.edu.ar

RESUMEN
Este trabajo se encuentra en el marco de unproyecto de investigaciónacreditado, que es continuación de tres proyectos previos del área de Tecnología Informática Aplicada en Educación (TIAE).Estudiantes de grado del equipo de proyecto, han intervenido, en las actividades de revisión de la literatura, preparación, y dictado de un taller,desarrollado en febrero-marzo de 2020. Se presentan en este artículo algunas líneas de investigación y resultados del proyecto, y se describen las características principales del taller de extensión universitaria desarrollado en la UNPSJB, que ayuda a propiciar el desarrollo del *pensamiento computacional (PeC)* de potenciales ingresantes o alumnos iniciales de Informática, con una propuesta basada en la Teoría de aprendizaje constructiva, motivando al alumno a desarrollar un videojuego en forma colaborativa empleando trello, exportando luego el juego, a Android. A través del ejemplo dado por los alumnos disertantes, se intenta que los participantes, identifiquen la gran importancia de participar en proyectos de investigación (PI) y congresos, a medida que avanzan en sus carreras universitarias.

Palabras claves:
Constructivismo. Videojuegos. Trello.

CONTEXTO
El presente trabajo se enmarca en unproyecto de investigación acreditado, reconocido por el programa de Incentivos, en el Sistema de Ciencia y Técnica de la UNPSJB, proyectotitulado: "Tecnología Informática Aplicada en Educación, y Aprendizaje Significativo", UNLP-UNPSJB, es continuación de tres PI del área de TIAE, iniciados en 2003, que apoyaron la formación de recursos humanos, generaron mejoras en procesos de enseñanza aprendizaje, originaron publicaciones nacionales e internacionales, y transferencias a la comunidad educativa y científica.

1. INTRODUCCION

El ***constructivismo*** es una teoría del aprendizaje, según la cual las personas construyen nuevos conocimientos de manera activa al combinar las experiencias con los conocimientos previos. El estudiante no recibe el conocimiento, sino que lo construye en su cabeza. El nuevo conocimiento es resultado del proceso de búsqueda de sentido de las situaciones nuevas conciliando las nuevas experiencias a información con lo que el estudiante ya sabe o ha experimentado. [1]
El construccionismo de Seymour Papert (1986) enuncia que "el aprendizaje es más eficaz cuando es parte de una actividad que el sujeto experimenta como la construcción de un *producto significativo*". [1][2]
Entre los indicadores que nos permiten detectar si una *actividad es significativa, encontramos:* debe estar de acuerdo a las necesidades o intereses de los destinatarios, losalumnosdisfrutan lo que hacen, se concentran en la tarea, participan con interés, interactúan con agrado entre ellos, y se muestran seguros y confiados. [3]
Se orientó al alumno a la construcción de un producto significativo, un videojuego. Examinamos ahora, las nociones de juego, y videojuego. El ***juego*** es una actividad esencial y permanente del ser humano, que le

ha permitido asimilar la cultura, fortalecer habilidades o destrezas para conocer, comprender y actuar sobre el mundo; es una fuente educativa, de diversión y de placer, que brinda vivencias y situaciones que favorecen el desarrollo del ser humano. [4]

Dempsey define un ***juego*** como un conjunto de actividades que afecta a uno o más jugadores, tiene metas, limitaciones, beneficios y consecuencias, es orientado por las normas, y es artificialen algunos aspectos. Incluyeelementos de competencia, incluso si esa es una competencia con uno mismo. [5]

Con los avances tecnológicos de los siglos XX y XXI, el juego ha evolucionado, producto de la convergencia de lo lúdico y lo tecnológico (Cabañes, 2012). Lo lúdico, lo asociamos a aquellas dinámicas relacionadas con los juegos tradicionales o convencionales, y lo tecnológico, al medio e instrumento donde los diseñadores desarrollan los juegos y los jugadores juegan. A la anterior convergencia, se la denomina ***videojuego***. ***Videojuego*** es una palabra compuesta, vídeo, indica que el soporte o salida de datos fundamental es la imagen, y juego, que es lo que confieredificultad a su discriminación, pero también es su potencial diferencial respecto a otras tecnologías. Inicialmente, los videojuegos buscaban el factor lúdico, y actualmente, han comenzado a cobrar importancia los "***juegos serios***". Calvo Ferrer (2012), los define como: "juegos cuya finalidad principal no es la de divertir al usuario, sino la de formarle o instruirle en determinadas áreas". [6]

Una de las iniciativas actuales de docentes y alumnos de diferentes países y niveles educativos, que proponen involucrar a las personas en el aprendizaje de la programación, consiste en el usode juegos orientados al aprendizaje de la programación: de entornos como Scratch, Alice, ***PilasEngine*** (https://pilas-engine.com.ar/), para acercarse a la programación, y al PeC. Varios autores afirman que aprender a programar es una llave que puede abrir la puerta al desarrollo de habilidades y competencias, entre ellas la capacidad de abstracción, de resolución de problemas, y de trabajo en equipo. [7]

Wing (2006) expresa que "el ***PeC*** es el proceso de pensamiento involucrado en la formulación de problemas y expresión de las soluciones en una forma que pueda ser efectivamente llevada a cabo por un agente de procesamiento de información". Brennan y Resnick (2012) destacan que el PeC abarca tres dimensiones: (conceptos, prácticas y perspectivas computacionales), que ayudan a evaluar si el alumno adquiere la capacidad de aplicarlas, al resolver problemas. En la dimensión 1, encontramos los procedimientos, alternativa condicional, e interactividad. En la dimensión 2, la habilidad de aplicar los conceptos computacionales siguiendo criterios de buenas prácticas en Programación, como legibilidad y reutilización. En la dimensión 3, trabajar colaborativamente para encontrar la solución, depreguntar si es posible dar solución a otros problemas desde un punto de vista computacional. [8]

La propuesta contempla el desarrollo del PeC de participantes, acercándolos a la construcción colaborativa de un videojuego, empleando la metodología Scrum, la herramienta colaborativa trello (https://trello.com/), rúbricas y pilas engine.

Una metodología de desarrollo es una forma de organizar el trabajo en función de un objetivo. Las metodologías ágiles de desarrollo promueven la evolución y adaptación del software, una las más usadas es Scrum, un marco de trabajo de procesos, empleado para gestionar el desarrollo deproductos complejos desde principios de losaños 90, en el cual se pueden usar varias técnicas y procesos. [9][10][11]

Investigaciones educativas, muestran que ***Trello*** ayuda a promover la tolerancia por las ideas de los demás, el respeto, aprender a escuchar, a convencer, a cambiar de opinión, a compartir; como así también ayuda a mejorar la capacidad de organización de los alumnos, y de trabajo en equipo. [12]

Trellousa el paradigma de gestión de proyectos Kanban. La palabra Kanban deriva del japonés, kan (significa visual), y ban

(significa tarjeta o tablero). Es un sistema de información que controla de modo armónico un proceso de producción, como puede ser la fabricación de un producto o la realización de un proyecto. Trello, permite tener una lista de tareas por hacer, que se están haciendo y hechas. Las listas están dentro de tableros, y dentro de ellas, se ubican las tarjetas, que van cambiando de lista, conforme se van realizando, o pasan por diferentes fases. Las tarjetas soportan diversos tipos de documentación, se puede adjuntar vídeos y archivos, sirven también como instrumento de comunicación en grupos de trabajo, acerca deun determinado tema. Permite compartir los materiales necesarios para llevar a cabo la tarea asignada y los resultados obtenidos. Es gratuita y se puede usar en computadora, Tablet con conexión a Internet y en dispositivos móviles inteligentes. [13] [14]

Pilas Engine es una herramienta para construir videojuegos de manera sencilla y divertida, dirigida a personas que comienzan a programar videojuegos, es libre y gratuita, multiplataforma e interactiva. Su documentación está en castellano, incluye actores y ejemplos prediseñados. [15]

Las ***rúbricas*** son guías precisas que valoran los aprendizajes y productos realizados. Desglosan los niveles de desempeño de los estudiantes en un aspecto determinado, con criterios específicos sobre rendimiento. [16]

2. LINEA DE INVESTIGACION Y DESARROLLO

Entre las principales líneas de investigación y desarrollo vinculadas al proyecto en curso, del área TIAE, encontramos: el modelo constructivista, aprendizaje significativo, metacognición, desarrollo móvil y aprendizaje móvil. El objetivo es el de contribuir a la formación de personas competentes, fomentar el desarrollo de la sociedad y aumentar el conocimiento, analizando y confeccionando propuestas educativas que propicien el aprendizaje significativo, diseñando y desarrollando sistemas interactivos de apoyo al aprendizaje con tecnología móvil, con el objeto de incentivar a los alumnos a investigar, y a participar en PI, y mejorar las competencias de nuevos profesionales de Informática

3. RESULTADOS OBTENIDOS/ESPERADOS

Entre los antecedentes encontramos publicaciones y trabajos aceptados en WICC de 2015 a 2018 inclusive, TE&ET 2015, y CONAIISIde 2016 a 2019 inclusive[17][18][19][20][21][22][23][24].

El "Evento de Transferencia: DeCoDev: desarrollo colaborativo de un videojuego", que se llevó a cabo en el laboratorio Jorge Ardengui, del departamento de Informática, FI de la UNPSJB, es un taller de extensión universitaria. Desde el proyecto, en 2019, se solicitó el aval institucional (resolución Res. DFI 1097/2019). Los disertantes son estudiantes de grado del equipo de proyecto y la codirectora. El taller propicia el desarrollo del PeC y propone una metodología de trabajo colaborativo; sus objetivos son:

- Motivar a potenciales alumnos ingresantes de Informática, a iniciarse en el desarrollo colaborativo de videojuegos.
- Mejorar las competencias de nuevos profesionales de Informática, de trabajo en equipo con colaboración, despertando en ellos el interés, por participar en PI al avanzar en la carrera.

Se preparó un formulario diseñado en Google Forms, a través del cual se realizó la inscripción al evento, conaval institucional, el código QR y el link https://forms.gle/DQykGTHHXhNfv9Gp9.

La puesta en práctica del evento se enfrentó a dificultades, la difusión del evento quedó visible en la página de Ingeniería, el 27-12-19, y el 3-01-20, se realizaron mejoras en el sitio y quedó inaccesible. Algunos alumnos iniciaron el taller, el segundo día.

El evento se desarrolló de 17 a 21 hs. en forma presencial los días 3, 4 y 5 de febrero de 2020. Trabajan en forma virtual los interesados en obtener el certificado de aprobación. Se elaboró un tablero de trello para cada grupo de alumnos, incorporando en los tableros, como administradores, a los disertantes. En cada tablero se compartió el

material con las presentaciones y proyectos de pilas engine con ejemplos preparados por los disertantes. Los alumnos, a medida que realizaban las actividades propuestas y videojuegos, recortaban la imagen de pantalla del videojuego desarrollado, e incluían en el tablero de trello, tanto la imagen de portada como el proyecto en pilas engine. La figura 1 muestra el tablero del grupo 1.

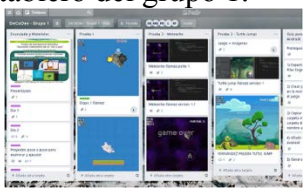


Figura 1. Tablero de trello del grupo 1

Se idearon los juegos meteorito, y Tutle Jump. El 5-2-20 se explicó cómo exportar un juego de pila engine en Android. Se generaron intercambios colaborativos entre los participantes, a través del tablero de trello y notificaciones en correos electrónicos. En los encuentros se facilitó y explicó a los alumnos las rúbricas de evaluación para la obtención de certificados. Con tres escalas, analizando: contenidos (uso de trello y funcionalidad), colaboración y creatividad. Se contemplaron algunos elementos de las tres dimensiones del PeC. Algunos alumnos realizaron la primera parte de la autoevaluación, que genera una instancia más de aprendizaje, cada pregunta incluye retroalimentación de apoyo a la comprensión de: diferencia entre equipo de trabajo, y grupo, características básicas de los proyectos, la referencia this, mensaje y objeto. Se incluyó en el tablero, el ahorcado móvil curso de ingreso, sobre teoría de conjuntos, desarrollado dentro del proyecto, de apoyo para alumnos ingresantes. Al finalizar el trabajo final, los participantes podrán terminar la autoevaluación con distintas dimensiones, una orientada a propiciar metacognición. La figura 2 muestra fotos del evento.



Figura 2. Fotos del evento

Se realizó un nuevo evento de transferencia en marzo. La difusión de otros resultados se realizará mediante futuras publicaciones.

4. FORMACIÓN DE RECURSOS HUMANOS

La codirectora avanza en su proyecto de tesis de "Magister en Tecnología Informática Aplicada en Educación", de la UNLP, aprobó cursos de postgrado, que originaron transferencias del proyecto, entre ellos, los titulados: "Bases y fundamentos para la configuración del proceso de enseñanza", "Sistemas de producción inteligente", y "Formación por competencias". La organización y las actividades del proyecto, fortalecen las habilidades investigativas y profesionales de los autores de este artículo, y la generación de transferencias en la comunidad educativa y científica. Varios integrantes adeudan un examen final y la tesina, para obtener el título de grado.

5. BIBLIOGRAFIA

[1] Sylvia Libow Martínez. Gary Stager. Inventar para aprender. Guía práctica para instalar la cultura maker en el aula. Siglo Veintiuno editores. Colección educación que aprende. 2019.

[2] Sanz Cecilia, y otros. (2016). Escenarios educativos mediados por Tecnologías de la Información y la Comunicación. Actas WICC 2016, ISBN: 978-950-698-377-2, págs. 983-988, 2016.

[3] Calero Pérez, M. Constructivismo pedagógico. Teorías y aplicaciones básicas. Alfaomega. 2008.

[4] Boude F., O. y Sosa, E. (2016) Juego Serio: Modelo teórico para su diseño y producción. En C. Parra (Ed.), Doctorado en Educación: Temas y conceptos (pp. 197-221). Chía, Colombia: Universidad de La Sabana.

[5] Apezteguía Matías, Rapetti Darío. (2014). Juego Educativo Móvil Colaborativo. Tesina de grado. Director: Dra. Gordillo Silvia Codirector: Dra. Challiol Cecilia. Facultad de Informática UNLP. Disponible en: http://sedici.unlp.edu.ar/handle/10915/47078. Acceso: 01-11-19

[6] Rocío Serna Rodrigo (2016.). El papel del videojuego en la formación de universos

transmedia. En Rosabel Roig-Vila (Ed.), Tecnología, innovación e investigación en los procesos de enseñanza-aprendizaje. Barcelona: Octaedro. ISBN 978-84-9921-848-9. Págs. 3041-3048, 2016.

[7] Sanz, Cecilia. (2016). Aprender a programar en tiempos digitales. Revista Bit & Byte. ISSN: 2468-9564. Págs. 12-14, 2016.

[8] Ahumada, Hernán y otros. (2019). Programación por Bloques en cátedras de Ingeniería en Informática. Actas WICC 2019, ISBN: 978-987-3619-27-4. Págs. 636-640. 2019.

[9] Loto, Matías.Durán, Elena. (2015). Diseño de una aplicación móvil personalizada de apoyo al aprendizaje de Redes de Computadoras. Actas de TE&ET 2015. ISBN: 978-950-656-154-3. Págs. 325-333. 2015.

[10] Nazareno, Roberto y otros. (2015). Trazabilidad de Procesos Scrum. Actas del Simposio Argentino de Ingeniería de Software (ASSE 2015). ISSN: 2451-7593. Págs. 284-298. 2015.

[11] Gálvez Diaz, María del Pilar. Tolaba, Carolina. (2019). Como Alcanzar Competencias Usando Scrum. Actas SAEI 2019. ISSN:2683-8958. Págs: 62-70. 2019.

[12] Delgado, Adelaida. Mesquida, Antoni Lluís. Mas, Antonia. (2014). Utilización de Trello para realizar el seguimiento del aprendizaje de equipos de trabajo. JENUI 2014. Disponible en: http://hdl.handle.net/2099/15518 Acceso: 18-11-19

[13] Rodriguez, Facundo y otros. (2016). Flip-Flop. Aplicación de Buenas Prácticas a partir de la Gamificación. CACIC 2016. Disponible en: http://sedici.unlp.edu.ar/handle/10915/56649. Acceso: 13-12-19

[14] Moreno Arias, Juan Carlos. (2016) Programación del videojuego "llumpak" para dispositivo móvil de sistema operativo Android. Universidad Central del Ecuador. Facultad de Ingeniería, Ciencias Físicas y Matemática. Carrera de Ingeniería en Computación Gráfica. Tutor: Fis. Bayardo Campuzano Nieto, Gonzalo. Disponible en: http://www.dspace.uce.edu.ec/bitstream/25000/6554/1/T-UCE-0011-23.pdf Acceso: 06-12-19

[15] Rueda, Sonia y otros. (2014). Herramientas para apoyar el descubrimiento de vocaciones en Ciencias de la Computación. CACIC 2014. Disponible en: http://sedici.unlp.edu.ar/handle/10915/42122 Acceso: 21-11-19

[16] Gatica-Lara, F. y otros. ¿Cómo elaborar una rúbrica? Elsevier. (2013). Disponible en: https://www.elsevier.es/es-revista-investigacion-educacion-medica-343-articulo-como-elaborar-una-rubrica-S200750571372684X Acceso 22-11-19.

[17] Belcastro, Angela, y otros. (2015).Recursos de la Web 2.0, y juegos interactivos, alternativas viables para propiciar el aprendizaje en Informática. WICC 2015.

[18] Belcastro, Angela y otros. (2015). Método para mejorar un software educativo en desarrollo. TE&ET 2015.

[19] A. Belcastro, Rodolfo Bertone.(2016). Modelos, Estrategias, y Recursos para la Enseñanza Mediada, en una Asignatura de una Carrera de Grado. CONAIISI 2016.

[20] A. Belcastro y otros. (2017). Juegos interactivos en ARDUINO y Java, para motivar y despertar el interés en Informática. WICC 2017.

[21] A. Belcastro; Rodolfo Bertone. (2017). Tarea Auténtica Mediada por Tecnología. CONAIISI 2017.

[22] A. Belcastro y otros. (2018).Vinculación con docentes de nivel secundario y con ingresantes de Informática, a través de talleres y juegos educativos desarrollados, uno de los cuales, es un juego móvil. WICC 2018.

[23] A. Belcastro; Rodolfo Bertone. (2018). Apoyando el Ejercicio de Metacognición en el Ámbito Universitario. CONAIISI 2018.

[24] A. Belcastro; Rodolfo Bertone; (2019). Experiencia de Acercamiento al Enfoque de Formación por Competencias que Intenta Propiciar Aprendizaje Significativo. CONAIISI 2019.

ESDEU: SISTEMA DE GESTIÓN TUTORIAL
CASO DE PRUEBA EN COMISIONES DE INGRESANTES 2020 DE UTN LA PLATA

ISTVAN, Romina [1]; LASAGNA, Valeria [2]

Universidad Tecnológica Nacional, Facultad Regional La Plata
[1] Grupo de I&D Aplicado a Sistemas Informáticos, GIDAS UTN FRLP
[2] Secretaría de TIC - UTN FRLP

Av. 60 s/n° esquina 124, CP 1900, La Plata, Buenos Aires, Argentina.
ristvan@frlp.utn.edu.ar; valerial@frlp.utn.edu.ar

RESUMEN

ESDEU es el Sistema de Gestión Tutorial desarrollado desde el año 2016 por la UTN La Plata dentro del marco del Proyecto de Investigación y Desarrollo (PID): «Estudio Sistemático de Deserción Estudiantil Universitaria», el cual continúa su línea de trabajo a partir del año 2020 con un nuevo PID: «Determinación de perfiles de riesgo de deserción estudiantil en UTN-FRLP utilizando técnicas de minería de datos».

La función principal del Sistema ESDEU se basa en calcular automáticamente los factores de riesgo de deserción, presentar alertas tempranas y gestionar de manera eficiente el seguimiento de los estudiantes. Asimismo potencia la comunicación entre tutores, coordinador tutores, ayudantes y docentes, facilitando el conjunto de tareas de gestión tutorial.

La primera versión del sistema se implementa en una prueba piloto para Ingresantes 2020 de la UTN La Plata. Los resultados obtenidos se exponen en el presente trabajo.

Palabras clave: Sistemas de Gestión Tutorial, Deserción Estudiantil, Deserción Universitaria, Tutorías, Indicadores de Deserción.

CONTEXTO

Argentina se caracteriza por ser uno de los países con menor graduación en proporción a la cantidad de estudiantes universitarios (CEA, 2018). Es por ello, que dentro de la Secretaría de Políticas Universitarias (SPU) surge el Plan Estratégico de Formación de Ingenieros (PEFI, 2012) cuyo fin es aumentar los egresados de las distintas carreras de Ingeniería.

Dentro de este contexto, la UTN La Plata desde principios del año 2016 se encuentra trabajando en el PID: «Estudio Sistemático de Deserción Estudiantil Universitaria» (PID ESDEU) homologado por Rectorado UTN con código TEINNLP0003786, continuando su desarrollo a partir del año 2020 en un nuevo PID: «Determinación de perfiles de riesgo de deserción estudiantil en UTN-FRLP utilizando técnicas de minería de datos» homologado por Rectorado UTN con código TEUTNLP0007653.

Dichos proyectos se complementan y presentan relevante importancia institucional ya que formalizan el primer estudio sobre deserción en la UTN La Plata y responden a la necesidad actual de elevar la tasa de graduación efectiva en cada una de las especialidades que brinda la Regional.

1. INTRODUCCIÓN

El producto final de los proyectos es un software de gestión tutorial basado en un modelo de indicadores de riesgo de deserción estudiantil obtenido mediante Minería de Datos. Dicho sistema identifica a aquellos alumnos próximos a desertar y de esta manera, posibilita la determinación de estrategias de retención ajustadas a las particularidades de cada estudiante y de la institución.

Presenta como objetivo general elevar la tasa de promoción efectiva de Ingenieros, en apoyo a las Políticas Públicas y en concordancia con el Plan PEFI, siendo sus objetivos específicos:

(i) Consolidar la primera base de datos unificada para el estudio sistemático de la deserción, la cual recopila información de diversas fuentes: académica, encuestas y asistencias.

(ii) Identificar a posibles desertores mediante la visualización de alertas tempranas.

(iii) Registrar el estado de avance y seguimiento de los alumnos.

(iv) Brindar un modelo de indicadores que represente los principales factores causales de la deserción.

Las actividades de investigación llevadas a cabo profundizan en las características y funcionalidades de herramientas existentes en Latinoamérica, obteniendo puntos de referencia sobre el potencial del nuevo software, enmarcándolo en sistemas de tutorías con impacto en la deserción estudiantil.

En función de lo anterior, se optimizan los objetivos del sistema contextualizando los módulos y fuentes de datos necesarias, identificando los actores con sus roles y perfiles, modelando así la interactividad con el sistema.

Seguidamente se obtiene el Modelo de Indicadores de Deserción, mediante técnicas de Minería de Datos. Y por último se realiza la incorporación de dicho modelo al sistema para realizar la predicción de deserción de cada estudiante.

El sistema ESDEU en su primer versión cuenta con los siguientes módulos:



Módulo Alertas: este módulo permite identificar el grado de riesgo de cada estudiante, generar listas de alumnos según los riesgos e identificar los alumnos que necesitan atención prioritaria. El sistema muestra a través de distintos colores alertas visuales con el nivel de prioridad asignada a cada alumno mediante un semáforo de cuatro colores, en relación a la deserción o posible deserción asociada: verde (sin riesgo), amarillo (riesgo leve), naranja (riesgo moderado) y rojo (crítico).

Módulo Tutorandos / Tutor: este módulo presenta el listado de estudiantes asignados a un tutor con la posibilidad de filtro, resalta las alertas de estudiantes posibles desertores identificando con distintos colores cuáles de sus tutorados se encuentran en una situación de deserción crítica, permite el acceso al perfil del estudiante: datos básicos de contacto y perfil académico, junto a su seguimiento.

El tutor puede visualizar el indicador y los factores de deserción ya calculados, con la posibilidad de ordenar descendentemente según

el riesgo para priorizar la atención en los alumnos más críticos.
El tutor genera cada vez que lo requiera reportes de la situación del alumno y la actividad tutorial llevada a cabo.
Carrera / Materia / Comisión: visualiza el listado de alumnos por materia/comisión, a fin de poder evaluar la cantidad de desertores y posibles desertores para un curso.
Seguimiento de Alumnos: en este módulo el Tutor es el encargado de registrar la información de seguimiento en la cual se planifican las acciones para mejorar la situación del estudiante.
El tutor durante su gestión puede registrar: fecha y hora de inicio y fin del contacto establecido con el alumno, sus resultados, observaciones y conclusiones.
Módulo de Alumno: permite al Coordinador de Tutores, Profesores, Jefes de Cátedra, Jefes de Departamento y Equipo Interdisciplinario visualizar los datos personales, de contacto y perfil académico del alumno, junto con su perfil de riesgo y el acceso al historial de seguimientos o entrevistas realizadas.
Módulo Comunicación / Mensajería Interna: permite de manera rápida comunicarse tanto a docentes, alumnos y tutores entre sí. Actualmente en etapa de desarrollo.
Módulo de Reportes y Estadísticas: busca fortalecer los procesos de permanencia académica proporcionando información en tiempo real. Los reportes son generados en archivos Excel, con filtros avanzados para los usuarios.
Módulo Configuración / Administrador: gestiona los accesos y permisos por rol a cada módulo del sistema.

Estimación de Riesgo de Deserción
Con el fin de determinar la Estimación de Riesgo de Deserción de cada estudiante la UTN La Plata utiliza datos recabados en Encuestas. Las mismas recogen información complementaria a los datos académicos (Laborales, Familiares y Económicos) como ser: Horas de Trabajo, Relación del Trabajo con la Carrera, Situación Económica de los Padres, Nivel de Instrucción de los Padres, Distancia a la institución y Becas, entre otros y aplica la técnica de Explotación de Datos (Minería de Datos o Data Mining) con el fin de detectar patrones y relaciones entre variables de manera automática.
A este modelo se suman dos enfoques complementarios: las Inasistencias y el Rendimiento Académico. De esta manera, se constituyen tres ejes de análisis:
(I) Modelo de Indicadores de Riesgo de Deserción Estudiantil. Perfil de Riesgo.
(II) Inasistencias (Cursadas Libres y Faltas para el ciclo lectivo en curso).
(III) Rendimiento Académico (Promedio, Finales y Parciales).
Cada uno de estos ejes se asocian con una medida de importancia relativa o peso, que representa la incidencia sobre la estimación. A mayor peso, mayor incidencia. Dichas medidas se encuentran en etapa de análisis.

La Estimación de Riesgo de Deserción obtenida para cada alumno en particular se visualiza en el sistema mediante un semáforo de colores: verde (sin riesgo), amarillo (riesgo leve), naranja (riesgo moderado) y rojo (crítico), junto a su valor numérico representativo.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

El avance de la tecnología permitió generar volúmenes de datos cada vez más grandes y difíciles de comprender y analizar. El área educativa no escapa a esta realidad.
En líneas generales, las instituciones registran datos personales, académicos, familiares y económicos de cada alumno pero carecen de modelos que les permitan describir de manera general a sus estudiantes.

En el caso puntual del fenómeno de deserción estudiantil, caracterizar a los estudiantes de una institución académica aporta información no trivial y de utilidad para la gestión, posibilitando el establecimiento de políticas tendientes a mejorar el desempeño académico de los alumnos (Lanzarini, Charnelli, Baldino y Díaz, 2015) promoviendo asimismo aportes sustanciales en el área de *Tecnología Informática Aplicada a la Educación.*

Distintas áreas han tratado de dar soluciones a este problema. Las técnicas de visualización a través de representaciones gráficas, algunas de las cuales son sumamente sofisticadas, han contribuido significativamente a la exploración y entendimiento de estos conjuntos de datos (Koutek, 2003; Nielson y Shriver, 1990). Por su parte, la Minería de Datos reúne un conjunto de técnicas capaces de modelizar y resumir la información, facilitando su comprensión y ayudando a la toma de decisiones (Charnelli, Lanzarini, Baldino y Diaz, 2014; Witten, Frank y Hall, 2011). Surge como un campo de la Estadística y las Ciencias de la Computación dando respuesta al intento de descubrir patrones en conjuntos inmanejables de datos.
Tanto el Proyecto ESDEU como el Proyecto Indicadores se ubican en el Programa Tecnología Educativa y Enseñanza de la Ingeniería de la Secretaría de Ciencia y Tecnología de la UTN.

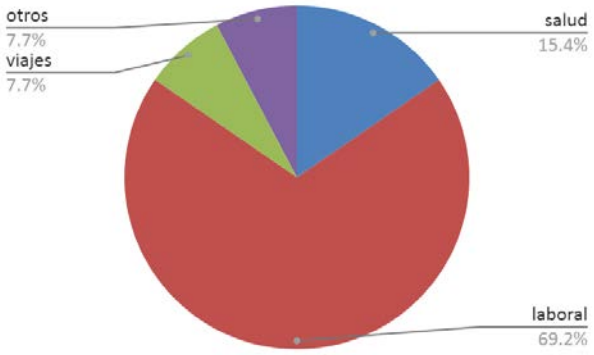
3. RESULTADOS OBTENIDOS / ESPERADOS

La primera versión del Sistema ESDEU se implementó en una prueba piloto para Ingresantes 2020 de la UTN La Plata. Para la misma se tomaron dos comisiones de prueba, una del turno mañana y una del turno noche.
Las mismas posibilitaron la identificación temprana de:

- alumnos con reiteradas inasistencias
- alumnos ausentes en las fechas de parciales
- alumnos que no aprobaron parciales

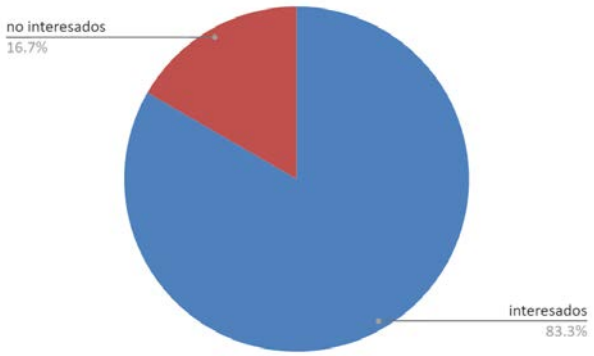
Esto permitió el contacto con los estudiantes antes de que pierdan contacto con la institución, posibilitando conocer las causas y situación particular de cada alumno, ofreciendo a su vez acompañamiento y apoyo académico.

En relación a las inasistencias reiteradas se identificaron las siguientes causas:



En donde se pone de relieve por sobre el resto, los temas laborales como causal sobresaliente.

Para estos estudiantes, que manifestaron su imposibilidad de concurrir a clases, se brinda la opción de continuar las cursadas a través de las aulas virtuales.
La relación interesados / no interesados se refleja en el siguiente gráfico:



Cabe destacar además, que la prueba posibilitó caracterizar a las poblaciones de estudiantes según el turno al que concurren, donde se concluyó que:

- Los alumnos del turno mañana presentan una edad promedio de 20 años frente a los 28 de la comisión vespertina.
- Los estudiantes de la mañana en general no trabajan durante el tiempo de cursada del Ingreso.
- Los alumnos del turno noche si lo hacen, con una carga horaria laboral igual o mayor a 20 horas semanales.
- Los estudiantes de la noche a diferencia de los de la mañana, poseen hijos o familiares a cargo.

Además de los contactos anteriores por inasistencias y rendimiento académico, el sistema alertó sobre perfiles de riesgo de deserción; esto permitió brindar apoyo y seguimiento presencial a alumnos que concurrieron a clases y rindieron oportunamente los exámenes pero presentaron un perfil socioeconómico de riesgo.

4. FORMACIÓN DE RECURSOS HUMANOS

Actualmente forman parte del equipo de trabajo un Director, un Coordinador, dos Docentes Investigadores de la Carrera de Ingeniería en Sistemas de Información (DISI) que se encuentran cursando sus carreras de postgrado, un tesista de postgrado y dos alumnos becarios de investigación.

Cuenta con una tesis de Maestría en Tecnología Informática Aplicada en Educación en curso y dos desarrollos de Prácticas Supervisadas (PS) de los estudiantes, necesarias para la obtención del título de grado de la carrera de Ingeniería en Sistemas de Información.

5. BIBLIOGRAFÍA

[1] CEA Centro de Estudios de la Educación Argentina. (2018). Universidad de Belgrano. Recuperado en junio de 2019 de http://boletin.ub.edu.ar/comunicaciones/flyers/cea_junio_2018.pdf

[2] Charnelli, E. Lanzarini, L. Baldino, G. Diaz, F. Determining the profiles of young people from Buenos Aires with a tendency to pursue computer science studies. XX Congreso Argentino de Ciencias de la Computación CACIC, 2014.

[3] Koutek, M. Scientific Visualization in Virtual Reality: Interaction Techniques and Application Development. Computer Graphics & CAD/CAM group, Faculty of Information Technology and Systems (ITS), Delft University of Technology (TU Delft), 2003.

[4] Lanzarini L, Charnelli E, Baldino G., Díaz J. (2015); "Selección de atributos representativos del avance académico de los alumnos universitarios usando técnicas de visualización. Un caso de estudio"; Revista TE&ET; no. 15. ISSN: 1850-9959. p. 42-50. 2015.

[5] Nielson, G. M.; Shriver, B. Visualization in scientific computing. IEEE Computer Society Press.United States of America, 1990.

[6] PEFI Plan Estratégico de Formación de Ingenieros 2012/2016; Secretaría de Políticas Universitarias, Ministerio de Educación, Presidencia de la Nación Argentina; http://pefi.siu.edu.ar/, 2012. Accedido en abril de 2018.

[7] Witten H., Frank E., Hall M. Data Mining: Practical Machine Learning Tools and Techniques (3er.edition). Morgan Kaufmann Series in Data Management Systems, Elsevier, 2011.

Visualización de datos en un Tablero de Comando aplicado a plataformas de educación a distancia en el nivel Superior

Tagarelli Sandra[1,2], Salinas Sergio[1], Caymes-Scutari Paola[3,4],Bianchini Germán[3], Ontiveros Patricia[2], Rotella Carina[2], Chirino Pamela[1], Galdamez Mariela[1]

[1]Laboratorio de Analítica de Datos, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza
[2]Laboratorio de Gobierno Electrónico, Departamento de Ingeniería en Sistemas de Información. Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza
[3]Laboratorio de Investigación en Cómputo Paralelo/Distribuido
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579
[4]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

stagarelli@frm.utn.edu.ar, s4salinas@gmail.com, pcaymesscutari@frm.utn.edu.ar, gbianchini@frm.utn.edu.ar, pontiveros@frm.utn.edu.ar, crotella@frm.utn.edu.ar, pamelaachirino@gmail.com, mariela.galdamez.16@gmail.com

RESUMEN

En la actualidad, la disposición masiva de datos en todos los campos del conocimiento representa un desafío muy grande para los tomadores de decisiones, tanto estratégicas como operativas. En particular, en muchas instituciones educativas de nivel superior se han desarrollado y aplicado plataformas de educación a distancia. Hay distintos paradigmas que se han ido aplicando para posibilitar esta relación distinta entre alumno y docente. Se requiere evaluar con indicadores definidos y bien diseñados en un tablero de comando. Un Tablero de comando, en este contexto del proceso de enseñanza- aprendizaje, captura y visualiza el seguimiento de las actividades de enseñanza, a través de indicadores, con el objetivo de conocer, reflexionar y tomar decisiones. Permite a los usuarios definir objetivos y analizar en qué medida los valores obtenidos de estos indicadores se acercan a los objetivos prefijados de rendimiento en esta nueva forma de interacción alumno-docente.(*Dashboard*) [9]. En este contexto, adquiere particular relevancia cómo se muestran los datos, es decir cuál es la mejor representación para los datos posible. Es claro que una representación visual de datos no apropiada puede obstaculizar su interpretación e inducir a errores. Por ello, en ocasiones es necesario recurrir a los conceptos que se extraen de las ciencias cognitivas sobre percepción humana. Encontrar la mejor

representación visual posible de los datos de un tablero de comando representa un desafío, y requiere un abordaje complementario entre los conceptos de percepción además de conocimientos de diseño visual.

Palabras clave: Tablero de Comando, Diseño visual, Percepción Humana, Proceso educativo, Visualización de datos.

CONTEXTO

La presente propuesta se trata de una tesis en desarrollo de Maestría en Inteligencia y Analítica de Negocios. El Programa de Maestría es una colaboración entre la Facultad Regional Mendoza de la Universidad Tecnológica Nacional y la Universidad de Ciencias Aplicadas de Neu-Ulm – Alemania.

Además, el presente trabajo se contextualiza en el Proyecto PID TEUTIME0007658TC, homologado y acreditado por la Universidad Tecnológica Nacional, debido a que los tesistas se encuentran trabajando en el proyecto mencionado.

Las universidades de todo el mundo han invertido recursos para mejorar la educación a través de *Learning Management Systems* (LMS, [4]). Un LMS es un software que admite el proceso de automatización de la enseñanza y el aprendizaje más allá de los límites geográficos.

La información generada durante el aprendizaje contribuye a mejorar los procesos de toma de decisiones tanto como a mejorar la experiencia general de aprendizaje a través de *Learning Analytics*.[10]

Learning Analytics es un proceso que comprende la medición de las variables del proceso de aprendizaje, para que recopilando estos datos, pueda realizarse el análisis de los mismos referidos a alumnos, docentes y todas las variables en donde se produce el proceso de aprendizaje.

Es posible mostrar los datos de LMS, en un tablero de comando adecuado (*Dashboard,* [9] eligiendo las medidas que representan el grado de logro de los objetivos.

En los últimos años, la tecnología ha puesto a disposición grandes volúmenes de datos y es urgente darle sentido tanto a dichos datos como a la historia subyacente. La visualización de datos y el hilo conductor de la visualización son cruciales para un mejor proceso de toma de decisiones.

1. INTRODUCCIÓN

Los seres humanos tienen arquitecturas y sistemas mentales que determinan cómo procesan la información que perciben; de igual modo, también las computadoras tienen su propia arquitectura de procesamiento de datos. Los avances en el área de psicología cognitiva han revelado esta arquitectura que subyace a las tareas de procesamiento de información humanas. Frecuentemente la comunicación entre un usuario y una computadora es visual, por lo que la información gráfica adquiere suma importancia en esta interacción. Incluso antes de las computadoras, el ser humano procesa información visual. [1].

A lo largo de los años, se han utilizado muchos enfoques para representar datos y estas diversas representaciones han revelado muchas de sus características que de otro modo no podrían apreciarse.

“Una interpretación efectiva e inequívoca de los datos puede depender críticamente de una representación visual cuidadosamente elegida”. [5]

En el pasado, los métodos para elegir la representación visual de los datos se basaban principalmente en la experiencia de un experto en los datos [5]. En cambio, los métodos actuales proponen enfoques que consideran la experiencia y asesoramiento de un "experto en visualización de datos".

Han surgido algunos paradigmas de visualización de datos, cada uno considerando variables relacionadas tanto con los interesados (*stakeholders*, [10]) como con los tipos de representaciones visuales apropiadas a las medidas o datos que deberían representar. [3]

Las visualizaciones pueden mostrar el contexto de los datos, la orientación, la tendencia, permiten realizar selecciones y proporcionar comentarios dinámicos y en tiempo real, por ejemplo, para analizar cambios, como datos multidimensionales. Algunos fenómenos científicos, ya sean atómicos, cósmicos y también comunes, se hacen visibles con la visualización científica. Además, es posible mostrar información abstracta, como datos estadísticos o financieros, incluso descubrir patrones de comportamiento en los datos [8].

Existe un principio considerado como básico en representación visual, expuesto hace muchos años y conocido como: *Visual Seeking Mantra*: Muestra el resumen primero, después zoom y filtro, y luego detalles [8]. Este principio constituyó un punto de partida para tratar de caracterizar las múltiples innovaciones de visualización de información que ocurren en los laboratorios de investigación, así como en la industria.

Otro paradigma para la visualización de datos ha propuesto agrupar taxonomía de tareas de visualizaciones de información. Este enfoque supone que los usuarios están viendo colecciones de elementos, donde los elementos tienen múltiples atributos.

La investigación relativamente reciente en el aprendizaje también introduce la necesidad de considerar objetivos y orientaciones, efectividad y eficiencia de los procesos de aprendizaje y, en consecuencia, el diseño del tablero de comando para facilitar la retroalimentación de esos objetivos referidos al análisis del aprendizaje en su conjunto. [6]. Un tablero (*Dashboard*) de estas características (LAD) puede ser utilizado tanto por estudiantes como por docentes. De acuerdo a [9], se pueden distinguir las siguientes partes que interactúan: alumno, docente, administrador, e investigador. En [2] generalizan a los usuarios en: mega nivel (gobernanza), macro nivel (institución), meso nivel (plan de estudios, maestro / tutor) y micro nivel (alumno).

De acuerdo con los trabajos e investigaciones recientes, hay investigadores que afirman no haber encontrado una publicación que brinde suficiente soporte teórico para la elección de representaciones de visualización para guiar tanto el diseño como el proceso de desarrollo [7].

En los últimos años, han surgido enfoques más integrales en representaciones visuales que proponen darle un hilo conductor a las representaciones visuales de datos.

"La visualización de datos, y la comunicación con datos en general, se encuentra en la intersección de la ciencia y el arte. Ciertamente hay algo de ciencia: mejores prácticas y pautas a seguir". [3]

Por todo lo anteriormente expuesto, una representación visual explota efectivamente la capacidad del sistema visual humano para reconocer la estructura espacial y los patrones de los datos. Y, por lo tanto, puede proporcionar la clave para la apreciación crítica e

integral de los datos, beneficiando así el posterior análisis, procesamiento y toma de decisiones.

Analizaremos el caso de UTN-FRM, una institución de educación superior que ha implementado Moodle, un LMS particular, y que en este caso constituye un caso de estudio. Proporcionaremos la representación visual más adecuada para *Moodle Dashboard*, (sistema de medición de desempeño de Moodle) en otras universidades e instituciones de nivel superior.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

El presente trabajo se enmarca en la rama de Analítica de Datos. Como disciplina, la Analítica de datos puede aplicarse a todas las disciplinas del desempeño humano. En este caso en particular, hablamos de Learning Analytics, debido a que se analizan variables relacionadas al proceso educativo en Educación Superior.

A su vez, también se contemplan ramas de los sistemas de educación a distancia (referidos en el texto como LMS), y por último conceptos relativos a control y medición de desempeño y seguimiento de variables. Tales variables se agrupan en Tableros de Comando, que proporcionan información tanto reactiva como proactiva, ya que además de analizar el pasado, nos permiten "anticiparnos" y tomar decisiones oportunas.

Otra línea de investigación considerada es la de Visualización de Información.

3. RESULTADOS ESPERADOS

Los *stakeholders* (docentes, alumnos y directivos) de la UTN-FRM requieren la definición de representaciones visuales adecuada para el Tablero de Educación a Distancia sobre Moodle. Se pretende proporcionar a la institución el diseño visual adecuado a los indicadores de desempeño definidos. Tanto la definición de indicadores como su representación visual respetarán las características distintivas de la Universidad, aun cuando se pueda aplicar a otras instituciones de nivel superior.

Representaciones visuales adecuadas del Tablero de Comando contribuyen a la mejora del proceso de toma de decisiones relacionadas a la plataforma de educación a distancia implementada (Moodle).

Cabe destacar también que la contribución principal que se busca es la transferencia de formación por parte de docentes-investigadores hacia docentes y alumnos de grado y al medio educativo (y de manera directa a las cátedras de la carrera Ingeniería en Sistemas de Información vinculadas como Sistemas de Gestión, Análisis de Sistemas, Modelado de procesos de negocio, Gobierno Electrónico y Computación Paralela).

4. FORMACIÓN DE RECURSOS HUMANOS

En el grupo de docentes participantes del proyecto se encuentran en desarrollo dos tesis de maestría directamente vinculadas a la temática. Una es la referida en el presente trabajo y la otra se refiere a la determinación de indicadores para medición del desempeño de objetivos para tablero de comando de la misma institución de educación superior y también referida a la plataforma Moodle.

Esto contribuye a la formación de los tesistas, pero también a los alumnos-becarios de grado participantes.

Entendemos además, que el desarrollo relativo a Analítica de Datos e Inteligencia de Negocios, en particular *Learning Analytics*, contribuirá a la

aplicación de conocimientos en una Institución real. Otro aporte será la formación de los usuarios y *stakeholders* de la Plataforma de Educación a distancia: permitirá a tomadores de decisiones apreciar relaciones no conocidas que muestran los datos o incluso datos que no se conocía. El análisis y organización de los datos constituye hoy un desafío en cualquier ámbito, dado que es muy grande la cantidad de información disponible, por lo que su organización e interpretación constituye un gran desafío para cualquier disciplina.

5. BIBLIOGRAFÍA

[1] Haber, R. N., & Wilkinson, L. (1982). PerceptualComponents-mcg1982030023.pdf, (May).

[2] Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. Technology, Knowledge and Learning, 19(1–2), 221–240.

[3] Nussbaumer Knaflic, C. (2015). Storytelling with data". A data visualization guide for business professionals. Hoboken, New Jersey: John Wiley & Sons

[4] Gorshenin, A. (Nov. 2018). "Toward Modern Educational IT-ecosystems: from Learning Management Systems to Digital Platforms". In: 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), pp. 1–5. DOI:10.1109/ICUMT.2018.

[5] Robertson, P. K. (1990). A methodology for scientific data visualisation: Choosing representations based on a natural scene paradigm, 114–123.

[6] Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2018). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. Computers in Human Behavior, 105512.

[7] Sedrakyan, G., Mannens, E., & Verbert, K. (2019). Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. Journal of Visual Languages and Computing, 50, 19–38.

[8] Shneiderman, B. (1996). Eyes have it: a task by data type taxonomy for information visualizations. IEEE Symposium on Visual Languages, Proceedings, 336–343.

[9] Klerkx, J., Verbert, K., & Duval, E. (2017). Learning Analytics Dashboards. Handbook of Learning Analytics, 143–150. https://doi.org/10.18608/hla17.012

[10] Chatti, Mohamed Amine, Dyckhoff, Anna Lea, Schroeder, U. Thüs, H., "A reference model for learning analytics". Int. J. Technology Enhanced Learning, Vol. 4, Nos. 5/6, 2012.

POR UN DISEÑO INCLUSIVO.

CASO DE USO EN UN MOOC DE ACCESIBILIDAD WEB

Javier Díaz, Alejandra Schiavoni, Paola Amadeo, Ivana Harari,
Alejandra Osorio, Federico Carrilao Ávila
LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.
Facultad de Informática. Universidad Nacional de La Plata
Calle 50 esq. 120, 2do Piso. Tel: +54 221 4223528
jdiaz@unlp.edu.ar , ales@info.unlp.edu.ar , pamadeo@linti.unlp.edu.ar, iharari@info.unlp.edu.ar,
aosorio@cespi.unlp.edu.ar, federicoca95@gmail.com

RESUMEN

En este artículo se presentan las líneas de investigación basadas en accesibilidad web y cursos masivos en línea (MOOC – Massive Open Online Courses) que se desarrollan en el LINTI, Laboratorio de Investigación en Nuevas Tecnologías Informáticas de la Facultad de Informática de la UNLP. Este tema está en consonancia con los recursos abiertos usados en educación y el uso de plataformas abiertas de e-learning, en los que se viene investigando desde hace varios años.

En esta etapa se está trabajando en la modificación y adaptación del MOOC sobre el tema de Accesibilidad Web, que se comenzó a impartir hace dos años. Las modificaciones surgen del resultado de etapas de análisis de diferentes aspectos de usabilidad y accesibilidad de la plataforma utilizada y de la estructura y organización del curso. Se realizan en forma continua evaluaciones y testeos automáticos con herramientas específicas y manuales a partir de usuarios con distintas clases de discapacidades. Las evaluaciones que se vienen realizando tienen en cuenta no sólo aspectos intrínsecos de la plataforma y del curso implementado, sino también la interacción de los participantes con las actividades, el contenido y los foros propuestos en el curso.

Palabras clave: MOOC, Recursos educativos abiertos, accesibilidad web.

CONTEXTO

El proyecto descripto en este artículo se desarrolla en el Laboratorio de Investigación en Nuevas Tecnologías Informáticas, LINTI de la Facultad de Informática de la UNLP y está enmarcado en el proyecto 11-F020 “Internet del futuro: Ciudades digitales inclusivas, innovadoras y sustentables, IoT, ciberseguridad y espacios de aprendizaje del futuro”, acreditado en el marco del Programa de Incentivos, bajo la dirección del Lic. Javier Díaz.

En la Facultad de Informática se utilizan plataformas de código abierto para las gestiones académicas desde hace más de 15 años, incluyendo sistemas de de gestión de aprendizaje como Moodle, repositorios abiertos y sistemas de gestión administrativa, como SIU Guaraní. A partir del surgimiento de los cursos abiertos masivos en línea, MOOC, se comenzó a utilizar la plataforma edX para implementar el curso sobre Accesibilidad Web, que va por su tercera edición.

Es importante mencionar que el tema de accesibilidad web se viene trabajando en la Facultad desde el año 2002, y dada su importancia se incorporó esta temática en el plan de estudios de las carreras que se dictan en la institución, a través de la asignatura Diseño Centrado en el Usuario. Sobre el tema, en el año 2013 se escribió el libro de cátedra "Guía de recomendaciones para diseño de software centrado en el usuario", disponible en el SEDICI, Repositorio Institucional de la UNLP, que tiene actualmente 10.000 descargas [1]. También se institucionalizó su abordaje mediante la creación de una Dirección de Accesibilidad, se desarrollaron tesinas, trabajos de cátedra al respecto y diferentes proyectos de extensión acreditados por la Universidad Nacional de La Plata. Los proyectos de los últimos años son "Trabajando por una Web Inclusiva" (2017) y "Por una Web Inclusiva" (2018), [2].

1. INTRODUCCIÓN

Los MOOCs han provocado un cambio significativo en el aprendizaje en línea, haciendo que los recursos educativos sean abiertos y estén disponibles y puedan ser accedidos por un gran número de personas. Si bien pueden ser considerados como una extensión de los enfoques de aprendizaje electrónico existente, los MOOCs ofrecen la oportunidad de repensar acerca de los nuevos modelos de educación superior en línea y gratuita, en donde los cursos están abiertos a cualquier usuario de internet [3].

Por lo tanto, la esencia de los MOOCs en relación al uso de recursos abiertos y a la alta disponibilidad de ellos, debe facilitar el aprendizaje a estudiantes que tienen algún tipo de discapacidad. Los MOOCs llegan a audiencias diseminadas en todo el mundo, por lo que es necesario considerar los potenciales estudiantes que accedan a ellos. Es por esto que enfocar la atención en la diversidad del espectro educativo se convierte en un imperativo moral y social [4].

El diseño de un MOOC es un proceso complejo debido a características propias, que muestran divergencias en comparación a los cursos tradicionales. Independientemente de su carácter abierto, el acceso a los cursos y sus plataformas pueden llegar a ser una dificultad adicional [5] [6].

Al crear un curso on-line, no basta con digitalizar los contenidos, y volcarlos en una plataforma, ya que los potenciales alumnos tienen diferentes necesidades y motivaciones, que los conducen por distintas trayectorias en el proceso de aprendizaje [7]. En muchos casos, hay suficientes materiales y ejercicios a disposición de los alumnos desde que ingresan en el curso, que les permiten gestionar su aprendizaje según su propio ritmo. En muchos casos, la incorporación de elementos interactivos (tests o evaluaciones) y contenido audiovisual agrega desafíos a los propios requerimientos de accesibilidad [6].

Resultan muy importantes los aspectos motivacionales para evitar el abandono de los cursos y favorecer el aprendizaje. Esto le exige a los docentes habilidades tecnológicas, didácticas y tutoriales que se deben poner en juego al momento de diseñar el curso en forma integral. Los MOOCs imponen cambios metodológicos, diseños colaborativos e interactivos, materiales ubicuos y atractivos que faciliten y promuevan la navegación y el descubrimiento, en entornos diseñados para tal fin.

No sólo las personas con discapacidades permanentes son las que se benefician con un curso accesible. Cualquiera de nosotros podemos experimentar alguna discapacidad temporaria o ambiental en algún

momento de nuestras vidas. Por ejemplo dificultad en distinguir los colores debido a ciertas condiciones de luminosidad (poca visión), o impedimentos para escuchar claramente en ambientes ruidosos, si no contamos con auriculares [8].

Para promover la accesibilidad, el Consorcio W3C creó la Web Accessibility Initiative (WAI) para crear los lineamientos para el contenido web, las herramientas de autoría y los navegadores y otras plataformas. Estos lineamientos son un buen punto de partida para entender las necesidades de accesibilidad de los usuarios. Si bien existen herramientas específicas para evaluar la accesibilidad, los comentarios y opiniones de los usuarios también tienen un valor muy importante para el establecimiento de pautas de diseño de MOOCs accesibles.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Las tareas de investigación, desarrollo e innovación que se vienen llevando a cabo en este proyecto incluyen los temas relacionados a e-learning y accesibilidad web. Respecto a e-learning se está trabajando con plataformas de código abierto, tanto sistemas de gestión de aprendizaje (LMS) como herramientas para el soporte de MOOCs. También se trabaja lo relacionado a la integración de distintas plataformas que se utilizan en la gestión académica. El tema de accesibilidad web se trata en forma transversal, evaluando el cumplimiento de las normas respectivas en cada uno de los casos.

En esta etapa se está trabajando con la plataforma OpenEdX [9] para el soporte de MOOCs. Se eligió dicha plataforma por las facilidades de instalación y configuración, por los módulos disponibles y la amplia comunidad de usuarios que contribuyen al proyecto. Es una de las más populares en el mercado, actualmente es el segundo proveedor en el mundo en cantidad de cursos ofrecidos. Provee toda la funcionalidad básica de las herramientas de MOOCs, en lo relativo a la creación de contenido, gestión de usuarios, confección de evaluaciones y utilización de foros y wikis.

Usando esta plataforma se implementó un MOOC sobre Accesibilidad Web, que se puso en producción en septiembre de 2017, y se encuentra actualmente disponible en https://actividades.linti.unlp.edu.ar. La del 2019 fue la tercera edición del curso, que va teniendo mayor cantidad de alumnos a lo largo de los años. Sobre la plataforma en cuestión se aplican los trabajos de investigación que se vienen realizando sobre el tema de accesibilidad Web, estudio de las normas de accesibilidad y su aplicación en el desarrollo de sistemas y sitios Web. OpenEdx cuenta con una línea de trabajo específica sobre la temática [10].

Continuando con las líneas planteadas en WICC 2019 [11], en esta etapa el foco de estudio se relaciona con la evaluación y adaptación del caso de uso desarrollado a partir de las experiencias llevadas a cabo durante las tres ediciones. La evaluación se basa en analizar el impacto del contenido y de las actividades y ejercitaciones planteadas a través de las experiencias y devoluciones de los estudiantes participantes, en algunos casos fueron personas con alguna discapacidad. A partir de la evaluación realizada, se procede a realizar las modificaciones correspondientes, tanto en la estructura y contenido del curso, como en la funcionalidad de la plataforma. El objetivo es establecer un punto inicial en

la tarea de diseñar y crear cursos masivos de calidad.

Los criterios a tener en cuenta para testear la accesibilidad se aplican a la naturaleza del contenido, estructuración del curso, la interfaz del usuario, herramientas sociales hasta analíticas de aprendizaje y estrategias de evaluación [12].

Se consideran diversas técnicas que incluyen herramientas automáticas y evaluaciones manuales a través de usuarios. En la evaluación del contenido se consideran varios aspectos para lograr que el estudiante se sienta involucrado y motivado en el aprendizaje. Para lograr esto se incluyen diferentes opciones que reflejen los intereses de los alumnos, estrategias para afrontar las tareas, alternativas de auto evaluaciones y actividades adecuadas según los objetivos. Además, se trata de proveer múltiples formas de representar el contenido, ya que varía la forma en que cada estudiante percibe y comprende el material propuesto. Se realiza una reestructuración de la organización de las unidades, planteando diferentes secuencias de acuerdo al perfil del alumno. Esto permite personalizar el trayecto de aprendizaje según sus necesidades. Se analizan en detalle las actividades planteadas, teniendo en cuenta que deben estimular su capacidad de relacionar directamente el contenido con su contexto personal, social o laboral, creando casos concretos de aplicación.

El uso de espacios de comunicación entre los alumnos representa una de las actividades más destacadas y es importante para fomentar la interacción y la generación de conocimiento compartido, logrando una forma de aprendizaje colaborativo.

En relación a las evaluaciones, se considera importante la incorporación de auto evaluaciones, que le permite al alumno tener respuestas inmediatas de su desempeño por parte del sistema. También se incluyen actividades a realizar por parte del estudiante para aplicar los conceptos aprendidos. En este punto se analizan las tareas incluidas en cada unidad, en cuanto a la comprensión de las consignas, el grado de cumplimiento en tiempo y forma por parte de los alumnos y los resultados obtenidos, tanto en las evaluaciones por pares como en las tradicionales corregidas por el profesor.

Respecto a las características de la plataforma, se evalúan cuestiones de usabilidad y facilidad de uso en cuanto a la localización y acceso a los contenidos; aspectos de instalación y configuración y cuestiones de accesibilidad. Esto requiere aplicar diferentes estrategias pedagógicas para poder, en el proceso de enseñanza a distancia, concientizar a los estudiantes, proponer el ponerse en el lugar del otro, transmitir empatía y valores éticos, incentivar la investigación de la problemática de los usuarios con discapacidad y trabajar por y para garantizar una experiencia del usuario inclusiva, promover la colaboración para la búsqueda conjunta de soluciones accesibles y formas de aplicarlas en las producciones informáticas que se desarrollen e instar a la evaluación y comprobación continua de la accesibilidad. Además, en la evaluación de los aspectos de accesibilidad se considera tanto las expectativas de los alumnos como el punto de vista y la experiencia de los docentes. Los enfoques utilizados al evaluar cada aspecto de la plataforma se traducen en modificaciones de módulos internos de manera de adaptarlos a los requerimientos necesarios con el objetivo de cumplir con los criterios y normas establecidas.

La implementación del curso nos sirve como experiencia y punto inicial para

transformar en MOOCs diversos cursos en línea que hoy se encuentran en un LMS tradicional. Este proceso involucra analizar y evaluar también nuevas plataformas de base, nuevos medios de comunicación, contenidos y formatos. En esta línea se realiza un trabajo interdisciplinario con diseñadores visuales, expertos en comunicación, contenidos y educadores que permite generar una experiencia y un conocimiento enriquecedor, extensible a otras temáticas y líneas de investigación del LINTI.

3. RESULTADOS Y OBJETIVOS

Según las líneas de trabajo descritas, se plantean los siguientes objetivos:

- Estudiar los distintos módulos de la arquitectura de la plataforma de base para realizar las modificaciones necesarias.
- Analizar metodologías de diseño de MOOCs aplicadas en universidades del mundo, analizando sus ventajas y posibles adecuaciones al ámbito local.
- Realizar encuestas y entrevistas individuales con diferentes personas con discapacidad que interactúen con la herramienta.
- Evaluar el curso creado desde la perspectiva docente, institucional y técnica, a partir del estudio de distintas experiencias, analizando las opiniones de los alumnos involucrados.
- Como base de la planificación estipulada, analizar los resultados obtenidos en cada etapa, corrigiendo posibles falencias en un ciclo de mejora continua.
- Redefinir secuencias de aprendizaje para que se adapten a diferentes perfiles de alumnos.
- Llevar a cabo estudios comparativos entre la experiencia del MOOC y los cursos en un LMS tradicional, que permitan aportar nuevos conceptos a los entornos de aprendizaje y enseñanza que se generan con estas herramientas.
- Realizar un análisis continuo de las pautas de accesibilidad, a través de evaluaciones heurísticas utilizando herramientas específicas y empíricas teniendo en cuenta el perfil del usuario, ya sea estudiante, docente, tutor o administrador. Evaluar la posibilidad de intervención y participación de una persona con discapacidad en los diferentes roles.
- Definir técnicas de diseño inclusivo, a partir de las evaluaciones realizadas y de la información recolectada.
- Utilizar las técnicas definidas en futuros cursos, a partir del caso de uso desarrollado.
- Replantear componentes del proceso de aprendizaje, relacionadas a la naturaleza del contenido y al diseño temporal, a fin de adaptar el curso según las evaluaciones realizadas.
- Medir el grado de aceptación del curso por parte de los participantes y realizar análisis de datos considerando distintas variables como formación, edad, género, procedencia, participación y rendimiento, entre otros.
- Definir y establecer pautas de diseño y construcción de MOOCs que se utilicen como punto de partida para la creación de cursos

masivos sobre temas que se investigan y sobre los cuales se viene trabajando en el LINTI.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo se encuentra formado por tres profesoras de amplia trayectoria en el campo de la investigación, que trabajan en el área de ambientes virtuales de aprendizaje y accesibilidad web. Además, dos alumnos de la carrera de Licenciatura en Informática de la Facultad de Informática.

La participación en eventos de la especialidad, ha permitido seguir estableciendo canales de comunicación con otros investigadores que trabajan en las mismas áreas.

5. REFERENCIAS

[1] Díaz, J., Amadeo, P., Harari, I. "Guía de recomendaciones para diseño de software centrado en el usuario". SEDICI, Repositorio Institucional de la UNLP, EDULP, ISBN: 978-950-34-1030-1, 2013.

[2] Dictamen Proyectos de Extensión Acreditados en la UNLP. https://unlp.edu.ar/proyectosext/proyectos_de_extension_acreditados_y_subsidiados-4708

[3] González, C., Collazos, C., García, R. "Desafío en el diseño de MOOCs: incorporación de aspectos para la colaboración y la gamificación". RED - revista de Educación a Distancia. Núm. 48. Art. 7. 30-Ene-2016. DOI: 10.6018/red/48/7

[4] Iniesto, F, Covadonga, R. "YourMOOC4all: a MOOCs inclusive design and useful feedback research project". Publicado en 2018 Learning With MOOCs (LWMOOCS). Madrid, España 26-28 Sept. DOI: 10.1109/LWMOOCS.2018.8534644. IEEE Xplore Digital Library.

[5] Sánchez Gordeon, S., Luján-Mora, S. "Research challenges in accesible MOOCs: a systematic literatura review 2008-2016". Universal Access in the Information Society, 1-15, 2017.

[6] Littlejohn, A., Hood, N., Milligan, C., Mustain, P. "Learning in MOOCs: Motivations and self-regulated learning in MOOCs". Internet High. Educ., vol. 29: 40–48, 2016.

[7] Montoro G., Muruzábal O., Sandoval G., Wee C. "7 pasos para diseñar un MOOC de calidad: Una propuesta para la colaboración entre profesores y diseñadores de aprendizaje". Actas de la Jornada de MOOCs en español en EMOOCs 2017 (EMOOCs-ES).

[8] Sánchez Gordon, S., Luján.Mora, S. "Adaptive Content Presentation Extension for Open edX. Enhancing MOOCs Accessibility for Users with Disabilities". ACHI 2015: The Eighth International Conference on Advances in Computer-Human Interactions, 2015

[9] Open edX Portal | Open Source MOOC Platform https://open.edx.org

[10] Accessibility Features | Open edX Portal | Open Source MOOC Platform https://open.edx.org/features-roadmap/accessibility/all

[11] Díaz, J., Schiavoni, A., Amadeo, P., Harari, I., Osorio, A., Carrilao Ávila, F. "Experiencia de un MOOC: el desafío de contar con aspectos de usabilidad y accesibilidad de la plataforma". WICC 2019, XXI Workshop de Investigadores

en Ciencias de la Computación, San Juan, Argentina, 25-26 Abril, 2019.

[12] Yousef, A. M. F., Chatti, M. A., Schroeder, U., Wosnitza, M. "What Drives a Successful MOOC? An Empirical Examination of Criteria to Assure Design Quality of MOOCs," 2014 IEEE 14th International Conference on Advanced Learning Technologies, Athens, 2014, pp. 44-48. doi: 10.1109/ICALT.2014.23

COMPUTACIÓN UBICUA APLICADA AL APRENDIZAJE: IMPLEMENTACIÓN EN EL CURSO DE INGRESO UNIVERSITARIO

Álvarez Margarita; Durán Elena, Únzaga Silvina, Fernández Reuter Beatriz, González Gabriela, Montoto Sebastián y Quintana Cancinos, Fernando
Instituto de Investigaciones en Informática y Sistemas de Información
Universidad Nacional de Santiago del Estero (UNSE)
e-mail: { alvarez; eduran, sunzaga, bfreuter, ggonzalez}@unse.edu.ar, sebastian.montoto@gmail.com, f.quintana88@yahoo.com.ar

RESUMEN

Integrar tecnologías en los procesos de aprendizaje de los estudiantes que ingresan a la universidad se vuelve desafiante para favorecer la colaboración, conectar espacios formales, no formales e informales; la ubicación del aprendizaje, tanto dentro como fuera del aula; adaptar los contenidos y actividades a las características de los alumnos y su contexto, etc. Es así, que se considera como alternativa válida el uso del modelo educativo de aprendizaje ubicuo. En este modelo, los ambientes de aprendizaje ubicuo superan las limitaciones de un ambiente de aprendizaje tradicional, y lo extienden haciendo realidad la idea de aprendizaje en cualquier lugar y momento.

Por lo tanto, los contenidos educativos pueden accederse y las interacciones pueden concretarse donde los estudiantes lo necesiten sin restricción de espacio ni de tiempo.
En este trabajo se presenta la línea de investigación: Desarrollos de sistemas para el aprendizaje ubicuo; en particular uno de los principales resultados obtenidos. Se trata de una aplicación ubicua para el curso de ingreso a la universidad para carreras de informática. La propuesta se basa en aprendizaje ubicuo con el uso de objetos de aprendizaje (OA) que se ofrecen de forma personalizada al estudiante. La aplicación ofrece servicios de apoyo para el aprendizaje.

***Palabras clave*:** Aprendizaje Ubicuo, Objetos de Aprendizaje, Personalización.

CONTEXTO

En este trabajo se presenta la línea de investigación: Desarrollos de sistemas para el aprendizaje ubicuo, que forma parte del proyecto "Métodos y Técnicas para desarrollos de Aplicaciones Ubicuas", correspondiente a la convocatoria 2016 de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SICYT - UNSE). El período de ejecución del proyecto es 2017-2020. La línea de investigación presentada en este trabajo, es una continuación de otra iniciada en el 2012, en el proyecto "Sistemas de información web personalizados, basados en ontologías, para soporte al aprendizaje ubicuo". La misma tiene como finalidad realizar propuestas de técnicas, métodos y estrategias para el diseño y construcción de aplicaciones de aprendizaje ubicuo.

1. INTRODUCCIÓN

Con el paso del tiempo han surgido diversas modalidades de enseñanza y aprendizaje, como son la modalidad presencial, a distancia y aprendizaje en línea (e-Learning). Con la aparición de la Web 2.0 y las Tecnologías de la Información y la Comunicación se destacan otras modalidades de enseñanza como es el aprendizaje que combina lo presencial con lo virtual (b-Learning), aprendizaje móvil (m-Learning) y el aprendizaje ubicuo (u-Learning), que surge como una alternativa para mejorar los procesos de aprendizaje ya que las personas acceden a mejores experiencias de aprendizaje en sus ambientes de la vida diaria (Hidalgo Pérez et al., 2015).
El aprendizaje ubicuo designa al conjunto de actividades formativas, apoyadas en tecnología, que están accesibles en cualquier lugar y desde cualquier dispositivo (Saadiah y Kamarularifin, 2010). El desarrollo de aplicaciones de u-learning requiere tener en cuenta, las necesidades del usuario y considerar el conjunto complejo de múltiples formas de movilidad, diferentes tecnologías móviles, diversidad de transportistas, así como distintas situaciones de aprendizaje que

se pueden presentar. En consecuencia, se hace necesario que estas aplicaciones manejen información de las características diferenciadas de los usuarios, de la tecnología que se utiliza y los contextos en los que se concreta el aprendizaje.

Para que las aplicaciones informáticas, que dan apoyo al aprendizaje ubicuo, gestionen eficientemente las problemáticas mencionadas se requiere de una arquitectura adecuada, para lo cual, en la línea de investigación que aquí se presenta, hemos propuesto una Arquitectura dirigida por modelos ontológicos para aplicaciones de aprendizaje ubicuo (Durán et al, 2014a). Sobre la base de esa arquitectura hemos diseñado servicios personalizados para ofrecer a los estudiantes en contextos ubicuos. Estos servicios han sido implementados en una aplicación de apoyo al aprendizaje de los estudiantes en el curso de ingreso a la universidad.

La Universidad Nacional de Santiago del Estero (UNSE) y, en particular la Facultad de Ciencias Exactas y Tecnologías, realiza el curso de ingreso a las diferentes carreras que ofrece. Entre las actividades de capacitación se encuentra el Taller de Ambientación Universitaria y los cursos disciplinares de cada carrera. El objetivo del taller es presentar a los estudiantes aspectos organizativos relacionados con la dinámica peculiar de la institución universitaria. Se aborda información general de la UNSE y de la facultad; los datos académicos y administrativos de la vida universitaria: sistema de gestión de estudiantes, planes de estudio, sistema de correlativas, horarios de clases, etc. Estas instancias de aprendizaje se caracterizan por la gran cantidad de temas que se abordan, la gran cantidad de estudiantes, los niveles de conocimiento, la capacidad de estudiar y la gran cantidad de secciones / sectores de la universidad donde se realizan diferentes funciones y actividades.

Dadas estas características, hemos considerado relevante apoyar el aprendizaje de los ingresantes desde un nuevo modelo y poner a prueba los resultados obtenidos en la línea de investigación.

2. DESCRIPCIÓN DE LA LINEA DE INVESTIGACION

Con esta línea de investigación se busca realizar contribuciones teóricas y metodológicas en el campo de los sistemas de información de apoyo al aprendizaje con el fin de favorecer el desarrollo de conocimiento científico-tecnológico de relevancia, realizando propuestas de técnicas, métodos y estrategias para el diseño y construcción de aplicaciones de aprendizaje ubicuo.

Para alcanzar lo expuesto anteriormente, y sobre la base de la Arquitectura dirigida por modelos ontológicos para aplicaciones de aprendizaje ubicuo (Durán et al, 2014a), hemos trabajado en el diseño de los diferentes módulos de software propuestos por la arquitectura: de Registración de Usuario, de Interfaz de Usuario, de Adquisición de Contexto Ubicuo, de Análisis de Petición de Usuario, de Personalización y de Mantenimiento de los Modelos Ontológicos. Estos módulos interactúan con diferentes ontologías (Durán, et al., 2016) que representan e integran la información del estudiante, del dominio y del contexto.

Una vez diseñados los módulos de los diferentes servicios, se crearon métodos para optimizar el trabajo de estos (Método para la generación de caminos de aprendizaje; Método para recomendar expertos al estudiante, Método para la formación de grupos, etc.)

Como una forma de validar la arquitectura se diseñó una aplicación de apoyo al curso de ingreso universitario. Esta aplicación ofrece los siguientes servicios al estudiante: recomendación personalizada de los OA de un Puntos de Interés (PI) sensado por el estudiante; recomendación de caminos de aprendizaje de acuerdo al objetivo pedagógico seleccionado por un estudiante; recomendación de los PI más cercanos a la ubicación del estudiante de acuerdo a sus conocimientos previos y disponibilidad horaria; recomendación de expertos para que el alumno realice consultas considerando la cercanía, el grado de experiencia del experto en consultas anteriores o la disponibilidad para realizar consultas online; recomendación de un grupo de compañeros para realizar una tarea de aprendizaje ubicua y colaborativa (Únzaga et al., 2015).

Desarrollamos, también, una estrategia para generar las recomendaciones personalizadas en el módulo de personalización que forma parte de la arquitectura mencionada (Durán et al., 2014b; Álvarez et al., 2015)

3. RESULTADOS OBTENIDOS

Los resultados obtenidos se detallan a continuación:

a. *Diseño y desarrollo del módulo de personalización que recomienda OAs de un PI específico:* el módulo infiere los OAs para el PI sobre el cual el estudiante quiere aprender. Para ello, el estudiante sensa el código QR asociado

al PI usando su dispositivo móvil. El módulo de acuerdo al estilo de aprendizaje del alumno infiere los OAs. La aplicación despliega en la pantalla del dispositivo la lista personalizada de OAs para que el estudiante seleccione el de su interés (Únzaga et al., 2015).

b. *Método para la generación de caminos de aprendizaje ubicuos:* el aprendizaje ubicuo permite a los estudiantes aprender en cualquier momento y lugar. Sin embargo, sin la asistencia adecuada, es posible que los estudiantes no logren sus objetivos de aprendizaje. En tal sentido, se está desarrollando un método híbrido para la generación de caminos de aprendizaje ubicuos personalizados utilizando técnicas de planificación, ontologías y agentes inteligentes. El método permite que el estudiante seleccione un objetivo de aprendizaje vinculado con el curso de ingreso y la aplicación lo guía en su proceso de aprendizaje mediante la generación de rutas de aprendizaje. El método permite que no existan demoras para determinar la siguiente mejor acción que el alumno debe tomar en el camino del aprendizaje. Además, a través de un mecanismo de monitoreo, el camino generado siempre se adapta al contexto del alumno. La replanificación de la ruta es necesario solo cuando el contexto del alumno cambia y se dificulta el logro de una meta de aprendizaje dentro del tiempo establecido **(**Fernández Reuter, et al., 2017).

c. *Diseño y desarrollo de recomendación de PIs cercanos de acuerdo a la ubicación actual del estudiante:* el Módulo detecta que el estudiante se encuentra dentro de la sede de la Universidad a partir de evaluar las coordenadas de su ubicación, que éste lee de su dispositivo. Luego, infiere los PIs más cercanos al estudiante y aplicando personalización basada en el perfil del usuario, recomienda una lista final de PIs candidatos de acuerdo con los conocimientos previos que tiene el estudiante y con los PI no realizados hasta ese momento (Montoto et al., 2017; Montoto, 2018).

d. *Método para recomendar expertos a los estudiantes:* se desarrollo un método que genera, por un lado, recomendaciones de expertos en el tema considerando las experiencias de estos expertos con otros estudiantes, disponibilidad y cercanía física con el estudiante para realizar consultas presenciales, y por otro lado, recomienda expertos en el tema considerando si están disponibles para realizar consultas online. La aplicación le muestra al estudiante las dos listas obtenidas para que escoge con quien comunicarse (Durán y Álvarez, 2017).

   El método fue desarrollado mediante un modelo multiagente. Genera recomendaciones de tutores sobre el tema basado en las experiencias de los tutores con otros estudiantes, su disponibilidad y su proximidad física, monitorea la interacción del alumno en el entorno de aprendizaje, detecta problemas de aprendizaje y ofrece ayuda personalizada. (Fernández Reuter et al., 2018).

   Para el método descripto se realizaron experimentos para validarlo mediante técnicas de simulación. El valor de precisión obtenido demuestra que el método produce resultados que se ajustan a su propósito (Durán y Álvarez, 2018).

e. *Método para la formación de grupos en ambientes ubicuos cuando el estudiante requiere realizar una tarea en forma colaborativa*: la formación de grupos de aprendizaje eficaces representa uno de los principales factores de éxito en el Aprendizaje Colaborativo. Sin embargo, las características distintivas de los alumnos hacen que la operación de formar grupos de aprendizaje adecuados sea una tarea difícil. Además, con el surgimiento del aprendizaje ubicuo, se suma la necesidad de considerar las condiciones contextuales que pueden influir en el éxito de un aprendizaje colaborativo y situado. Para ello, se ha desarrollado un método para generar recomendaciones de compañeros con los que un estudiante puede conformar un grupo de trabajo para desarrollar una tarea colaborativamente en un ambiente de aprendizaje ubicuo. El método tiene como fortaleza que contempla en el agrupamiento los parámetros situacionales: la ubicación del estudiante y de sus compañeros, nivel de conocimiento de los estudiantes, tiempo disponible para realizar la tarea y nivel de interactividad de cada estudiante. Esto permite obtener grupos homogéneos en cuanto al nivel de conocimiento de sus integrantes y heterogéneos en cuanto al nivel de interactividad. Otra fortaleza del método es considerar la cercanía de los integrantes para el agrupamiento. Además, se impone la restricción de recomendar un grupo que cumpla con la condición de ser pequeño (no más de cuatro integrantes) para eficientizar la tarea, y heterogéneo en cuanto al nivel de interactividad (Álvarez, et al., 2017; Quintana Cancinos, 2019).

f. *Diseño y construcción de objetos de aprendizaje:* para el Taller de Ambientación Universitaria y Contenidos Disciplinares específico se han desarrollado OAs, que incluyen principalmente texto, imágenes, vídeos y, actividades prácticas y de evaluación que incorporan vídeos interactivos, preguntas de selección múltiple, verdadero-falso, ordenación de listas, etc. (Durán, et al., 2019).

4. FORMACIÓN DE RECURSOS HUMANOS

El desarrollo de la línea de investigación presentada facilita la formación de recursos humanos de la Facultad de Ciencias Exactas y Tecnologías de la UNSE. En este sentido, dos integrantes están desarrollando su Tesis de Doctorado en Ciencias de la Computación sobre temáticas del proyecto.

Otra integrante del proyecto, ha elaborado su Trabajo Final Integrador de la carrera de postgrado Especialización en Enseñanza de la Tecnología de la UNSE y está desarrollando su tesis de la Maestría en Informática Educativa de la UNSE. También, tres alumnos de la carrera de Licenciatura en Sistemas de Información están cumpliendo con becas de investigación y desarrollando sus trabajos finales de grado.

Además, con el desarrollo de este proyecto se está afianzando el grupo de investigación en las temáticas del proyecto: aprendizaje ubicuo, personalización y ontologías, lo que contribuye a una mejora en el fondo de conocimiento disciplinar disponible, no sólo a nivel local sino también regional y nacional.

5. BIBLIOGRAFÍA

Álvarez, M., Únzaga, S. y Durán, E. (2015). Recomendaciones Personalizadas para Aplicaciones de Soporte al Aprendizaje Ubicuo. XXI Congreso Argentino de Ciencias de la Computación. CACIC 2015. Junín. Argentina. 2015.

Álvarez, Margarita M; Únzaga Silvina I. y Durán Elena B. (2017). Método para generar recomendaciones personalizadas para integrar grupos de aprendizaje ubicuo y colaborativo. XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017). Libro de actas. Pag. 363- 372. ISBN 978-950-34-1539-9. Compiladores: De Giusti, Armando Eduardo y Pesado, Patricia Mabel. Disponible en: http://sedici.unlp.edu.ar/handle/10915/63019 La Plata. Prov. Buenos Aires. 9 al 13 de octubre.

Durán, E.; Álvarez, M. y Únzaga, S. (2014a). Ontological model-driven architecture for ubiquitous learning applications. EATIS 2014 - 7th Euro American Association on Telematics and Information Systems. Valparaíso (Chile). 2 al 4 de Abril de 2014.

Durán, E.; Alvarez, M. y Unzaga, S. (2014b). Design of a Personalization Module for U-learning application. Fifth International Conference on Advances in New Technologies, Interactive Interfaces and Communicability (ADNTIIC 2014), ALAIPO, Huerta Grande, Córdoba, Argentina, ISBN 978.88.96.471.37.1.

Durán, E., Álvarez, M. y Únzaga S. (2016). Modelo ontológico para personalizar aplicaciones de aprendizaje ubicuo. EATIS 2016. Colombia. 27-29 de abril de 2016.

Durán Elena y Álvarez Margarita (2017). Method for Generating Expert Recommendations to Advise students on Ubiquitous Learning Experiences. Conferencia Internacional de la Sociedad Chilena de Ciencia de la Computación (SCCC 2017). Arica, Chile. 16 al 20 de Octubre de 2017.

Durán, Elena y Alvarez, Margarita (2018). Recommendations of experts to advise students on ubiquitous learning experience: method and validation. IEEE Latin America Transaction. ISSN: 1548-0992. Vol 16 No 9. pp: 2314-2320 (2018): Special Issue on New Trends in Electronics. Disponible en: https://www.inaoep.mx/~IEEElat/index.php/transactions/article/view/ 5

Durán, E., Álvarez, M., Únzaga, S. y Montoto, S. (2019). Pedagogical proposal with ubiquitous learning for an admission course of careers computer. ICERI2019 Proceedings. 12th annual International Conference of Education, Research and Innovation Seville, Spain. 11-13 November, 2019. ISBN: 978-84-09-14755-7 / ISSN: 2340-1095 doi: 10.21125/iceri.2019. Publisher: IATED. Pages: 8641-8651.

Fernández Reuter, Beatriz; Durán, Elena y Amandi, Analía (2017). Designing a Hybrid Method for Personalized Ubiquitous Learning Paths Generation. Conferencia Internacional de la Sociedad Chilena de Ciencia de la Computación (SCCC 2017). Arica, Chile. 16 al 20 de Octubre de 2017.

Fernández Reuter, Beatriz, Álvarez Margarita, Duran Elena y González Gabriela (2018). Multi-agent system model for tutor recommendation in ubiquitous learning

environments. Proceedings First Workshop on Advanced Virtual Environments and Education - WAVE2 2018. Pag. 10-17. Florianópolis, Brasil, del 04 al 05 de Octubre de 2018. DOI: 10.5753/wave.2018.1. Disponible en https://mafiadoc.com/wave2-2018_5c1e5c4f097c4776518b464c.html.

Hidalgo Pérez Sandra Elizabeth, Orozco Aguirre María del Sol y Daza Ramírez Marco Tulio (2015). Trabajando con Aprendizaje Ubicuo en los alumnos que cursan la materia de Tecnologías de la Información. Revista Iberoamericana para la Investigación y el Desarrollo Educativo. Vol. 6, Núm. 11 Julio - Diciembre 2015. ISSN 2007 - 7467

Montoto Sebastián, Únzaga Silvina I. y Durán Elena B. (2017). Diseño de una aplicación ubicua personalizada de apoyo a estudiantes del curso de ingreso. Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA. San Fernando del Valle de Catamarca. 10 y 11 de Agosto. ISSN 1853-6662. pp: 191-198.

Montoto Daniel Sebastian (2018). Desarrollo de una Aplicación Ubicua Personalizada de Apoyo al Curso de Ingreso de la FCEyT-UNSE. 47 JAIIO 2018. 3 al 7 de Septiembre de 2018. CABA, Buenos Aires. pp: 37- 55.

Quintana Cancinos, Fernando D. (2019). Método de formación de grupos colaborativos en ambientes de aprendizaje ubicuos. 3rd International Virtual Conference on Educational Research and Innovation (CIVINEDU 2019). Octubre 9 - 10, 2019. ISBN 978-84-09-17174-3.

Saadiah Y., Erny A. y Kamarularifin A. J. (2010). The definition and characteristics of ubiquitous learning: A discussion. International Journal of Education and Development using Information and Communication Technology. (IJEDICT), Vol. 6, Issue 1, pp. 117-127.

Únzaga S., Álvarez M. y Durán E. (2015). Modelo de Requerimientos de una Aplicación de Apoyo al Aprendizaje Ubicuo para el Ingreso Universitario. TE&ET'15: X Congreso de Tecnología en Educación y Educación en Tecnología 2015. Argentina.

Avances y Uso de las Tecnologías Informáticas en la Educación

Adair Martins, Carina Fracchia, Claudia Allan, Susana Parra,
Natalia Baeza, Carolina Celeste, Nahuel Mamani, Kevin Isaias Pascual,
Ana Alonso de Armiño, Roberto Laurent

Departamento de Computación Aplicada / Facultad de Informática
Departamento de Electrotecnia / Facultad de Ingeniería
Universidad Nacional del Comahue

Dirección: Buenos Aires 1400, 8300 – Neuquén
Teléfono: 0299 - 4490300 int. 429

e-mails: {adair.martins, carina.fracchia, claudia.allan, susana.parra }@fi.uncoma.edu.ar,
{baeza.natalia, celeste.carolina.s, anaalonso}@gmail.com,
{nahuel.mamani, kevin.isaia}@est.fi.uncoma.edu.ar, {roberto_laurent}@yahoo.com.ar

Resumen

Los docentes hoy en día continúan enfrentando al desafío de construir el conocimiento con sus estudiantes de la forma más significativa, para lo cual siguen recurriendo al uso y al desarrollo de nuevas tecnologías informáticas para mejorar y fortalecer a los procesos de enseñanza y aprendizaje en los distintos niveles de educación. En este contexto el grupo de trabajo continúa avanzando en el uso y desarrollo de aplicaciones de Realidad Aumentada y Realidad Virtual, y de Objetos de Aprendizaje. Se presentan algunos avances alcanzados en las principales líneas de investigación que se está realizando.

Palabras Claves: Tecnologías Informáticas, Realidad Aumentada, Realidad Virtual, Objetos de aprendizaje, Métodos Computacionales.

Contexto

Las líneas de investigación que se presentan están enmarcadas en el Proyecto de Investigación 04/F016: "Computación Aplicada a las Ciencias y Educación" de la Facultad de Informática (FAIF), Universidad Nacional del Comahue (UNCo). El proyecto se inició en el año 2018 y su finalización está prevista para fines del año 2021. Participan docentes y estudiantes avanzados de la carrera de Licenciatura en Ciencias de la Computación (FAIF), docentes de la Facultad de Ingeniería (FI) e investigadores de la Universidad Católica de Brasilia (UCB), Brasil.

1. Introducción

El uso de las tecnologías Realidad Aumentada (RA) y Realidad Virtual (RV) se ha incrementado en las últimas décadas. Sus ventajas principales son la posibilidad de enriquecer un contexto real con información adicional multimedial y su flexibilidad, lo que permite que los recursos digitales desarrollados puedan ser utilizados en diferentes contextos y áreas de conocimiento. En el ámbito educativo favorece la enseñanza activa facilitando la creación de escenarios artificiales seguros para estudiantes, como pueden ser los laboratorios y los simuladores [8]. Para el trabajo con RA se requiere de dispositivos tales como PC, celular o tablet con cámara web, y para los dispositivos

móviles en algunos casos se necesita contar con giroscopio y acelerómetro. Para el trabajo con recursos RV es necesario contar con un visor y en algunos casos se requiere un dispositivo móvil, como sucede con el Cardboard de Google. La tecnología RA en un contexto educativo posibilita generar un escenario donde los estudiantes pueden manipular objetos virtuales multimediales con el propio cuerpo, sin la necesidad de una mediación a través de elementos tales como el teclado o mouse, como si es requerido en otros recursos TIC usados en la actualidad. Por otra parte, a través de recursos RV se logra generar actividades inmersivas, simuladas pero casi reales, en las que los estudiantes de los distintos niveles educativos pueden implicarse logrando así una mayor capacidad de retener conocimientos. Se podría recurrir a recorridos virtuales a través del uso de videos 360° por ejemplo para la enseñanza de geografía e historia, permitiendo así a los estudiantes realizar visitas a museos, reconstruir monumentos históricos, logrando de esta forma que puedan involucrarse en las actividades de aprendizaje de una manera intuitiva e divertida, mejorando por lo tanto su capacidad de retención de los temas [1-3].

La enseñanza de los contenidos del cálculo diferencial en varias variables continúa en la actualidad siendo un desafío para los docentes. Según referencias bibliográficas se describen las dificultades encontradas por los estudiantes en el entendimiento de los conceptos teóricos, puntualmente dificultades en la visualización e interpretación geométrica de funciones de dos variables [4-5]. La representación gráfica y la visualización juega un rol fundamental en el proceso de enseñanza y aprendizaje de funciones de dos variables y normalmente presentan mayores dificultades que en el trabajo con conceptos del cálculo en una variable, ya que se debe trasladar del plano al espacio. Existen distintos software matemáticos utilizados para la enseñanza de estos contenidos, pero en general trabajan con representaciones parciales aisladas, o sea en forma analítica, numérica o gráfica, sin que se pueda lograr una integración de las tres representaciones para una mejor comprensión de los conceptos involucrados. Surge entonces la importancia de investigar el desempeño de nuevas metodologías, creando recursos didácticos utilizando la tecnología para mejorar las prácticas docentes. Lo mencionado anteriormente motivó realizar la propuesta para el diseño e implementación de OA para la enseñanza de contenidos de Cálculo Diferencial e Integral de funciones de dos variables, con el objetivo puesto en la visualización de los conceptos en el espacio 3D y en la interactividad. Los OA están conformados por materiales de soporte digital y carácter educativo. Son diseñados con el propósito de su reutilización en distintos ámbitos educativos. Se analizaron las características de distintas metodologías, y se utilizó para el diseño y construcción del OA la metodología CrOA [6]. Se trata de una metodología argentina, que ofrece una guía de trabajo atendiendo a conceptos tecnológicos y pedagógicos de los OA. Se trabaja con representación en el espacio tridimensional, permitiendo la manipulación de figuras, su rotación y transformación. La finalidad del OA diseñado es que los estudiantes se familiaricen con las representaciones gráficas de los objetos matemáticos tridimensionales, animarlos a que reinterpreten gráficamente aquello que calculan algebraicamente y puedan explorar propiedades de estos objetos [7-9].

2. Líneas de Investigación y Desarrollo

Se está trabajando en tres líneas de investigación: Uso y Desarrollo de Recursos TIC, Realidad Aumentada y Realidad Virtual, Métodos Computacionales y Simulación. Las líneas están interrelacionadas, persiguiendo como propósito general acercar la Universidad al medio.

3. Resultados Obtenidos/Esperados

En la línea "Realidad Aumentada (RA) y Realidad Virtual (RV)" se continúa avanzando en la investigación con el objetivo de realizar el desarrollo de nuevas aplicaciones para ser utilizadas en diferentes ámbitos, tales como: turismo, medicina, educación, patrimonio cultural, entre otros. En un contexto educativo no formal, por ejemplo: un museo o caminatas, donde se puede aprender sobre la flora y fauna de una región, el uso de RA brinda la posibilidad del desarrollo de escenarios para explorar e interactuar con el entorno generando recursos y conocimientos de manera creativa y lúdica. Esto implica la voluntad de centrar el aprendizaje en la participación activa del estudiante/visitante, en sus intereses, en situaciones relevantes y directamente relacionadas con su vida real. Mediante la difusión de experiencias realizadas y resultados de investigaciones en la temática RA [10,11], se logró un contacto con la subsecretaría de Turismo de Neuquén para el desarrollo de software RA y RV para dar a conocer las aves que habitan en la Isla 132 de la ciudad de Neuquén. Se realizaron tres aplicaciones que además de utilizarse en caminatas dentro de la isla, se mostraron en el stand de la subsecretaría de Turismo en la Fiesta de la Confluencia, realizada en febrero del año 2019 [12,13]. La aplicación RA Aves permite identificar a 20 de las aves que habitan en la isla 132, algunas de ellas son migratorias. Esta aplicación está disponible actualmente en Google Play. Las aplicaciones RV permiten navegar en las caudalosas aguas del río Limay y hacer un recorrido de Eco-turismo en bicicleta. Todas las imágenes y videos utilizados fueron proporcionados por la subsecretaría de Turismo de la ciudad de Neuquén. Para el desarrollo de la app Aves se utilizó el software Unity y la librería Arcolib. Se puede mencionar también otro trabajo realizado por el grupo de RA, que consistió en el diseño de modelos 3D, escaneado y reconstrucción 3D, de dos modelos anatómicos para la enseñanza de anatomía en la carrera de medicina de la UNCO. Se desarrolló también una aplicación móvil de RA para el trabajo con los modelos diseñados. Este trabajo fue presentado en el Encuentro sobre Experiencias Mediadas por Tecnología en la Facultad de Ciencias Médicas en diciembre de 2019. También en esta línea de investigación se ha realizado exposiciones en varias jornadas de investigación y extensión, dictado de cursos de capacitación, entre otros. Se pueden mencionar algunos: Conferencia: "Realidad Aumentada y Realidad Virtual: Innovando en la Promoción y Difusión del Turismo Regional", Hackatón Innovación, Banco Provincia de Neuquén (BPN), julio 2019. Conferencia: "Innovar en educación: Recursos Educativos Abiertos + Realidad Aumentada", I WorkShop sobre Prácticas Educativas Abiertas, RED ISEDU, Universidad Nacional de San Luis, abril 2019. Curso: "Recursos para la Enseñanza en Escenarios Digitales". Centro Regional Universitario Bariloche, UNCO, octubre, 2019. Exposición en las Jornadas de Investigación y Extensión, FAIF, agosto 2019 y en la Expo Vocacional, UNCO, agosto 2019.

En la línea "Uso y desarrollo de TIC" se continúa con el trabajo realizado en investigaciones anteriores en Objetos de Aprendizaje (OA), estudiando el potencial de las herramientas de geometría dinámica tridimensional para desarrollar OA que faciliten la práctica docente y la comprensión de los temas del cálculo en varias variables. De acuerdo a lo descrito anteriormente se realizó el diseño e implementación de un OA para la enseñanza de derivadas direccionales, el mismo está siendo utilizado por los docentes en el marco de las clases teóricas y prácticas de la asignatura Métodos Computacionales para el Cálculo de las carreras de Licenciatura en Ciencias de la Computación, Licenciatura en Sistemas de Información y Profesorado en Informática de la FAIF. Los estudiantes lo usan a través de la Plataforma de Educación del Comahue (PEDCO), utilizada en la UNCO como

repositorio de material didáctico. Fue publicado también en el repositorio online de GeoGebra [14]. De esta manera puede ser accedido desde cualquier PC o dispositivo móvil. El OA presenta pantallas divididas en vistas: una vista teórica, con conceptos desarrollados en el sistema de representación algebraico y otra vista gráfica, donde se irán representando en el espacio las figuras correspondientes. Se incluyen pantallas donde se desarrolla una animación de algunos objetos utilizados, con la intención de favorecer la comprensión de los conceptos. Se ofrece al final del recorrido la opción de interactuar con el objeto introduciendo distintas funciones para su representación. Para la implementación se utilizó el software de geometría dinámica GeoGebra [15], que ofrece diversos objetos y funciones que permiten un desarrollo de manera dinámica, interactiva y manipulable. Al implementar el OA se genera un archivo de GeoGebra (.ggb). Para transformar el mismo en OA, es necesario empaquetarlo bajo un cierto formato y agregar los metadatos correspondientes, que permitan su catalogación, reuso y publicación. Se trabajó con el estándar de metadatos LOM y la herramienta de autor eXeLearning para editar los metadatos y generar un OA con formato SCORM. Para la validación y testeo se realizaron test de unidad, módulo y sistema y prueba de aceptación llevada a cabo a través de encuestas a los estudiantes. En todos los casos los resultados fueron satisfactorios, permitiendo verificar que el OA desarrollado cumple con los requerimientos planteados y permitiría mejorar la comprensión por parte de los alumnos de los tema teóricos y prácticos mencionados anteriormente [16].

Siguiendo con las líneas de investigación, se ha desarrollado también una experiencia para la creación de un modelo utilizando las tecnologías de modelado 3D, RA y trabajo interdisciplinario entre las áreas de matemática e Informática en un Colegio Industrial de la ciudad de Neuquén. La motivación y estimulación de los procesos creativos cobran especial importancia en la enseñanza y aprendizaje en varios niveles educativos. El modelado en 3D es fundamentalmente una práctica multidisciplinar que permite integrar diferentes áreas de conocimiento, como puede ser las matemáticas, dibujo técnico e informática, posibilitando por ejemplo para una escuela técnica la resolución de problemas propios mediante el diseño y construcción de objetos reales. La experiencia fue realizada con estudiantes de dos cursos de primer año del Instituto Tecnológico del Comahue (ITC), teniendo acceso a una computadora o a un celular. Existen varios software para el modelado en 3D, para la experiencia se utilizó el software libre online TinkerCad [17] por su interfaz de trabajo muy simple e intuitiva y para la aplicación se utilizó AClass!. Se realizaron revisiones previas por las docentes de informática (FAIF/ITC) y de matemática (ITC) sobre conceptos de unidades de medida, figuras geométricas, perímetros, volúmenes, entre otros, y sobre software para la construcción del modelo. Se trabajó en esta primera construcción del objeto con conceptos de dimensión, posición, espacio, unidades de medidas, y se realizaron cálculos de perímetros y volumen visualizando las diferentes caras del cubo y analizando también las vistas del plano. Con herramientas de duplicación y el uso de paletas de colores se logró la construcción del modelo del cubo de Rubik. El modelo 3D fue exportado a un formato objeto (obj.) reconocido por la aplicación de RA y donde fue posible trabajarlo desde la aplicación AClass! para crear un nuevo proyecto cubo. En esta actividad los estudiantes usaron como marcador una fotografía de la pantalla de la computadora con el diseño del modelo en 3D. Con el modelo de RA creado se permitió visualizar e interactuar con el modelo a través de la herramienta utilizada y finalmente se imprimió el modelo desarrollado en 3D. Se realizaron trabajos prácticos y un cuestionario a los participantes para evaluar logros

obtenidos. A través de los resultados fueron realizadas comparaciones, análisis de ventajas y desventajas en el uso de las herramientas para la visualización y análisis de problemas matemáticos. Los resultados fueron satisfactorios, y puede concluirse que los objetivos fueron alcanzados en cuanto al uso de tecnologías de modelado en 3D, RA y trabajo interdisciplinario, mostrando que la experiencia fue muy motivadora tanto para los estudiantes como para los docentes. Se propone continuar con nuevas experiencias fomentando el trabajo colaborativo entre docentes de la universidad y del nivel técnico en el uso y desarrollo de nuevos recursos digitales para mejorar y fortalecer la calidad educativa.

4. Formación de Recursos Humanos

El grupo de investigación viene trabajando y formándose en proyectos anteriores a través de la realización de cursos de postgrado, extensión, entre otros. Se encuentran en desarrollo dos tesis de grado de la carrera de Licenciatura en Ciencias de la Computación. Dos docentes investigadores realizaron todos los cursos de las maestrías: "Tecnología Aplicada en Educación", UNLP, y "Enseñanza de las Ciencias Exactas y Naturales", UNCo. Ambas se encuentran en etapa de tesis.

5. Bibliografía

[1] Becker, S. A., Freeman, A., Hall, C. G., Cummins, M., & Yuhnke, B.. NMC/CoSN horizon report: (pp.1-52), 2016. The New Media Consortium.

[2] Reporte Edu Trends | Realidad Aumentada y Realidad Virtual, Tecnológico de Monterrey, 2017. https://observatorio.itesm.mx/edu-trends-realidad-virtual-y-realidad-aumentada/

[3] Azuma R. T. , Survey of Augmented Reality In Presence: Teleoperators and Virtual Environments, Vol 6, N°4,(pp. 355- 385), 1997.

[4] Silva Santana, B., Alves da Silva, M. Aprendizagem de Cálculo: A partir do uso de software matemáticos. III CONADU, ISSN 2358-8829, 2016.

[5] Oye, N., Shallsuku, Z. and Iahad, N., The role of ict in education: Focus on university undergraduates taking mathematics as a course. International Journal of Advanced Computer Science and Applications (IJACSA), 2016.

[6] Sanz, C., Barranquero, F., Moralejo, L., Metodología para la creación de objetos de aprendizaje CrOA. http://croa.info.unlp.edu.ar

[7] Fierro, W; Bosquez, V, Design and production of a learning object for university teaching: An experience from theory to practice.ISBN 978-1-5090-6149-5, 2016.

[8] Jimenez Lopez, E. Luna Cámara, M. Lopez Cuevas, S. and Peraza Arrollo, R., Desarrollo de un objeto de aprendizaje para la enseñanza de las matemáticas: el caso de las funciones, 2013. http://www.laccei.org/LACCEI2013-Cancun/RefereedPapers/RP198. pdf.

[7] W. R. Fierro and V. A. Bosquez. Design and production of a learning object for university teaching: An experience from theory to practice. ISBN 978-1-5090-6149-5, 2016.

[9] C. Allan, S. Parra, and A. Martins. Objetos de aprendizaje para la interpretación geométrica de métodos numéricos: Uso de geogebra. Revista TE&ET, (20):51_56, ISSN 1850-9959, 2017.

[10] Realidad aumentada en la enseñanza primaria: diseño de juegos de mesa para las áreas ciencias sociales y matemáticas. DOCREA, (6),89-104. https://drive.google.com/file/d/1zlDnTzEP-NsLQbIOv7hV-hX8PG94sIMp/view

[11] Fracchia, C. C., Alonso de Armiño, A. C., & Martins, A.. Realidad Aumentada aplicada a la enseñanza de Ciencias Naturales. TE&ET, 2015. http ://s edici.unlp.edu.ar/handle/10915/50745

[12][http://www.barinoticias.com.ar/index.php?option=com_content&task=view&id=63795&Itemid=2

[13] https://www.lmneuquen.com/lo-que-no-te -podes-perder-la-fiesta-los-neuquinos-n622676

[14] https://www.geogebra.org/m/efys2w39

[15] https://www.geogebra.org

[16] Allan C., Parra S., Martins A., Objetos de Aprendizaje para la enseñanza de Derivadas Direccionales: diseño, implementación y evaluación. XIV (TE & ET), ISBN 978-987-733-196, San Luís, 2019.

[17] www.tinkercad.com

Análisis y detección de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico - Fase 2.

María Alejandra Paz Menvielle, Calixto Maldonado, Karina Ligorria, Analía Guzmán, Martín Casatti, Nicolás Horenstein

Departamento de Ingeniería en Sistemas de Información
CIDS – Centro de Investigación Desarrollo y Transferencia
de Sistemas de Información
Facultad Regional Córdoba – Universidad Tecnológica Nacional
Maestro Marcelo López esq. Cruz Roja Argentina – Córdoba
0351 – 4686385

pazmalejandra@gmail.com, calixtomaldonado@hotmail.com, karinaligorria@hotmail.com, aguzman@frc.utn.edu.ar, mcasatti@frc.utn.edu.ar, nicolashorenstein@gmail.com

Resumen

En este trabajo se presenta una línea de Investigación que busca ampliar la detección de patrones asociados a la evolución de una base de conocimiento representada en una base de datos orientada a grafos, la misma contiene respuestas de exámenes en formato de texto de redacción libre relacionadas a un dominio específico, utilizada para realizar el análisis de texto en respuestas a preguntas de exámenes en la cátedra de Paradigmas de Programación, con el objetivo de detectar el grado de acierto de las respuestas de los alumnos. Dentro de los patrones que se pretenden descubrir se encuentran aquellos asociados a las respuestas de los alumnos, a la forma de representación de las preguntas de los docentes, entre otras. Es por ello que el presente proyecto busca avanzar en la línea de investigación relacionada a la detección de patrones a partir de grafos dirigidos, tanto en sus aspectos teóricos como prácticos y en sus aplicaciones.

Palabras clave: grafos – patrones – métricas - evaluación.

Contexto

El presente trabajo forma parte del proyecto de investigación y desarrollo que ha sido homologado por la Secretaría de Investigación, Desarrollo y Posgrado de la Universidad

Tecnológica Nacional, reconocido con el código SIUTICO0007786TC, el mismo forma parte del Centro de Investigaciones, Desarrollo y Transferencia de Sistemas de Información – CIDS.

Para su desarrollo se utilizará como caso testigo a la cátedra de Paradigmas de Programación, perteneciente a la carrera Ingeniería en Sistemas de Información, dictada en la Facultad Regional Córdoba, de la Universidad Tecnológica Nacional. Se respetarán los contenidos mínimos fijados para esta asignatura tal cual figuran en la ordenanza 1150 de la carrera, los cuales pertenecen al bloque de tecnologías básicas dentro del área programación, que están principalmente referidos a los paradigmas lógicos, funcional y de orientación a objetos. Además se cumplirá con los descriptores y criterios de intensidad de formación práctica de la Resolución Ministerial 786/09, los que se encuentran definidos en el área de tecnologías básicas, sub-área programación que incluyen a los paradigmas y lenguajes de programación.

El trabajo que aquí se presenta es la fase 2 del proyecto denominado "Análisis y detección de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico", PID SIUTNCO0004812 [1]. El cual tenía como finalidad analizar, detectar y evaluar patrones topológicos frecuentes en un grafo conceptual. Durante el transcurso del mismo se trabajó principalmente en dos etapas, la adquisición y preprocesamiento de datos, y la extracción de características y agrupamiento con el fin de obtener algunas métricas y patrones relacionados al grafo conceptual obtenido.

El presente proyecto propone extender la base de datos agregando información en nuevos grafos para cotejarlos con el grafo original, a fin de identificar y mostrar una variedad más amplia de patrones para la gestión del conocimiento de la materia Paradigmas de Programación.

Introducción

Análisis de patrones

Existen muchos fenómenos diferentes que pueden ser representados con grafos y debido a que estos pueden tener grandes dimensiones se hace importante poder extraer información implícita de estas representaciones para una mejor comprensión. La minería de grafos [2] se enfoca en encontrar elementos o patrones representativos dentro de grafos. Una de las tareas de la minería de grafo es extraer subestructuras repetitivas o subgrafos frecuentes dentro de grafos [3]. La minería de subgrafos frecuentes es una rama dentro de la minería de patrones frecuentes. La minería de subgrafos frecuentes sobre colecciones de grafos ha demandado el diseño de métodos especializados [4]. Las colecciones utilizadas en los mismos se han denominado bases de grafos o conjuntos de grafos. Mientras que la minería de datos se enfoca principalmente en los valores de los datos que se están buscando, en los esquemas semi-estructurados y en los grafos, el enfoque se encuentra en etiquetas frecuentes y topologías comunes [5].

Patrones y grafos

Un patrón es una entidad a la que se le puede dar un nombre y que está representada por un conjunto de propiedades medidas y las relaciones entre ellas (vector de características) [6]. En el dominio utilizado como caso testigo, por ejemplo, un patrón puede ser la ruta resultante de una respuesta de un alumno, de las cuales se extrae el vector de características formado por un conjunto de valores numéricos que pueden representar nivel de exactitud de la respuesta, la puntuación de la misma, la cantidad utilizada de conceptos y de relaciones, etc. El reconocimiento automático, descripción, clasificación y agrupamiento de patrones son actividades importantes en una gran variedad de disciplinas científicas, como la biología, la psicología, la medicina, la visión por computador, la inteligencia artificial, la teledetección, etc. Lo importante de detectar patrones en los datos es que se pueden inferir causas para la agrupación de los mismos (en el caso de que estemos detectando patrones ya conocidos y ya estudiados). Este trabajo, tiene paralelos con el llamado SNA (Social Network Analysis, Análisis de Redes Sociales) que es una disciplina cuyo objetivo es "Analizar la estructura de una red social para 'inferir conocimiento' de un individuo, un grupo, o las relaciones entre ellos" [7]. Un grafo conceptual [8] es un sistema de notación simbólica y de representación del conocimiento. Presentado por John F. Sowa, se basa en los gráficos existenciales [9] de Charles Sanders Pierce, en las estructuras de redes semánticas y en datos de la lingüística, la filosofía y la psicología.

Patrones en grafos

El reconocimiento o detección de patrones dentro de grafos busca detectar un subgrafo (patrón) en un grafo (objetivo). Debemos considerar que esta búsqueda de coincidencias se puede descomponer en dos partes: una concordancia estructural, en donde los nodos y relaciones del patrón conforman una estructura existente en el grafo objetivo; una concordancia a nivel de elementos, en donde los nodos y relaciones, a nivel de sus atributos particulares, tiene los mismos valores que en la estructura encontrada en el grafo objetivo. Muchas veces la búsqueda de estas dos concordancias se ejecuta de forma separada para optimizar los algoritmos o reducir el espacio de búsqueda [10]. En el dominio bajo estudio, la detección de un subgrafo (patrón), se realizará en los grafos (objetivos) que representan los contenidos de la materia y las respuestas de los alumnos en las instancias de exámenes.

Métricas en grafos

Una herramienta ampliamente utilizada para describir grafos y que muchas veces se utiliza para iniciar el análisis de patrones existentes en los mismos, es el cálculo de métricas [11], ya sean locales o globales, que permiten caracterizar el grafo objetivo o el grafo patrón. También las métricas se pueden dividir en dos grandes grupos: Métricas estáticas y Métricas dinámicas

Líneas de Investigación, Desarrollo e Innovación

- Estudio de los diferentes patrones topológicos de grafos que puedan ser relevantes en la búsqueda de información en el dominio elegido, analizando si dichos patrones tienen comportamientos recurrentes o subyacentes.
- Estudio de algoritmos que permitan detectar patrones conocidos en la teoría de grafos como son las "comunidades", "pares", "rutas principales" y otros patrones comunes por medio del análisis de las métricas de la base de conocimientos.
- Detección de patrones que, aún no siendo comunes en otras áreas de la teoría de grafos, si lo son recurrentes en el dominio bajo estudio.
- Descripción de algunas características importantes que se relacionan con el aprendizaje a partir de la aplicación de los patrones encontrados sobre las respuestas elaboradas por los alumnos en un examen.
- Estudio de herramientas de visualización y análisis de grafos para realizar los análisis preliminares y la determinación de los parámetros y métricas de la base de datos [9].
- Automatización de algunos de estos análisis para incluirlos en una herramienta ad-hoc.

Resultados y Objetivos

El objetivo del presente estudio es "Detectar patrones subyacentes en la base de conocimientos asociada a la materia Paradigmas de Programación, relacionados a los contenidos y a las respuestas a preguntas de examen, que permitan caracterizar diversos comportamientos de docentes y alumnos y obtener información relevante para mejorar el dictado y la gestión de dichos contenidos" Para ello se han identificado los siguientes objetivos particulares:

1. Identificar patrones topológicos en diferentes grafos; que permitan brindar información relevante para la gestión del conocimiento de la materia Paradigmas de Programación.
2. Proponer un conjunto de herramientas que permitan ampliar la base de conocimiento; con nuevos grafos, para representar los contenidos de la materia y las respuestas dada por los alumnos en diferentes instancias de evaluación, tanto las respuestas correctas y las incorrectas.
3. Actualizar el software de preprocesamiento de textos existente para mejorar los resultados del análisis ortográfico, sintáctico y semántico, como paso previo a la carga de información textual en el grafo conceptual para su análisis.
4. Analizar los patrones identificados con el fin de elaborar informes relacionados al comportamiento de los mismos en el dominio de aplicación.

Formación de Recursos Humanos

Dentro del desarrollo de este proyecto de investigación se desarrollará el trabajo de Tesis de Maestría de dos integrantes docentes del presente proyecto. Se incorporan al equipo de trabajo docentes-investigadores de la

carrera Ingeniería en Sistemas de Información como investigadores de apoyo con la finalidad de que inicie su formación en investigación científica y tecnológica, se incorpora un becario graduado BINID y becarios alumnos, quienes colaborarán en la recolección, manipulación y desarrollo de este marco metodológico. En el marco del proyecto los estudiantes tendrán la posibilidad de hacer la Práctica Supervisada de quinto año. Los avances, propuestas y herramientas construidas, estarán disponibles para su transferencia y aplicación en el Centro de Investigaciones, Desarrollo y Transferencia de Sistemas de Información - CIDS. Del mismo modo la detección de patrones sobre el dominio de conocimiento de la materia Paradigmas de Programación continuará beneficiando a los integrantes de la cátedra y a los estudiantes.

Referencias

[1] PAZ MENVIELLE, A. y col. (2018) "Análisis y detección de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico", XX Workshop de Investigaciones en Ciencias de la Computación, Corrientes, Argentina.

[2] CHAKRABARTI, D., FALOUTSOS (2012), "Graph Mining Laws, Tools and Case Studies". C. Morgan & Claypool Publishers.

[3] BERZAL,F. (2007). "Graph Mining". Department of Computer Science and Artificial Intelligence E.T.S Ingeniería Informática, Universidad de Granada, Granada, España.

[4] HAN, J., CHENG, H., DONG, X., YAN, X. (2007). "Frequent Pattern Mining: Current Status and Future Directions". Data Mining and Knowledge Discovery, 2007, vol 15, núm. 1, pp. 55-86.

[5] VANETIK, Natalia; GUDES, Ehud; SHIMONY, Solomon Eyal. (2002). "Computing frequent graph patterns from semistructured data. En Data Mining". ICDM 2003. Proceedings. 2002 IEEE International Conference on. IEEE, 2002. p. 458-465.

[6] WATANABE S. (1985). Pattern Recognition: Human and Mechanical. Wiley, New York.

[7] SCOTT, JOHN; CARRINGTON, PETER J. (2011). The SAGE handbook of social network analysis. SAGE publications.

[8] SOWA, JOHN F. (1992). Conceptual graph summary, Conceptual Structures: Current Research and Practice. Ellis Horwood, New York London Toront, pp. 3-66.

[9] PEIRCE, CHARLES S. (1909). Existential graphs, Collected Papers of Charles Sanders Peirce 4, pp. 1-7.

[10] FAN, WENFEI. (2012). Graph pattern matching revised for social network analysis. Proceedings of the 15th International Conference on Database Theory. ACM, p. 8-21.

[11] VAN STEEN, MAARTEN. (2010). An Introduction to Graph Theory and Complex Networks. Copyrighted material.

MINERÍA DE DATOS EDUCACIONAL PARA DETERMINAR PERFILES DE ALUMNOS RECURSANTES EN CARRERAS DE INGENIERÍAS

C. Claudio, E. Jramoy, F. Barrera, F. Cardona, R. Navarro Peláez, S. Racca
Grupo de Investigación Base de Datos
Universidad Tecnológica Nacional - Facultad Regional San Francisco
Av. de la Universidad 501 - San Francisco, Córdoba, Argentina
{cjcarrizo77, emilianolucas, facubarrera123, ferdcardona, ingrnavarropelaez, raccasofi}@gmail.com

RESUMEN

Las instituciones universitarias tienen el desafío de asegurar y mantener el nivel de calidad académica con el fin de proveer profesionales altamente capacitados que respondan a las demandas del mercado laboral actual, especialmente en el área de las TICs. Uno de los aspectos en donde las universidades deben poner mayor énfasis es en el rendimiento académico, ya que generalmente un bajo rendimiento académico está asociado con una alta tasa de deserción de alumnos. Para evaluar el rendimiento académico de un alumno es necesario conocer si existen patrones o perfiles comunes a grupos de alumnos, esto es de significativa importancia para definir acciones que permitan mejorar el desempeño de los alumnos.

En este trabajo se propone el uso de minería de datos educacional para la construcción de modelos que permitan identificar perfiles de alumnos que recursan materias básicas en el primer año de las carreras de ingenierías.

Los resultados de este proyecto serán un aporte para el área de gestión académica, ya que podrán contar con un instrumento objetivo que les permitirá definir acciones a futuro en pos de lograr la mejora en el rendimiento académico de los alumnos en materias básicas del primer año.

Palabras clave: minería de datos educacional – modelos – patrones – alumnos recursantes - ingeniería

CONTEXTO

La Unidad Científico Tecnológica donde se enmarca el presente proyecto es el Grupo de Investigación "Base de Datos". Esta Unidad desarrolla sus líneas de investigación en concordancia con las áreas prioritarias correspondientes a la carrera de Ingeniería en Sistemas de Información, que se encuentran formalizadas a través de la Resolución de Consejo Directivo N 353/2016 de la Facultad Regional San Francisco. El presente proyecto de investigación y desarrollo, de tipo sin incorporación en programa de incentivos, se encuentra homologado y financiado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional bajo el código UTN7854 y según la Disposición SCTyP Nº 221/19, el periodo de ejecución es desde Enero de 2020 hasta Diciembre de 2021, el mismo está incluido en el Programa I&D + i de Tecnología de las Organizaciones de la Universidad Tecnológica Nacional.

1. INTRODUCCIÓN

Las universidades tienen el constante desafío de asegurar y mantener el nivel de calidad académica en pos de proveer a la sociedad profesionales que estén altamente capacitados para responder a las demandas actuales del mercado laboral, especialmente en el área de las tecnologías de la información y las comunicaciones

(TICs). Para ello, es necesario sistematizar procesos que permitan evaluar en forma permanente aspectos relacionados a la calidad académica de la Universidad (Briand et al., 1999). Uno de los aspectos sobre los cuales se debe realizar un constante seguimiento y monitoreo es en el rendimiento académico, debido a que comúnmente el bajo rendimiento académico está asociado a una alta tasa de deserción.

El rendimiento académico se puede definir como la productividad del sujeto, matizado por sus actividades, rasgos y la percepción más o menos correcta de los cometidos asignados (Maletic et. al., 2002). Generalmente el rendimiento académico se evalúa teniendo en cuenta aspectos que influyen en el desempeño, pudiendo mencionar entre ellos: factores socioeconómicos, la amplitud de programas de estudio, las metodologías de enseñanza, conocimientos previos del alumno (Marcus, 2003). Según datos estadísticos brindados por el Área de Gestión Académica de la Universidad Tecnológica Nacional Facultad Regional San Francisco, en las carreras de ingenierías existe históricamente una alta tasa de alumnos con bajo rendimiento académico en instancias de evaluación en materias básicas del primer año (Álgebra y Geometría Analítica, Análisis Matemático I y Física I), esto provoca que los alumnos cursen varias veces este tipo de materias con el fin de alcanzar la regularidad o la aprobación directa. Esto ocasiona no sólo el atraso del alumno en la carrera, sino que además muchos alumnos optan por abandonar la carrera al sentirse frustrados por no poder regularizar o aprobar materias que pertenecen al área básica.

El rendimiento académico suele estar afectado por múltiples factores de tipo socio-educativos, socio-económicos, demográfico, actitudinales, que sin duda condicionan el desempeño del alumno frente al cursado de las materias. Para evaluar el rendimiento académico de un alumno es necesario conocer si existen patrones o perfiles comunes a grupos de alumnos, de esta manera la determinación de perfiles se convierte en una estrategia de valor significativo a la hora de tomar acciones que permitan mejorar el desempeño de los alumnos (González, 1988; Di Gresia, 2007). El establecimiento de perfiles consiste en el proceso de determinación y clasificación de patrones.

En el área de Inteligencia Artificial y del Aprendizaje de máquinas (Mitchell, 1997), existen muchos métodos que permiten determinar y clasificar patrones, los cuales permiten devolver información de valor para la toma de decisiones, uno de ellos es la minería de datos o DM (Data Mining). DM es el uso consistente de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre-procesados (fayyad et al., 2001; hand et al., 2000; frawley et al., 1992). El conjunto de técnicas de análisis de datos que permiten extraer tendencias, patrones y regularidades para describir y comprender mejor los datos forman parte de la DM. Además, permite extraer patrones y tendencias para predecir comportamientos futuros (Simon, 1997; berson & Smith, 1997; White, 2001).

Para poder extraer conocimiento usando minería de datos, se requiere la aplicación de uno o varios modelos o algoritmos. Estos modelos pueden ser descriptivos o predictivos (L.-D. Chen, T. Sakaguchi, and M. N. Frolick, 2000). Los modelos descriptivos usan modelos estadísticos como, por ejemplo: distribución probabilística, correlación, regresión, análisis de clusters y análisis de discriminación. En los modelos descriptivos se conocen las características principales del conjunto de datos a analizar. Por otra parte, los modelos predictivos se basan mayormente en

técnicas de aprendizaje automático (machine learning) y emplean inteligencia artificial (J. L. Hung and K. Zhang, 2006). Estos modelos se crean para predecir valores desconocidos de variables llamadas variables objetivo.

La variedad de aplicaciones para análisis de información en sectores comerciales, financieros, médicos, educativos o científicos, ha permitido el desarrollo de gran cantidad de métodos y algoritmos para predecir tendencias y descubrir patrones. Entre los más importantes se encuentran: clustering, clasificación, regresión, análisis de desviación y análisis de cesta de mercado (M. J. Jafar, 2010). Cuando se aplica minería de datos en instituciones educativas la disciplina se conoce como minería de datos educacionales (MDE). La MDE es una disciplina en evolución que usa tecnologías informáticas como son almacenes de datos y herramientas de inteligencia de negocios para descubrir tendencias y patrones sobre datos educacionales. El conocimiento que DM genera actúa como soporte para las autoridades de centros de educación superior en la toma de decisiones oportunas y a los profesores para analizar el comportamiento y aprendizaje de sus alumnos (R. Jindal and M. D. Borah, 2013). La disciplina se enfoca en el diseño de modelos para mejorar las experiencias del aprendizaje y la eficiencia organizacional (R. A. Huebner, 2013).

Al ser una disciplina nueva, no existe una definición ampliamente aceptada de MDE. Una definición que se ajusta a los objetivos de nuestra investigación es proporcionada por la International Society of Educational Data Mining que define a MDE como "una disciplina en evolución, que tiene que ver con el desarrollo de métodos para explorar los tipos únicos de datos que provienen de ambientes educativos y por medio de la aplicación y uso de estos métodos permite una mejor comprensión de los estudiantes y el entorno en el que aprenden".

La MDE usa métodos, herramientas y algoritmos de DM para investigar datos de interacciones de estudiantes y docentes con el sistema educativo, colaboración entre estudiantes, datos administrativos y datos demográficos. La MDE integra métodos, algoritmos y técnicas con los cuales se puede realizar diferentes experimentos y diseñar modelos para predecir u obtener patrones de datos educacionales.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los principales ejes de trabajo de esta línea de investigación son los siguientes:

- Obtener y explorar los datos académicos de alumnos de ingenierías de UTN FR San Francisco.

- Identificar y seleccionar los alumnos que han recursado materias básicas en el primer año de acuerdo a la muestra definida.

- Encuestar a alumnos recursantes de materias básicas para obtención de datos socioeconómicos, socioeducativos, demográficos y actitudinales.

- Seleccionar características del alumno recursante basándose en aspectos personales, académicos, socioeducativos, socio-económicos, socio-demográficos y actitudinales.

- Caracterizar la disciplina de modelos descriptivos.

- Caracterizar y comparar los diferentes métodos, herramientas y algoritmos de minería de datos educativos.

- Analizar y seleccionar el método y la tipología más acorde para la construcción del modelo descriptivo.

- Construir el modelo descriptivo de acuerdo al método y tipología de MDE seleccionada.

- Determinar el perfil del alumno de ingeniería recursante en materias básicas en el primer año.

3. RESULTADOS OBTENIDOS/ESPERADOS

Dentro de los resultados esperados/obtenidos para esta línea de investigación, se han logrado los siguientes:

Resultados Obtenidos

- Se han obtenido los datos académicos de alumnos de ingenierías de UTN FR San Francisco y se ha realizado una exploración de los mismos a través de la herramienta RapidMiner.

- Se ha identificado la población de alumnos y se han seleccionado, de acuerdo a una muestra, aquellos que han recursado materias básicas en el primer año.

- Se ha comenzado con el diseño de la encuesta a realizar a alumnos recursantes de materias básicas para obtener datos socioeconómicos, socioeducativos, demográficos y actitudinales.

Resultados Esperados

- Obtener los resultados de las encuestas y realizar el registro correspondiente.

- Obtener el modelo descriptivo que permita determinar el perfil del alumno de ingeniería recursante en materias básicas en el primer año.

- Difundir y divulgar los resultados obtenidos.

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de investigación se trabaja desde el Grupo de Investigación "Base de Datos" perteneciente a la UTN Facultad Regional San Francisco, Córdoba. El equipo de trabajo está constituido por un Director de Proyecto, un Co-Director, tres docentes investigadores, un graduado y dos alumnos de la especialidad Sistemas de Información. A su vez, de los docentes que participan, 2 de ellos están llevando adelante sus tesis de maestrías, a través de la Universidad Nacional de San Luis.

5. BIBLIOGRAFÍA

Berson, A. & Smith, S. J. (1997). Data Warehouse, data Mining & OLAP. MC GRAW HILL. USA.

Briand, L., Daly,J. & Wüst, J. (1999). A unified framework for coupling measurement in objectoriented systems. IEEE Transactions on Software Engineering, 25 (1), 91-121.

Data Mining and OLAP", Tata McGraw – Hill Berson & Smith, 1997. "Data Warehousing, Edition.

Di Gresia, L. (2007). Rendimiento académico universitario. Asociación Argentina de Economía Política. Extraído el 5 de marzo de 2012 desde: http://www.aaep.org.ar/anales/works/works2007/digresia.pdf.

Fayyad, U.M.; grinstein, g. & Wierse, A. (2001) Information Visualization in data Mining and Knowledge discovery. Morgan Kaufmann. harcourt Intl.

Frawley, W., Piatetsky-Shapiro, g. & Matheus, C. (1992) Knowledge discovery in databases: An Overview. AI magazine, 13(3), 57.

González, A. J. (1998) Indicadores del rendimiento escolar: relación entre pruebas objetivas y calificaciones. Revista de Educación, 287, 31-54. España.

Hand, d.J.; Mannila, h. & Smyth, P. (2000) Principles of data Mining. The MIT Press. USA.

International Society Educational Data Mining. "Educational Data Mining", 2015. http://www.educationaldatamining.org/.

J. L. Hung and K. Zhang, "Data Mining Applications to Online Learning," World Conf. E-Learning Corp. Gov. Heal. High. Educ., pp. 2014–2021, 2006.

L.-D. Chen, T. Sakaguchi, and M. N. Frolick, "Data Mining Methods, Applications, and Tools," Inf. Syst. Manag., vol. 17, no. 1, pp. 65–70, Jan. 2000.

Maletic, J., Collard, M. & Marcus, A (2002). Source Code Files as Structured Documents. Proceedings 10th IEEE International Workshop on Program Comprehension (IWPC'02), 289- 292.

Marcus, A. (2003). Semantic Driven Program Analysis, Kent, OH, USA, Kent State University Doctoral Thesis.

Mitchell, T. (1997) Machine Learning. Mc Graw Hill.

M. J. Jafar, "A Tools-Based Approach to Teaching Data Mining Methods.," J. Inf. Technol. Educ., vol. 9, pp. 1–29, 2010.

Oswaldo Moscoso-Zeam, Sergio Luján-Mora - Minería de Datos Educacionales: una visión holística Educational Data Mining: an holistic view

R. A. Huebner, "A Survey of Educational DataMining Research," Res. High. Educ. J., pp. 1–13, 2013.

R. Jindal and M. D. Borah, "A Survey on Educational Data Mining and Research Trends," Int. J. Database Manag. Syst., vol. 5, no. 3, pp. 53–73, 2013.

S. A. Kumar and Dr.Vijayalakshmi.M.N, "A Novel Approach in Data Mining Techniques for Educational Data.

S.Anupama," 3rd Int. Conf. Mach. Learn. Comput. (ICMLC 2011) A, no. Icmlc, pp. 152–154, 2011.

Siemens, G., Baker, R.S.J.d. (2012) Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. Proceedings of the 2nd International Conference on Learning Analytics and Know.

Simon, A. (1997) data Warehouse, data Mining and OLAP. John Wiley & Sons. USA.

White, C. J. (2001) IbM Enterprise Analytics for the Intelligent e-business. IbM Press. USA.

Entornos y herramientas digitales para el aprendizaje y la colaboración

Sanz Cecilia[1,3], Madoz Cristina[1], Gorga Gladys[1], Gonzalez Alejandro[1], Zangara Alejandra[1], Iglesias Luciano[1], Ibáñez Eduardo[1], Violini Lucía[1,2], Fachal Adriana[1], Archuby Federico[1,2], Manresa-Yee Cristina[4], Pesado Patricia[1]

[1]Instituto de Investigación en Informática LIDI (III-LIDI). Centro Asociado CIC.
Facultad de Informática – Universidad Nacional de La Plata
[2]Becario/a UNLP
[3]Investigador Asociado de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)
[4] Departamento de Ciencias Matemáticas e Informática. Universidad de las Islas Baleares, España

{csanz, cmadoz, ggorga, agonzalez, li, eibanez, lviolini, farchuby}@lidi.info.unlp.edu.ar,
alejandra.zangara@gmail.com, afachal@hotmail.com, cristina.manresa@uib.es

RESUMEN

En este trabajo se presentan algunas de las líneas de investigación y desarrollo en el marco del subproyecto "Metodologías y herramientas para la apropiación de tecnologías digitales en escenarios educativos híbridos", correspondiente al Instituto de Investigación en Informática LIDI. En particular, se analizan los resultados obtenidos durante 2019 e inicios de 2020, y aquellos que se esperan alcanzar en el transcurso del año.

El foco del proyecto se centra en la investigación, desarrollo e innovación en el área de tecnologías digitales para escenarios educativos. Específicamente, se abordan como ejes temáticos: los materiales educativos digitales, los juegos serios, los entornos digitales, y la colaboración mediada con dichas tecnologías. Se investiga sobre las metodologías de diseño y creación de estas herramientas y entornos, y su aplicación en diferentes contextos educativos con evaluación de diversas variables de impacto. Además, se participa en la formación de recursos humanos en el área y en la cooperación con otras universidad el país y del exterior.

Palabras clave: materiales educativos digitales, entornos digitales para el aprendizaje, escenarios educativos, juegos serios, colaboración

CONTEXTO

Este subproyecto llamado "Metodologías y herramientas para la apropiación de tecnologías digitales en escenarios educativos híbridos" forma parte de un proyecto más general titulado: "Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso" (período 2018-2021), perteneciente al Instituto de Investigación en Informática LIDI, de la Facultad de Informática de la Universidad Nacional de La Plata y acreditado por el Ministerio de Educación de la Nación.

1. INTRODUCCION

El subproyecto, foco de esta presentación, cuenta con una serie de ejes temáticos que guían y viabilizan el alcance de sus objetivos de investigación, desarrollo e innovación.

La variedad de herramientas y entornos digitales que se han combinado con diferentes estrategias de enseñanza y de aprendizaje, han generado nuevos entramados en los escenarios educativos (Osorio, 2010). La integración y el uso de materiales hipermediales, píldoras formativas y otros tipos de recursos en las propuestas de enseñanza se vuelven más frecuentes. Sin embargo, aún es necesario profundizar en la investigación de las metodologías para su diseño, atendiendo a fundamentos desde el punto de vista del diseño tecnológico e instruccional. Consideraciones acerca de la granularidad de estos materiales, su estandarización (Allen & Mugisa, 2010) para la integración en distintos entornos tecnológicos, y su ensamblaje para la conformación de itinerarios (Astudillo, Sanz & Santacruz, 2017), son temas que constituyen la agenda de investigación actual. En este sentido, en el proyecto que aquí se presenta, se está investigando sobre metodologías y *frameworks* para el diseño y creación de objetos de aprendizaje (OA) (Violini, Sanz, & Pesado, 2019), considerando a su granularidad y estandarización. Los resultados 2018, 2019, y 2020 del proyecto en cuestión avanzan sobre estos temas, y abren nuevas líneas que los profundizan.

En los últimos años también los juegos serios digitales, que son aquellos cuyo objetivo caracterizante va más allá del entretenimiento (Dörner, Effelsberg, Göbel & Wiemeyer, 2016), han sido foco de múltiples investigaciones. En (Del Moral Pérez, Duque, & Fernández García, 2018), se cita a autores tales como Squire y Jan (2007), que

consideran que los juegos serios pueden ser favorecedores del desarrollo de habilidades relacionadas con la argumentación científica y la resolución de problemas. También mencionan que Filsecker & Hickey (2014), indican que estos juegos generan un mayor grado de implicación de los estudiantes en las tareas. En otros trabajos, se valora el uso de juegos serios para aumentar la motivación de los/as alumno/as (Kiili & Ketamo, 2018). En el proyecto que aquí se describe se trabaja en el estudio de metodologías para el diseño de juegos serios, que combinan técnicas de la Ingeniería de Software, el diseño instruccional, y el modelado del perfil de los jugadores (estudiantes), utilizando instrumentos provenientes de las Ciencias de la Educación y la Psicología (Archuby, Sanz & Pesado, 2020).

Finalmente, otra de las líneas en las que se investiga y que se presentarán resultados en este trabajo, se vincula con el trabajo colaborativo mediado por tecnologías digitales. En el proyecto se estudian sistemas que posibilitan el seguimiento y soporte de procesos colaborativos en Educación. Se ha avanzado en el estudio de sistemas y estrategias de *mirroring* que permiten reflejar el proceso de un grupo colaborativo (Phielix, Prins & Kirschner, 2010).

Cabe señalar que las líneas de I+D+I que se abordan se vinculan con las temáticas de la Maestría y Especialización en Tecnología Informática Aplicada en Educación de la misma Facultad, por lo que esta vinculación favorece el desarrollo de tesis y trabajos finales en estos temas.

2. LINEAS DE INVESTIGACION / DESARROLLO/ INNOVACIÓN

Se mencionan aquí las principales líneas de investigación, desarrollo e innovación abordadas en el marco del proyecto:

- Entornos digitales para la mediación de procesos educativos. Integración de nuevas funcionalidades y formas de acceso a estos espacios, trazabilidad de las actividades, estándares. Estrategias para el diseño e implementación de estos tipos de entornos.
- Materiales educativos digitales. Metodologías para su diseño y producción. Objetos de aprendizaje. Multimedia e hipermedia en escenarios educativos. Nuevos entramados de medios, soportes y lenguajes.
- Integración de TIC en procesos educativos. Análisis de las actitudes y percepciones de los docentes. Hibridación de las modalidades educativas.
- Juegos Serios con diferentes paradigmas de interacción.
- Trabajo colaborativo mediado por TICs. Autorregulación y capacidades metacognitivas como factores claves para su desarrollo. Conceptualización, análisis y desarrollo de software y metodologías. Actividades colaborativas aprovechando dispositivos móviles.

En el proyecto participan investigadores formados, en formación, becarios, tesistas y alumnos de grado.

3. RESULTADOS OBTENIDOS

Se presentan aquí los principales resultados que se han alcanzado en el período 2019 - inicios de 2020.

En relación a la línea vinculada a entornos digitales se continuó avanzando en el proyecto IDEAS[1] (entorno virtual de enseñanza y aprendizaje - EVEA). Se ha profundizado en la integración de la herramienta de autoevaluación de este entorno con el juego serio Desafiate, que toma las preguntas de las autoevaluaciones creadas en IDEAS y las presenta como desafíos del juego. Desafiate está disponible para dispositivos móviles con Android en Google Play[2]. Además, se llevó adelante una investigación sobre la percepción de los estudiantes sobre su motivación intrínseca cuando resuelven evaluaciones en IDEAS, en Desafiate y con lápiz y papel. Los resultados muestran una preferencia por el uso de Desafiate (Archuby, Sanz, & Pesado, 2019).

Se continuó trabajando en la integración de IDEAS con una herramienta específicamente diseñada para la comunicación sincrónica coordinada (llamada *Infomeeting*). La herramienta está en prueba aún (Digiani, Sanz, & Gorga, 2018).

Se presentó una propuesta de tesis doctoral centrada en el estudio de modelos de integración de estrategias de gamificación en EVEAs. La gamificación es definida como el uso de elementos de diseño de juegos en contextos no relacionados con juegos (Contreras Espinosa, 2016). Algunos autores consideran a la gamificación como el diseño de experiencias de aprendizaje que puedan ser vividas como un juego (Cornellà Canals, & Estebanell Minguell, 2018). El foco de esta tesis se vincula con la búsqueda de modelos que permitan integrar estas estrategias en EVEAs y que puedan ofrecer indicadores de si estos modelos aplicados afectan el nivel de presencia de lo/as estudiantes (Garrison, Anderson, & Archer, 2003). La tesis es

[1] Proyecto IDEAS: https://proyectoideas.info.unlp.edu.ar/

[2] Desafiate en Google Play: https://play.google.com/store/apps/details?id=com.Archuby.Piratas

dirigida en forma conjunta con investigadores de la Universidad de Zaragoza con quienes se trabaja en cooperación en estas temáticas (Vera-Mora, Sanz, Coma-Roselló, & Baldassarri, 2019).

Al mismo tiempo, se continuó avanzando en la tesis doctoral referida a entornos 3D para el aprendizaje de personas con discapacidad auditiva. En este sentido, se evaluaron y analizaron metodologías didácticas para el trabajo con personas con discapacidad. Se indagaron los modelos didácticos M-Free y B-Free y se aplicaron en un estudio de casos. Los resultados fueron presentados en (Fachal, Abásolo, & Sanz, 2019). También se avanzó en el diseño de un espacio en OpenSim[3] para llevar adelante el trabajo de campo de esta tesis.

Otro de los estudios que se realizaron entre 2018 y 2019, se relacionaron con la autorregulación del aprendizaje en situaciones educativas mediadas por tecnologías. En estos temas, tanto las herramientas que se utilizan para fomentar la autorregulación, como las estrategias, y sus impactos, son foco de estudio. Por un lado, se avanzó en un trabajo de tesis doctoral específicamente enfocada en la autorregulación en procesos educativos en EVEA, y se realizó la primer revisión sistemática de literatura al respecto (Fierro-Saltos, et al., 2019). Por otro lado, se finalizó un trabajo de especialización que permitió estudiar un corpus de artículos referidos a la autorregulación, y el rendimiento académico en procesos mediados por tecnologías de la información y la comunicación (Dieser, Sanz, & Zangara, 2019).

También se continuó con la investigación en los temas de trabajo colaborativo mediado por tecnologías digitales y estrategias de seguimiento (Zangara & Sanz, 2019; Zangara & Sanz 2018). Se dio difusión a los resultados de la tesis doctoral, en la que se creó MetSCIN (Metodología de Seguimiento del Trabajo Colaborativo mediado por Tecnología Informática). MetSCIN fue aplicada en las cohortes de 2015 a 2019 inclusive de un curso de postgrado, en el marco de una actividad colaborativa. Actualmente, se está elaborando una propuesta de tesis doctoral para dar continuidad a estas temáticas, en la que se abordará el diseño de herramientas que den soporte a la metodología MetSCIN. También se ha desarrollado una tesis que propone un procedimiento metodológico para planificar las salidas educativas a centros interactivos de ciencia y tecnología que plantea momentos de intervención y situaciones de trabajo que desencadenan acciones colaborativas, donde docentes y alumnos participan en la elaboración y realización de la visita (Dávila, Fernández, Gorga, 2019).

[3] http://opensimulator.org/wiki/Main_Page

En relación a la línea de materiales educativos digitales se avanzó en incorporar nuevos objetos de aprendizaje y otros materiales educativos hipermediales al repositorio creado en 2018. El sitio está disponible en http://roa.info.unlp.edu.ar. Se está desarrollando una tesis que propone un material interactivo hipermedial para el aprendizaje de los conceptos básicos de la Química General, que se encuentra en etapa de evaluación por parte de docentes y alumnos (García, Bertone, & Gorga, 2019).

En el tema de objetos de aprendizaje también se continúa con el desarrollo de una tesis doctoral vinculada al desarrollo de un *framework* para la creación de OA (Sanz, Barranquero, & Moralejo, 2016; Violini, Sanz, Pesado, 2019). Durante este año se evolucionará el prototipo actual para someterlo a una nueva evaluación con expertos.

A partir de una tesis de maestría sobre materiales educativos hipermediales para Matemática, se difundieron resultados (Del Río, Sanz, & Buccari, 2019), y se elaboró una propuesta doctoral, ya aprobada, para avanzar sobre estos temas. La tesis se centra en el modelado de usuarios y el etiquetado libre de recursos educativos matemáticos para su recomendación. Es dirigida de forma compartida con investigadores de la Universidad de Zaragoza.

En cuanto a los juegos serios, se investigaron sus posibilidades para el desarrollo de competencias digitales en docentes (Sandí Delgado & Sanz, 2020). En este estudio, se analizaron las competencias tecnológicas esperadas en el profesorado en distintos países de Iberoamérica, y se propuso una metodología para integrar juegos serios para el desarrollo de competencias tecnológicas. Se trabajó con el juego serio Astrocódigo (Sanz, Artola, Miceli, & Bione, 2018) con docentes de la Universidad de Costa Rica (Sandí Delgado, Sanz & Lovos, 2018).

Este juego serio que forma parte de uno de los resultados previos del proyecto, ha sido utilizado también, durante los últimos 3 años, con estudiantes y docentes de escuelas secundarias de la región de La Plata con el fin de llevar adelante acciones de articulación entre escuela media y universidad, a partir del Proyecto Nexos y Proyecto Reforticca[4] (Abásolo Guerrero, et al., 2019).

En cuanto a los proyectos vinculados con la temática y los acuerdos de cooperación, el III- LIDI participa en los siguientes:

- Se cuenta con un acuerdo de colaboración en estos temas con la Universidad de Zaragoza y se trabaja en forma conjunta.
- Se tienen un acuerdo de cooperación con la Universidad de Islas Baleares, en particular se

[4] https://reforticca.info.unlp.edu.ar/

trabaja con el Departamento de Ciencias Matemáticas e Informática. En este contexto se dirigen tesis y se participa de proyectos conjuntos.

- Se participa del proyecto REFORTICCA: Recursos para el Empoderamiento de FORmadores en TIC, Ciencias y Ambiente. Proyecto que se desarrolla en el marco de los Proyectos de Innovación y Transferencia en Áreas Prioritarias de la Pcia. de Buenos Aires (PIT-AP-BA). Período: 2017-2019.
- Se participa del Proyecto Nexos de articulación entre la escuela y la universidad, a partir del cual se llevan a cabo acciones con diferentes escuelas de la región de La Plata y se participa de ferias y exposiciones de ciencia, tecnología y educación.

4. FORMACION DE RECURSOS HUMANOS

Se participa en la formación de recursos humanos a través de la dirección de tesis de doctorado, maestría y trabajos finales de especialización y tesinas de grado.
En 2019, se cuenta como resultado con 2 trabajos de Maestría y 2 de Especialización en el área de Tecnología y Educación, dirigidos por miembros de este proyecto. Se han abordado 1 tesina de grado dirigida por miembros del proyecto en relación a los temas aquí presentados. Se continúa con la dirección de tesis doctorado, y de maestría en curso.

5. BIBLIOGRAFIA

- Abásolo Guerrero, M. J., Castro, M. L., Santos, G., Sanz, C. V., Bouciguez, M. J., Garcimuño, M., Cenich, G., Miranda, A., and Papini, C. (2019). Reforticca: Recursos para el empoderamiento de formadores en tic, ciencias y ambiente. In XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de San Juan).
- Allen, C & Mugisa, E. (2010). Improving Learning Object Reuse Through OOD: A Theory of Learning Objects. In Journal of Object Technology, vol. 9, no. 6, 2010, pages 51–75, doi:10.5381/jot.2010.9.6.a3
- Archuby, F., Sanz, C., Pesado, P. (2019). Análisis de la experiencia de utilización del juego serio Desafiate para la autoevaluación de los alumnos. XXV Congreso Argentino de Ciencias de la Computación. Octubre, Córdoba. ISBN: 978-987-688-377-1. PP. 206-217. Actas del Congreso.
- Archuby, F., Sanz, C., & Pesado, P. (2020). Tesis de Maestría en Tecnología Informática Aplicada en Educación en desarrollo.
- Astudillo, G., Sanz, C. & Santacruz-Valencia, L. (2017). Proceedings of the Twelfth Latin American Conference on Learning Technologies (LACLO), 2017, ISBN: 978-1-5386-2376-3, págs. 1-4, doi. 10.1109/LACLO.2017.8120939.
- Contreras-Espinosa, R. S. (2016). Juegos digitales y gamificación aplicados en el ámbito de la educación Digital games and gamification applied to education. Revista Iberoamericana de Educación a Distancia, 19(192), 27–33. https://doi.org/10.5944/ried.19.2.16143
- Cornellà Canals, P., & Estebanell Minguell, M. (2018). GaMoodlification Moodle al servicio de la gamificación del aprendizaje. Campus Virtuales, 7(2), 9–25. Tomado de https://dialnet.unirioja.es/servlet/articulo?codigo=6681868
- Dávila, A.; Fernández, A.;Gorga, G. (2019) La colaboración mediada por tecnología como estrategia para potenciar visitas guiadas. El caso del Centro Interactivo de Ciencia y Tecnología Abremate, desde la mirada de sus actores. Propuesta de Tesis de Magister aprobada. Facultad de Informática. UNLP.
- Del Río,L.; Sanz, C. & Buccari, N. (2019) Incidence of a hypermedia educational material on the Teaching and Learning of Mathematics. Journal New Approaches In Educational Research. Vol. 8. No. 1. January 2019. 2254-7339 DOI: 10.7821/naer.2019.1.334
- Dieser, P; Sanz,C. & Zangara, A. (2019) Estrategias de autorregulación del aprendizaje en escenarios educativos mediados por tecnologías de la información y la comunicación. Una revisión y análisis en la Educación Superior Iberoamericana. Trabajo final de Especialización en TIAE. Aprobado y disponible en: http://sedici.unlp.edu.ar/handle/10915/85104
- Del Moral Pérez, M. E., Guzman Duque, A., Fernández García, L, C. (2018). Journal Of New Approaches In Educational Research. Vol. 7. No. 1. Enero 2018. pp. 34–42 ISSN: 2254-7339 DOI: 10.7821/naer.2018.1.248
- Digiani, P.; Sanz,C. & Gorga, G. (2018) *InfoMeeting*: una herramienta de comunicación sincrónica moderada. Aprobada en 2018. Facultad de Informática. UNLP.
- Dörner, R., Effelsberg, W., Göbel, S., Wiemeyer, J. (2016). Serious Games. Foundations, concepts and practice. Springer. Alemania
- Fachal A., Abásolo M.J., Sanz, C. (2019). Sanz C., Experiencias en el uso de TIC y rampas digitales en la enseñanza de informática a alumnos de educación terciaria con

discapacidad visual o auditiva. XXV Congreso Argentino de Ciencias de la Computación (CACIC) (Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019). PP. 1110-1120. Disponible en: http://sedici.unlp.edu.ar/handle/10915/90556

- Fierro-Saltos, W., Sanz, C., Zangara, A., Guevara, C., Arias-Flores, H., Castillo-Salazar, D., Varela-Aldás, J., Borja-Galeas, C., Rivera, R., Hidalgo-Guijarro, J., et al. (2019). Autonomous learning mediated by digital technology processes in higher education: A systematic review. In *International Conference on Human Systems Engineering and Design: Future Trends and Applications*, pages 65-71. Springer, Cham.
- Filsecker, M., & Hickey, D. T. (2014). A multilevel analysis of the effects of external rewards on elementary students' motivation, engagement and learning in an educational game. Computers & Education, 75, 136-148. doi:10.1016/j.compedu.2014.02.008
- García, R.; Bertone, R. & Gorga, G. (2019) Producción de material multimedia interactivo con contenido de Química General. Propuesta de Tesis de Magister aprobada. Facultad de Informática. UNLP.
- Garrison, D. R., Anderson, T., & Archer, W. (2003). A theory of critical inquiry in online distance education. (& W. G. A. In M. G. Moore, Ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Kiili, K., & Ketamo, H. (2018). Evaluating Cognitive and A_ective Outcomes of a Digital Game-Based Math Test. IEEE Transactions on Learning Technologies, 11 (2), 255{263. doi: 10.1109/TLT.2017.2687458
- Osorio G. Luz Adriana (2010). Ambientes híbridos de aprendizaje: elementos para su diseño e implementación. SISTEMAS: Ambientes Educativos Modernos Basados en Tecnología (117) (2010), pp. 70-79. http://www.acis.org.co/fileadmin/Revista_117/Uno.pdf
- Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection. Computers in Human Behavior, 26(2), 151-16.
- Sandí Delgado, J.C.; Sanz, C. & Lovos, E. (2018). Juegos serios para la indagación de competencias tecnológicas que puedan integrarse en la práctica pedagógica del profesorado. Una propuesta de aplicación en la Sede del Atlántico de la Universidad de Costa Rica (UCR). Tesis de Maestría en TIAE. Disponible en: http://sedici.unlp.edu.ar/handle/10915/71063
- Sandí Delgado, J. C.; Sanz, C. (2020). Serious Games to Enhance Digital Competencies Acquisition for Training Faculty. Revista Educación. Vol. 44, Nro. 1, pp. 471-489.
- Sanz, C., Barranquero, F. & Moralejo, L. (2016). CROA: a learning object design and creation methodology to bridge the gap between educators and reusable educational material creation. EDULEARN 2016 - *8th annual International Conference on Education and New Learning Technologies*. Barcelona (Spain). 4th - 6th of July, 2016. Pages: 4583-4592. ISBN:978-84-608-8860-4 ISSN:2340-1117. **DOI:** 10.21125/edulearn.2016.2101
- Sanz, C.; Artola, V.; Miceli, P., & Bione, J. (2018) Astrocode in the wild, EDULEARN18 Proceedings, pp. 7542-7548. Pages: 7542-7548 Publication year: 2018 ISBN: 978-84-09-02709-5 ISSN: 2340-1117 doi: 10.21125/edulearn.2018.1759. Conference name: 10th International Conference on Education and New Learning Technologies. Palma, Spain
- Squire, K. D., & Jan, M. (2007). Mad City Mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. Journal of Science Education and Technology, 16(1), 5-29. doi:10.1007/s10956- 006-9037-z
- Violini, L.; Sanz, C. & Pesado, P. (2019). Learning Objects. Case Studies. 24th Argentine Congress, Tandil, Argentina, October 8–12, 2018, Revised Selected Papers. Springer.
- Vera-Mora, G., Sanz, C., Coma-Roselló, T. & Baldassarri, S. (2019) Propuesta de un modelo de gamificación para entornos virtuales de enseñanza y aprendizaje en Instituciones de Educación Superior y su impacto en la presencia social, cognitiva y docente en dichos entornos. Propuesta de Tesis Doctoral aprobada. Facultad de Informática. UNLP.
- Zangara, A. & Sanz, C. (2018) Interacción e Interactividad en el trabajo colaborativo mediado por tecnología informática. Tesis doctoral. Faculta de Informática. UNLP. Disponible en: http://sedici.unlp.edu.ar/handle/10915/67175
- Zangara, A. & Sanz, C. (2019). Del trabajo grupal al colaborativo. antecedentes, conceptualización y propuesta de abordaje didáctico. *Signos Universitarios*, (54). Referato Nacional

Tecnologías emergentes y modelos de interacción avanzados para contextos educativos

Sanz Cecilia[1,4], Artola Verónica[1,3], Salazar Mesía Natalí[1], Iglesias Luciano[1], Archuby Federico[1,2], Nordio Mauricio[1,4], Buffarini Abril[1], Baldassarri Sandra[5]

[1]Instituto de Investigación en Informática LIDI (III-LIDI). Centro Asociado CIC.
Facultad de Informática – Universidad Nacional de La Plata
[2]Becaria TIPO A UNLP
[3]Becaria Doctoral CONICET
[4] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

[5]GIGA AffectiveLab, Universidad of Zaragoza, España

{csanz, vartola, nsalazar, farchuby, mnordio}@lidi.info.unlp.edu.ar,
li@info.unlp.edu.ar, abrilbuffarini@gmail.com, sandra@unizar.es

RESUMEN

El foco del proyecto que aquí se describe, se centra en la investigación, desarrollo e innovación en el área de tecnologías digitales para escenarios educativos. Una de las líneas en las que se trabaja es el estudio las potencialidades de diferentes paradigmas de interacción persona-ordenador emergentes para su integración en actividades educativas, y para el diseño y desarrollo de materiales y entornos físico-digitales orientados a la enseñanza y el aprendizaje. En particular, se investiga sobre las posibilidades de la interacción tangible, la realidad virtual, la realidad aumentada y la computación afectiva en el marco de procesos educativos.
Como parte del proyecto, se participa en la formación de recursos humanos en el área a través de tesis de postgrado, trabajos finales de grado, becarios de investigación y proyectos con alumnos. En el trabajo se describen los principales resultados alcanzados durante 2019 e inicios de 2020, en relación a las temáticas de estas líneas de I+D+I.

Palabras clave: interacción persona-ordenador, interacción tangible, realidad aumentada, realidad virtual, computación afectiva escenarios educativos

CONTEXTO

Este subproyecto llamado "Metodologías y herramientas para la apropiación de tecnologías digitales en escenarios educativos híbridos" forma parte de un proyecto más general titulado: "Metodologías, técnicas y herramientas de Ingeniería de Software en escenarios híbridos. Mejora de proceso" (período 2018-2021), perteneciente al Instituto de Investigación en Informática LIDI, de la Facultad de Informática de la Universidad Nacional de La Plata y acreditado por el Ministerio de Educación de la Nación.

1. INTRODUCCION

El avance en el uso de las tecnologías digitales por toda la sociedad, y sus constantes evoluciones, ha incrementado la búsqueda y la investigación de nuevas formas de interacción entre las personas y las computadoras. Al mismo tiempo, la creencia de que la combinación entre las habilidades ya adquiridas y lo que le resulta familiar a las personas, beneficia y enriquece esta interacción (De Russis, 2015), ha dado lugar a llamadas interfaces naturales (Galeano Echeverri, 2014) y su integración en procesos educativos (Osorio, 2010). Así una tendencia actual es combinar el uso de las superficies y los entornos físicos del escenario real, como medio para interactuar con la tecnología. En este sentido, la interacción tangible (IT) y la realidad aumentada (RA) se proponen aprovechar parte del contexto real para ser enriquecido o para ser utilizado como parte de la interacción. La interacción tangible se centra en volver manipulables a los bits. La IT, a diferencia de la RA que considera integrar objetos digitales visuales al mundo real, incorpora un fuerte vínculo con los objetos físicos del entorno que se pueden utilizar como entrada y control de sus contrapartes digitales e integran a los sucesos del contexto (Ishii & Ullmer, 1997). Estos autores identifican a los objetos tangibles como entidades físicas y reales que se pueden tocar y funcionan como puente de interacción con aplicaciones digitales. Los objetos tangibles se pueden clasificar en objetos pasivos y activos (Alvarado, Sanz, & Baldassarri, 2019). En este sentido en 2019 se ha avanzado en la investigación sobre ambos tipos de objetos para su

integración en aplicaciones educativas basadas en IT y en particular, para juegos serios IT.
También es interés de este subproyecto investigar y desarrollar aplicaciones de RA innovadoras que puedan enriquecer procesos educativos. En este caso, se busca a través de la RA diseñar actividades donde el rol del alumno sea activo, y que le permita explorar y conocer información del contexto real, que sea aumentada digitalmente. En 2019 se han estudiado metodologías para crear juegos educativos basados en RA. Se ha trabajado en equipos interdisciplinarios con el fin de diseñar y desarrollar este tipo de aplicaciones en el marco del proyecto aquí presentado.
Se investigan también las posibilidades de la realidad virtual para el aprendizaje de temáticas en las que la experiencia, mediante diversos canales sensoriales, resulte el eje de la situación educativa. Algunos estudios han demostrado que aquellos usuarios que experimentan con entornos de aprendizaje inmersivos de RV demuestran mayor curiosidad y un mayor disfrute del aprendizaje (Moro et al., 2017, citado en Chirinos, 2020). Lograr aplicaciones educativas basadas en RV, es un desafío actual, que convoca el estudio de técnicas para la generación de historias, de estéticas, mecánicas de juego y estrategias vinculadas al diseño instruccional para poder alcanzar los objetivos educativos deseados, y al mismo tiempo, atender a la inmersión, el sentido de presencia propios de la RV (Liu, Dede, Huang, & Richards, 2017). En este proyecto se estudian estas temáticas y se trabaja en el desarrollo y su puesta en contexto de diversos micro-juegos educativos de RV.
Finalmente, también se trabaja en una línea sobre computación afectiva. Se emplean sus técnicas para la recomendación de recursos educativos meta-anotados con emociones, y otros tipos de aplicaciones en el escenario educativo.
Cabe señalar que el subproyecto se vincula con las temáticas de la Maestría y Especialización en Tecnología Informática Aplicada en Educación de la misma Facultad, por lo que se desarrollan tesis y trabajos finales en estos temas.

2. LINEAS DE INVESTIGACION / DESARROLLO

Se mencionan aquí las principales líneas de investigación y desarrollo abordadas en el marco del proyecto:

- Juegos educativos basados en Realidad Aumentada, Interacción Tangible y Realidad Virtual. Juegos pervasivos.
- Metodologías para el diseño de actividades educativas basadas en estos paradigmas.
- Interacción Tangible. Marcos que fundamentan este paradigma, relación entre objetos físicos y digitales, aplicaciones en educación.
- Objetos Pasivos y Activos para IT. Tipos de *feedback* en objetos activos.
- Realidad Aumentada para juegos educativos
- Entornos inmersivos e interactivos basados en realidad virtual.
- Computación afectiva en entornos digitales para el escenario educativo.

Cabe señalar que los distintos miembros del proyecto participan en vinculación a las diferentes líneas de investigación del proyecto. Se desarrollan acciones con becarios, tesistas, pasantes y alumnos. Al mismo tiempo, se realizan acciones de cooperación y vinculación con otras universidades del país y del exterior.

3. RESULTADOS OBTENIDOS

Se presentan aquí los principales resultados que se han alcanzado en el subproyecto, y corresponden al período 2019 e inicios de 2020.
En este último año se ha avanzado en la investigación y aplicación de marcos teóricos y descriptivos de interacción tangible. Estos han sido utilizados para dar sustento al diseño e implementación de una herramienta de autor, llamada EDIT, que posibilita a personas no expertas en el área de Informática crear sus propias aplicaciones basadas en Interacción Tangible (Artola, Pesado & Sanz, 2019). En particular, EDIT ha sido planificada para que docentes y terapeutas puedan crear actividades educativas basadas en IT. Para ello, posibilita realizar un proyecto integrado por una secuencia de actividades que se pueden diseñar a partir de plantillas. Además, la herramienta EDIT forma parte de una tesis doctoral en el tema (Artola & Sanz, 2020), y toma en consideración una revisión sistemática sobre editores y herramientas de autor existentes destinadas a facilitar la creación de aplicaciones IT, otorgando diferentes capas de abstracción.
Durante el 2019 se han llevado a cabo sesiones con docentes de escuelas y universidades (alrededor de 38) para presentar el concepto de IT y utilizar EDIT. Durante estas sesiones se indagó el nivel de aceptación de esta herramienta por parte de los docentes, utilizando el modelo TAM (*Technology Acceptance Model*) (Davis, Bagozzi, & Warshaw, 1989). Este modelo ha sido utilizado en numerosas experiencias (Teo, 2009; Sandí Delgado, Sanz & Lovos, 2018), en la que se consideran variables tales como utilidad y disfrute percibido, y facilidad de uso, de un producto como anticipadoras de la intención de uso de éste. Los resultados de las sesiones han mostrado un alto nivel de aceptación de EDIT por parte de los docentes. Al mismo tiempo, a través de

un *focus group* (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2010), se analizaron las reflexiones y percepciones de los docentes sobre la IT y sus posibilidades en los contextos educativos.
Se trabajó también en el marco de un proyecto con alumnos de la Facultad de Informática de la UNLP, en el diseño y desarrollo de una aplicación IT, llamada Murales, para el desarrollo de la creatividad (Pirondo, Ibañez, Corro, Artola, & Sanz, 2019). Esta aplicación combina el uso de objetos físicos del entorno para la creación de murales digitales sobre una mesa interactiva. Se integró también a Murales, una extensión para el trabajo con una Kinect, que permite lograr la inmersión de las personas dentro de un mural creado (Iglesias & Sanz, 2020). Se ha utilizado esta aplicación con estudiantes y docentes de diferentes niveles educativos en el marco del proyecto Reforticca[1] (se puede ver un video en: http://163.10.22.82/OAS/Videos/muralesvideo.mp4). Se desarrolló también un avance en la aplicación El conquistador, un juego de preguntas y respuestas con juguetes activos que ha sido planificado para el trabajo sobre la mesa interactiva VisionAR. La aplicación se desarrolló en el marco de un proyecto con alumnos, y los juguetes activos forman parte de una investigación de tesis de maestría finalizada en 2019 (Sanz, Cruz, Nordio, Artola, & Baldassarri, 2019). También esta aplicación se utilizó en el marco de sesiones con estudiantes y alumnos de escuelas de la región (puede accederse a un video del juego en: http://163.10.22.82/OAS/juegoconquistador/elconquistador2.mp4).
Se continúa con la investigación sobre la utilización de materiales educativos y actividades de realidad aumentada en procesos educativos. Se finalizó un trabajo de especialización sobre el estudio y comparación de librerías de realidad aumentada y sus posibilidades para el desarrollo de actividades educativas (Salazar, Sanz & Gorga, 2019 a y b). Este trabajo se profundiza a través de una tesis de maestría en desarrollo (Salazar, Sanz & Gorga, 2019 c). En 2019 se ha completado el desarrollo de un juego serio basado en realidad aumentada sobre las experiencias y vivencias de Charles Darwin a bordo del Beagle. El juego fue desarrollado interdisciplinariamente (Lizarralde, Sanz, Gorga, Buffarini, Beltrán & Kraselsky, 2019) y está disponible en Google Play[2]
En forma complementaria a esta aplicación, se desarrolló otro micro-juego basado en RA, para dispositivos móviles, en el marco de un proyecto con alumnos (Naiouf, Loza Bonora, Salazar Mesía, Archuby, Sanz, 2019). En esta última aplicación también sobre Darwin se exploró la detección de objetos 3D, construidos con una impresora 3D (Nordio & Sanz, 2019).
También se profundizó en una revisión sistemática sobre aplicaciones móviles para el aprendizaje de personas con discapacidad auditiva, se indagó como parte de ella si existían aplicaciones con RA para este contexto (Herrera, Sanz, & Manresa, 2019).
En cuanto a la línea de realidad virtual, durante 2019 se avanzó en un trabajo de maestría sobre una aplicación de realidad virtual, que permite acercar patrimonios de la Argentina a niños en situación de vulnerabilidad social. El trabajo se relaciona con un proyecto de extensión de la Facultad de Ciencias Económicas, donde participan docentes y estudiantes de la carrera de Turismo (Comparato & Sanz, 2019). La aplicación presenta un recorrido por el Parque Nacional Iguazú y permite a través de pequeños desafíos educativos, ir ganando huellas. La aplicación integra pequeños fragmentos de videos 360, que aportan al objetivo de acercar (Rucci, Chirinos, Sanz, Ravea & Comparato, 2019).
Actualmente, la aplicación continúa en desarrollo para completar otro recorrido focalizado en la fauna del Parque Nacional Iguazú. Se espera en este año avanzar en esto, y en su utilización en talleres con alumnos. En 2019 se llevaron a cabo 2 talleres con estudiantes de escuela primaria y secundaria, y muestras en exposiciones de ciencia, tecnología y educación.
Actualmente, se está desarrollando una nueva aplicación de realidad virtual que presenta una historia relacionada con personajes e hitos de la Ciencias de la Computación (Mazza, Sanz, & Artola, 2020).
En cuanto a los temas de computación afectiva, se avanzó en dos tesis de doctorado vinculadas a la recomendación de recursos (música, y videos educativos) basado en emociones (Ospitia-Medina, Beltrán, Sanz & Baldassarri, 2019; Astudillo, Sanz & Baldassarri, 2020). También se trabaja en el modelado de perfil de usuario para recomendación de recursos matemáticos en el marco de otra tesis doctoral cuya propuesta ya ha sido aprobada (Del Río, Alvarez, Sanz & Baldassarri, 2020).
También se viene participando en estas temáticas a través de otros proyectos con universidades del país (Lovos & Sanz, 2019) y del exterior (Baldassarri, Sanz, Coma, Aguelo-Arguis, & Alvarez, 2019).
En cuanto a los proyectos vinculados con la temática y los acuerdos de cooperación, el III- LIDI participa en los siguientes:

- Se ha firmado un acuerdo de colaboración en estos temas con la Universidad de Zaragoza y la

[1] Reforticca: Recursos para el Empoderamiento de FORmadores en TIC, Ciencias y Ambiente. Proyecto PITAP – CIC. URL:

[2] Ruta Darwin con Realidad Aumentada: https://play.google.com/store/apps/details?id=com.RutaDarwin.AgustinLizarraldeTesis

Universidad de Islas Baleares para cooperar en estas temáticas. Durante el 2019 se realizó una estancia en la UZ por parte de uno de los investigadores formados del proyecto, a partir de la cual se profundizaron los lazos de cooperación. Se participa en el proyecto "*Pervasive Gaming Experiences For @ll* (Pergamex)", en particular en el subproyecto (RTI2018-096986-B-C31) de la UZ. Además, la Dra. Sanz es miembro colaborador del grupo de Investigación en Interfaces Avanzadas (AffectiveLab),

- Se participa en la Red constituida por universidades de Iberoamérica en el marco del programa "Pablo Neruda" dentro del Espacio Iberoamericano del Conocimiento (EIC) y de la Organización de los Estados Iberoamericanos (OEI) orientada a la movilidad de estudiantes y docentes de doctorado.
- Se participa en la RedAUTI: Red temática en Aplicaciones y Usabilidad de la Televisión Digital Interactiva. En este ámbito se analizan materiales educativos para la TVDI.
- REFORTICCA: Recursos para el Empoderamiento de FORmadores en TIC, Ciencias y Ambiente. Proyecto que se desarrolla en el marco de los Proyectos De Innovación y Transferencia en Áreas Prioritarias de la Pcia. de Buenos Aires (PIT-AP-BA). Período: 2017-2019.
- Se participa en un proyecto con la Universidad Nacional de Río Negro en relación a estas temáticas y como asesor en un proyecto de la Universidad Nacional de Santiago del Estero.

4. FORMACION DE RECURSOS HUMANOS

En este proyecto se realizan acciones para la formación de recursos humanos a través de la dirección de tesis de doctorado, maestría y trabajos finales de especialización y tesinas de grado. Al mismo tiempo participan del proyecto, becarios que están realizando sus tesis de postgrado y alumnos de grado que se vinculan en estas investigaciones.
En 2019, se han aprobado 3 trabajos de Maestría y 2 de Especialización en el área de Tecnología y Educación, dirigidos por miembros de este proyecto. Se desataca la tesis de maestría de Mainor Alvarado Cruz, titulada: "Diseño e implementación de juguetes interactivos para actividades educativas basadas en interacción tangible", dirigida por la Dra. Cecilia Sanz, la Dra. Sandra Baldassarri y asesorada por la Lic. Verónica Artola[3]. Esta tesis corresponde al título de Magister en Tecnología Informática Aplicada en Educación de la Facultad de Informática de la Universidad Nacional de La Plata.

[3] La tesis se encuentra disponible en: http://sedici.unlp.edu.ar/handle/10915/87674

También la tesis doctoral de un miembro del proyecto que se encuentra en evaluación.
Además se ha finalizado 1 tesina de grado dirigidas por miembros del proyecto y 3 proyectos con alumnos.

5. BIBLIOGRAFIA

- Alvarado, M. C., Sanz, C., & Baldassarri, S. (2019). Analyzing experiences with active objects in interaction-based educational activities. In Proceedings of the XX International Conference on Human Computer Interaction. ACM. No.: 29 https://doi.org/10.1145/3335595.3335616
- Artola, V.; Pesado, P. & Sanz, C (2019). EDIT: una herramienta de autor para la creación de actividades educativas basadas en interacción tangible. Actas del XIV Congreso sobre Tecnología en Educación & Educación en Tecnología (TE&ET 2019), ISBN: 978-987-733-196-7, págs. 115-124. Disponible en: http://sedici.unlp.edu.ar/handle/10915/90735
- Artola, V. & Sanz, C. (2020). Informe de Tesis Doctoral. Interacción tangible en escenarios educativos. Diseño de una herramienta de autor para la creación de Actividades educativas basadas en interacción tangible. Entregado para su evaluación en diciembre 2019. Esperando evaluación.
- Astudillo, G.; Sanz, C. & Baldassarri, S. (2020). Definición de un modelo de relaciones entre tipos de videos educativos, perfiles de usuarios y emociones. Propuesta de tesis de doctorado en elaboración.
- Baldassarri, S.; Sanz, C.; Coma, T.; Aguelo, A. & Alvarez, P. Involving students in the generation of automatic assessment tests. 12th annual International Conference of Education, Research and Innovation. ICERI2019 Proceedings, isbn 978-84-09-14755-7, issn 2340-1095, doi: 10.21125/iceri.2019.1687, url: http://dx.doi.org/10.21125/iceri.2019.1687. Publicado por IATED. Sevilla, España, Noviembre, 2019. Pp. 7105-7113
- Chirinos, Y., Sanz, C. & Dapoto, S. (2020) La realidad virtual como mediadora de aprendizajes. Desarrollo de una aplicación móvil de realidad virtual orientada a niños. Informe de tesis de Maestría en Tecnología Informática Aplicada en Educación en desarrollo.
- Comparato, G. & Sanz, C. (2019). Proyecto de Extensión aprobado por UNLP: Huellas

Patrimoniales. Viajes inclusivos y nuevas experiencias sensoriales por medio de realidad virtual.

- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. Management science, 35(8), 982–1003
- Del Río, L.; Alvarez, X.; Sanz, C. & Baldassarri, S. (2020). Etiquetado libre y análisis del comportamiento de los usuarios para el diseño de un sistema recomendador de recursos educativos matemáticos. Propuesta de Tesis doctoral aprobada y en desarrollo.
- De Russis, L. (2015). Interacting with Smart Environments: Users, Interfaces, and Devices. IOS Press. Vol7, Nro.1, ISSN: 1876-1364. J. Ambient Intell. Smart Environ, pp 115-116.
- Galeano Echeverri, O. J. (2014). Consideraciones en el Desarrollo de Interfaces Naturales Gestuales. *Revista CINTEX*, *19*, 183-193. Recuperado a partir de https://revistas.pascualbravo.edu.co/index.php/cintex/article/view/46
- Hernández Sampieri, R.; Fernández Collado, C.; & Baptista Lucio, M. del P. (2010). Metodología de la investigación. Mc Graw Hill (5th ed.). México D. F: Mc Graw Hill.
- Herrera, S.; Sanz, C. & Manresa-Yee, C. (2019). Mobile Computing for Hearing-Impaired Children. Artículo enviado a revista para su evaluación. Desarrollado como parte de la cooperación con la Universidad de Islas Baleares y la Universidad Nacional de La Plata.
- Iglesias, L. & Sanz, C. (2020). Aplicación de inmersión en Murales con Kinect. Informe de Tareas.
- Ishii H. & Ullmer B. (1997). Tangible bits: towards seamless interfaces between people, bits and atoms. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (p. 241).
- Liu, D.; Dede, C.; Huang, R. & Richards, J. (editors) (2017). Virtual, Augmented, and Mixed Realities in Education. Springer.
- Lizarralde, A.; Sanz, C.; Gorga, G.; Buffarini, A.; Beltrán, E. & Kraselsky, R. (2019). Ruta Darwin: un juego con realidad aumentada para conocer las experiencias de Charles Darwin en su travesía a bordo del Beagle. XIV Congreso Nacional de Tecnología en Educación y Educación en Tecnología (TE&ET 2019). ISBN: 978-987-733-196-7. Pp. 262-264.
- Lovos, E.; & Sanz, C. (2019) Herramientas de autor para enriquecer materiales de lectura: análisis comparativo. XXV Congreso Argentino de Ciencias de la Computación (CACIC2019). ISBN: 978-987-688-377-1. Pp. 299-313.
- Mazza, M., Sanz, C., & Artola, V. (2020). Juego serio con realidad virtual para jóvenes orientado a conocer hitos de la historia de la Informática Propuesta de trabajo de grado en desarrollo.
- Naiouf, T.; Loza Bonora, L.; Salazar Mesia, N.; Archuby, F. & Sanz, C. (2019). RD3D – Ruta Darwin 3D una aplicación educativa con Realidad Aumentada. Informe del proyecto con alumnos.
- Nordio, M. & Sanz, C. (2019). Impresión de objetos 3D. Informe de tareas de personal de apoyo CIC.
- Ospitia-Medina, Y.; Beltrán, J. R.; Sanz, C. & Baldassarri, S. (2019). Dimensional emotion prediction through low-level musical features. In Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound …, pages 231--234.
- Osorio G. Luz Adriana (2010). Ambientes híbridos de aprendizaje: elementos para su diseño e implementación. SISTEMAS: Ambientes Educativos Modernos Basados en Tecnología (117) (2010), pp. 70-79. http://www.acis.org.co/fileadmin/Revista_117/Uno.pdf
- Pirondo, F.; Ibañez, B.; Corro, B.; Nordio, M.; Artola, V. & Sanz, C. (2019). Murales: una aplicación para la creación de murales digitales usando objetos físicos. Informe del proyecto con alumnos.
- Rucci, A.C.; Chirinos, Y.; Sanz, C.; Ravea, N. & Comparato, G. (2019) "HuVi", aplicación móvil de Realidad Virtual sobre turismo en los Sitios Patrimonios de la Humanidad (UNESCO) de Argentina. Presentada en el Congreso Turismo, Tecnologías y Discapacidad. España. Disponible en: http://www.ttd-congress.com/sites/default/files/actas2019esv2.pdf
- Salazar, N.; Sanz, C. & Gorga, G. (2019 a). Análisis comparativo de librerías de realidad aumentada. Sus posibilidades para la creación de actividades educativas. Especialización en Tecnología Informática aplicada en Educación. Aprobada en 2019.
- Salazar Mesía, N.; Sanz, C. & Gorga, G. (2019 b). Diseño de plantillas para la creación de actividades educativas con Realidad Aumentada en AuthorAR. XXV Congreso Argentino de Ciencias de la Computación. Octubre, Córdoba. ISBN: 978-987-688-377-1.
- Salazar Mesia, N.; Sanz, C. & Gorga, G. (2019 c). Posibilidades de las librerías de Realidad Aumentada en el desarrollo de actividades educativas. XIV Congreso Tecnología en Educación y Educación en Tecnología (TEYET

2019). Universidad Nacional de San Luis (San Luis). Julio de 2019.

- Sandí Delgado, J. C.; Sanz, C. & Lovos, E. (2018). Juegos serios para la indagación de competencias tecnológicas que puedan integrarse en la práctica pedagógica del profesorado. Una propuesta de aplicación en la Sede del Atlántico de la Universidad de Costa Rica (UCR). Tesis de Magister en Tecnología Informática Aplicada en Educación. UNLP.
- Sanz, C., Cruz, M., Nordio, M., Artola, V. & Baldassarri, S. The Conqueror: an educational game based on tangible interaction. 12th annual International Conference of Education, Research and Innovation. Publicado en ICERI2019 Proceedings, isbn 978-84-09-14755-7, issn 2340-1095, doi: 10.21125/iceri.2019.2023, url: http://dx.doi.org/10.21125/iceri.2019.2023. Publicado por IATED. Sevilla, España. Noviembre, 2019. Pp. 8498-8505
- Teo, T. (2009). Modelling technology acceptance in education: A study of pre-service teachers. Computers & Education, 52(2), 302–312.

Diseño e implementación de juguetes interactivos para actividades educativas basadas en interacción tangible

Tesis de Maestría en Tecnología Informática Aplicada en Educación
Mainor Alvarado Cruz
mainorcruz@gmail.com
Directoras: Cecilia Sanz (Universidad Nacional de La Plata) y Sandra Baldassarri (Universidad de Zaragoza)
Asesora: Verónica Artola (Universidad Nacional de La Plata)

Resumen

Las teorías de aprendizaje y la innovación en los procesos de enseñanza y aprendizaje se presentan como un fundamento para la integrar la interacción tangible (IT) en procesos educativos. Uno de los aspectos en los que se ha centrado la atención en la IT es en cómo el uso de objetos tangibles (objetos físicos de uso cotidiano o conocidos por los usuarios) en vinculación con entornos digitales pueden apoyar o mejorar el aprendizaje de determinados temas y/o habilidades. En los últimos años se han desarrollado diferentes proyectos de investigación en donde se diseñan objetos tangibles e interfaces de usuario basadas en IT para apoyar el aprendizaje. Sin embargo, las nuevas tecnologías permiten generar propuestas novedosas en las que los objetos tangibles, a través de alguna automatización permiten extender las capacidades de interacción de las personas con su entorno, así como generar nuevas formas para controlar y representar información digital. En este trabajo de tesis se realizó una investigación con el objetivo de indagar las posibilidades de los objetos llamados activos (con capacidades de actuación o *feedback* al interactuar con información digital). En particular, se aborda el trabajo con juguetes interactivos en actividades educativas basadas en IT. Los juguetes interactivos son desarrollados con objetos tangibles que permiten generar interacciones por medio de actuadores y sistemas de control. Para alcanzar el objetivo, se llevó adelante un proceso de búsqueda y revisión de material bibliográfico relacionado con la temática de estudio. Se aplicaron criterios de inclusión y exclusión para definir el cuerpo de trabajos a considerar para el análisis de experiencias de IT en el escenario educativo, en donde se emplean objetos activos. Se seleccionaron 10 casos para analizar, aplicando una serie de criterios definidos por el tesista. Así se establecieron 9 criterios para analizar las experiencias. Estos criterios se orientaron a caracterizar los antecedentes en 4 categorías: (i) descripciones generales, (ii) interacción tangible, (iii) interfaces de usuario tangible y (iv) metodológico educativo. Como resultado del análisis se evidenció el escaso desarrollo de experiencias de IT con objetos activos en Latinoamérica. Además, se constató que los proyectos se enfocan en niveles educativos de educación superior, que proponen actividades de aprendizaje, tanto exploratorias como expresivas, las modalidades de interacción se dan mayormente entre objetos activos y *tabletops*, y se emplea principalmente *feedback* relacionado con las tareas y la autorregulación. Finalmente, en general las experiencias analizadas remiten a actividades grupales y colaborativas. Se pudo identificar que los objetos tangibles activos tienen las características para integrar más y mayores capacidades que permiten crear enlaces entre lo virtual y real, lograr una mayor flexibilidad, abordar la generación de metáforas y enfocar la atención de los estudiantes. También, posibilitan la creación de interacciones con diversos canales de comunicación, de esta forma se pueden crear actividades educativas multimodales que fomenten el aprendizaje. A través del análisis, se identificaron situaciones de la IT con objetos activos que permitieron diseñar y planificar el estudio de caso. Este estudio permitió analizar algunas variables en relación con el impacto que tienen los juguetes interactivos en una actividad específica con estudiantes y/o docentes. El estudio de caso consistió en una actividad educativa basada en un juego de IT utilizando juguetes interactivos. Para ello, se desarrolló un juego de IT en el marco de un proyecto con alumnos del Instituto de Investigación en Informática LIDI de la Universidad Nacional de La Plata (UNLP), como parte de esta tesis se diseñaron y desarrollaron dos juguetes interactivos para la dinámica del juego, y se organizaron sesiones de trabajo con estudiantes y docentes. Estas sesiones mostraron que los juguetes interactivos son considerados como un aspecto atractivo para los estudiantes y docentes; posibilitan crear propuestas innovadoras para integrar contenidos educativos con tecnología, y para brindar nuevas experiencias de juego y aprendizaje. Asimismo, se observó el disfrute de los estudiantes durante la experiencia, y buenos resultados respecto con la motivación intrínseca de los estudiantes durante la participación en las actividades educativas planificadas.

Plataforma Web de entrenamiento en el uso de pulsadores virtuales para personas con alteraciones en el desarrollo del lenguaje

Aldo Matamala Jorge Rodríguez Laura Cecchi

email: aldo.emanuel.matamala@gmail.com {j.rodrig, lcecchi}@fi.uncoma.edu.ar

Grupo de Investigación en Lenguajes e Inteligencia Artificial
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE

Resumen

El lenguaje oral es la forma básica que utilizan las personas para comunicarse, sin embargo existe un colectivo que tiene serias dificultades para dominar en forma fluida el lenguaje.

Los Sistemas Alternativos Aumentativos de Comunicación(SAAC), son un conjunto de dispositivos y formas de expresión no verbales destinadas a atenuar las dificultades de comunicación y lenguaje de las personas con discapacidad. Se observa que la disponibilidad de este tipo de recurso es limitada en la Escuelas Especiales.

En este trabajo se introducen inicialmente, las características formales, que requieren los dispositivos educativos para implementar SAAC.

A partir de estos atributos, se plantea como objetivo el desarrollo de una herramienta Web, que permita el aprendizaje y entrenamiento para uso de los kits adaptativos que poseen las escuelas. Así, personas con alteraciones en el desarrollo del lenguaje podrán aprender a hacer uso de pulsadores para expresar ideas, pensamientos, necesidades y deseos.

En la implementación de esta herramienta los pulsadores serán simulados sobre un dispositivo móvil, de modo que los aprendices no requieran del recurso físico para su aprendizaje.

La línea de investigación que se presenta se desarrolla en forma colaborativa entre investigadores de dos proyectos de investigación financiados por la Universidad Nacional del Comahue.

Palabras Clave: TECNOLOGÍA INCLUSIVA - EDUCACIÓN - PULSADORES VIRTUALES - SISTEMAS ALTERNATIVOS AUMENTATIVOS DE COMUNICACIÓN .

Contexto

Esta línea de investigación se enmarca en los Proyectos de Investigación “Agentes Inteligentes. Modelos Formales y Aplicaciones para la Educación (04/F015)” y “Agentes Inteligentes y Web Semántica (04/F014)”, de la Facultad de Informática, ambos financiados por la Universidad Nacional del Comahue y con una duración de cuatro años, a partir de enero de 2017. Estos proyectos se encuentran incluidos en el Grupo de Investigación en Lenguajes e Inteligencia Artificial(GILIA), de la Facultad de Informática, en donde se desarrollan sus líneas de investigación, desarrollo y transferencia.

Asimismo, el trabajo está en el contexto del Convenio Marco de Colaboración, firmado entre la Facultad de Informática y el Ministerio de Educación de la Provincia del Neuquén, en 2016. Particularmente, se trabaja a través del Consejo Provincial de Educación, de la Provin-

cia de Neuquén, con la Escuela Especial N°3 y la Escuela Integral de Adolescente y Jóvenes N°4. En este sentido, el grupo de investigación cuenta con el asesoramiento de especialistas en discapacidad múltiple y en comunicación aumentativa-alternativa, que son docentes de estas escuelas.

1. Introducción

Un mismo mensaje hablado puede significar algo completamente distinto, incluso lo contrario, variando sólo la entonación de voz del mismo. Podemos expresar una idea en palabras, a la vez que la negamos con el gesto o la entonación y, en tal caso, es mucho más apropiada la interpretación que se desprende de los elementos paralingüísticos, que de los propiamente lingüísticos. La evolución de la comunicación en el niño se inicia con el gesto, las vocalizaciones, la entonación y, cuando incorpora el habla, ésta acaba por dominar todas las posibilidades anteriores, constituyéndose la principal forma de comunicación.

Pero existe un colectivo de personas que nunca llega a conseguir un dominio suficiente del lenguaje como para, basados en esta capacidad, lograr hacerse entender [3].

Los Sistemas Alternativos Aumentativos de Comunicación(SAAC), son un conjunto de dispositivos y formas de expresión no verbales destinadas a atenuar las dificultades de comunicación y lenguaje de las personas con discapacidad. Estos sistemas se organizan en dos grandes categorías: sistemas sin ayuda y con ayuda. Los sistemas sin ayuda son aquellos en los cuales el individuo usa sólo los movimientos de las manos o corporales para comunicarse. En contraste, los sistemas con ayuda o asistidos son los que requieren del uso de un sistema especial, tal como un pictograma o tablero de palabras, cuaderno o ayuda computarizada [7].

En los sistemas asistidos, un método utilizado es el escaneo, el cual involucra pasar por una serie de opciones en orden, hasta encontrar el mensaje deseado [11]. El estudiante a menudo tiene acceso a la ayuda mediante un dispositivo, activando una llave o un pulsador sensible a la presión, usando una palanca de mano o de brazo, o empujando un pedal con un pie o rodilla.

En nuestro país, bajo el programa de "Conectar Igualdad"(decreto N°459/10 [6]), se hizo entrega de kits de tecnología adaptativa a las escuelas de educación especial. Estos incluyen una impresora multifunción, un brazo articulado, un mouse activado por botones, un "switch" con pulsador grande y uno con pulsador pequeño. Las escuelas que han recibido el kit, poseen 2 pulsadores para una matrícula de 20 a 30 personas. Por otro lado, existe poca oferta nacional de pulsadores y adquirirlos desde el exterior resulta oneroso. Desde la página de "Conectar Igualdad"[5] se ofrecen algunos sistemas, sin embargo, éstos presentan algunos inconvenientes relacionados con las limitaciones pedagógicas y de idioma extranjero.

En este contexto, se propone la definición de un modelo teórico que tenga la capacidad de orientar procesos de desarrollo de dispositivos tecnológicos accesibles a la población destino.

Asimismo, se plantea como objetivo el diseño e implementación de una herramienta Web que permita el entrenamiento sobre dos pulsadores virtuales. La herramienta contará con un plan de entrenamiento basado en un conjunto de actividades interactivas, que serán propuestas por especialistas en discapacidad múltiple y en comunicación aumentativa-alternativa, con los que se trabaja en colaboración. En este trabajo, se utilizarán dispositivos móviles, como smartphones y tablets, en reemplazo de los pulsadores físicos. La comunicación se realizará mediante la ejecución de un pulsador virtual en el dispositivo.

El trabajo presentado está estructurado como sigue. En la siguiente sección se describe la línea de investigación y desarrollo. En la sección 3 se detallan los resultados obtenidos y las líneas de investigación futuras. Finalmente, se comentan aspectos en relación a la formación de recursos humanos.

2. Líneas de investigación y desarrollo

Esta línea de investigación surge en el contexto de dos proyectos de investigación y propone el desarrollo de una plataforma Web, que permita a personas con alteraciones en el desarrollo del lenguaje, aprender a hacer uso de pulsadores, para expresar ideas, pensamientos, necesidades y deseos.

Por un lado, el proyecto de investigación *Agentes Inteligentes. Modelos Formales y Aplicaciones para la Educación* tiene entre sus objetivos, tanto la producción de modelos teóricos, como el desarrollo de dispositivos computacionales destinados a personas con alteraciones en el desarrollo del lenguaje. Se presta especial atención a favorecer los procesos de enseñanza y de aprendizaje que mejoren las posibilidades comunicativas.

Estos objetivos se complementan con los del proyecto de investigación *Agentes Inteligentes y Web Semántica*, que busca, entre otras cosas, el desarrollo de sistemas multiagentes inteligentes, que se adapten a los usuarios, utilizando técnicas de Inteligencia Artificial.

Asimismo, el trabajo se realiza en el contexto de las actividades que se desarrollan en colaboración con la Escuela Especial N°3 y la Escuela Integral de Adolescente y Jóvenes N°4, dependientes del Consejo Provincial de Educación de la Provincia de Neuquén.

Desde la perspectiva metodológica, esta iniciativa se ubica el en ámbito de la Investigación Acción Participativa aplicada a los proceso de investigación, desarrollo y transferencia de resultados. La participación comunitaria se expresa en este caso en la acción desplegada por las Escuelas Especiales y su equipo de especialistas.

Los trabajos de investigación y desarrollo que se realizan en el marco de esta línea consideran un conjunto de características específicas para la elaboración de modelos teóricos y la construcción de este tipo de dispositivos educativos. En [9, 10] se establecen algunas características de referencia. En base a éstas y a las interacciones establecidas con especialistas en el área de Comunicación Aumentativa-Alternativa, se definen los siguientes atributos:

- Fácil acceso: no requerir de dispositivos de última generación, ni de instalación de software o hardware específico.
- Dispositivo conocido y familiar: es importante que la persona pueda tolerar y aceptar el posicionamiento del dispositivo alrededor de ellos [4].
- Conectividad inalámbrica: ofrecer conectividad inalámbrica desde el dispositivo móvil a la interfaz Web.
- Retroalimentación multisensorial: permitir un conjunto de reacciones o respuestas, lumínica, auditiva o vibratoria, desde el dispositivo móvil al ejecutarse una pulsación.
- Entrenamiento de “escaneo formal”: posibilitar el uso de dos pulsadores virtuales en el dispositivo móvil para su implementación [4].
- Compatibilidad: asegurar la compatibilidad con pulsadores estándar.
- Gestión de usuarios: se deberá diferenciar a las diversas clases de usuarios con sus respectivos privilegios. Por ejemplo, estudiante, docente y administrador.
- Gestión de actividades: permitir la administración de un conjunto de actividades asociadas a uno o varios usuarios y al tipo de entrenamiento.

La plataforma propuesta cumple los atributos antes definidos. Por el contexto de uso de la misma, la compatibilidad respecto de los pulsadores será con aquellos provistos por el programa “Conectar Igualdad”.

La importancia de esta línea radica en la necesidad de un colectivo de personas en mejorar sus posibilidades comunicativas. Disponer de dispositivos como los descriptos en esta sección, resulta de un valor significativo para las escuelas especiales de la región.

3. Resultados obtenidos y esperados

Del proceso de análisis basado en entrevistas con la asesora profesional y de un relevamiento sobre estudios referidos a metodologías y herramientas de Comunicación Aumentativa-Alternativa (CAA), se pudo determinar que los tipos de actividades de entrenamiento, requieren de un ambiente en tiempo real, similar al de los juegos multijugador.

El diseño de la herramienta plantea una arquitectura cliente- servidor que tenga como características principales la escalabilidad, para facilitar la incorporación de nuevas actividades como métodos de interacción con el usuario y la reutilización de componentes, para implementar las mismas actividades de entrenamiento con diferentes métodos de interacción.

En esta herramienta Web se requiere de un pulsador que realice el escaneo y otro para la selección de una imagen, implementados como clientes en dispositivos diferentes, comunicados entre sí y accediendo al servidor en tiempo real. Es de sumo interés que la herramienta brinde una experiencia fluida al usuario final, por lo que parte de la arquitectura debe ser dirigida por eventos.

Actualmente, se está trabajando en el patrón de diseño Modelo-Vista-Plantilla (MVT: Model-View-Template)[8, 2], que ofrece la ventaja de separar el Modelo y la Vista. Esto logra independencia entre los datos y la representación visual, facilita el manejo de errores y permite que el sistema sea escalable.

Una capa asíncrona se define por debajo del patrón MVT para manejar las conexiones y sockets. Esta capa permite que la comunicación entre dos clientes sea persistente, bi-direccional y con baja latencia comparado con las técnicas de *polling*, que es lo que se necesita para llevar a cabo la interacción entre la actividad de entrenamiento, alojada en el servidor, y los pulsadores virtuales.

Para la implementación de esta arquitectura se seleccionaron las siguientes herramientas:

- *Django*: Es la pieza fundamental para implementar el servidor. Es un framework de desarrollo de aplicaciones Web escrito en Python y mantenido por Django Software Corporation[2]. Permite construir aplicaciones escalables respetando el patrón *MVT*.
- *Channels*: Siguiendo los líneamientos de la arquitectura dirigida por eventos se eligió Channels[1], que es un módulo desarrollado para Django. Este provee un marco de trabajo, en el que se pueden gestionar conexiones con WebSockets y HTTP2(conexiones permanentes), así como el manejo de tareas asíncronas. Esto se suma al manejo de HTTP que ofrece Django y cuyo comportamiento no se ve variado con la inclusión de Channels.

Se espera que a partir del desarrollo de la herramienta propuesta, que será de licencia libre, escuelas de todo el país hagan uso más activo de los recursos tecnológicos provistos por el plan “Conectar Igualdad”, como así también que profesionales de la informática en conjunto con docentes y/o especialistas, desarrollen futuros módulos de comunicación y de diversos métodos de interacción con el usuario, con el fin de mejorar la calidad de la educación y, por ende, la vida de personas con trastornos en el lenguaje.

4. Formación de Recursos Humanos

El primer autor de este trabajo se encuentra desarrollando su tesis de grado de la Licenciatura en Ciencias de la Computación, en la temática de esta línea de investigación.

Por otra parte, otro de los autores de este trabajo está inscripto en la Maestría en Enseñanza en Escenarios Digitales, que desarrollan de manera conjunta la Universidad Nacional de Cuyo, Universidad Nacional del Comahue, Universidad Nacional de la Patagonia Austral, Universidad Nacional de la Patagonia San Juan Bosco, Universidad Nacional San Luis, Universidad Nacional Chilecito y Universidad Nacional La Pampa.

Referencias

[1] Channels Homepage. Django Software Foundation. https://channels.readthedocs.io/en/latest/. Online; accedido 20 Marzo 2020.

[2] Django Framework. Django Software Foundation. https://docs.djangoproject.com/es/3.0/faq/general/. Online; accedido 20 Marzo 2020.

[3] Carles Augé and Jordi Escoin. Tecnologías de ayuda y sistemas aumentativos y alternativos de comunicación en personas con discapacidad motora. *F. Alcantud y FJ Soto (coords.), Tecnologías de ayuda en personas con trastornos de comunicación*, pages 139–160, 2003.

[4] Ian Bean. *Switch Progression Learning Journeys Road Map*. Inclusive Technology Ltd, Riverside Court Huddersfield Road Delph Oldham OL3 5FZ, 1 edition, 2011.

[5] Homepage de Conectar Igualdad. https://www.argentina.gob.ar/educacion/aprender-conectados/conectar-igualdad.

[6] Presidencia de la Nación. Decreto 459/10. *Boletín Oficial de la República Argentina*, 2010.

[7] Donald Fuller, Lyle Lloyd, and Ralf Schlosser. Further development of an augmentative and alternative communication symbol taxonomy. *Augmentative and Alternative Communication*, 8(1), 1992.

[8] JavaTPoint. Django MVT. https://www.javatpoint.com/django-mvt. Online; accedido 20 Marzo 2020.

[9] Janice Light, David McNaughton, David Beukelman, Susan Koch Fager, Melanie Fried-Oken, Thomas Jakobs, and Erik Jakobs. Challenges and opportunities in augmentative and alternative communication: Research and technology development to enhance communication and participation for individuals with complex communication needs. *Augmentative and Alternative Communication*, 35(1):1–12, 2019.

[10] David McNaughton and Janice Light. The iPad and Mobile Technology Revolution: Benefits and Challenges for Individuals who require Augmentative and Alternative Communication. *Augmentative and Alternative Communication*, 29(2):107–116, 2013.

[11] Joan M Miller. Comunicación aumentativa y alternativa. *Mount Saint Mary College*, page 9, 2018.

JUEGOS SERIOS MÓVILES. DISEÑO, DESARROLLO E INTEGRACIÓN EN ESCENARIOS EDUCATIVOS

Cecilia Sanz[2]; Edith Lovos[1]; Martin Goin[3], Mónica Ricca[1], Claudio Molina[3], Evangelina Gil[1], Ivan Basciano[1], Manuel Gastaminza [1]

[1] Universidad Nacional de Río Negro, Sede Atlántica, Centro Interdisciplinario en Derechos, Inclusión y Sociedad (CIEDIS),

[2] Investigador asociado de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires. Instituto de Investigación en Informática LIDI (III-LIDI), Facultad de Informática, Universidad Nacional de la Plata

[3] Universidad Nacional de Río Negro, Sede Andina, Centro Interdisciplinario en Derechos, Inclusión y Sociedad (CIEDIS),

csanz@lidi.info.unlp.edu.ar - {elovos, mgoin,mricca, cmolina}@unrn.edu.ar - {EvaEG19,ivaanbas19,Manuelgastaminza}@gmail.com

RESUMEN

Las tecnologías móviles, principalmente celulares inteligentes tienen cada vez más presencia en los espacios educativos, transformando de diversas formas las dinámicas de las aulas. Esto ha dado lugar a diversos estudios que presentan resultados sobre su aporte para el desarrollo de entornos de aprendizajes menos rígidos, personalizados y ubicuos (García, 2018, De la Torre Cantero et al; 2015, Demir & Akpinar, 2018). Sumado a esto, los dispositivos móviles posibilitan el diseño de actividades educativas que incluyan tecnologías emergentes, como es el caso de la realidad aumentada (RA) y su integración a través de estrategias de juegos. En esta línea, desde el proyecto de investigación (PI-40C-750) acreditado por la Universidad Nacional de Río Negro (UNRN) se busca generar conocimiento respecto al diseño, desarrollo y aplicación de juegos educativos móviles (JEM) con RA en espacios educativos de nivel medio y superior, atendiendo fundamentalmente al contexto de la UNRN, y a través de una metodología de investigación-acción colaborativa.

Se ha puesto el énfasis en la exploración de tecnologías emergentes como la realidad aumentada y dinámicas de juegos, en particular con el uso de dispositivos móviles, analizando las posibilidades y barreras que se presentan en su inclusión en procesos educativos. Se abordaron distintos aspectos tales como metodologías para el diseño de juegos serios y herramientas para implementación de los mismos, así como también las posibilidades de inclusión, y evaluación de juego serios móviles en propuestas pedagógicas. En este trabajo se presentan detalles de la investigación realizada, que se encuentra transitando su primer año de desarrollo.

Palabras clave: Dispositivos móviles, Juegos Serios, Realidad Aumentada, Diseño.

CONTEXTO

Esta investigación, se lleva adelante en el marco del proyecto de investigación bianual

denominado “Juegos educativos móviles con realidad aumentada. Aspectos de diseño, desarrollo e integración en escenarios educativos”, acreditado y financiado por la UNRN, e iniciado en Mayo de 2019. Asimismo, resulta importante destacar que el mismo busca dar continuidad a los avances en el área de tecnología informática aplicada en Educación, que se vienen desarrollando desde el año 2016, en el contexto de la Sede Atlántica de la UNRN y como parte de las actividades de la unidad ejecutora Centro Interdisciplinario de Estudios sobre Derechos Inclusión y Sociedad (CIEDIS).

1. INTRODUCCIÓN

El reporte Horizon (2017), presentaba al aprendizaje móvil (m-learning) como una de las tecnologías a ser adoptadas a corto plazo, en los ámbitos educativos de nivel superior, dado que posibilita la interacción entre docentes y estudiantes, y permite que éstos últimos utilicen los dispositivos móviles como medio para poner en práctica competencias necesarias para insertarse en la sociedad actual, tales como: comunicación, colaboración y producción de contenido. A nivel nacional, se observa un grado creciente de penetración de la telefonía celular en los espacios de educación formal e informal particularmente en el nivel superior. En este sentido, y en el contexto de la UNRN los dispositivos móviles se presentan como un recurso accesible, que permite trabajar en el diseño de propuestas educativas innovadoras, en particular aquellas que permitan aprovechar entornos reales en combinación con elementos virtuales y recursos lúdicos que puedan sumar al aspecto motivacional tan necesario en los procesos de enseñanza y aprendizaje.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo principal del proyecto se enfoca en desarrollar conocimiento sobre el diseño, desarrollo y aplicación de juegos educativos móviles con realidad aumentada, en espacios de enseñanza y aprendizaje de nivel medio y superior en el contexto de la UNRN.

Los objetivos específicos son:

- Revisar marcos teóricos y analizar antecedentes sobre interacción persona ordenador (HCI) y el uso de juegos educativos móviles con RA en contextos de enseñanza y aprendizaje.
- Estudiar metodologías de diseño de juegos educativos usando RA y dispositivos móviles.
- Aportar al desarrollo de metodologías para el diseño, desarrollo e integración de juegos educativos en situaciones específicas, haciendo uso de dispositivos móviles y tecnologías emergentes como la realidad aumentada.
- Diseñar, desarrollar y evaluar experiencias de uso de juegos educativos móviles con RA en situaciones educativas concretas.

En relación a los objetivos, las principales líneas de investigación abordadas que se abordan son:

- Marcos teóricos sobre el diseño de juegos serios (JS), en particular aquellos que incluyen realidad aumentada (RA).
- Exploración de herramientas para el desarrollo de juegos móviles (JM) o Juegos Serios móviles (JSM). Profundizar en aquellas que posibiliten la inclusión de RA y/o puedan ser consideradas y herramientas de autor.
- Diseño de propuestas didácticas que incorporen JS.
- Implementación y evaluación de experiencias educativas que incluyan JSM.
- Instrumentos para recolección de datos y evaluación de experiencias de aprendizaje usando JSM.

3. RESULTADOS

El proyecto está culminando su primer año de desarrollo, y hasta aquí se ha logrado avanzar acorde a lo planificado y a los

objetivos establecidos. Se sigue una metodología de investigación-acción participativa (IAP), buscando como sostienen Ruiz et al; (2002) hacer converger teoría y práctica entre quienes llevan adelante la investigación y quienes buscan poner en práctica los aportes de ésta, siguiendo un proceso de cíclico que en forma resumida se compone de las etapas:

- *Diagnóstico*: dónde se llevó adelante una revisión bibliográfica que abordó temas como: metodologías de diseño de JSM, JSM en contextos de enseñanza y aprendizaje, en particular en el nivel medio y superior, herramientas para el desarrollo de JMs que incluyan RA, incluídas aquellas que puedan ser utilizadas por perfiles no informáticos. Así mismo, se indagó sobre prácticas propias del contexto, en particular en relación a la inclusión e integración de actividades y materiales basados en RA que hagan uso de dinámicas de juego.
- *Planificación*. a partir del análisis de la etapa previa, se avanzó en el diseño acciones, de acuerdo a las vacancias y/o necesidades encontradas en el contexto del PI. Así es posible mencionar:

1. Diseño de una propuesta didáctica de inclusión e integración de JSM para abordar temas del área de matemáticas, que resultan transversales a las carreras de la UNRN, y destinada a estudiantes ingresantes que participen del Ciclo Introductorio a Carrera (CIC) dispuesto por la UNRN.
2. Diseño de un estudio comparativo entre herramientas que permiten el diseño de juegos para dispositivos móviles e incluir actividades de realidad aumentada usando diversos disparadores como QR, geoposicionamiento o imágenes.
3. Diseño de una propuesta formativa en modalidad curso virtual, denominada "Juegos y tecnologías emergentes como recursos pedagógicos" destinada a los docentes de nivel medio de la orientación informática de la Escuela Secundaria Río Negro. Se pretende que la implementación de la misma permita a los participantes, avanzar en el diseño de propuestas pedagógicas que incluyan JSM y RA para abordar y trabajar temas específicos de su práctica docente. Esta propuesta cuenta con el aval de la UNRN y del Ministerio de Educación y Derechos Humanos de la Provincia de Río Negro (Res.0851/2019) y está previsto llevarla adelante durante el primer semestre de 2020.
4. Co-diseño de un JSM que incluya aspectos de RA, destinado a la exploración de aspectos históricos y culturales de la comarca Viedma-Patagones.

- *Acción*: en esta etapa se desarrollan las experiencias de las propuestas didácticas antes mencionadas y se busca avanzar con el diseño y desarrollo de los JSM. Durante febrero 2020, se concretó la experiencia destinada a los estudiantes que cursan el ingreso establecido por UNRN, en carreras específicas de las Sede Andina y Atlántica. La misma se apoya en los aportes de Sandí Delgado y Sanz (2019) y Massa (2018), e incluye el JSM denominado MAHI (Padilla et al; 2017), que permite trabajar el proceso de resolución de ecuaciones algebraicas de forma lúdica. Se llevó adelante una recolección de datos para conocer la experiencia de usuario y su percepción de aprendizaje, siguiendo los aportes del modelo MEEGA+(Petri et al; 2018). Así mismo se recolectaron datos sobre aprendizajes, con la intención de conocer el impacto de la propuesta desde esa perspectiva y poder realizar una comparación con resultados de experiencias previas.
- *Reflexión,* aquí se busca discutir y analizar los resultados obtenidos en las

diferentes experiencias. Así en esta etapa se plantean las diferentes situaciones que surgieron de las experiencias realizadas, se identifican nuevas necesidades y/o modificaciones a los procesos, dando lugar a la forma cíclica de la IAP. Así, en base a las experiencias implementadas en el proyecto, se presentó para su evaluación un trabajo sobre la experiencia de JSM en el ingreso, a las VII Jornadas Nacionales y III Latinoamericanas de Ingreso y Permanencia en Carreras Científico-Tecnológicas (IPECYT, 2020). Asimismo, y en función del desarrollo y reflexión sobre las experiencias señaladas en la fase anterior, está previsto realizar publicaciones que den cuenta de las metodologías de diseño de JSM exploradas, las fortalezas y debilidades observadas así como recomendaciones y líneas de trabajo futuras. De esta forma, se busca realizar un aporte a la temática contextualizado.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigadores del proyecto está conformado por dos docentes de la Sede Atlántica (Viedma) y dos docentes de la Sede Andina (San Carlos de Bariloche) de la UNRN, donde tres de ellos tienen dedicación completa y otro es concurrente, y completan el equipo, dos estudiantes de las carreras Lic. en Sistemas y Lic. en Comunicación Social respectivamente, con pertenencia a la Sede Atlántica. Los docentes se vinculan a carreras de ingeniería, comunicación social y sistemas y específicamente a asignaturas del área de matemáticas, literatura y programación de computadoras. La dirección del proyecto, está a cargo de una investigadora externa, con pertenencia institucional a un laboratorio de investigación de la Facultad de Informática de la Universidad Nacional de La Plata, cuya experiencia y antecedentes en la temática del proyecto, resultan un aporte importante a la formación del equipo de trabajo.

Asimismo, una de las integrantes del proyecto, está participando de los cursos de postgrado del Doctorado en Ciencias Informáticas (UNLP), y finalizando la Especialización en Docencia Universitaria de la UNRN, mientras que algunos de los estudiantes han tomado cursos específicos sobre desarrollo de juegos móviles usando el motor Unity y testing de software .

También, se realizan aportes a la formación de nuevos investigadores, a través de los dispositivos previstos por UNRN, como son las Becas BEITTA y BT.

Por último, y hasta aquí se han diseñado y llevado adelante actividades de extensión y transferencia, que consideramos son un aporte a la formación de recursos humanos, tales como:

- Proyecto de Extensión: ”Juegos Analógicos para el Aprendizaje de Algoritmos en el Nivel Primario” (Res UNRN. 589/19), cuyo objetivo es la introducción de actividades educativas lúdicas unplugged, que permitan desde edades tempranas, promover la construcción de saberes y el desarrollo de competencias para la resolución de problemas usando algoritmos. El mismo está destinado a docentes experimentados y en formación, y estudiantes de 5º y 6º grado del nivel primario.
- Diseño e implementación del taller denominado “*El uso de realidad aumentada para la divulgación de resultados de investigación*” destinado a investigadores nóveles, becarios y tesistas de la unidad ejecutora CIEDIS en el mes de noviembre de 2019.

5. REFERENCIAS

Becker, S. A., Cummins, M., Davis, A., Freeman, A., Hall, C. G., & Ananthanarayanan, V. (2017). NMC horizon report: 2017 higher education edition (pp. 1-60). The New Media Consortium.

De la Torre Cantero, J., Martín-Dorta, N., Pérez, J. L. S., Carrera, C. C., & González, M.

C. (2015). Entorno de aprendizaje ubicuo con realidad aumentada y tabletas para estimular la comprensión del espacio tridimensional. Revista de Educación a Distancia, (37)

Demir, K., & Akpinar, E. (2018). The Effect of Mobile Learning Applications on Students' Academic Achievement and Attitudes toward Mobile Learning. Malaysian Online Journal of Educational Technology, 6(2), 48-59.

García, N. J. L. (2018). Políticas transnacionales sobre aprendizaje móvil y educación: una selección de textos relevantes. Edutec. Revista Electrónica de Tecnología Educativa, (65), 93-109.

Massa, S. M. Educación Con Videojuegos: Nuevos Desafíos. El videojuego en el aula deficiencias y humanidades, 69.

Padilla, Y. A. M., Vázquez-Reyes, S., González, A. M., & Hernández, A. G. (2017, October). Mahi: Support tool for practicing first-degree algebraic equations. In 2017 6th International Conference on Software Process Improvement (CIMPS) (pp. 1-7). IEEE.

Petri Giani, Gresse von Wangenheim Christiane y Borgatto Adriano F.. MEEGA+, Systematic Model to Evaluate Educational Games. En: Newton Lee (eds) Encyclopedia of Computer Graphics and Games. Springer, Cham, 2018.

Padilla, Y. A. M., Vázquez-Reyes, S., González, A. M., & Hernández, A. G. (2017, October). Mahi: Support tool for practicing first-degree algebraic equations. In 2017 6th International Conference on Software Process Improvement (CIMPS) (pp. 1-7). IEEE.

Petri Giani, Gresse von Wangenheim Christiane y Borgatto Adriano F.. MEEGA+, Systematic Model to Evaluate Educational Games. En: Newton Lee (eds) Encyclopedia of Computer Graphics and Games. Springer, Cham, 2018.

Ruiz, F., Polo, M., Piattini, M., & Alarcos, G. (2002). Utilización de Investigación-Acción en la Definición de un Entorno para la Gestión del Proceso de Mantenimiento del Software. In 1er. Workshop en: Métodos de Investigación y Fundamentos Filosóficos en Ingeniería del Software y Sistemas de Información.(MIFISIS'2002). Madrid

Sandí, J. C., & Sanz, C. V. (2018). Análisis comparativo de juegos serios educativos. Indagación sobre sus posibilidades para la adquisición de competencias tecnológicas en la formación del profesorado. Trabajo Final presentado para obtener el grado de Especialista en Tecnología Informática Aplicada en Educación. Universidad Nacional de La Plata (UNLP), Buenos Aires, Argentina. Retrieved from http://hdl.handle.net/10915/65653

Realidad Aumentada y Realidad Virtual: experiencias en diferentes ámbitos de aplicación

Claudia Banchoff, Laura Fava, Alejandra Schiavoni, Sofía Martin

LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.
Facultad de Informática. Universidad Nacional de La Plata
Calle 50 esq. 120, 2do Piso. Tel: +54 221 4223528
{cbanchoff, lfava, ales}@info.unlp.edu.ar, smartin@linti.unlp.edu.ar

RESUMEN

Este artículo describe la línea de investigación, desarrollo e innovación vinculada al desarrollo de aplicaciones usando Realidad Aumentada (RA) y Realidad Virtual (RV) que se viene desarrollando en el LINTI, Laboratorio de Investigación y Nuevas Tecnologías Informáticas de la UNLP (Argentina). Dentro de esta línea se abordan aspectos relacionados al desarrollo de aplicaciones educativas interactivas, materiales didácticos, objetos de aprendizaje, así como también, aplicaciones destinadas a la recreación de niños durante tratamientos terapéuticos y/o de rehabilitación prolongados o para acompañar las visitas en edificios históricos.

En este artículo, se continúa con las líneas de I+D presentadas en WICC 2019 (Banchoff, C., Fava, L., Schiavoni, A., Martin, S., 2019), donde se analizaron y evaluaron entornos de desarrollo para aplicaciones de gamificación usando RA y RV. En esta etapa se presentan los avances en los desarrollos de aplicaciones usando RA y RV destinadas a ámbitos educativos, recreativos y al aŕea de arquitectura.

Palabras clave: juegos serios, aprendizaje basado en juegos, realidad aumentada, realidad virtual, gamificación.

CONTEXTO

La línea de investigación centrada en el desarrollo y aplicación de juegos serios se viene trabajando en el LINTI desde hace una década. Los desarrollos y experiencias realizados están vinculados a distintas problemáticas sociales. En los últimos años se ha focalizado en la incorporación de tecnologías de RA/RV como un elemento innovador y motivador para complementar las actividades que se trabajan en el aula de la escuela.

En este sentido se ha consolidado el uso de videojuegos en el aula tales como RAÍCES (Fava, L., Banchoff, C., Nomdedeu, L., Martin, S., 2017), y realizado experiencias de RA utilizando herramientas como Scratch con niños y niñas de primaria en dos jornadas realizadas en la Facultad de Informática. Estas actividades dan el marco propicio para el desarrollo conjunto de proyectos de extensión e investigación aplicada, en los cuales se trabaja con distintas instituciones educativas.

La línea de investigación que se presenta en este trabajo, "Realidad Aumentada y Realidad Virtual:

experiencias en diferentes ámbitos de aplicación" está enmarcada en el proyecto: "De la Sociedad del Conocimiento a la Sociedad 5.0: un abordaje tecnológico y ético en nuestra región" del Programa Nacional de Incentivos a docentes-investigadores, que se desarrolla en el LINTI .

1. INTRODUCCIÓN

Es sabido que los métodos tradicionales de enseñanza como los libros o la disertación oral, muchas veces fallan al momento de motivar a sus destinatarios y capturar su atención. Los videojuegos y las aplicaciones interactivas que incorporan aspectos de realidad aumentada (RA) y/o realidad virtual (RV) son instrumentos muy poderosos para ser usados en múltiples contextos, en especial, en el ámbito educativo.

Existen muchas experiencias con distintos niveles de aceptación. En nuestra región el desafío es generar actividades que puedan ser llevadas adelante en los establecimientos de gestión pública que sólo cuentan, en el mejor de los casos, con netbooks entregadas en el marco de distintos programas gubernamentales, pero que no tienen acceso a dispositivos móviles y cascos de RV.

El Observatorio de Innovación Educativa del Tecnológico de Monterrey, en su reporte de diciembre 2017, analiza el rol que juega en el ámbito educativo la inclusión en los procesos de enseñanza-aprendizaje de tecnologías de RA y RV. (Edu Trends 2017). Si bien los datos son interesantes, en nuestra región no pueden aplicarse las mismas estrategias. No hay duda, que estas tecnologías pueden resultar altamente atrayentes a docentes y estudiantes, pero no se cuenta con estudios y resultados registrados en nuestra región que sirvan de modelo que retroalimenten nuevas experiencias.

Existe una amplia variedad de aplicaciones disponibles, pero en la gran mayoría de los casos, las mismas son utilizadas por docentes "curiosos" que demuestran interés en el tema y están dispuestos a enfrentar el desafío de aplicarlas. En muchos casos se trata de aplicaciones en idiomas extranjeros o con licencias privativas que limitan su uso.

En base a experiencias previas, se realizan nuevos desarrollos en conjunto con los equipos docentes que trabajan en los distintos establecimientos con los que la Facultad interactúa.

La aplicación de tecnologías de RA en el ámbito de las escuelas es, desde el punto de vista de acceso a los recursos físicos necesarios, más simple que la incorporación de RV. Es común encontrar acceso a dispositivos móviles, tanto en los equipos docentes como en los mismos estudiantes. El uso de RA, a través de dispositivos móviles ha mostrado nuevas posibilidades para diferentes audiencias y contextos (Das P., Zhu Mo, McLaughlin, L., Bilgrami, Z., Milanaik, R., 2017). Sin embargo, para utilizar aplicaciones de RV, las soluciones comerciales que podemos encontrar son los Oculus Rift, HTC Vive o Sony PlayStation RV, las cuales ofrecen una experiencia bastante lograda, pero requieren el apoyo de un computadora muy potente que no está al alcance de la comunidad educativa. Afortunadamente, también existen opciones más económicas, como Samsung Gear VR, o, incluso, las Cardboard de Google, que funcionan con un celular pero con una experiencia menos lograda. Estos recursos limitan notablemente el tipo de desarrollo y actividad que se puede realizar. En el

ámbito de la escuela, es importante que los desarrollos realizados puedan ser usados con recursos de bajo costo.

Para el desarrollo de aplicaciones, tanto de RA como RV, hay un gran número de opciones disponibles que van desde opciones privativas a libres, y que ofrecen mayor o menor compatibilidad con plataformas. Algunas de ellas son Unreal Engine 4 de Epic Games, Unity de Unity Technologies, y CryEngine de Crytek. Algunas de éstas proveen licencias que habilitan su uso en ámbitos académicos y, si bien tienen restricciones respecto al sistema operativo del editor, es posible desarrollar aplicaciones para casi todas las plataformas disponibles.

Por lo analizado se puede concluir que el uso de técnicas de RA brinda un mayor potencial a las aplicaciones. En el caso de la educación conducen a un mejor rendimiento en el proceso de aprendizaje promoviendo un mayor compromiso de los estudiantes y una mayor motivación en el proceso. En este mismo análisis, se detectó que las técnicas de RA basadas en el reconocimiento de imágenes son más utilizadas que las basadas en ubicación, y que el aula es el entorno más utilizado para su aplicación. En otros contextos, por ejemplo en edificios históricos, el uso de RA permite integrar objetos del mundo real con animaciones e información adicional, favoreciendo su difusión, y preservando el patrimonio cultural tangible e intangible (Shengli Xu, 2018)

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Los ejes en los que se está investigando están relacionados con el desarrollo de videojuegos y aplicaciones interactivas usando RA en ámbitos educativos de nivel inicial y primario, ámbitos educativos no formales y recreación en general. Asimismo se está trabajando con RV para crear entornos que faciliten el uso de modelos arquitectónicos para su visualización y manipulación antes de la construcción.

A continuación se detallan los ejes de investigación:

- Estudio y análisis de nuevas versiones de software específicos con el fin de utilizar nuevas funcionalidades provistas.
- Análisis de los contextos reales en los cuales se están desarrollando las aplicaciones para realizar los ajustes necesarios que mejor se adapten a cada uno de ellos.
- Extensión de los prototipos desarrollados como base para la implementación de los productos finales.
- Aplicación de los prototipos desarrollados en contextos reales, a través del uso en instituciones de distintos ámbitos con el objetivo de analizar su impacto.
- Realización de pruebas y testeos continuos con el objetivo de ajustar y adaptar los prototipos desarrollados que servirán de base para el desarrollo de las aplicaciones futuras.

3. RESULTADOS Y OBJETIVOS

Para esta línea de desarrollo de aplicaciones con RA y RV se ha avanzado en la realización de varias aplicaciones. A continuación se describen los resultados de cada uno de los proyectos en los que se ha trabajado.

Una de las aplicaciones creadas es **Sinfonía Terrestre**, a desarrollada para el

Museo de Ciencias Naturales de La Plata, en la cual se aplican técnicas de RA en el recorrido de tres salas del museo.

Para la primera sala, la Tierra, una historia de cambios, una de las actividades desarrolladas consistió en modelar las placas tectónicas y animarlas para simular los choques entre ellas. La Fig. 1 muestra esa animación que aparece sobre la marca, la cual es acompañada por texto, sonido y vibraciones del celular.



Fig. 1. Actividad en Sala 1

Para la sala 2, Paleontología-Paleozoico, se creó un personaje representado por un diplodocus en un ambiente a escala real con el cual se puede interactuar. Finalmente, para la sala 3, Materia laberintos de la evolución, se simuló un experimento logrando la mayor interactividad a través de RA.

Sinfonía Terrestre fue testeada en las salas del museo, y ha tenido muy buena aceptación por parte de los visitantes.

En el año 2019 se presentaron dos trabajos de alumnos en el marco del Proyecto de Desarrollo de Aplicaciones e Innovación de la Facultad de Informática de la UNLP[1] denominado "Otras realidades en la Escuela". Uno de ellos tuvo como resultado el desarrollo un juego que propone la utilización de conceptos de RA para interactuar con los recursos físicos con que cuentan los jardines de gestión estatal[2]. El juego desarrollado permite elegir, a través de un dado mostrado con RA, un destino donde se debe dirigir el robot mediante su programación.

Otro de los trabajos presentados propone recorrer un laberinto realizado a través de RV, para salir del mismo se deben utilizar una serie de comandos direccionales, los cuales se deben ordenar a través un algoritmo. En la Fig. 2 se pueden ver algunas capturas de la aplicación desde el dispositivo.

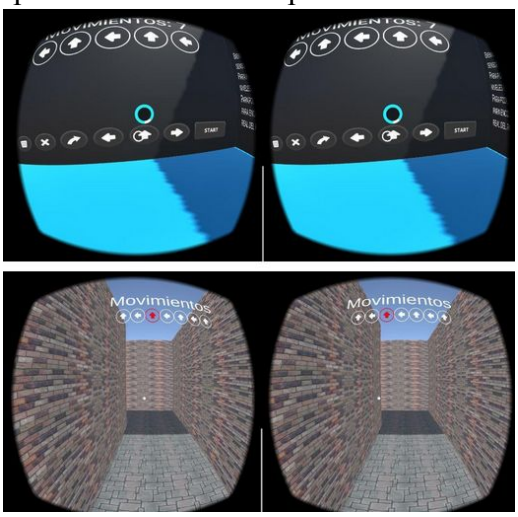


Fig. 2: Juego de laberinto en RV

El objetivo general de esa línea de trabajo es aplicar técnicas de RA y RV en desarrollos concretos, en especial, para el ámbito educativo.

Para cumplir con el objetivo general, se proponen los siguientes objetivos específicos:

- Analizar los distintos contextos de uso de las aplicaciones, los cuales direccionarán las tecnologías a utilizar.

1 https://www.info.unlp.edu.ar/proyectos-con-alumnos/

2 https://www.educ.ar/noticias/200288/nbspse-presentoacute-aprender-conectados

- Desarrollar aplicaciones interactivas que puedan utilizarse en el ámbito de la escuela, el museo de Ciencias Naturales y el Hospital de Niños de La Plata.
- Construir casos de prueba de las aplicaciones desarrolladas con el fin de aplicarlos en los entornos para los cuales fueron desarrolladas y comprobar su adecuación y usabilidad.
- Continuar promoviendo esta temática en el marco del desarrollo de las tesinas de grado.

Los trabajos presentados en este artículo se llevan a cabo de manera interdisciplinaria entre docentes y alumnos de la Facultad de Informática y directivos del Museo de Ciencias Naturales.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de la línea de I+D+i presentada en este artículo está formado por docentes investigadores categorizados del LINTI y estudiantes de la Facultad de Informática. Asimismo en algunos de los proyectos se trabajará con un equipo interdisciplinario de profesionales del Museo de Ciencias Naturales de la UNLP y un equipo de docentes de la cátedra Diseño multimedial de la Facultad de Bellas Artes de la UNLP. A través de la generación permanente de conocimiento por medio de líneas de investigación y desarrollo de aplicaciones vinculadas al sector productivo y su aplicación en el ámbito social, el LINTI promueve el uso innovador de las tecnologías informáticas en la región.

5. BIBLIOGRAFÍA

Banchoff C, Fava, L., Schiavoni, A., Martin S. (2019). *Aplicaciones interactivas usando realidad aumentada y realidad virtual*. XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019). Abril, 2019.

Das, P.; Zhu, M.; McLaughlin, L.; Bilgrami, Z.; Milanaik, R.L. (2017). *Augmented reality video games: new possibilities and implications for children and adolescents*. Multimodal Technologies and Interaction, 2017; 1:8.

Edu Trends, (2017). *Observatorio de Innovacióń Educativa del Tecnológico de Monterrey*. Disponible en http://bit.ly/EduTrendsRARV Último acceso, marzo de 2020.

Fava, L., Banchoff, C., Nomdedeu, L., Martín, S. (2017). *Las culturas originarias a través de Raíces Un videojuego como recurso didáctico-cognitivo*. LACLO 2017. ISBN: 978-1-5386-2376-3.

Shengli Xu (2018). *Intangible Cultural Heritage Development Based on Augmented Reality Technology*. 2018 International Conference on Robots & Intelligent System (ICRIS). DOI: 10.1109/ICRIS.2018.00094

XXII Workshop de Investigadores en Ciencias de la Computación

El Calafate, 2020

MATERIALES EDUCATIVOS DIGITALES: MACRO NARRATIVAS Y MICRO NARRATIVAS PARA LA CONSTRUCCIÓN DE CONOCIMIENTOS EN LA UNIVERSIDAD.

González, Liliana M[1]; Ureta, Laura; Marcovecchio, María José; Margarit, Viviana; Pontoriero, Francisco; Rossetti, Gabriela; Cuadros Patricia; Gallardo Vanesa.

[1] e-mail: lilianamirna@gmail.com

Grupo de investigación perteneciente al Gabinete de Tecnología e Innovación Educativa.
Dpto. de Informática - Dpto. de Geofísica y Astronomía. - Dpto. de Biología
Facultad de Ciencias Exactas, Físicas y Naturales
Dpto. de Matemáticas
Facultad de Ingeniería
Universidad Nacional de San Juan

Resumen

No es novedad que las tecnologías de la información y comunicación han invadido de diferentes maneras y con distintos formatos nuestra vida cotidiana y especialmente la educación. Los jóvenes tienen el mundo en las yemas de los dedos a través de las tecnologías; ellas atraviesan sus modos de pensar, aprender y conocer, razón por la cual es de relevancia que las instituciones educativas, docentes y responsables de su formación identifiquemos y reconozcamos las tendencias culturales. La inmediatez, lo instantáneo, lo nuevo, las diversidades predominan en sus actividades.

Aquí es pertinente considerar la propuesta de Maggio (2012), quién expresa "lo primero que debemos hacer es reconocer lo que los alumnos hacen" (p.153), preguntarnos qué son capaces de hacer, cuáles son sus preferencias tecnológicas, en qué actividades las usan y cómo usan las tecnologías, qué valor les asignan en sus procesos de estudio y producción. Esto significa reconocer el valor de los "usos culturales y formas cognitivas del mismo modo como deberíamos tener en cuenta los estilos cognitivos propios de los sujetos culturales que son nuestros alumnos" (p.153)

Resulta evidente la necesidad de empoderar a nuestros alumnos con capacidades que los ayuden a mejorar su autonomía, la toma de decisiones, la identificación de sus propias necesidades.

Ante esta realidad, cabe preguntarnos, ¿Cuáles son las tendencias culturales de nuestros alumnos en relación al uso de las tecnologías? ¿Qué materiales educativos digitales prefieren? ¿se inclinan por lo hipertextual, por el texto plano, por lo multimedial? ¿eligen los grandes relatos o las micro narrativas?

¿La intencionalidad pedagógica de nuestras prácticas de enseñanza atiende a funciones cognitivas? ¿Cuándo utilizamos tecnologías

de diferente complejidad priorizamos el "efecto de la tecnología"?

A lo largo de la historia las generaciones han ido relatando las experiencias vividas, las diferentes formas de actuar, de pensar, de vivir con otros. Se han utilizado y siguen utilizando diferentes lenguajes y hoy en la actualidad lenguajes con diferentes formatos y soportes. Las narraciones se cristalizaron a través de pintura rupestres y diferentes formas de arte y cultura de los pueblos hasta la integración de lenguajes que hoy facilitan las tecnologías más complejas.

En las instituciones educativas han sido y siguen siendo creadoras de relatos validados en un lenguaje propio de la época, a su vez el currículum se convirtió en el regulador de esas narraciones producidas en un determinado contexto socio-histórico; tales como tradiciones, historias de vida, saberes populares, moral, entre otras. Las instituciones sociales y específicamente las educativas privilegiaron la oralidad y la escritura en sus formas de narrar. Así lo visual, lo audiovisual, el sonido y otras formas de representación se fueron incorporando paulatinamente como complementos de lo oral y lo escrito.

En la actualidad con el avance de las tecnologías en sus diferentes formatos y soportes, el lenguaje multimedial trae consigo un nuevo tipo de narración, aquella en la que se interrelacionan distintos medios como el vídeo, el texto, la fotografía, la infografía, la animación y la ilustración; son las nuevas narrativas digitales.

La lógica de la narración unidireccional fue tomando otros formatos, aunque persiste, dando lugar a la lógica de lo hipertextual y lo multimedial y así a la hipermedia.

Palabras clave: MATERIALES EDUCATIVOS- NARRATIVAS- UNIVERSIDAD

Contexto

El Proyecto mencionado –en proceso de evaluación por CICITCA- tiene como objetivo principal generar conocimientos que aporten a la problemática: qué efectos cognitivos dejan en alumnos universitarios los materiales educativos digitales. Se encuentra enmarcado en el "Gabinete de Tecnología e Innovación Educativa creado por Resolución 03/18, dependiente de Decanato de la Facultad de Ciencias Exactas, Físicas y Naturales. El Gabinete se constituye en un espacio académico y de investigación cuyo es generar estrategias, metodologías, medios y materiales, así como instrumentos tecnológicos, que contribuyan a la inserción de innovaciones tecnológicas en pos de la mejora de los procesos educativos de la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan

1. Introducción

La intencionalidad pedagógica de las prácticas de enseñanza es el eje vertebrador de los procesos de aprendizajes que desarrollan los alumnos. El docente es quién define cómo desarrollar e implementar las prácticas de enseñanza contemplando la realidad socio-educativa, el contexto socio-histórico en el que se da la interacción entre los componentes del triángulo interactivo –docente/alumnos/contenidos-

El alumno realiza los procesos de construcción de conocimiento mediados por la ayuda, guía y orientación del docente y de sus compañeros. Estos procesos están mediados por la acción del docente, de otros alumnos, el contenido y -en la actualidad, en muchos casos- por las nuevas tecnologías. La concepción constructivista considera que los procesos de construcción de conocimientos no

se realizan en el vacío sino en un contexto socio-cultural determinado y en un tiempo particular, orientado por objetivos plasmados en un diseño educativo.

Los iniciadores de la escuela histórico-cultural (Leontiev, 1932; Luria, 1928; Vigotsky, 1929) en sus primeros escritos aluden a la tesis de que "el hombre difiere de los animales por el hecho de que puede hacer y emplear herramientas" y cómo ellas operan sobre el hombre y en su "condición psíquica". Esto trae consigo un cambio sustancial, el alumno "pone, entre la función y la tarea, un determinado medio auxiliar (…) gracias al cual logra realizarla"(Luria, 1928:495 citado en Salomon, 1993:27).

Salomon (2001) expresa "La estructura básica de la cognición humana que resulta de la mediación instrumentalmente ha sido tradicionalmente representada en un triángulo [...]" (p.27-28).

Cole y Engeström (en Salomon, 2001) refieren al triángulo como una "...descripción de los condicionamientos estructurales fundamentales de la cognición humana individual" (29). Estos autores consideran que ese triángulo no contempla lo colectivo de las actividades humanas o de los sistemas de actividad (Leontiev, 1978,1981). A partir de allí, Engeström (1987) presenta el Triángulo fundamental de la mediación ampliada (op. cit. 2001) que permite incorporar la comunidad, las reglas, y la división del trabajo.

En este nuevo triángulo el sujeto no está aislado, es parte de una comunidad; las relaciones entre ellos están mediadas por "artefactos mediadores" y por "reglas" que normativizan y regulan los comportamientos e interacciones entre los miembros de la comunidad. Asimismo, "...las comunidades implican una "división del trabajo", la distribución, constantemente negociada, de tareas, poderes y responsabilidades entre los que participan en el sistema de actividad."(Salomon, 2001:30)

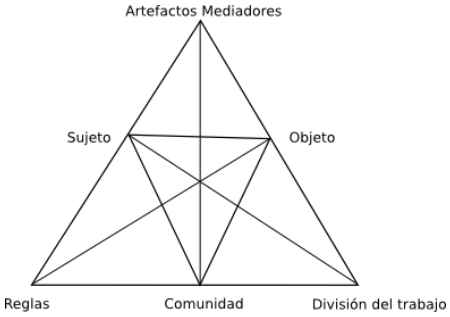


Figura 1 Triángulo fundamental de la mediación ampliada. Engeström (1997, citado en Salomón 2001.31)

Salomon, Perkins y Globerson (1992) distinguen dos tipos de efectos cognitivos:

> los efectos que se obtienen en conjunción CON la tecnología en el curso de la colaboración intelectual con ella, y los efectos procedentes DE la tecnología, en términos del residuo cognitivo transferible dejado por la colaboración, tras la forma de un mayor dominio de habilidades y estrategias. (1992:7)

Siguiendo a los autores mencionados, los efectos CON la tecnología "influyen en lo que los alumnos hacen, en la calidad y cuando lo hacen". En cambio, refieren a efectos DE la tecnología cuando se dan cambios duraderos en las capacidades cognitivas generales de los alumnos como resultado de a ver operado con tecnologías; cambios en el "dominio del conocimiento, de la habilidad, o bien de la

profundidad de la comprensión" después de dejar el uso de la tecnología. (Salomon et al,1992:8)

Recapitulando, el conocimiento es el resultado de un proceso permanente de construcción a través de la interacción y del diálogo con "otros", situado en un contexto socio-histórico. Desde esta perspectiva los alumnos aprenden por la interacción con el objeto según sus estructuras cognitivas, conocimientos previos y su experiencia; "aprenden haciendo", investigando, analizando ejemplos y casos concretos de la realidad.

Por otro lado, a lo largo de la historia las generaciones han ido relatando las experiencias vividas, las diferentes formas de actuar, de pensar, de vivir con otros. Se han utilizado y siguen utilizando diferentes lenguajes y hoy en la actualidad lenguajes con diferentes formatos y soportes. Las narraciones se cristalizaron a través de pintura rupestres y diferentes formas de arte y cultura de los pueblos hasta la integración de lenguajes que hoy facilitan las tecnologías más complejas.

Las instituciones sociales y específicamente las educativas privilegiaron la oralidad y la escritura en sus formas de narrar. Así lo visual, lo audiovisual, el sonido y otras formas de representación se fueron incorporando paulatinamente como complementos de lo oral y lo escrito.

En la actualidad con el avance de las tecnologías en sus diferentes formatos y soportes, el lenguaje multimedial trae consigo un nuevo tipo de narración, aquella en la que se interrelacionan distintos medios como el vídeo, el texto, la fotografía, la infografía, la animación y la ilustración; son las nuevas narrativas digitales.

La lógica de la narración unidireccional fue tomando otros formatos, aunque persiste, dando lugar a la lógica de lo hipertextual y lo multimedial y así a la hipermedia.

Ambrosino (2017) expresa:

> Las narraciones establecen un contexto las mismas nos otorgan una estructura que posibilita construir una trama más elaborada. […] son fundamentalmente un lenguaje que se construye a partir de relatos que ofrecen significación y realización. Se refieren al tejido de la acción e intencionalidad docente. […]…la narrativa es la que otorga sentido y significación al contenido, la sociabilidad y estructura los escenarios de aprendizaje para la construcción de conocimiento.

Teniendo en cuentas estas conceptualizaciones nos proponemos llevar adelante una investigación de "diseño" que nos permita construir macro/micro narrativas que orienten a nuestros alumnos a desarrollar conocimiento como sujetos culturales en un paradigma socio-tecnológico.

2. Líneas de Investigación, Desarrollo e Innovación

- ❖ Efectos cognitivos que dejan los materiales educativos digitales en estudiantes universitarios

3. Resultados Obtenidos/Esperados

- Diseño de materiales para un "Curso de capacitación de Diseño Básico de Aula Virtuales en Moodle". SIED-UNSJ

- Materiales educativos para la Diplomatura "Educación y nuevas

tecnologías en tiempos de convergencia tecnológica".

4. Formación de Recursos Humanos

El equipo de trabajo está conformado de la siguiente manera:

- Directora: González, Liliana Mirna
- Co-directora: Ureta Laura
- Investigadores:

Pontoriero, Francisco, Margarit Viviana, Marcovecchio, María J., Rossetti Beiram, Gabriela, Cuadros Patricia, Gallardo Vanesa

Se propone volcar los resultados en actividades de Postgrado, tales como:

- Diseño e implementación de la Diplomatura "Educación y nuevas tecnologías en tiempos de convergencia tecnológica". Aprobada por Ordenanza Nº 5 CD-FCEFN, Prof. Responsable: Mgter Liliana Mirna González. 1º cohorte mayo -- Diciembre 2015. 2º cohorte mayo -- diciembre 2016. 3º cohorte mayo -- diciembre 2017. 4° cohorte agosto 2018. 5° cohorte setiembre 2019.
- Co-dirección/Asesoramiento de Tesis de Maestría en Informática. F.C.E.FyN de la UNSJ.
- Docente de la Maestría en Informática. F.C.E.FyN de la UNSJ. Materia: Educación y Tecnologías.

5. Bibliografía

- Ambrosino, María Alejandra (2017) Docencia y NARRATIVAS Transmedia en la educación superior. Trayectorias Universitarias. Volumen 3, N° 4. Recuperado de http://revistas.unlp.edu.ar/TrayectoriasUniversitarias
- Casablancas S. (2012) Tecnologías digitales en las aulas: roles docentes, de los estudiantes y del conocimiento" para las IV Jornadas de TIC e Innovación en el Aula - UNLP Recuperado de https://jornadas.ead.unlp.edu.ar/ consultado 1/12/17.
- Lion Carina (13 de setiembre de 2012) Las narrativas docentes en entornos digitales. Ibertí. OEI. Recuperado de https://www.youtube.com/watch?v=8TJQFeP51rg

  —(26 de octubre de 2012) Aprender con tecnologías. Escenarios e interrogantes para la Educación Superior. UBA Tic + I Encuentro virtual sobre tic y enseñanza en el nivel superior. Recuperado de https://www.youtube.com/watch?v=yZivS_vJ-n8
- Maggio M. (2012) Enriquecer la Enseñanza. Paidós. Argentina

  —(2018) Reinventar la clase en la Universidad. Paidós. Argentina
- Salomon, Perkins y Globerson (1992) Coparticipando en el conocimiento: la ampliación de la inteligencia humana con las tecnologías inteligentes. En Revista Comunicación, lenguaje y educación. CL&E, n° 13, Madrid.
- Scolari, C. (2014). Narrativas transmedia: nuevas formas de comunicar en la era digital. Anuario AC/E de cultura digital. Recuperado de https://www.accioncultural.es/media/Default%20Files/activ/2014/Adj/Anuario_ACE_2014/6Transmedia_CScolari.pdf
- Odetti, V. (2017). El diseño de materiales didácticos hipermediales: El caso del PENT FLACSO. Buenos Aires: Teseo.

  — (2016). Materiales didácticos hipermediales: lecciones aprendidas y desafíos pendientes. En García, J. M., y Báez Sus, M. (Comp.) Educación y tecnologías en perspectiva. Diez años de FLACSO Uruguay. Disponible en http://flacso.edu.uy/publicaciones/libro_educacion_tecnologia_2016/Baez_Garcia_Educacion_y_tecnologias_en_perspectiva.pdf

  — (2018) Narrativas Trasmedia. Educación y TIC. EL ABROJO. Montevideo. Uruguay.

REFORTICCA: Recursos para el Empoderamiento de FORmadores en TIC, Ciencias y Ambiente

María José Abásolo[123], Maria L. Castro[4], Graciela Santos[3], Cecilia Sanz[1], Andrea Miranda[3], Gabriela Cenich[3], Mayra Garcimuño[4], Cecilia Papini[3], Mauro Natale[5], María José Bouciguez[3]

[1] Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata (UNLP)
{mjabasolo, csanz}@lidi.info.unlp.edu.ar

[2] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA)

[3] Educación en Ciencias con Tecnología (ECienTec)
Facultad de Ciencias Exactas - Universidad Nacional del Centro Pcia. Bs. As. (Fac. Cs. Ex. ,UNICEN)
{nsantos, mlcastro, amiranda, gabcen, mcpapini}@exa.unicen.edu.ar

[4] Centro de Investigaciones en Física e Ingeniería del Centro de la Provincia de Buenos Aires (CIFICEN) - CONICET/CIC/UNICEN

[5] Núcleo Consolidado de Matemática Pura y Aplicada (NUCOMPA)-Fac. Cs. Ex. ,UNICEN

Resumen

La línea de investigación y desarrollo que aquí se presenta tiene por objetivo brindar recursos educativos y herramientas digitales para ser usadas como soporte del proceso de enseñanza y aprendizaje de ciencias físicas, matemáticas, pensamiento computacional y medio ambiente. En particular, se estudian, desarrollan y evalúan aplicaciones de realidad virtual, realidad aumentada, simulaciones, juegos e interacción tangible, para el soporte de procesos de enseñanza y aprendizaje. Además, se propone crear una comunidad virtual de docentes para la construcción colaborativa de propuestas de enseñanza que utilicen TIC.

Palabras Clave: Enseñanza, Aprendizaje, Formación Docente, TIC, Realidad Virtual, Realidad Aumentada, Interacción Tangible

Contexto

En esta línea se lleva a cabo el proyecto "REFORTICCA Recursos para el Empoderamiento de FORmadores en TIC, Ciencias y Ambiente" Proyecto de Innovación y Transferencia en Áreas Prioritarias de la Provincia de Buenos Aires de la CICPBA (2017-2019) llevado a cabo entre los grupos de investigación III-LIDI (UNLP), EcienTec y CIFICEN (UNICEN) [1-3].

Introducción

En los últimos años han surgido nuevas TIC - incluyendo simulaciones, juegos, Realidad Aumentada (RA), Realidad Virtual (RV), Interacción Tangible (IT)- que se presentan como herramientas que pueden apoyar los procesos de enseñanza y aprendizaje tanto en los diferentes niveles educativos de la educación formal como en educación no formal, y educación especial. La integración genuina de las TIC a las prácticas educativas exige nuevos escenarios de formación que promuevan aprendizajes significativos en los docentes, además de la implementación de metodologías flexibles de desarrollo y acceso al conocimiento sustentadas en la producción colaborativa y el acceso abierto a los recursos. Se hace necesario que los

docentes se apropien de la tecnología y adquieran las competencias necesarias para enriquecer sus prácticas docentes con la tecnología para incidir en una mejora de la educación en ciencias en el ámbito de la Provincia de Buenos Aires.

El objetivo general de investigación es brindar recursos educativos y herramientas TIC para ser usadas como soporte del proceso de enseñanza y aprendizaje de ciencias físicas, matemáticas y medio ambiente. En particular se orienta a empoderar a los docentes mediante el desarrollo de competencias TIC, ofreciendo dispositivos para la formación contínua de docentes mediante cursos de capacitación sobre el uso de las TIC en la enseñanza, y poniendo a disposición recursos que integren los conocimientos disciplinares, didácticos y tecnológicos. Estos dispositivos tiene por objetivo promover la formación de una comunidad de docentes de diferentes niveles educativos, incluidos investigadores, para trabajar colaborativamente en el diseño de propuestas de enseñanza con TIC.

Realidad Aumentada (RA)

Recientemente la RA se popularizó con el uso de aplicaciones móviles que se ejecutan en un celular o tablet de bajo costo, siendo las más fáciles de implementar y utilizar los siguientes dos tipos que no requieren equipo especial:

- Códigos QR (QRCode) y códigos de barra, para mostrar información relacionada o visitar sitios web
- Imágenes impresas, como los marcadores de RA o imágenes en general, para mostrar información y modelos 3D o videos registrados espacialmente con los marcadores

Realidad Virtual (RV)

Si bien existen diferentes dispositivos de visualización utilizados en RV, la tecnología más popularizadas son las gafas. Dado que el precio de estos dispositivos puede ser una limitación para su uso en educación, pueden emularse mediante el uso de celulares en soportes de cartón del tipo Google Cardboard[1]. Entre las aplicaciones educativas se resalta la realización de visitas virtuales. La toma de fotografías panorámicas puede hacerse mediante una cámara 360º o mediante una aplicación móvil que permita obtenerla a partir de un paneo con la cámara incorporada. Para la creación de paseos virtuales se utiliza la aplicación en la nube Google Tour Creator[2], que permite marcar puntos de interés con audio, imágenes y texto asociados en una secuencia de fotos panorámicas. Para visualizar el paseo virtual puede simplemente hacerse en un navegador, o puede utilizarse Google Expeditions[3], una aplicación Android que permite utilizar las gafas Google Cardboard.

RA y RV para comunicación de las ciencias

RA y RV resultan herramientas atractivas tanto para la educación en las aulas como también en el diseño de exposiciones y otras actividades vinculadas con la comunicación pública de la ciencia y de patrimonio histórico y natural. Castro et al presentan ejemplos de incorporación de RA en un museo y un sendero natural. En particular en el museo se incorporó el uso de posters aumentados, en particular con el vídeo de una excavación reciente donde se halló un gliptodonte y también con un modelo 3D del gliptodonte. En relación al paseo “Sendero Pampa”, el cual es un recorrido de un ambiente natural pampeano para observar la flora y fauna autóctonas, se incorporan códigos QR en cartelería existente de forma de vincular con la información existente en el sitio web.

[1] https://vr.google.com/cardboard/
[2] https://arvr.google.com/tourcreator/
[3] https://edu.google.com/expeditions/

Simulaciones y Juegos

Los videojuegos y simulaciones se emplean en la enseñanza de las ciencias, en particular en la enseñanza de la Física [5]. Bouciguez et al [6] presentan el desarrollo del videojuego "SpaceEscape: The F.E.M." utilizado para una secuencia didáctica para enseñar la inducción electromagnética. El proceso de desarrollo de un videojuego educativo implica integrar componentes lúdicos, tecnológicos y educativos, lo cual requiere un trabajo articulado y continuo en equipos interdisciplinarios.

Geogebra

Geogebra[4] es un software de uso y distribución gratuita, para el modelado y simulación en Matemática y en Física, con gran versatilidad y potencialidad educativa que permite relacionar distintas representaciones de un problema.
Natale et al [7] realizaron una experiencia colaborativa con un grupo de docentes de nivel secundario y universitario con la doble intención de modificar prácticas de enseñanza incluyendo las TIC en un aula de nivel universitario e iniciar un camino de estudio colaborativo en temas de enseñanza de la geometría en este nivel. Colaborativamente se realiza el diseño, estudio y puesta en el aula de una secuencia de dos problemas de geometría para trabajar con GeoGebra. Se registran las escenas del aula y las de discusión y reflexión a cargo del grupo colaborativo de docentes e investigadores.
Esta experiencia de formación e investigación se sostiene especialmente en cuatro supuestos de referencia:
- La clase se concibe como una comunidad que produce conocimientos matemáticos a partir de la interacción de los alumnos con problemas.
- La enseñanza de la geometría debe superar situaciones que ponen énfasis solamente en aspectos perceptivos de las representaciones de los objetos geométricos.
- El software GeoGebra es un programa concebido desde su diseño como herramienta didáctica, busca favorecer la exploración y la investigación como medios para aprender matemática.
- La constitución de grupos colaborativos de docentes e investigadores posibilita la producción de conocimientos relacionados con la propia práctica profesional. Esta producción se sostiene con dos lógicas y objetivos diferentes: la de la formación docente continua y la de la investigación.

Comunidad virtual para la formación continua

Los docentes necesitan apropiarse de modos de hacer con tecnología, adquirir las competencias necesarias para diseñar estrategias de enseñanza que integren las tecnologías tanto al aprendizaje de ciencias como a la gestión pedagógica de las prácticas de enseñanza.
Miranda et al [8] definen un objeto de Enseñanza (OE) como una propuesta para el aula que integra TIC e incluye posibles actividades, estrategias para su implementación, registros de las experiencias de clase y sus análisis, además posee una estructura que posibilita su publicación en un repositorio digital abierto. A este fin se propone el desarrollo de una metodología activa elaborada a partir de experiencias de formación continua situada gestadas en torno a la construcción colaborativa de propuestas de enseñanza con TIC .

[4] https://www.geogebra.org/

Líneas de investigación y desarrollo

- RA y RV aplicada en el salón de clases y en exposiciones de ciencia y patrimoni
- Modelado y simulación en Matemática y Física con Geogebra
- Videojuegos educativos para la enseñanza de Matemática y Física
- Dispositivos de formación continua de docentes para adquisición de competencias digitales para la integración de tecnologías a la enseñanza.

Resultados

Durante el año 2019 se han alcanzado los siguientes resultados:

- Se actualizó el sitio web del proyecto para la difusión de materiales y recursos educativos que se generen durante el desarrollo.[5]
- Se realizó el Tour Virtual Darwin en Punta Alta[6]. Autor: M.J.Abásolo. Fue presentado en la muestra itinerante Ruta Darwin del Ministerio de Ciencia, Tecnología e Innovación de la Prov. de Buenos Aires.
- Se inició un proyecto de investigación en colaboración con investigadores del área de Arqueología y Comunicación Social de la Facultad de Ciencias Sociales de la UNICEN, para generar contenidos y recursos que pongan en valor el patrimonio del Museo Mulazzi y el sitio arqueológico Arroyo Seco de la ciudad de Tres Arroyos. Se está trabajando en la creación de un tour virtual con imágenes 360 del sitio arqueológico y en el desarrollo de distintas infografías, videos y audios, a los cuales los visitantes del museo podrán acceder usando marcadores de RA. Directora: Massigoge, A.; Co-directora: Abásolo, M. J. Integrantes: Castro, M. L.; Garcimuño, M. Inicio: Octubre 2019
- Se creó un espacio online dinámico, en formato de repositorio, que posibilita la publicación de las producciones de los docentes y permite que las propuestas puedan ser utilizadas y comentadas o enriquecidas por otros docentes que se interesen en las mismas.
- Se procesó la información registrada de la implementación de los OE creados por los equipos docentes. Se analizó la información y se definieron categorías para su comunicación en el repositorio que permitieran transmitir lo más claramente posible la experiencia de aula para su potencial réplica por otros docentes en otros contextos áulicos
- Se avanzó en el ajuste y mejora de la metodología para generar OE con el objetivo de profundizar las propuestas desde las dimensiones disciplinar, tecnológica y pedagógica. Se pudo observar un mayor compromiso con la tarea y la valorización de las prácticas docentes tanto en los resultados obtenidos como en el reconocimiento de la comunidad. Este avance fue presentado en las V Jornadas de Enseñanza e Investigación Educativa en el campo de las Ciencias Exactas y Naturales de la Facultad de Humanidades y Ciencias de la educación de la UNLP [9].
- Se conformó un grupo colaborativo de docentes de la Facultad de Ciencias Exactas (UNICEN) para realizar estudios e investigaciones sobre la enseñanza de geometría con GeoGebra.
- Se completó el sitio web del circuito de reconocimiento de especies nativas conocido como Sendero Pampa.[7] , con la incorporación de información de nuevas estaciones a las que se les incorporó RA en formato de códigos QR.
- Se realizó un vídeo “Serie

[5] https://reforticca.info.unlp.edu.ar/

[6] https://poly.google.com/u/0/view/fwResn04qIP

[7] ttps://www.senderopampa.com.ar/

Instantes Naturales Bom-Bus"[8], referido a los polinizadores locales con la técnica *stop-motion* que se incorporó al sitio web Sendero Pampa. Se diseñaron secuencias didácticas que utilizan el vídeo Bom-bus: "Guía Complementaria para trabajar con Bom-bus" y "Encuesta de Proceso de Aprendizaje", para el docente.

- Se está trabajando en la implementación de un cuadricóptero (DRON) utilizando los sensores y la potencia computacional de un teléfono inteligente (celular). Con este proyecto se busca propiciar instancias de alfabetización digital de docentes y estudiantes, a través de la integración de las áreas de conocimiento emergentes, como la programación y la robótica, y facilitar un recurso digital y propuestas pedagógicas, que favorezcan el aprendizaje de campos tradicionales del saber, como las ciencias naturales.
- Se desarrolló el taller de "Realidad Virtual y Realidad Aumentada" a cargo de M.J.Abásolo (III-LIDI) y L. Castro (UNICEN) dirigido a docentes de nivel primario, 21 de mayo 17hs, Escuela 53, Tandil
- Se desarrolló el taller de Realidad Virtual y Realidad Aumentada, impartido por M.J.Abásolo (UNLP), M.L.Castro(UNICEN), M.Garcimuño (UNICEN) 19º Foro de Enseñanza de Ciencias y Tecnologías realizado en Feria Internacional del Libro de Buenos Aires, que tendrá lugar en el Predio La Rural entre el 25 de abril y el 13 de mayo de 2019.
- Se desarrolló el taller sobre Realidad Aumentada, Realidad Virtual e Interacción tangible en escenarios educativos para docentes de todo el país con más de 35 inscriptos a partir de una invitación de la Red de Universidades con Carreras en Informática (RedUNCI). El taller estuvo a cargo de la Dra. Cecilia Sanz y acompañada por la Lic. Verónica Artola y Lic. Natalí Salazar Mesía. Duración: 10 hs. intensivas distribuidas en dos encuentros.
- Se desarrolló el taller "Espacios Interactivos de Ciencia: alternativa para la Educación en Ciencias", en el 4to Encuentro de Física del Noroeste del Chubut, organizado por IES 813, Puelo, y IES 804, Esquel. Esquel, 17 y 18 de octubre de 2019. María Luján Castro.
- Se realizó la Jornada de trabajo de docentes de Matemática y Física de escuelas secundarias. 16 de febrero 2019 de 9 a 16 hs, UNICEN, SUM de Aulas Comunes II Campus Universitario Tandil. Objetivo: exponer las propuestas de enseñanza con GeoGebra y sus respectivas puestas en aula. Asistieron diez docentes, miembros de los cuatro equipos conformados, quienes compartieron producciones, recorridos, logros y dificultades. El trabajo de cada equipo y los objetivos que fueron alcanzando fueron diversos. Dos de los equipos presentaron y publicaron sus trabajos en diferentes jornadas [10-11]. A partir del acercamiento inicial al estudio de la experiencia y las reflexiones que surgieron en los intercambios tomaron la decisión de profundizar en el estudio de las mismas. Para ello se les brindó apoyo en la sistematización de los registros y en los aspectos teórico y metodológicos que les permitieron avanzar en el análisis de los datos.
- Se desarrolló el taller "Pensamiento computacional en la Enseñanza de las Ciencias" a cargo de la Dra. Graciela Santos y la participación de Dra Andrea Miranda, Gabriela Cenich, Prof Andrea Vergara, Prof. Milagros Paoletti y estudiante Camila Pérez Castillo. Destinado a docentes de las ciudades de Ayacucho (20 docentes), Azul (26 docentes) y Benito Juárez (11 docentes), a partir de una convocatoria de Ministerio de Ciencia, Tecnología e innovación de la Provincia Bs As. Cada ciclo del curso taller se realizó en tres

[8] https://www.youtube.com/watch?v=Cd6CG8nQj50

encuentro de 12 hs. Se requirió elaboración de material didácticos y la compra de componentes de Arduino y 3 pendrive. Este curso taller fue la primera instancia para formar comunidades de aprendizaje colaborativo mediante a partir de equipos de docentes de Física, Química y Biología en las ciudades de Ayacucho, Azul y Benito Juárez, con el objeto de integrar el pensamiento computacional al currículo .

- Se desarrolló el taller de AstroCódigo en Colegio Corazón Eucarístico de La Plata (8 alumnos), ExpoUNLP 2019 (30 alumnos), Colegio Nacional Rafael Hernández (4 docentes y 30 alumnos), UNLP en VIVO (50 alumnos y 6 tutores).[9]
- Se desarrolló el taller con HuVi, aplicación de Realidad Virtual, orientada a acercar patrimonios de la Argentina a niños con vulnerabilidad social. Se realizó en 2019 en EGB nº 17 Ignacio Gorriti, Berisso (25 niños y 1 maestra), Asociación El Roble (alrededor de 16 niños y 6 tutores), ExpoUNLP (más de 100 estudiantes recorrieron el stand).
- Se desarrolló el taller de Autorregulación de los Aprendizajes. Escuela Secundaria Nro. 51 de Los Hornos (alumnos alrededor de 30 y 6 docentes)
- Se desarrolló el taller con docentes sobre interacción tangible. Los docentes de diferentes escuelas de la región utilizaron una herramienta de autor para crear aplicaciones basadas en interacción tangible sobre una mesa interactiva. Como parte del taller crearon: una aplicación de línea de tiempos para asociar objetos físicos con su año de invención; otra para clasificar animales (juguetes) en las categorías invertebrados y vertebrados, y una para abordar el tema de tiempos de degradación de diversos materiales, utilizando el material real (ejemplo: una lata, una botella plástica, etc.) ubicándola según su tiempo de degradación.

[9] https://www.nacio.unlp.edu.ar/2019/11/una-aventura-de-programacion/

Formación de recursos humanos

El desarrollo de algunos recursos educativos fue realizado por alumnos dentro del marco de la carrera Profesorado de Informática de la Facultad de Cs. Exactas de la UNICEN y alumnos del posgrado de la Facultad de Informática de la UNLP.
Se llevan a cabo las siguientes direcciones de recursos humanos relacionadas con estas líneas de investigación:

- María José Bouciguez "Ambientes virtuales altamente interactivos basados en videojuegos y simulaciones para la educación en ciencias" Directores: Santos, G. y Abásolo, M.J. (tesis de Doctorado en Ciencias, Fac. Inf. de la UNLP, en curso)
- Marisa Salerno "Herramienta colaborativa para la creación de Objetos de Enseñanza" Directores: Miranda, A. y Jonás, I. (tesis de grado de Ing. de Sistemas, Fac. Cs. Exactas de la UNICEN, en curso).
- Nicolas Laugas. "Utilización de sensores de smartphone con Android para controlar los motores de un cuadricóptero". Directores: Castro, M. L. y Marone, J. Becario EVC-CIN 2018 Inicio: Junio de 2019.

Referencias

[1] Abásolo M.J., Castro M., Santos G., Sanz C., Bouciguez M.J., Garcimuño M., Cenich G.,Miranda A., Papini C. (2019) *REFORTICCA: Recursos para el Empoderamiento de FORmadores en TIC, Ciencias y Ambiente.* XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019), RedUNCI, ISBN 978-987-3619-27-4

[2] Abásolo, M.J.; Sanz, C.; Santos, G.;

Castro, M.; Miranda, A.; Cenich, G.; Bouciguez, M.J. ; Papini, C. (2018) *REFORTICCA: Recursos para el Empoderamiento de FORmadores en TIC, Ciencias y Ambiente.* XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018), pp. 1197-1202, RedUNCI, ISBN 978-987-3619-27-4

[3] Abásolo, M.J.; Sanz, C.; Naiouf, M.; De Giusti, A.; Santos, G.; Castro, M.;Bouciguez, M.J. G. (2017) *Realidad Aumentada, Realidad Virtual e Interacción Tangible para la Educación.* XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017), pp. 1312-1316, RedUNCI, ISBN 978-987-42-5143-5

[4] Maria Lujan Castro, Maria José Abasolo, Mayra Garcimuño, Cecilia Ramirez (2019) *"Realidad Aumentada como complemento en la Comunicación Pública de la Ciencia"* RED POP 2019 - Panamá 22 al 25 de Abril de 2019.

[5] Bouciguez, M. J.; Santos, G. ;Abásolo, M. J. (2013) *Potencialidad de los videojuegos en el aprendizaje de Física.* Actas de WEFA 2013 I Workshop de Enseñanza de Fisica en Argentina, Universidad Nacional del Centro de la Pcia. de Bs. As., ISBN 978-950-658-342-2.

[6] Bouciguez, M.J., Braunmüller, M., Bravo, B., Santos, G.y Abásolo, M. J. (2019) *Desarrollo del videojuego "SpaceEscape: The F.E.M."para una secuencia didáctica de inducción electromagnética*. X Congreso Iberoamericano de Educación Científica Enseñanza y Aprendizaje de las Ciencias en Debate. Montevideo, Uruguay, 25, 26-28 de Marzo del 2019

[7] Mauro Natale; María Cecilia Papini (2019) *Producir geometría con GeoGebra. Una experiencia colaborativa en el nivel universitario*. Actas V Jornadas de Enseñanza e Investigación Educativa en el campo de las Ciencias Exactas y Naturales, realizadas en la Facultad de Humanidades y Ciencias de la Educación. Universidad Nacional de La Plata. Ensenada, mayo de 2019.

[8] Miranda, A.; Cenich, G.; Santos, G.; Papini, C.; Bouciguez, M. J., (2019) *Construcción colaborativa de Objetos de Enseñanza. Formación continua para promover PEA.* Revista digital Docentes Conectados, 2(4). San Luis: Centro de Informática Educativa. ISSN 2618-2912

[9] Miranda, A.; G.; Santos, G.; Cenich, Papini, C.; Bouciguez, M. J., (2019), Experiencia de formación continua para la enseñanza de Ciencias con TIC. En Actas V Jornadas de Enseñanza e Investigación Educativa en el campo de las Ciencias Exactas y Naturales Facultad de Humanidades y Ciencias de la Educación. Universidad Nacional de La Plata. La Plata: UNLP. ISSN 2250-8473

[10] Borrego, E., Ciappina, D., Médico, C. (2019) *Construcciones imposibles de cuadriláteros con GeoGebra. Una experiencia en el aula.* Actas V Jornadas de Enseñanza e Investigación Educativa en el campo de las Ciencias Exactas y Naturales, Fac. de Humanidades y Cs. de la Educación, UNLP.

[11] Atela, M. A.; Fernández; J. P., Vila, M. (2019). *Articulación entre nivel primario y secundario. Una experiencia alrededor de la proporcionalidad mediada por TIC.* Actas V Jornadas de Enseñanza e Investigación Educativa en el campo de las Ciencias Exactas y Naturales, Fac. de Humanidades y Cs. de la Educación, UNLP.

Proyectos de Innovación y Transferencia de Tecnologías de la región del Noroeste de Buenos Aires (PRITT NOBA)

Claudia Russo[1], Hugo Ramón[1], Sandra Serafino[2], Benjamin Cicerchia[2], Mónica Sarobe[2], Agustín Balmer[2], Sebastian Adorno[2], Gabriel Perez[2], Marcelo Guiguet[2], Damian Montes de Oca[2], Guido Italiano[2]

Instituto de Investigación y Transferencia en Tecnología (ITT) [5] – CIC

Escuela de Tecnología

Universidad Nacional del Noroeste de la Provincia de Buenos Aires

Sarmiento y Newbery, Junín (B) – TE: (0236) 4477050

{claudia.russo, hugo.ramon, sandra.serafino, lucas.cicerchia, monica.sarobe, agustin.balmer, sebastian.adorno, gabriel.perez, marcelo.guiguet, damian.montesdeoca, guido.italiano
}@itt.unnoba.edu.ar

Resumen

Los productos derivados del agro requieren de una gestión adecuada de la producción y de los factores involucrados en ella, dónde la Agricultura de Precisión (AP) es la respuesta más adecuada para lograr los requerimientos actuales de la producción [1]. Ésta se trata de la administración del campo basado en la observación, la medición y la respuesta frente a la variabilidad, temporal y espacial, inter o intra cultivo [2]. La AP propone un manejo de la tierra y el cultivo de forma selectiva de acuerdo a las necesidades del mismo, junto con una correcta administración de las semillas y los agroquímicos con el fin de mejorar la producción, mejorar los costos y hacer un uso responsable del medio ambiente [3] [4] [5]. Desde hace algunos años la UNNOBA viene trabajando en diferentes proyectos en tecnología, particularmente en el área de robótica y en la aplicación de tecnología aplicada al agro. En Robótica con la implantación de microcontroladores en sistemas automatizados, la utilización de sensores de todo tipo y dotando de conectividad a sus desarrollos. Desde el año 2015 se trabaja en el desarrollo de una Plataforma de Fenotipado de Alto caudal con relevamiento de datos a campo, proyecto de trabajo conjunto con el INTA Estación Experimental Agropecuaria Pergamino en el marco de proyectos de Innovación y Transferencia de Tecnologías (PRITT). El objetivo fundamental es el desarrollo de una plataforma auto navegable de sensado de datos mediante uso de visión artificial y procesamiento digital de imágenes.

La relevancia de estos proyectos y su impacto regional los han llevado a integrar el Banco Nacional de Proyectos de Desarrollo Tecnológico y Social (PDTS) de la Secretaria de Gobierno de Ciencia,

[1] Investigador Asociado CIC (Adjunto sin Director)
[2] Investigador ITT-UNNOBA

Tecnología e Innovación Productiva del Ministerio de Educación, Cultura, Ciencia y Tecnología, con resolución RESOL-2019-89-APN-SECACT#MECCYT.

Palabras clave: Robótica, Imágenes, Visión artificial, Agricultura de Precisión, Big Data.

Contexto

La pertinencia de estos proyectos se enmarca en los objetivos de la Universidad y en su misión como institución académica y generadora de conocimiento científico atendiendo en forma permanente e interdisciplinaria las aspiraciones, problemas y necesidades de la sociedad, colaborando en su planteamiento, análisis y solución; y contribuyendo al desarrollo social sustentable de la región en la que se encuentra inmersa. Y también lo hace dentro de las actuales líneas de investigación que la institución prioriza como son los "Sistemas de producción regional | Diversificación de los sistemas productivos | Agregado de valor a las producciones regionales", y el "Mejoramiento genético y biotecnología". La UNNOBA aporta su experiencia y conocimiento en aspectos como ser la captura y posterior tratamiento de la información, el INTA por su parte, direcciona los esfuerzos indicando dónde buscar, qué buscar, en qué tiempo hacerlo y cuál es la mejor técnica para la extracción de los datos solicitados. La Estación Experimental Agropecuaria del INTA en Pergamino es un centro de referencia tecnológica regional en cuestiones agrarias. Productores e investigadores confluyen en la entidad volcando su trabajo y experiencia en temas tales como ser la mejora en los procesos agrícolas y la mejora genética de especies vegetales susceptibles de cultivos.

El equipo que lleva adelante estos proyectos está constituido por profesionales de ambas instituciones, UNNOBA e INTA. Por parte de la UNNOBA intervienen docentes e investigadores pertenecientes al ITT, y estudiantes de las carreras de Informática de la Escuela de Tecnología de la UNNOBA. Y por parte del INTA, intervienen mejoradores genéticos, ecofisiólogos e ingenieros agrónomos.

Introducción

La aplicación conjunta de las técnicas de *Mejora Genética* y la *Agricultura de Precisión* son la respuesta más adecuada a las limitaciones que impone el cambio climático y al aumento de la demanda mundial en alimentos y productos derivados como ser biocombustibles. El fenotipado a campo es una herramienta esencial en mejora genética, ya que posibilita la toma de datos en forma rápida, masiva y precisa. A estos fines se piensa en el desarrollo de Plataformas de Fenotipado de Alto Rendimiento [5].

El desarrollo tecnológico alcanzado en ese sentido por la UNNOBA en colaboración con el INTA dentro del marco de la primera Convocatoria 2014 del PRITT, puso a nuestra disposición una plataforma de fenotipado con la variedad de sensores necesarios para evaluar cultivos de maíz

hasta el estadío v5/v6, que dispone de alta capacidades de procesamiento y trabaja en la modalidad stand-alone. Esta plataforma es tripulada y dispone de dispositivos de visión inteligente que utilizan algoritmos de tracking invertido para determinar en forma automática el momento de disparo de las cámaras encargadas de captar las imágenes de los surcos de cultivos. Existen otras tantas universidades dedicadas a generación de plataformas terrestres de sensado, de las que se han analizado experiencias para volcar a nuestro trabajo, como por ejemplo el Centro Australiano de Robots a Campo de la Universidad de Sydney [6] que es una de las pioneras.

Dentro del área de sensores, se destaca la importancia de la captura de información por reconocimiento de imagen. Las imágenes que esta plataforma releva corresponden a las bandas del visible e infrarojo cercano del espectro electromagnético.

La información a recolectar incluye conteo de plantas, emergencia, volumen de follaje del cultivo y NVDI. Algunos de estos aspectos pueden obtenerse utilizando un mismo tipo de sensor, otros no. El tipo de procesamiento de imagen requerido, está fuertemente ligado a la naturaleza del rasgo buscado. Es necesario el posterior procesamiento digital de las imágenes relevadas. Respecto del procesamiento de los datos, imágenes en particular, se aplican técnicas de visión artificial, preprocesado y procesado digital dependiendo de los resultados a obtener por cada rasgo particular relevante para el proceso de fenotipado. Entre las técnicas utilizadas podemos mencionar la aplicación de filtrado espacial y frecuencial para el tratamiento de ruido y mejora en el caso de ser necesario; morfología matemática, crecimiento de regiones, y contornos activos entre otras para el proceso de segmentado; clasificadores estadísticos y redes neuronales para el proceso de clasificación y análisis de texturas, color y forma para el proceso de descripción e interpretación de datos.

Una segunda Convocatoria PRITT en el año 2017 nos ha permitido avanzar hacia una plataforma autonavegable mediante técnicas de visión artificial que admite además recorrida intra-surcos y el análisis de etapas de fenotipados que incluyen la floración (V7 en adelante). El objetivo de este nuevo desarrollo se basa en la necesidad de generar equipo de bajo costo, reutilizable, modular y con alto nivel de portabilidad, de alta capacidad de relevamiento de datos, de alta autonomía tanto energética como en el proceso de navegación, y con la posibilidad de interconexión con otras plataformas terrestres o drones. En los estudios realizados, los ensayos a campo recorridos y sensados ya por nuestro equipo y las publicaciones analizadas se detectó que los datos característicos para relevar sobre un determinado tipo de cultivo, por estar éstos sujetos a variabilidad fenológica, son diferentes e imposibles de captar de una única forma (único recorrido, única posición de las cámaras, única altura, etc.).

Líneas de Investigación, Desarrollo e Innovación

El desarrollo de estosproyectos han permitido al equipo de trabajo avanzar en varios aspectos de relevancia que incluyen el análisis, estudio y desarrollo de procesos de tecnología informática aplicada a un área de especial sensibilidad en nuestra región como es la agricultura de precisión. Debido a que la UNNOBA se encuentran ubicada dentro de la Región Pampeana, más específicamente en lo que se denomina la Pampa Húmeda, una de las regiones más relevantes en lo que respecta a producción agrícola, se espera que este trabajo sea el inicio de una serie de trabajos y proyectos de investigación que la universidad siga fomentando para la continuación y mejora de plataformas robóticas que sirvan de soporte a los especialistas del sector agrícola. Sumado al beneficio directo de los recursos humanos involucrados, con el objetivo fundamental de consolidar un equipo de trabajo que se especialice en la temática.

Resultados y Objetivos

Los resultados obtenidos por el momento nos han permitido generar dos plataformas de sensado a campo, que permiten relevar los datos en su totalidad durante el proceso evolutivo de un cultivo permitiendo recorrer los ensayos tanto intra como extra surcos y/o a través de caminos. También nos ha permitido desarrollar y mejorar un sistema de sensado de datos automatizado y portable a otros tipos de cultivos y/o frutos. Se ha desarrollado en el marco de estos proyectos además una placa que permite integrar los sistemas de navegación por visión artificial, de sensado, y de storage. Por otro lado, se ha generado un banco de datos relevante de videos e imágenes sobre ensayos de cultivos de diferentes variedades genéticas de maíz. También se han desarrollado un conjunto de librerías de procesamiento digital de imágenes que permiten resolver algunos de los problemas planteados por los especialistas sobre los conjuntos de datos relevados durante los ensayos, como por ejemplo el estado de emergencia. Por último, se espera poder lograr que la plataforma trabaje de modo conjunto con un vehículo aéreo no tripulado, al cual se necesita proveer de sistema de visión que permita realizar un mapeo del cultivo, pudiendo detectar parcelas, tanto para navegar el cultivo como para, además, a partir de las imágenes obtenidas, obtener información sobre el cultivo en otra escala.

Formación de Recursos Humanos

Estos proyectos han permitido la formación de varios recursos humanos en el área de Tecnología, mediante trabajos de Finalización de las Carrera de Ingeniería en Informática y Licenciatura en Sistemas, becas de la Comisión de Investigaciones Científicas (CIC) de la Provincia de Buenos Aires, becas del Consejo Interuniversitario Nacional (CIN), Becas de Investigación y Transferencia para Alumnos avanzados de las Carreras de Informática (PROMINF), y Becas de Investigación y Transferencia (BIAT).

Bibliografía

[1] D. Tilman, C. Balzer, J. Hill, and B. L. Befort, “Global food demand and the sustainable intensification of agriculture,” Proc. Natl. Acad. Sci. United States Am. United States Am., vol. 108, no. 50, pp. 20260–20264, 2011.

[2] E. R. and J. L. Y. Wang, K. Lee, S. Cui, “Agriculture robot and applications,” in Future information engineering and manufacturing science : proceedings of the 2014 International Conference on Future Information Engineering and Manufacturing Science (FIEMS 2014), 2015, pp. 43–46.

[3] C. Fernández-Quintanilla, “Agricultura de Precisión,” Cienc. y Medio Ambient. - Segundas jornadas científicas sobre medio Ambient. del CCMA-CSIC, pp. 189–19, 2002.

[4] S. Blackmore, “Developing The Principles of Precision farming,” Cent. Precis. Farming, Denmark, no. January 2002, 2002.

[5] García E. y Flego F., “Tecnología agropecuaria. Agricultura de precisión.” [Online]. Available: http://www.palermo.edu/ingenieria/downloads/pdfwebc&T8/8CyT12.pdf. [Accessed: 09-Mar-2018].

[6] “Sydney University-Australian Centre for Field Robotics.” [Online]. Available: https://sydney.edu.au/engineering/our-research/robotics-and-intelligent-systems/australian-centre-for-field-robotics/agriculture-and-the-environment.html [Accessed: 09-Mar-2018].

Tesis Doctoral

Robustez de las métricas de clasificación de cadencia de tecleo frente a variaciones emocionales

Doctorado en Ciencias Informáticas, Facultad de Informática, UNLP
Tesista: Enrique P. Calot
Jorge S. Ierache (director), Waldo Hasperué (codirector)
Defendida el 18 de junio de 2019.

Laboratorio de Sistemas de Información Avanzados,
Facultad de Ingeniería, Universidad de Buenos Aires, Argentina
ecalot@lsia.fi.uba.ar

Resumen La dinámica de tecleo permite la identificación de una persona por la forma en que escribe. La tesis aquí resumida se enfoca en analizar la robustez de los algoritmos de análisis de cadencia de tecleo frente a variaciones en los registros biométricos mediante electroencefalografía y cuestionario de autoinforme, utilizando el enfoque dimensional para modelar estados afectivos. Se realizó un experimento para capturar patrones de tecleo en diferentes estados afectivos. Los resultados sugirieron que la tasa de aciertos para ciertas distancias de clasificación, como las métricas A y R, la distancia de Camberra, Manhattan y una distancia basada en Minkowski se ven influenciadas negativamente por los cambios en las respuestas de excitación y valencia. La distancia euclídea fue la menos afectada de las seis evaluadas.

Keywords: Keystroke dynamics, Biometrics, EEG, Sentiment analysis

1. Introducción

El ritmo de escritura es único para cada persona, y las variaciones entre las personas son tan profundas que permiten la identificación del mecanógrafo. En el teclado, el estudio de los patrones de tecleo así como los ritmos en sí se denominan *keystroke dynamics*. En las computadoras, el inicio de esta área particular de investigación se remonta a Gaines et al. [1], quien en 1980 realizó el primer análisis estadístico con textos libres y palabras fijas. Cada estilo de escritura es exclusivo de su propietario y no puede ser robado, perdido u olvidado[2].

El análisis de pulsación de teclas, como característica biométrica, se puede aplicar sobre un texto estático o un texto libre continuo. Los textos estáticos, como las contraseñas, con frecuencia se escriben utilizando la misma secuencia; mientras que los textos libres continuos comparan estilos de escritura usando secuencias de teclas comunes entre dos textos para determinar si pertenecen a la misma persona. La longitud de estos textos puede diferir considerablemente. Los textos libres requieren cadenas de teclas más largas para lograr la misma precisión que en textos estáticos[3].

2. Motivación

La necesidad de brindar una mayor capacidad de autenticación de usuarios con el empleo de biometría fue en aumento en los últimos años. La cadencia de tecleo permite generar firmas de seguridad sin tener que usar dispositivos específicos, siendo completamente transparente y no invasiva para el usuario. Teniendo en cuenta la hipóstasis de que la variabilidad emocional influiría en la calidad de la autenticación, es esencial construir lotes de datos representativos de la variabilidad emocional; además, determinar qué métrica es la más adecuada para la cadencia de tecleo, cuál adoptar, qué mejor técnica usar y qué tan robusta es cada una.

3. Temas de investigación

La tesis[4] se centró en los cambios emocionales a corto plazo, principalmente la relajación y la emoción. Sin pretender incursionar en el área de la neurociencia, solo se enfocó en analizar la robustez de los algoritmos de keystroke dynamics frente a las variaciones en los registros biométricos a través de la electroencefalografía (EEG), en particular las ondas alfa, beta o cerebrales que, según la literatura [5–15], están asociados con estados de relajación y excitación.

El desarrollo abarca tres preguntas, las dos primeras son secundarias y necesarias para avanzar sobre la tercera, que es la central; a saber

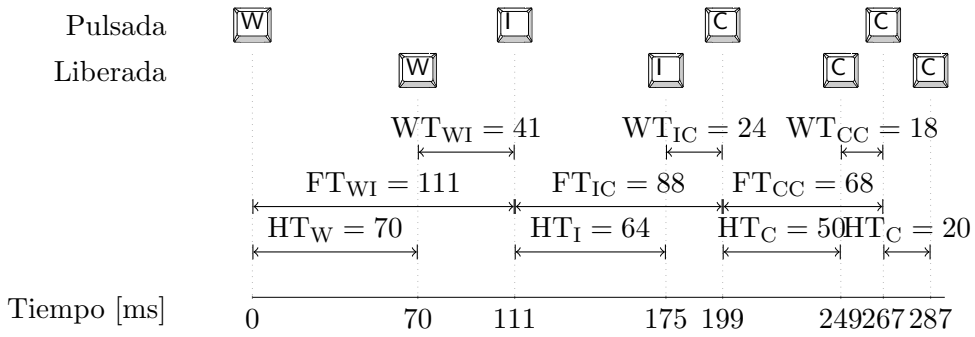


Figura 1: Tiempo interteclas.

1. ¿qué métricas se deben utilizar en el análisis de los resultados?
2. ¿qué cantidad de keystrokes es necesaria para que las pruebas sean estadísticamente significativas?
3. pregunta principal: ¿pueden los estados emocionales afectar la calidad de los algoritmos de clasificación mediante métricas utilizando cadencia de tecleo?

En clasificación de cadencia de tecleo las métricas son formas de medir la distancia entre dos patrones para sí decidir si pertenecen o no a una misma clase (por ejemplo, la misma persona).

3.1. Keystroke dynamics

La Figura 1 presenta un ejemplo de la cadencia de para la palabra "wicc". Hold time (HT) y wait time (WT) son el tiempo que una tecla está presionada y el tiempo interteclas respectivamente. Si se presiona una tecla dos veces, generará un solo evento HT, que es la media de ambos tiempos; lo mismo ocurrirá cuando ocurra un evento WT entre el mismo par de teclas.

Combinando los valores compartidos entre dos muestras nos permite formar un vector de números y calcular su distancia implica encontrar una medida de cuán distintos son los patrones de ambas muestras. Los vectores pueden ser contrastados con varias métricas, como por ejemplo Manhattan ($\mathcal{L}^1$) y Euclídea ($\mathcal{L}^2$), aunque la tesis presenta otras varias.

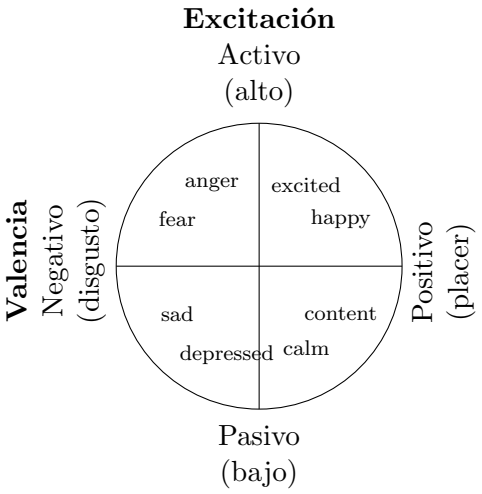
3.2. Modelos emocionales

Se utiliza el *Circumplex of Affectivity* de James Russell [16, 17], también conocido como *circumplejo* o modelo de Excitación–Valencia. Este modelo es bidimensional y cada eje es excitación (relajado *vs.* exaltado) y valencia (placer *vs.* disgusto). La Figura 2a ejemplifica la ubicación de 8 emociones utilizando este modelo [18]. La Figura 2b muestra las fases (con sus nombres en inglés) empleadas en la experimentación de esta tesis y como las personas fueron estimuladas: fase 1, high arousal positive valence (HAPV); fase 2, low arousal positive valence (LAPV); fase 3, low arousal negative valence (LANV); y fase 4, high arousal negative valence (HANV).

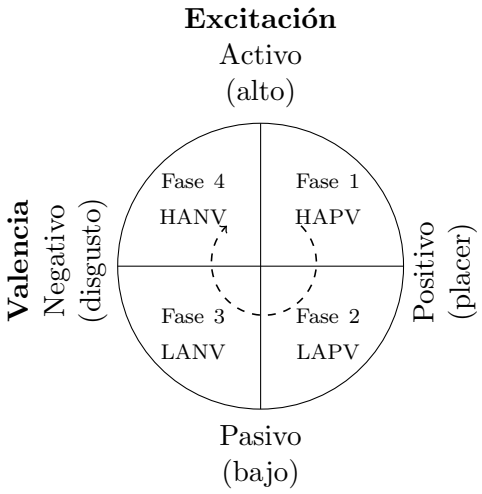
3.3. Electroencefalografía (EEG) como interfaz cerebro–máquina (BCI)

Una *interfaz cerebro-computadora* (BCI) permite a una persona controlar o comunicarse con el mundo externo sin la necesidad de utilizar las rutas de salida estándar del cerebro, los nervios o los músculos. Los sistemas BCI proporcionan una nueva opción de comunicación y control esencial para las personas con discapacidad que deterioran los canales normales de comunicación y control. También podrían proporcionar a las personas sin discapacidades un canal de control suplementario o un canal de control en circunstancias especiales [19–21].

La experimentación de esta tesis utiliza una de las formas posibles de leer la actividad bioeléctrica del cerebro (EEG): los headsets *Emotiv EPOC*, un dispositivo de bajo costo compuesto por 14 electrodos más 2 electrodos de referencia (también utilizados para reducir el ruido), construidos para poder colocarlos en la cabeza. Los electrodos se disponen en el cuero cabelludo en posiciones ya especificadas según el *Sistema internacional 10–20* (ver Figura 3a) [22]. En este sistema, las distancias entre los electrodos suelen ser del 10 % o 20 % del tamaño de la cabeza (de ahí el nombre 10–20). La configuración del electrodo usando EPOC incluye electrodos intermedios (ver Figura 3b).

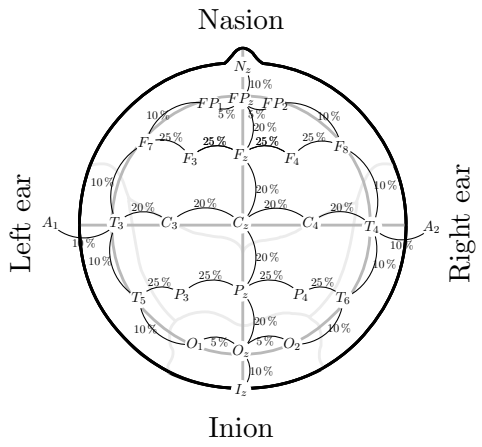


(a) Ubicación de ocho emociones. Se conserva el nombre de las mismas en el inglés, idioma original.

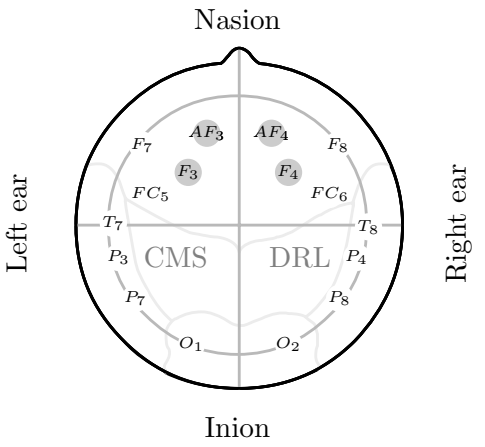


(b) Orden de las fases de adquisición de la experimentación.

Figura 2: Modelo Excitación–Valencia.



(a) Posicionamiento de electrodos en el sistema internacional 10–20.



(b) Ubicación de los electrodos del Emotiv EPOC. AF y F, 3 y 4 están en la zona de emociones.

Figura 3: Ubicación de los electrodos EEG.

3.4. Adquisición

La Figura 2b de la Sec. 3.2 describe el camino por el cual el participante es guiado a través del circumplejo de Russell. Para esto, el individuo fue estimulado con videos e imágenes obtenidas de IAPS —de los cuales se conocen sus valores aproximados en el circumplejo gracias al trabajo de Lang et al. [23]— y luego, el usuario ingresó textos libres, capturando sus patrones de keystroke dynamics. A lo largo del proceso, se midieron las ondas EEG. La Figura 4 describe la arquitectura del experimento.

OpenViBE se utilizó para recibir y almacenar, filtrar y procesar posteriormente las muestras en bruto. Se puede usar para calcular la posición del estado emocional en el modelo de Excitación–Valencia [24, 6].

3.5. Hipótesis secundarias

Se eligieron 10 oraciones de aproximadamente 50 caracteres cada una, con un total de 533 caracteres. Esto induce a los usuarios a generar un promedio de 739 pulsaciones de teclas (oscilando entre 666 y 823 con una desviación estándar de 37.7) y conjuntos de eventos de dimensión promedio 282 (oscilando entre 220 y 365 con una desviación estándar de 46.3) gracias a la alta repetición de pulsaciones de teclas. Aunque las oraciones se introducen sin clasificar, las cadencias son las mismas

en todas las fases y temas. Solicitar más pulsaciones de teclas pone en peligro todo el experimento a medida que el usuario comienza a aburrirse.

Las preguntas *¿qué métricas se deben utilizar en el análisis de los resultados?* y *¿qué cantidad de keystrokes es necesaria para que las pruebas sean estadísticamente significativas?* fueron resueltas tras analizarse cuatro conjuntos de datos (ver Figura 5): uno propio (LSIA, que contiene 68808 muestras de más de 300 keystrokes) y tres pertenecientes a PROSODY [25] (una base de datos de 4,6 millones de keystrokes tipeados por personas con contenido de opinión: revisión de restaurantes (REVIEW), matrimonio gay (GAY) y control de armas (GUN); diseñados para producir diferentes estados emocionales en los participantes, aunque no se han adquirido mediciones biométricas.

Las mejores métricas resultaron ser las Métricas R[3] y A[3], Manhattan ($\mathcal{L}^1$) y Minkowski con $p = 0{,}4$ ($\mathcal{L}^{0,4}$)[26]; y se decidió por un valor de 350 keystrokes por muestra ya que todos los casos fueron asintóticos no se observan diferencias significativas al superar este umbral[27].

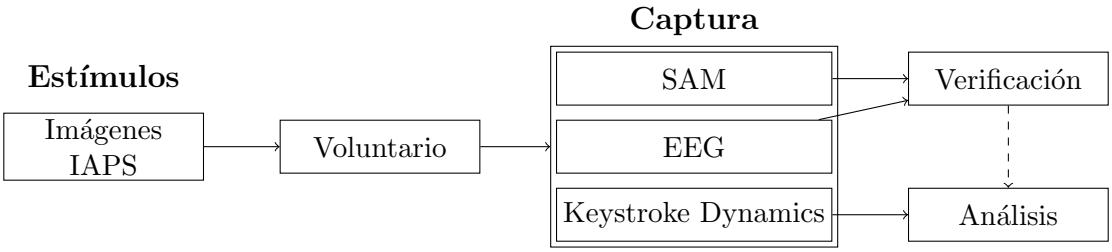
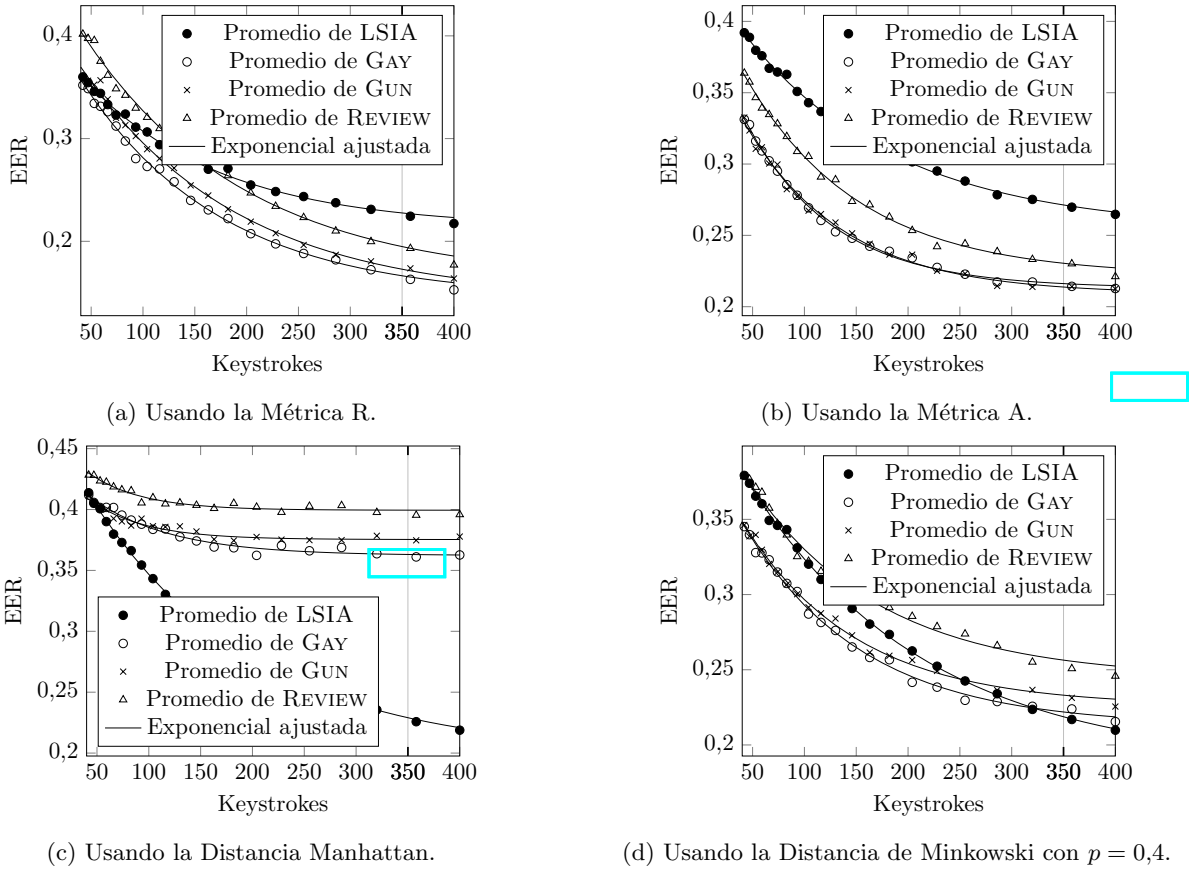


Figura 4: Descripción experimental.



(a) Usando la Métrica R.

(b) Usando la Métrica A.

(c) Usando la Distancia Manhattan.

(d) Usando la Distancia de Minkowski con $p = 0{,}4$.

Figura 5: Curvas exponenciales producidas por el EER para pulsaciones de teclado que varían en el rango 40–400 para diferentes conjuntos de datos y métricas.

Cuadro 1: Resultados del experimento para las métricas evaluadas.

		Métrica A		Métrica R		Camberra		Minkowski ($\mathcal{L}^{0,4}$)		Manhattan ($\mathcal{L}^1$)		Euclídea ($\mathcal{L}^2$)	
Emociones no combinadas	Usuario	EER	máx/mín	EER	máx/mín	EER	máx/mín	EER	máx/mín	EER	máx/mín	EER	máx/mín
	1	29.1±5	37/23	22.3±2	23/19	31±3	34/27	33.4±3	38/29	35.3±2	38/34	39.3±5	46/31
	2	29.2±4	35/23	30.5±4	35/23	31.2±4	35/25	9.1±6	19/4	11.3±6	19/4	13±3	19/11
	3	30.8±0	31/31	23.7±2	27/20	28.5±2	31/27	32.4±2	35/31	34.7±2	38/31	36.1±2	38/34
	4	26.5±3	31/22	18.9±3	22/15	22.8±3	27/19	27.7±2	31/27	30.6±3	35/27	32.9±3	39/31
	5	34.7±2	38/31	25.8±5	35/20	35.8±2	38/34	33.5±3	37/27	33.5±4	37/27	39.2±3	43/35
	6	31.2±3	37/29	27.8±4	35/23	33.8±3	38/31	37±2	40/35	39.2±3	42/34	41.4±5	50/37
	7	30.9±0	31/31	24.1±3	27/19	28.8±2	31/27	24.6±2	27/23	23.4±4	27/17	32.4±6	42/23
	8	29.7±4	35/23	20.9±3	26/17	29.5±3	31/23	32.3±2	35/31	33.8±4	38/27	39±3	42/35
	9	31.5±1	34/31	25.1±2	27/23	33.8±3	37/29	28.5±2	31/27	31.5±4	35/23	38.9±5	46/31
	10	30.8±0	31/31	26.4±3	31/23	32.6±3	38/29	27.3±3	31/23	28.8±2	31/27	31.9±4	37/27
	11	34.1±1	35/33	32.5±8	38/19	35.7±2	38/35	27.7±4	35/23	31.8±4	39/27	36.2±2	38/33
	12	30.3±4	38/27	21.9±4	29/19	26.5±3	31/23	30±6	35/23	31.8±3	35/27	36.7±4	42/31
	Total	30.7	38/22	25.0	38/15	30.8	38/19	28.6	40/4	30.5	42/4	34.7	50/11
SWT	p-value	4.302E-01		9.722E-01		7.190E-01		8.938E-03		1.077E-02		8.870E-04	
SWT	Statistic	0.9345		0.9776		0.9556		0.7982		0.8053		0.7023	
SWT	H_0	No Rechazada		No Rechazada		No Rechazada		Rechazada		Rechazada		Rechazada	
Emociones combinadas	Usuario	EER	máx/mín	EER	máx/mín	EER	máx/mín	EER	máx/mín	EER	máx/mín	EER	máx/mín
	1	28.5±5	37/22	18.6±1	20/17	27.3±2	31/25	32.6±2	35/29	34.8±3	40/31	38.7±3	43/35
	2	28±3	31/26	30.3±3	34/26	32±2	34/29	8.1±3	15/6	9.7±4	14/3	11.9±3	17/9
	3	29.4±1	31/29	22.3±1	23/20	26.3±2	29/23	30.9±2	34/29	34.2±3	40/31	38.2±2	40/34
	4	26.5±4	31/20	16.9±2	20/14	21.2±2	23/19	26.3±1	29/26	29±3	33/26	32.6±2	35/29
	5	33.7±1	34/31	23.7±5	31/20	33±1	34/31	30.7±3	33/26	33.9±3	38/29	39.3±2	43/37
	6	31.4±4	37/25	26.7±3	31/23	32.8±3	37/29	34.3±2	37/31	38.4±2	40/34	41.7±4	48/37
	7	29.1±1	31/29	23.3±2	26/20	27.7±2	31/26	23.6±2	26/21	22.5±3	26/17	31.8±5	40/25
	8	27.6±3	31/23	17.8±3	23/14	27.4±3	31/23	30.9±3	37/29	33.5±3	37/29	38.2±3	44/34
	9	29.1±1	31/29	22.9±2	26/20	31.6±3	35/26	28±2	31/26	30.3±4	34/23	36.1±4	43/29
	10	28.6±0	29/29	26.3±3	31/23	30.5±2	34/29	26.4±2	29/23	28.3±1	29/27	33.2±3	38/29
	11	34.4±2	37/32	32.8±7	40/23	35.4±1	37/34	27±3	31/23	30.9±4	37/26	36.5±2	38/33
	12	29.2±5	37/23	18.2±2	23/17	24.8±5	33/20	28.8±4	33/23	32±3	37/29	36.6±3	42/31
	Total	29.6	37/20	23.3	40/14	29.2	37/19	27.3	37/6	29.8	40/3	34.6	48/9
SWT	p-value	6.20E-02		4.89E-01		8.32E-01		3.14E-03		1.16E-02		6.62E-04	
SWT	Estad.	0.8682		0.9393		0.9635		0.7567		0.8081		0.6890	
SWT	H_0	No Rechazada		No Rechazada		No Rechazada		Rechazada		Rechazada		Rechazada	
PtT	p-value	1.658E-03		1.384E-03		5.779E-04		2.526E-03		8.590E-03		5.563E-01	
PtT	Estad.	4.13		4.24		4.77		78		73		47	
PtT	H_0	Rechazada		Rechazada		Rechazada		Rechazada		Rechazada		Not Rechazada	
	$\Delta\mu$ [‰]	11		17		16		13		7		2	

4. Resultados experimentales

El cuadro 1 muestra los resultados obtenidos para seis métricas diferentes, donde el p-value muestra la probabilidad de que la hipótesis sea rechazada. Todos los valores de equal error rate (EER, el punto donde el error de aciertos es igual al error de rechazos) [26] se expresan en % a menos que se especifique lo contrario. $\Delta\mu$ es el desvío obtenido entre dos emociones distintas: la diferencia entre los errores de evaluar muestras con emociones combinadas (muestras de personas con varias emociones) y no combinadas (muestras de personas con la misma emoción).

El desvío es visible, pero se debe probar que es estadísticamente significativo. Para ello fueron realizados tests de normalidad (Shapiro–Wilk, o SWT) [28]. Las muestras de las métricas A, R y Camberra no rechazaron la hipótesis de normalidad, por lo que fue posible realizar un t-test pareado [29]. Para las demás métricas se utilizó la prueba de los rangos con signo de Wilcoxon (conocida en inglés como *Wilcoxon signed rank test* o SRT) [30] que es válida para estos casos.

En las primeras cinco métricas, se observa un cambio cerca de $\Delta\mu = 1\,\%$. Para la norma euclídea, en cambio, se ve un valor de $\Delta\mu = 2‰$.

5. Conclusiones

Se examinaron y revisaron las métricas de clasificación que prevalecen actualmente. Además, se contribuye a la verificación de los algoritmos basados en distancias: la métrica A y su valor empírico del umbral $t = 1{,}25$ encontrado por Gunetti y Picardi [3]. Adicionalmente, se propone el uso de la métrica de Minkowski con $p = 0{,}4$, y se muestra que Minkowski se comporta de manera adecuada frente a un amplio conjunto de datos tomados de un entorno de producción. Tras haber recopilado información en un entorno hostil y no idealizado, la distancia de Minkowski puede disminuir el EER medio de 21,9 % en la métrica A o 17,4 % en la métrica R a una tasa media de 15,6 %, lo cual es una mejora sustancial. Esto es importante por su aplicabilidad en el área de la seguridad, exponiendo a posibles intrusos. Además, se hace una diferenciación de los tres tipos de agregación en las métricas de Minkowski especificando cuándo dividir por el número de eventos; en la literatura revisada, la mayoría de los artículos y tesis suelen pasar por alto esta aclaración.

Además, se confirmó empíricamente en dos set de datos, que para los métodos considerados basados en distancias —A, R, Manhattan, Minkwoski y euclídea—, el EER asintótico más bajo se alcanza, en promedio, de manera exponencialmente decreciente desde el EER máximo en función del número de keystrokes de la sesión. El ajuste demostró ser preciso, con un error cuadrático medio inferior al 3,5 % en el subconjunto de ajuste más desfavorable.

Con respecto a la pregunta hipótesis central, como prueba preliminar, el experimento, predice resultados futuros prometedores en conjuntos de datos más grandes. El EER entre una muestra que incluye un estado emocional varió alrededor del 1 % en cinco métricas. De esto, se puede concluir que hay indicios de que los estados emocionales influyen en la calidad de la clasificación de los individuos y que, si un patrón de la persona capturado durante ese estado emocional no está disponible, la precisión del experimento puede empeorar. La métrica euclídea, por otro lado, fue notablemente resistente al cambio emocional y es digna de más investigación, especialmente con muestras más grandes.

Para el desarrollo de la tesis y su adecuada experimentación fue necesaria la realización de un framework de postprocesamiento de datos biométricos en Python compatible con hojas de notas de Jupyter y, para cuando se requiere procesar datos de manera intensiva, en el contexto de esta tesis se portó el framework al lenguaje C y se agregó soporte de paralelización. Para la captura de keystrokes fueron diseñados cuatro *keystroke loggers* (keylogger que además registra los tiempos de las pulsaciones), los primero tres, desarrollados en ANSI C: uno para Mac OS X, uno para Microsoft Windows y uno para GNU/Linux. El cuarto, desarrollado en JavaScript, es capaz de adquirir cadencia de tecleo en un entorno web. Los cuatro capturadores son compatibles entre sí produciendo información en un formato estandarizado y creado en el contexto de esta tesis. Este formato resultó de gran aporte por permitir compatibilidad con los frameworks de postprocesamiento de los datos.

5.1. Contexto

La tesis para el doctorado en Ciencias Informáticas de la Facultad de Informática de la UNLP se radicó en el Laboratorio de Sistemas de Información Avanzados (LSIA) de la Facultad de Ingeniería

de la UBA, fue desarrollada en el marco del proyecto UBACyT 20020130200140BA y actualmente se extiende en el PDE-44-2019 realizándose transferencia tecnológica a empresa de comercio electrónico más grande de la región. El proyecto se desarrolló en colaboración con el ISIER (FICCTE-SECYT UM) que asistió en asuntos de BCI, bajo el PID 01-001/12/14. La dirección estuvo a cargo de Jorge Ierache (LSIA, UBA) y la co-dirección de Waldo Hasperué (LIDI, UNLP).

5.2. Producción científica relacionada con la tesis

Durante el desarrollo de esta tesis se han comunicado resultados parciales a través de diversas publicaciones que a continuación se detallan

Trabajos previos en congresos

- Enrique P. Calot, Juan Manuel Rodríguez, y Jorge S. Ierache. Improving versatility in keystroke dynamic systems. En *Proceedings del XIX Congreso Argentino de Ciencias de la Computación*, número 5606, 2013. ISBN 978-987-23963-1-2. URL http://hdl.handle.net/10915/32428
- Enrique P. Calot, Juan Manuel Rodríguez, y Jorge S. Ierache. Improving versatility in keystroke dynamic systems. En Jorge Raúl Finochietto y Patricia Mabel Pesado, editores, *Computer Science & Technology Series. XIX Argentine Congress of Computer Science, Selected papers*, págs. 289–298. Editorial de la Universidad Nacional de La Plata (EDULP), 2014. ISBN 978-987-1985-49-4. URL http://lsia.fi.uba.ar/papers/calot14b.pdf
- Enrique P. Calot, Francisco Pirra, Juan Manuel Rodríguez, Gustavo Pereira, Juan Iribarren, y Jorge S. Ierache. Métodos adaptativos de educción de dinámica de tecleo centrado en el contexto emocional de un individuo aplicando interfaz cerebro computadora. En *XVI Workshop de Investigadores en Ciencias de la Computación*, mar. 2014. ISBN 978-950-34-1084-4. URL http://lsia.fi.uba.ar/papers/calot14.pdf

Series internacionales y capítulos de libro

- Enrique P. Calot, Jorge S. Ierache, y Waldo Hasperué. Robustness of keystroke dynamics identification algorithms against brain-wave variations associated with emotional variations. En *Intelligent Systems and Applications*, vol. 1037 de *Advances in Intelligent Systems and Computing*. Springer, 2019. ISBN 978-3-030-29515-8. doi: 10.1007/978-3-030-29516-5_15
- Nahuel González, Enrique P. Calot, y Jorge S. Ierache. A replication of two free text keystroke dynamics experiments under harsher conditions. En *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, págs. 1–6, sep. 2016. doi: 10.1109/BIOSIG.2016.7736905
- Nahuel González y Enrique P. Calot. Finite context modeling of keystroke dynamics in free text. En *Biometrics Special Interest Group (BIOSIG), 2015 International Conference of the*, págs. 1–5, sep. 2015. ISBN 978-3-88579-639-8. doi: 10.1109/BIOSIG.2015.7314606

Congresos internacionales

- Enrique P. Calot, Jorge S. Ierache, y Waldo Hasperué. Document typist identification by classification metrics applying keystroke dynamics under unidealised conditions. En *Second International Workshop on Computational Document Forensics, IWCDF at ICDAR 2019, Sydney, Australia, September 22–25, 2019*, págs. 19–24. IEEE, 2019. ISBN 978-1-7281-5054-3. doi: 10.1109/ICDARW.2019.70136
- Enrique P. Calot, Nahuel González, Jorge S. Ierache, y Waldo Hasperué. A convergence model for distance-based keystroke dynamics classification methods, 2019. Comunicado el 7 de febrero de 2019

Congresos nacionales

- Enrique P. Calot y Jorge S. Ierache. Multimodal biometric recording architecture for the exploitation of applications in the context of affective computing. En *Proceedings del XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017)*, número 10529, págs. 1030–1039, 2017. ISBN 978-950-34-1539-9. URL http://hdl.handle.net/10915/63866

Además de siete comunicaciones a WICC[37–43].

Esta tesis doctoral

- Enrique P. Calot. *Robustez de las métricas de clasificación de cadencia de tecleo frente a variaciones emocionales*. Tesis doctoral, Universidad Nacional de La Plata, 2019. URL http://sedici.unlp.edu.ar/handle/10915/76652

5.3. Futuros trabajos

Como futuras líneas de investigación, se requiere un trabajo adicional sobre la correlación entre los patrones emocionales y las ondas cerebrales. Además, experimentar con un mayor volumen de personas fortalecería los resultados registrados por este trabajo. La diferencia en el comportamiento de la norma euclidiana merece un estudio más profundo, al igual que las distancias de Minkowski con valores de p cercanos a dos.

La incorporación de otros registros biométricos como la frecuencia cardíaca (HR) o el electrocardiograma (ECG) aplicando el EEG como un marcador robusto inicial se considerará como una futura línea de investigación. El conjunto de datos EEG actuará como ya clasificado; por lo tanto, para el análisis de recursos humanos, se explorará la predicción del conjunto de datos con audio y video (como marcadores del estado emocional) utilizando clasificadores previamente entrenados.

Bibliografía

[1] R. Stockton Gaines, William Lisowski, S. James Press, y Norman Shapiro. Authentication by keystroke timing: Some preliminary results. Technical report, DTIC Document, 1980.

[2] Despina Polemi. Biometric techniques: review and evaluation of biometric techniques for identification and authentication, including an appraisal of the areas where they are most applicable. *Reported prepared for the European Commision DG XIIIC*, 4, 1997.

[3] Daniele Gunetti y Claudia Picardi. Keystroke analysis of free text. *ACM Transactions on Information and System Security (TISSEC)*, 8(3):312–347, ago. 2005. ISSN 1094-9224. doi: 10.1145/1085126.1085129.

[4] Enrique P. Calot. *Robustez de las métricas de clasificación de cadencia de tecleo frente a variaciones emocionales*. Tesis doctoral, Universidad Nacional de La Plata, 2019. URL http://sedici.unlp.edu.ar/handle/10915/76652.

[5] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, y Xinyi Yang. A review of emotion recognition using physiological signals. *Sensors*, 18(7), 2018. ISSN 1424-8220. doi: 10.3390/s18072074.

[6] Danny Oude Bos. EEG-based emotion recognition-The Influence of Visual and Auditory Stimuli. *Capita Selecta (MSc course)*, 2006.

[7] Panagiotis C. Petrantonakis y Leontios J. Hadjileontiadis. Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on Affective Computing*, 1(2):81–97, jul. 2010. ISSN 1949-3045. doi: 10.1109/t-affc.2010.7.

[8] Gyanendra K. Verma y Uma Shanker Tiwary. Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals. *NeuroImage*, 102: 162–172, nov. 2014. doi: 10.1016/j.neuroimage.2013.11.007.

[9] Dongmin Shin, Dongil Shin, y Dongkyoo Shin. Development of emotion recognition interface using complex EEG/ECG bio-signal for interactive contents. *Multimedia Tools and Applications*, 76(9):11449–11470, dic. 2016. ISSN 1380-7501. doi: 10.1007/s11042-016-4203-7.

[10] Wei-Long Zheng, Jia-Yi Zhu, y Bao-Liang Lu. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Transactions on Affective Computing*, 2017. ISSN 1949-3045. doi: 10.1109/taffc.2017.2712143.

[11] Anas Samara, Maria Luiza Recena Menezes, y Leo Galway. Feature extraction for emotion recognition and modelling using neurophysiological data. En *2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS)*. Institute of Electrical and Electronics Engineers (IEEE), dic. 2016. ISBN 978-1-5090-5566-1. doi: 10.1109/iucc-css.2016.027.

[12] Xiang Li, Dawei Song, Peng Zhang, Guangliang Yu, Yuexian Hou, y Bin Hu. Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. En *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Institute of Electrical and Electronics Engineers (IEEE), dic. 2016. ISBN 978-1-5090-1611-2. doi: 10.1109/bibm.2016.7822545.

[13] Jianhai Zhang, Ming Chen, Sanqing Hu, Yu Cao, y Robert Kozma. PNN for EEG-based emotion recognition. En *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Institute of Electrical and Electronics Engineers (IEEE), oct. 2016. ISBN 978-1-5090-1897-0. doi: 10.1109/smc.2016.7844584.

[14] Hernan F. Garcia, Mauricio A. Alvarez, y Alvaro A. Orozco. Gaussian process dynamical models for multimodal affect recognition. En *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Institute of Electrical and Electronics Engineers (IEEE), ago. 2016. ISBN 978-1-4577-0220-4. doi: 10.1109/embc.2016.7590834.

[15] Jingxin Liu, Hongying Meng, Asoke Nandi, y Maozhen Li. Emotion detection from eeg recordings. En *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, págs. 2173–2178, 2016. URL http://bura.brunel.ac.uk/handle/2438/13138.

[16] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980. ISSN 0022-3514. doi: 10.1037/h0077714.

[17] Jonathan Posner, James A. Russell, y Bradley S. Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(03), sep. 2005. doi: 10.1017/s0954579405050340.

[18] Hatice Gunes, Björn Schuller, Maja Pantic, y Roddy Cowie. Emotion representation, analysis and synthesis in continuous space: A survey. En *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, págs. 827–834. IEEE, mar. 2011. doi: 10.1109/FG.2011.5771357.

[19] Jonathan R. Wolpaw, Dennis J. McFarland, Gregory W. Neat, y Catherine A. Forneris. An EEG-based brain-computer interface for cursor control. *Electroencephalography and clinical neurophysiology*, 78(3):252–259, mar. 1991. ISSN 0013-4694. URL http://view.ncbi.nlm.nih.gov/pubmed/1707798.

[20] Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, y Theresa M. Vaughan. Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002. doi: 10.1016/S1388-2457(02)00057-3.

[21] Jonathan R. Wolpaw y Dennis J. McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. En *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, págs. 17849–17854. National Academy of Sciences, dic. 2004. doi: 10.1073/pnas.0403504101.

[22] Erim Yurci. Emotion detection from eeg signals: Correlating cerebral cortex activity with music evoked emotion. *Universitat Pompeu Fabra, Barcelona*, 2014.

[23] Peter J. Lang, Margaret M. Bradley, y Bruce N. Cuthbert. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical report A-8*, 2008.

[24] Lisa A. Feldman. Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of personality and social psychology*, 69(1):153, 1995.

[25] Ritwik Banerjee, Song Feng, Jun Seok Kang, y Yejin Choi. Keystroke patterns as prosody in digital writings: A case study with deceptive reviews and essays. En *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, págs. 1469–1473, Doha, Qatar, oct. 2014. Association for Computational Linguistics.

[26] Enrique P. Calot, Jorge S. Ierache, y Waldo Hasperué. Document typist identification by classification metrics applying keystroke dynamics under unidealised conditions. En *Second International Workshop on Computational Document Forensics, IWCDF at ICDAR 2019, Sydney, Australia, September 22–25, 2019*, págs. 19–24. IEEE, 2019. ISBN 978-1-7281-5054-3. doi: 10.1109/ICDARW.2019.70136.

[27] Enrique P. Calot, Nahuel González, Jorge S. Ierache, y Waldo Hasperué. A convergence model for distance-based keystroke dynamics classification methods, 2019. Comunicado el 7 de febrero de 2019.

[28] Samuel Sanford Shapiro y Martin B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. doi: 10.1093/biomet/52.3-4.591.

[29] Cyril H. Goulden. Methods of statistical analysis. chapter 4, págs. 50–55. John Wiley & Sons, Inc., New York, 2 edition, mayo 1956.
[30] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945. doi: 10.2307/3001968.
[31] Enrique P. Calot, Juan Manuel Rodríguez, y Jorge S. Ierache. Improving versatility in keystroke dynamic systems. En *Proceedings del XIX Congreso Argentino de Ciencias de la Computación*, número 5606, 2013. ISBN 978-987-23963-1-2. URL http://hdl.handle.net/10915/32428.
[32] Enrique P. Calot, Juan Manuel Rodríguez, y Jorge S. Ierache. Improving versatility in keystroke dynamic systems. En Jorge Raúl Finochietto y Patricia Mabel Pesado, editores, *Computer Science & Technology Series. XIX Argentine Congress of Computer Science, Selected papers*, págs. 289–298. Editorial de la Universidad Nacional de La Plata (EDULP), 2014. ISBN 978-987-1985-49-4. URL http://lsia.fi.uba.ar/papers/calot14b.pdf.
[33] Enrique P. Calot, Francisco Pirra, Juan Manuel Rodríguez, Gustavo Pereira, Juan Iribarren, y Jorge S. Ierache. Métodos adaptativos de educción de dinámica de tecleo centrado en el contexto emocional de un individuo aplicando interfaz cerebro computadora. En *XVI Workshop de Investigadores en Ciencias de la Computación*, mar. 2014. ISBN 978-950-34-1084-4. URL http://lsia.fi.uba.ar/papers/calot14.pdf.
[34] Enrique P. Calot, Jorge S. Ierache, y Waldo Hasperué. Robustness of keystroke dynamics identification algorithms against brain-wave variations associated with emotional variations. En *Intelligent Systems and Applications*, vol. 1037 de *Advances in Intelligent Systems and Computing*. Springer, 2019. ISBN 978-3-030-29515-8. doi: 10.1007/978-3-030-29516-5_15.
[35] Nahuel González, Enrique P. Calot, y Jorge S. Ierache. A replication of two free text keystroke dynamics experiments under harsher conditions. En *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, págs. 1–6, sep. 2016. doi: 10.1109/BIOSIG.2016.7736905.
[36] Nahuel González y Enrique P. Calot. Finite context modeling of keystroke dynamics in free text. En *Biometrics Special Interest Group (BIOSIG), 2015 International Conference of the*, págs. 1–5, sep. 2015. ISBN 978-3-88579-639-8. doi: 10.1109/BIOSIG.2015.7314606.
[37] Enrique P. Calot y Jorge S. Ierache. Multimodal biometric recording architecture for the exploitation of applications in the context of affective computing. En *Proceedings del XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017)*, número 10529, págs. 1030–1039, 2017. ISBN 978-950-34-1539-9. URL http://hdl.handle.net/10915/63866.
[38] Germán M. Concilio, Enrique P. Calot, Jorge S. Ierache, y Hernán D. Merlino. Application of Keystroke Dynamics Modelling Techniques to Strengthen the User Identification in the Context of E-commerce. En Armando Eduardo De Giusti, editor, *VII Workshop de Seguridad Informática (CACIC 2018, Tandil, Buenos Aires, Argentina)*, 2018. In press.
[39] Gustavo Pereira, Facundo Nervo, Juan Iribarren, Enrique P. Calot, Jorge S. Ierache, y Norberto Mazza. Líneas de investigación aplicada al empleo de interfases cerebro-maquina. En *Proceedings del XVII Workshop de Investigadores en Ciencias de la Computación (Salta, 2015)*, pág. 5, jun. 2015. URL http://hdl.handle.net/10915/46146.
[40] Nahuel González, Enrique P. Calot, Ariel Liguori, Francisco Pirra, Juan Manuel Rodríguez, Gustavo Pereira, Facundo Nervo, Juan Iribarren, y Jorge S. Ierache. Educción de dinámica de tecleo centrado en el contexto emocional de un individuo. En *Proceedings del XVII Workshop de Investigadores en Ciencias de la Computación (Salta, 2015)*, pág. 5, jun. 2015. URL http://hdl.handle.net/10915/46260.
[41] Enrique P. Calot, Ezequiel L. Aceto, Juan Manuel Rodríguez, Ariel Liguori, María Alejandra Ochoa, Hernán D. Merlino, Enrique Fernández, Nahuel González, Francisco Pirra, y Jorge S. Ierache. Líneas de investigación del laboratorio de sistemas de información avanzados. En *Proceedings del XVII Workshop de Investigadores en Ciencias de la Computación (Salta, 2015)*, pág. 4, jun. 2015. ISBN 978-987-633-134-0. URL http://hdl.handle.net/10915/46107.
[42] Enrique P. Calot, Federico M. Rossi, Nahuel Gonzalez, Waldo Hasperué, y Jorge S. Ierache. Avances en educción de dinámica de tecleo y el contexto emocional de un individuo aplicando interfaz cerebro computadora. En *XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina)*, págs. 872–876, jun. 2016. ISBN 978-950-698-377-2. URL http://hdl.handle.net/10915/53247.
[43] Federico M. Rossi, Enrique P. Calot, y Jorge S. Ierache. Educción emocional de un individuo en contextos multimodales en computación afectiva. En *XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, Buenos Aires, Argentina)*, jul. 2017. ISBN 978-987-42-5143-5.

Empoderamiento de la Conciencia Situacional en operaciones militares usando Realidad Aumentada

Alejandro Mitaritonna [1,3]
[1] Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF)
Villa Martelli, Buenos Aires, Argentina
amitaritonna@citedef.gob.ar

María José Abásolo [2,3]
[2] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA)
[3] Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata (UNLP)
calle 50 y 120 (1900) La Plata, Buenos Aires, Argentina
mjabasolo@lidi.info.unlp.edu.ar

Francisco Montero[4]
[4] Universidad de Castilla-La Mancha
Sistemas informáticos, E.S. de Ingeniería Informática Albacete
Albacete, España
Francisco.MSimarro@uclm.es

I. DATOS

Lugar defensa: Universidad Nacional de La Plata, Facultad de Informática
Fecha defensa: 2 de diciembre de 2019

II. MOTIVACIÓN

La Conciencia Situacional (CS) es la percepción de nuestro entorno, la comprensión de su significado y la proyección de su estado. La falta de CS es un factor causal en muchos accidentes militares. Las técnicas de Realidad Aumentada (RA) se utilizan para apoyar la CS y la interacción en situaciones de crisis donde el tiempo de respuesta juega un rol central. En primer lugar, se revisan los proyectos militares existentes. Luego, el diseño de una arquitectura militar basada en RA como soporte para CS, llamada RAIOM (Realidad Aumentada para la identificación de Objetivos Militares), se introduce y valida con usuarios reales simulando escenarios militares tradicionales. Se presenta una propuesta de un modelo para identificación, selección y clasificación de los requisitos de la CS para ser utilizados en el campo militar; además de que este modelo también se centra en el diseño de soluciones basadas en RA. Este modelo, denominado 3D-SA, facilita el análisis, primero, y el diseño, más tarde, de los sistemas militares donde el concepto de CS es esencial. El modelo 3D-SA intenta facilitar el uso y la aplicación de soluciones basadas en RA así como también cubrir una deficiencia en la brecha entre los requisitos y las soluciones de diseño de la CS en sistemas militares. Se implementó un procesamiento distribuido en una arquitectura cliente-servidor utilizando lentes ópticos AR transparentes por el lado del cliente y una mini placa por el lado del servidor. Se implementó una aplicación de RA que utiliza la arquitectura propuesta como prueba de concepto integrada. La implementación se evaluó para medir la efectividad, la eficiencia y la usabilidad de la arquitectura de software en términos de tres niveles de CS, como son la percepción, la comprensión y la proyección. Se utilizaron diferentes técnicas en la evaluación, tales como User Testing, Thinking Aloud Protocol and SAGAT / SART. También se evaluó la experiencia del usuario (UX) utilizando el cuestionario UMUX. Los resultados fueron optimistas según el grado de cumplimiento de la CS obtenido por los participantes en el experimento.

Palabras Claves: Conciencia situacional; Framework de software; Diseño centrado en el usuario; Experiencia del usuario; SCRUM+DCU; SAGAT; SART; UMUX

III. INTRODUCCIÓN

Azuma, R [1] detalla que la RA permite al usuario ver el mundo real, con objetos virtuales superpuestos o compuestos con el mundo real. Por lo tanto, la RA complementa la realidad, en lugar de reemplazarla por completo.

Endsley, M. R [2,3] menciona que la CS es la percepción de los elementos en el entorno dentro de un volumen de tiempo y espacio, la comprensión de su significado y la proyección de su situación en el futuro cercano. La CS incluye la pérdida o percepción incompleta o cambio de los elementos presentes en el entorno operativo del soldado. Este factor puede estar relacionado con las percepciones limitadas de un individuo de su entorno operativo. Algunas de estas dificultades pueden superarse utilizando una tecnología que pueda organizar y mostrar la información al usuario automáticamente.

Durante los últimos años, se diseñaron varios proyectos militares basados en RA para mejorar las CS como apoyo para la toma de decisiones en entornos desconocidos. Los ejércitos de la mayoría de los principales países están llevando a cabo diversos programas de modernización destinados a mejorar las capacidades del soldado. Todos coinciden en un denominador común: aumentar la efectividad y la eficiencia del soldado proporcionando mayor seguridad y capacidad para sobrevivir en combate. Sin embargo, las funcionalidades implementadas en estos proyectos militares basados en RA deben analizarse para justificar si son apropiadas para mejorar la CS de los usuarios y, de esta manera, para determinar la necesidad concreta de construir un framework para desarrollar aplicaciones basadas en RA como apoyo; Además, es importante resaltar el tipo de tecnología respaldada por estos proyectos.

IV. PROYECTOS MILITARES BASADO EN RA Y HERRAMIENTAS DE RA

A. El soldado digital

Muchas actividades de investigación han explorado nuevas características utilizando técnicas basadas en RA. En [4] el concepto de Soldado del Futuro (SF) fue acuñado para identificar las capacidades de un soldado en el campo de batalla. Este guerrero se adaptará a cada área tecnológica con especial énfasis en el rendimiento cognitivo para mejorar la eficiencia del soldado. Hay siete áreas principales aplicables al soldado: rendimiento y entrenamiento humano; Protección; Letalidad; Movilidad y logística; Redes; Sensores; Poder de fuego y Energía. El programa SF ha centrado sus esfuerzos en la investigación y el desarrollo de componentes electrónicos aplicando la RA como tecnología para visualizar información del entorno donde se produce el combate y de esta manera facilitar la toma de decisiones en situaciones hostiles. Algunos ejemplos de estos proyectos militares se pueden encontrar en [5,6]. Es importante resaltar el área del desempeño humano. Esta área está relacionada con el desempeño de los soldados en el campo de batalla. El desempeño adecuado en una operación militar depende de las CS. La CS desempeña un papel fundamental en las operaciones militares en cualquier terreno donde se lleve a cabo una misión. El comando y el control dependen de la CS ya que las decisiones deben tomarse en función de la situación actual. Una de las principales tecnologías para mejorar la CS del soldado es enviar información a través de sensores geográficamente dispersos y luego visualizar esa información utilizando técnicas de RA [7].

B. Proyectos militares basados en RA

En el ámbito de los framework militares basados en RA, es importante considerar un conjunto de elementos que ayuden a tener una CS adecuada. Estos elementos son los siguientes [5]:

1. Reconocimiento del terreno: es importante tener una referencia precisa del lugar de operaciones para aumentar la percepción del entorno.
2. Reconocimiento de infraestructura: es importante identificar edificios e infraestructura en un terreno hostil, ya sea para usarlo como referencia de ubicación o para ejecutar una misión específica.
3. Reconocimiento del entorno geográfico: es importante obtener información contextual adicional, como referencias geográficas, piezas de artillería, etc.
4. Alerta de amenaza mediante símbolos descriptivos: el soldado debe comprender los símbolos de las amenazas para determinar el grado de relevancia de las mismas.
5. Ubicación aliada: todos los soldados deben conocer la posición de los aliados en el campo de batalla.
6. Seguimiento de ruta: para verificar en un mapa digital la ruta realizada por los soldados.
7. Comunicación: entre el comando central y el Jefe de Patrulla. La comunicación entre la cadena de comando es parte de las tareas de comando y control.
8. Filtrado de información: para filtrar la información principal del entorno contextual.

De acuerdo con la lista de elementos a tener en cuenta en los sistemas digitales militares para mejorar la CS en el campo de batalla, presentamos una tabla comparativa (Tabla I) de los elementos de CS de los principales proyectos militares detallados en [5].

Además, podemos detallar una lista de tecnologías que deben tener los proyectos militares para apoyar los elementos de las CS (Tabla II). Uno de los proyectos militares más avanzados es ULTRA-Vis [12]. Por lo tanto, la Tabla II toma las principales características tecnológicas y no tecnológicas (es decir, software de código abierto y componentes COTS) de este proyecto y otras características adicionales. Hemos agregado el proyecto RAIOM en la última columna de ambas tablas (Tabla I y Tabla II). El framework RAIOM debe proporcionar todos los elementos y características tecnológicas para apoyar a la CS.

TABLA I. **CARASTERÍSTICAS DE PROYECTOS MILITARES**

SA's element	Eyekon [9]	BARS [10]	iARM [11]	ULTRA-Vis [12]	RAIOM
Terrain operations recognition	X	X	X	X	X
Infrastructure recognition			X	X	X
Geographical environment. recognition	X		X	X	X
Alert using descriptive symbols	X		X	X	X
Allied location			X	X	X
Path tracking					X
Communication		X	X	X	X
Information filtering	X	X	X	X	X

TABLA II. **TECNOLOGIAS SOPORTADAS EN PROYECTOS MILITARES**

Technologies	Eyekon	BARS	iARM	ULTRA-Vis	RAIOM
AR visualization	X		X	X	X
Lightweight devices					X
Distributed processing					X
Sensors (GPS, IMU, etc)	X		X	X	X
Voice recognition		X	X	X	X
Gesture recognition		X	X		X
Target recognition			X	X	X
Terrain recognition				X	X
Collaborative communication		X	X	X	X
Open Source software					X
Security	X			X	X
COTS		X	X		X

C. Herramientas de RA

Es importante mencionar que existe una gran cantidad de frameworks basados en RA, pero no son suficientes para cubrir los requisitos funcionales militares para mejorar la CS. Sin embargo, vamos a detallar algunas características funcionales y no funcionales de los frameworks no comerciales para considerarlos cuando diseñemos nuestro propio framework militar basado en RA. Se identifica un conjunto de criterios funcionales tecnológicos [8]:

1. Visualización de objeto 3D
2. Reconocimiento de objetos
3. Geolocalización
4. Sin marcadores
5. Reconocimiento en línea

En aras de exponer las características funcionales en aplicaciones basadas en RA, hemos seleccionado y analizado cinco herramientas (ver Tabla III).

Además, las características no funcionales también deben tenerse en cuenta. Una lista de criterios no funcionales en sistemas que utilizan la RA son [13]:

1. Disponibilidad fuera de línea
2. Fácil de integrar
3. Fácil de expandir
4. Licencia de software
5. Procesamiento distribuido

En la Tabla IV se muestran las características no funcionales.

TABLA III. **CARACTERISTICAS FUNCIONALES DE HERRAMENTAS DE RA**

Functional feature	ARToolKit [15]	Vuforia [16]	EasyAR [17]	Layar [18]	DroidAR [19]
3D Object	-	X	-	-	X
Image recognition	X	X	X	X	X
Geolocation	X	X	X	X	X
Markerless	-	X	-	-	-
Online recognition	-	X	-	X	-

TABLA IV. **CARACTERÍSTICAS NO FUNCIONALES DE HERRAMIENTAS DE RA**

Non-functional feature	ARToolKit	Vuforia	EasyAR	Layar	DroidAR
Off-line availability	X	-	-	-	X
Easy to integrate	-	-	-	-	-
Easy to expand	X	-	-	-	X
Software license	Open Source	Commercial	Freeware	Freeware	Open Source
Distributed processing	-	-	-	-	-

V. RAIOM. SISTEMA MILITAR CON RA

A. Proceso de diseño de RAIOM

El diseño de una arquitectura militar basada en RA como soporte para la CS, llamada RAIOM (Realidad Aumentada para la identificación de Objetivos Militares), se introduce y se valida con usuarios reales en escenarios militares tradicionales simulados.

El segmento militar fue poco explorado y fue necesario diseñar, programar y mostrar a los usuarios clave el funcionamiento de cada parte del framework como unidades independientes.

Utilizamos el modelo 3D-SA [14] para la identificación, selección y clasificación de los requisitos de la CS junto con el diseño de soluciones basadas en RA. Este modelo 3D-SA facilita el análisis, primero y el diseño posterior de sistemas donde el concepto de CS es esencial. El modelo 3D-SA presentado intenta facilitar el uso y la aplicación de soluciones basadas en RA para mejorar la CS, así como para cubrir una deficiencia en la brecha entre los requisitos y las soluciones de diseño. Con el modelo 3D-SA, documentamos, identificamos y clasificamos los requisitos de la CS y proporcionamos prototipos relacionados con las historias de usuarios de personal militar que contribuyen a su realización a nivel de análisis y su diseño posterior.

El modelo 3D-SA considera tres elementos principales: *fase* de la CS, *características* de la CS y *modalidad de interacción* (Fig. 1). La primera dimensión, fase, incluye los tres niveles de la CS que propone Endsley que son muy útiles para identificar los requisitos en un entorno problemático o confuso [20]: percepción de los elementos en el entorno, comprensión de situación y proyección del estado futuro. La segunda dimensión, característica, incluye diez preguntas agrupadas por preguntas del tipo: "quién", "qué" y "dónde", propuesta por Gutwin y Greenberg [21] para ser respondidas para determinar si un componente o herramienta proporciona un apoyo adecuado a la CS. Finalmente, la tercera dimensión, la modalidad de interacción, está relacionada con el diseño de la solución de RA, que incluye interacción visual, auditiva, háptica, somatosensorial y olfativa.

Teniendo en cuenta los requerimientos militares, proponemos documentarlos como historias de usuarios. Las historias de usuarios son un concepto muy similar a los casos de uso tradicionales; son descripciones cortas y simples de una característica contada desde la perspectiva de la persona que desea la nueva capacidad, generalmente un usuario o cliente del sistema. Con el fin de identificar un listado de historias de usuarios relevantes, se utilizó un framewrok para la obtención de requisitos en un entorno militar, donde el concepto de caso de uso fue cambiado por el término de la historia del usuario. Al usar este framework, se identificaron, documentaron y crearon prototipos de un conjunto de historias de usuarios.

Siguiendo el modelo 3D-SA, identificamos requisitos y definimos prototipos relacionados con historias de usuarios del personal militar. Por ejemplo: "Como comandante militar, quiero saber mi ubicación y mi dirección para que se proporcione información de mis coordenadas GPS, una brújula y referencias de ubicación relevantes, como edificios o lugares". Esta es una historia de usuario que se puede clasificar y documentar siguiendo 3D-SA: (fase = percepción, característica = dónde y modalidad = visual) y una solución de diseño se puede mostrar en la Figura 2, donde se muestra un radar de 360 grados y una brújula que se encuentra en la esquina inferior derecha del prototipo. Ejemplos adicionales de prototipos se muestran gráficamente en la Figura 3, Figura 4 y Figura 5.

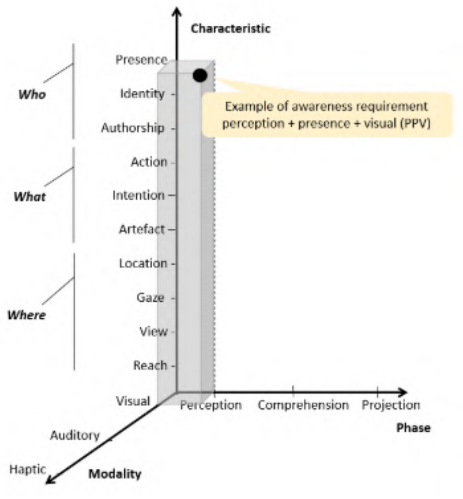


Fig. 1. Modelo 3D-SA [14], un modelo para la documentación de requisitos junto con el diseño de soluciones basadas en RA

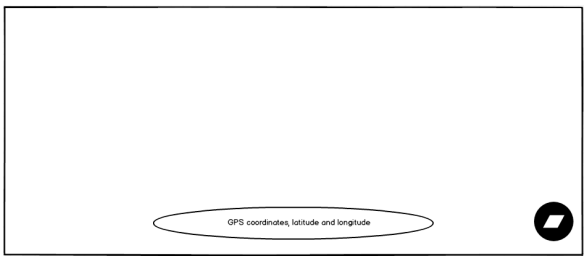


Fig. 2. Prototipo de solución de diseño para la historia del usuario de *location*

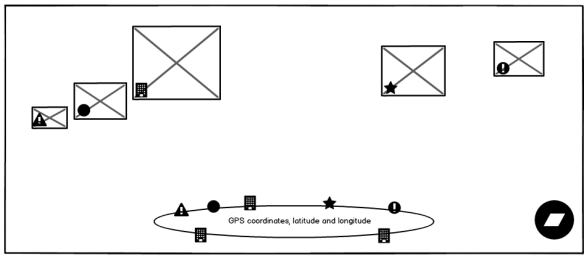


Fig 3. Prototipo de solución de diseño para la historia de usuario *landmark*

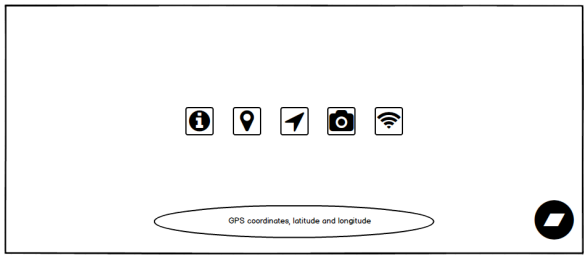


Fig. 4. Prototipo de solución de diseño para la historia de usuario *selection*

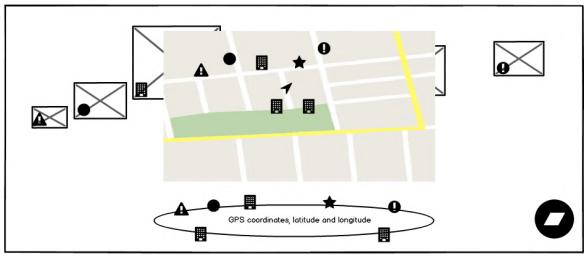


Fig. 5. Prototipo de solución de diseño para la historia de usuario *mapa interactivo*

El proceso de construcción del framework se basa en el desarrollo de pruebas individuales de conceptos como prototipos para demostrar el cumplimiento de los requisitos del usuario. Por lo tanto, la metodología de desarrollo de software utilizada combina un método ágil [22] y técnicas de diseño centrado en el usuario (DCU) [23]. El interés en la integración del DCU y los enfoques ágiles está creciendo [23]. SCRUM [24] fue elegido como un método ágil junto con los principios de diseño proporcionados por el DCU, lo que resultó en una metodología combinada SCRUM-DCU [25-26]. El DCU incluye entrevistas con usuarios principales (militares), identificación de tareas relevantes para los usuarios (considerando las preguntas Quién, Qué y Dónde), realización de actividades de creación de prototipos, evaluación e iteración. SCRUM incluye diseño iterativo, participación del usuario, pruebas continuas y creación de prototipos.

B. RAIOM. ARquitectura basada en capas

La arquitectura basada en capas está compuesta por aplicaciones, middleware y sistema operativo (Fig. 6). El Middleware, el software que se encuentra entre el sistema operativo y las aplicaciones que se ejecutan en él, facilita la integración tecnológica [27, 28] y está compuesto por la capa *Modules* y la capa *Libraries*.

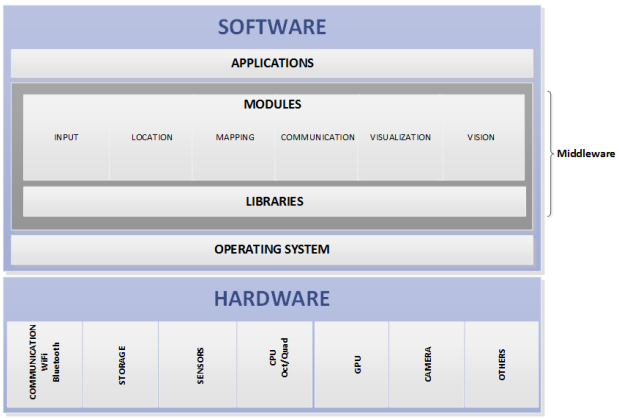


Fig. 6. RAIOM. Arquitectura basada en capas

La capa Modules está compuesta por componentes donde cada uno se comunica entre sí ofreciendo servicios. Los componentes que pertenecen a la capa Modules son los siguientes:

- Input: permite el ingreso de datos al sistema mediante reconocimiento de voz o gestual.
- Location: permite obtener información de georreferenciación del dispositivo. Además, permite rastrear al usuario utilizando los sensores del dispositivo.
- Mapping: permite cargar mapas en el sistema. Implementa una interfaz SIG (Sitema de Información Geográfica).
- Communication: permite la comunicación entre dispositivos mediante mensajes estructurados.
- Visualization: muestra una representación gráfica de los símbolos militares en el lado del Cliente y filtra la información de prioridad del entorno contextual.
- Vision: implementa procesamiento de imágenes, reconocimiento de objetos y algoritmos de tracking. Además, parte del reconocimiento gestual se ejecuta en este módulo

C. RAIOM. Arquitectura cliente-servidor

En primer lugar, comparamos el tiempo de rendimiento de la ejecución del algoritmo de reconocimiento de objetos ORB en diferentes dispositivos. El ODROID-XU3 demuestra una gran ventaja sobre el rendimiento de los teléfonos inteligentes, particularmente el Huawei G8 (Fig. 7)

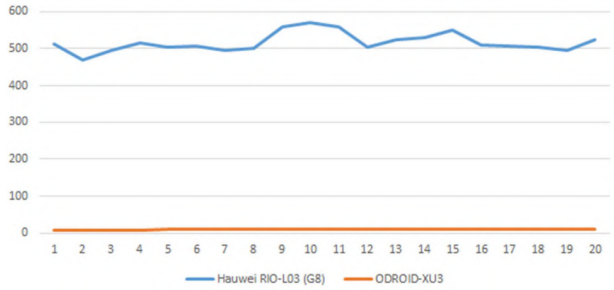


Fig. 7. Rendimiento de Huawei G8 y ODROID-XU3 para el algoritmo ORB de reconocimiento de objetos (en milisegundos)

Elegimos una arquitectura cliente-servidor. Se analizaron los cuatro niveles diferentes de tareas de procesamiento sugeridos en [29]. En el caso ideal, todo el trabajo lo realiza de forma nativa el cliente, lo que lo hace independiente del servidor y la infraestructura. En el otro extremo, todo el trabajo lo realiza el servidor. RAIOM toma un nivel intermedio donde los procesos que requieren menos computación se ejecutan en el lado del cliente, como alimentación de video, sensores, mapeo y representación de imágenes. Mientras tanto, los procesos más intensivos se realizan en el lado del servidor, como el procesamiento de imágenes, el reconocimiento de objetos y el tracking implementado en el módulo de *Vision* (Fig. 8).

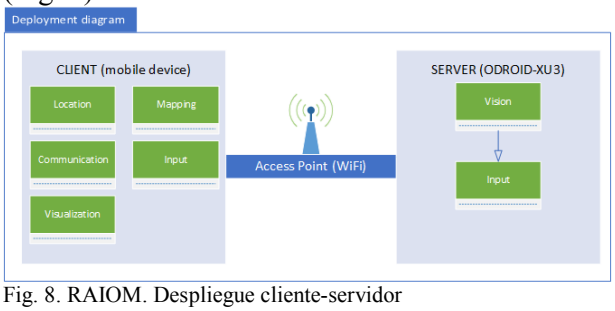


Fig. 8. RAIOM. Despliegue cliente-servidor

Se utilizó una mini placa ODROID-XU3 como servidor. El software instalado en el servidor fue Ubuntu Server versión 16.04, Python versión 3.0, OpenCV 3.1 y Node.js. Los componentes de software instalados en el lado del cliente fueron Android 6.0, OpenCV 3.1, Socket.io y librerías. La comunicación entre ambas partes fue implementada por Socket.io. La comunicación entre el cliente y el servidor se ejecuta enviando un frame (imagen) del cliente al servidor codificando los datos en formato JSON.

En el lado del cliente, comparamos entre lentes de RA del tipo optical see-through, Epson Moverio BT-300 (Fig. 9.a) y video see-through a través de un teléfono inteligente dentro de un carboard (Fig. 9.b). A partir de las pruebas de rendimiento de tiempo de adquisición de video (Fig. 10), concluimos que Huawei G8 funciona mejor. Las gafas Epson Moverio BT 300 tienen el segundo mejor tiempo de rendimiento.

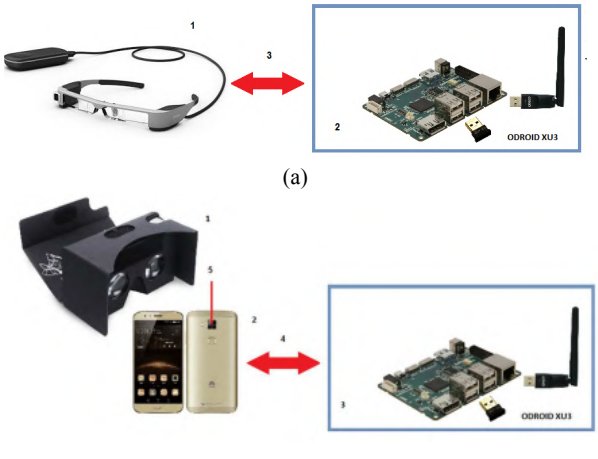


Fig. 9. Hardware RAIOM: (a) Gafas optical see-through Epson Moverio BT-300 como cliente y mini placa ODROID-XU3 como servidor. (b) Gafas video see-through con un teléfono inteligente dentro de un cardboard como cliente y la mini placa ODROID-XU3 como el servidor.

Aunque el dispositivo más rápido para la captura de video fue el Huawei, también comparamos el rendimiento de la gafa Epson Moverio BT 300 en el pipeline de RA que ejecuta el algoritmo ORB en el lado del servidor. La Fig. 11 muestra que la gafa Epson Moverio BT 300 tuvo el mejor rendimiento y también fue la opción con mejor ergonomía.

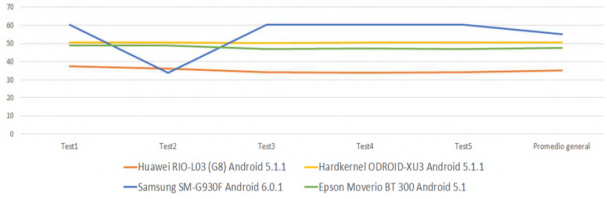


Fig. 10. Tiempo de adquisición de video de diferentes dispositivos medidos en milisegundos

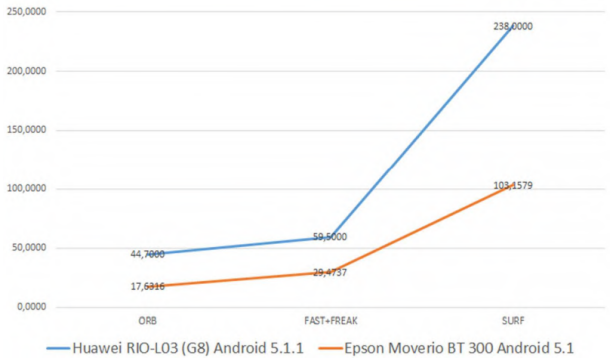
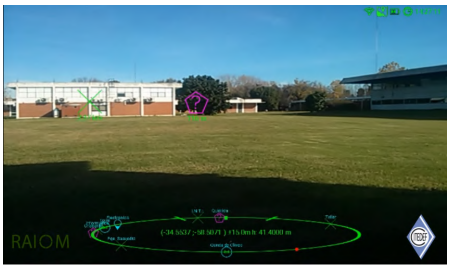


Fig. 11. Tiempo procesamiento utilizando los algoritmos ORB, FAST+FREAK y SURF ejecutándose en las gafas Epson Moverio BT 300 y el teléfono inteligente Huawei

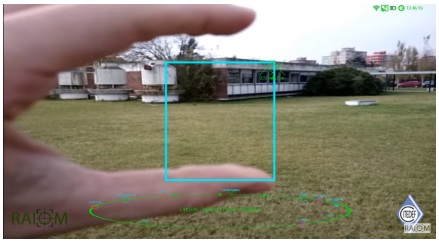
D. RAIOM. Funcionalidades

Todas las funcionalidades se integraron en una aplicación con las siguientes características:

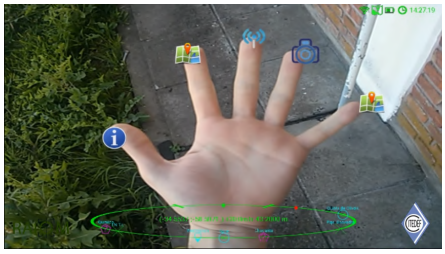
1. Radar de 360 grados: los objetivos tácticos y de amenaza se visualizan en el radar usando símbolos militares (Fig. 12.a).
2. Pose del operador: el norte magnético sirve para guiar al operador. La posición relativa del operador se calcula según sus movimientos.
3. Símbolos militares adaptables: los símbolos utilizados son adaptables para identificar las amenazas y los objetivos tácticos.
4. Reconocimiento de puntos de referencia: la aplicación es capaz de reconocer edificios y otros tipos de infraestructura (Fig. 12.b).
5. Menú por reconocimiento de gestos: los datos a pedido se pueden obtener mediante el reconocimiento de gestos (Fig. 12.c).
6. Mapas interactivos: el operador puede visualizar los mapas del terreno superpuestos sobre su mano mediante reconocimiento de gestos (Fig. 12.d.)



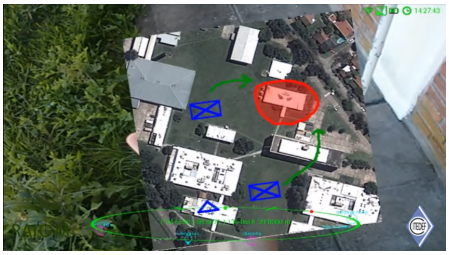
(a)



(b)



(c)



(d)

Fig. 12. Funcionalidades de RAIOM: radar de 360 grados (a), reconocimiento de puntos de referencia (b), menú por reconocimiento de gestos (c) y mapa interactivo (d)

VI. EVALUACIÓN Y RESULTADOS

A. Metodología de evaluación

El despliegue de RAIOM presentado se evaluó para medir la efectividad, la eficiencia y la experiencia del usuario para cumplir con el modelo 3D-SA propuesto.

Se utilizaron diferentes técnicas para llevar a cabo las actividades de evaluación, tales como:

- User Testing: se refiere a una técnica utilizada en el proceso de diseño para evaluar un producto, característica o prototipo con un usuario real.
- Thinking Aloud: es un protocolo utilizado para recopilar datos en pruebas de usabilidad en el diseño y desarrollo de productos.
- SAGAT / SART: SAGAT se desarrolló para ayudar en los diseños de interfaz al proporcionar una medida objetiva de la CS del usuario. SART permite a los operadores calificar un diseño de sistema [30, 31].

Las pruebas para evaluar la CS de los participantes se basaron en tareas que deben ejecutar. La Tabla V muestra ejemplos de

actividades propuestas en el experimento de evaluación en situaciones militares tradicionales. La taxonomía propuesta por Bloom [32] se utilizó para identificar acciones y elaborar las preguntas que se hicieron a cada uno de los participantes. La Tabla VI muestra los niveles cognitivos, ejemplo de acciones y preguntas realizadas a los participantes. Los evaluadores hicieron preguntas a cada participante sobre las situaciones recreadas siguiendo un protocolo Think-Aloud entre evaluador y participante. Los cuestionarios fueron respondidos en base a una escala Likert de 5 puntos (SART). Finalmente, las métricas de efectividad y eficiencia se tabularon documentando las tasas y tiempos de falla / éxito asociados.

Por otra parte, la UX se evaluó mediante el cuestionario UMUX [33]. El cuestionario se integró con las siguientes preguntas sobre RAIOM: cubre mis necesidades, es agradable de usar, es fácil de usar, no necesita configuraciones. También los cuestionarios fueron respondidos en base a una escala Likert de 5 puntos.

TABLA V. **EJEMPLO DE ACTIVIDADES PROPUESTAS EN EL EXPERIMENTO DE EVALUACIÓN**

Perception	*Comprehension*	*Projection*
Terrain type / capabilities of threats	Ability to support plan	Projected ability of plan to meet mission objectives
Resources available	Risk of mission failure/success	Projected risk of troops
Civilian disposition	Risk of casualties / loss of equipment	Projected ability to obtain information
Weather	Ability to counteract enemy actions	Projected ability to communicate
Enemy disposition	Ability to mitigate risk	Projected actions of enemy

TABLA VI. **PREGUNTAS PARA EVALUAR EL GRADO DE PERCEPCIÓN, COMPRENSIÓN Y PROYECCIÓN**

Cognitive level	*Example of Actions*	*Questions*
Perception	Describe Find Locate Identify Recognize	Q1. What is your geographical location? Q2. Can you recognize that building? Q3. Can you find the purpose of the mission? Q4. Can you identify the threats?
Comprehension	Classify Interpret Infer Explain	Q5. Could you explain what is happening around you? Q6. Could you classify the type of threat? Q7. What can you say about the distance to the target? Do you have that information? Q8. Could you say that the available information is enough to determine the dangerousness of the environment?
Projection	Calculate Use Perform To plan	Q9. How would you use the information provided by the system to preserve human and material resources? Q10. Could you plan an escape route? (taking into account the surrounding threats) Q11. How would you organize your human and material means in case of attack?

B. Participantes

El experimento fue realizado por diez participantes seleccionados del Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF) y el DIC 601 en noviembre de 2019. Todos los participantes tienen las siguientes características: entrenamiento militar en comandos, conocimiento de símbolos militares, conocimiento sobre de conciencia situacional en procesos militares y experiencia de más de tres años en funciones de comando. Los participantes conforman dos grupos diferentes: cuatro usuarios expertos con poder de decisión en operaciones militares (grupo A) y seis participantes con perfil operativo (grupo B). Ambos grupos realizan las tareas de evaluación relacionadas con los niveles de percepción y comprensión de la CS, mientras que las tareas relacionadas con el nivel de proyección solo las realizan los participantes del grupo A que tienen el poder de decidir en la cadena militar.

C. Resultados

Todos los participantes pudieron completar las tareas propuestas y responder las preguntas adecuadamente. La Tabla VII muestra los valores promedio de la percepción, comprensión y proyección. De los puntajes obtenidos que van desde 4,08 a 4,67, concluimos que el nivel de CS es alto y el sistema facilita la CS de los participantes en el experimento. Las puntuaciones del grupo A fueron más altas que las del grupo B. El grado de cumplimiento de la CS fue del 89% para el grupo A y del 86% para el grupo B.

TABLA VII. **PREGUNTAS PARA EVALUAR EL GRADO DE PERCEPCIÓN, COMPRENSIÓN Y PROYECCIÓN**

Level	**Participant Score**		
	Group A	**Group B**	**Total**
Perception	4,67	4,50	**4,58**
Comprehension	4,33	4,08	**4,21**
Projection	4,33	-	**4,33**

Los resultados de la experiencia del usuario (UX) se muestran en la Tabla VIII. De los puntajes obtenidos que van desde 3,67 a 5,00 podemos decir que los usuarios están satisfechos con la aplicación desarrollada utilizando la arquitectura de software RAIOM. La interfaz diseñada cumple con el 88% de cumplimiento (grupo A) y el 87% (grupo B), respectivamente.

TABLA VIII. **PREGUNTAS PARA EVALUAR LA EXPERIENCIA DEL USUARIO (UX) Y RESULTADOS POR GRUPOS DE PARTICIPANTES**

	Group A	Group B	Total
RAIOM covers my needs	4,00	4,00	4,00
RAIOM is pleasant in use	5,00	4,67	4,84
RAIOM is easy to use	5,00	5,00	5,00
RAIOM does not need settings	3,67	3,67	3,67
TOTAL	4,42	4,33	4,38

VII. CONCLUSIÓN

Los sistemas militares y las aplicaciones basadas en RA son cada vez más importantes en el mundo de los sistemas. Uno de los principales problemas del desarrollo de los sistemas militares es su nivel de cumplimiento de CS, que puede dar lugar a una falta de calidad. Nuestro objetivo es justificar, presentar y evaluar una arquitectura de software basada en RA para apoyar la CS en un entorno militar. La razón principal que ha motivado el diseño de la arquitectura de software RAIOM fue que los frameworks actuales basados en RA tienen deficiencias y limitaciones para cubrir algunos requisitos funcionales militares y estas limitaciones afectan al nivel de CS.

Se implementó un procesamiento distribuido en una arquitectura cliente-servidor basada en RA. Se implementó utilizando un dispositivo del tipo optical see-through y un mini board ODROID XU3. Los resultados de las pruebas mostraron que las actividades y tareas propuestas fueron altamente efectivas y eficientes, mejorando significativamente la CS de los usuarios mediante el uso de RAIOM. Además, la UX del sistema RAIOM fue alta.

VIII. APORTES

En cuanto al aporte de la presente tesis se pueden listar las principales contribuciones:

1. El uso de la RA como soporte a la CS para la toma de decisiones en el sector militar aprovechando el avance del software y de los dispositivos móviles actuales;
2. La utilización de una metodología combinada SCRUM-DCU para ello el diseño del framework y el desarrollo de las aplicaciones;
3. El diseño de un modelo tridimensional (3D-SA) para la identificación, selección y clasificación de requisitos de la CS y diseño de prototipos;
4. La consideración de una arquitectura basada en capas a fin facilitar la integración tecnológica;
5. La utilización de métodos válidos y confiables para evaluar la CS tales como SAGAT, SART, User Testing y Thinking Aloud y la experiencia del usuario mediante UMUX;
6. La implementación de un sistema distribuido utilizando componentes de hardware externo (ODROID-XU3) en donde se ejecuta el procesamiento de imágenes y reconocimiento de objetos

IX. LÍNEAS DE I/D FUTURAS

Existen diversas líneas de investigación que quedan abiertas luego de la finalización del informe, a saber:

- Capacidad de reconocimiento de objetos en modalidad nocturno: esta funcionalidad implicaría desarrollar un Módulo para operaciones nocturna implementando algoritmos específicos en el módulo de *Vision*.
- Deep Learning (DL): se propone la implementación de técnicas de clasificación basado en la utilización de redes neuronales implementada en el módulo de *Vision* o en su defecto en un módulo específico de DL diseñado separadamente.
- HMD militarizado: el laboratorio de Visión Aplicada en conjunto con el departamento de I+D+i de Tecnologías Informáticas del CITEDEF han diseñado y desarrollado, mediante el uso de impresora 3D, cada uno de los componentes que forman parte del soporte de visión (gafa monocular Optivent ORA-2[1]) y procesamiento (ODROID XU-4[2] de escala más pequeña al ODROID XU-3) en un casco militar de kevlar. En su fase inicial de prueba se están ajustando cada pieza que conformaran el casco militar de RA (HMD).

X. REFERENCIAS

[1] Azuma, R. T. A Survey of Augmented Reality. Teleoperators and Virtual Environments 6, 1997, 4, pp. 355-385.

[2] Endsley, M. R. Design and evaluation for situation awareness enhancement. Proceedings of the Human Factors Society Annual Meeting, 1988, 32(2), pp. 97-101

[3] Endsley, M. R. A taxonomy of situation awareness errors, human factors in aviation operations. In R. Fuller, N. Johnston, and N. McDonald (Eds.), Human Factors in Aviation Operations, 1995, pp. 287-292

[1] Optinvent ORA-2: http://www.optinvent.com/our_products/ora-2/

[2] ODROID-XU-4: https://www.hardkernel.com/main/products/prdt_info.php

[4] RDECOM, FUTURE SOLDIER 2030 Initiative, U. A. N. S. R. Center, Ed., USA, 2009

[5] A. Mitaritonna, M. J. Abásolo. Mejorando la conciencia situacional en operaciones militares utilizando la realidad aumentada. CACIC XIX Congreso Argentino de Ciencias de la Computación, 2013, pp. 356-365

[6] A. Mitaritonna Alejandro, M. J. Abásolo. Improving Situational Awareness in Military Operations using Augmented Reality. 23rd International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Pilsen, Czech Republic, 2015

[7] J. Timonen y J. Vankka, Enhancing Situational Awareness by Means of Information Integration of Sensor Networks and Visualization, S. 8756, Ed., Baltimore, Maryland, USA: SPIE 8756, 2013.

[8] F. Herpich, R. L. Martins Guarese y L. M. Rockenbach Tarouco. A Comparative Analysis of Augmented Reality Frameworks Aimed at the Development of Educational Applications. Creative Education, Vol.08, Article ID:77994, 2017.

[9] Hicks, Jeffrey; Flanagan, Richard; Dr. Petrov, Plamen; Dr. Stoyen, Alexander. Eyekon: Distributed Augmented Reality for Soldier Teams. © Copyright 21st Century Systems, Inc., 2003

[10] Livingston, Mark A.; Rosenblum, Lawrence J.; Julier, Simon J.; Brown, Dennis; Baillot, Yohan; Swan II, J. Edward; Gabbard, Joseph L.; Hix, Deborah. An Augmented Reality System for Military Operations in Urban Terrain. Proceedings of Interservice / Industry Training, Simulation & Education Conference (I/ITSEC), December 2 -5, Orlando, Florida, 2002, page 89 (abstract only)

[11] Julier, S.; Lanzagorta, M.; Baillot, Y.; Rosenblum, L.; Feiner, S.; Hollerer, T.; Sestito S. Information filtering for mobile augmented reality. In: Augmented Reality. (ISAR 2000).Proceedings.IEEE and ACM International Symposium, 2000

[12] D. Roberts, A. Menozzi, B. Clipp, P. Russler, J. Cook, R. Karl, E. Wenger, W. Church, J. Mauger, C. Volpe, C. Argenta, M. Wille, S. Snarski, T. Sherrill, J. Lupo, R. Hobson, JM Frahm, and J. Heinly. Soldier-worn augmented reality system for tactical icon visualization. Head- and Helmet-Mounted Displays XVII; and Display Technologies and Applications for Defense, Security, and Avionics VI, P. Marasco, P. Havig, D. Desjardins, K. Sarma, Editors, Proc. SPIE, Volume 8383, 838305, 2012

[13] V. Rautenbach y S. Coetzee. Results of an Evaluation of Augmented Reality Mobile Development Frameworks for Addresses in Augmented Reality. Conference: Conference: FOSS4G 2015, Seoul, Korea, 2015.

[14] A. Mitaritonna; M. J. Abásolo; F. Montero. Situational Awareness through Augmented Reality: 3D-SA Model to relate Requirements, Design and Evaluation. International Conference on Virtual Reality and Visualization (ICVRV), Shenzhen University Town Conference Center, Shenzhen, China, 2019

[15] ArToolKit, http://www.hitl.washington.edu/artoolkit/; 2020 [accessed 19 February 2020]

[16] Vuforia, https://www.vuforia.com/; 2020 [accessed 19 February 2020]

[17] EasyAR, https://www.easyar.com/; 2020 [accessed 19 February 2020]

[18] Layar, https://www.layar.com/; 2020 [accessed 19 February 2020]

[19] DroidAR, https://www.layar.com/; 2020 [accessed 19 February 2020]

[20] M. R. Endsley. Towards a theory of sa in dynamic systems. Human

Factors, 37:32–64, 1995.

[21] C. Gutwin and S. Greenberg. A descriptive framework of workspace

awareness for real-time groupware. Computer Supported Cooperative

Work (CSCW), 2002, 11(3): pp. 411–446.

[22] Agiles methologies, http://www.agiles.org/; 2020 [accessed 19 February 2020]

[23] Karel Vredenburg, Ji-Ye Mao, Paul W Smith, Tom Carey. A Survey of User-Centered Design Practice. Volume No, 4, Issue No. 1, April 20-25, Minneapolis, Minnesota, USA., 2002

[24] SCRUM, https://www.scrum.org/; 2020 [accessed 19 February 2020]

[25] C. Ardito, M. T. Baldassarre, D. Caivano y R. Lanzilotti. Integrating a SCRUM-based process with Human Centred Design: an Experience from an Action Research Study. IEEE/ACM 5th International Workshop on Conducting Empirical Studies in Industry (CESI), 2017

[26] Muñoz, A. A., Helander, K. N., de Gooijer, T., Ralph, M. Integrating Scrum and UCD: Insights from Two Case Studies. Integrating User-Centred Design in Agile Development. Part of the Human-Computer Interaction Series book series (HCIS), 2016, pp. 97-115.

[27] S. Krakowiak. Middleware Architecture with Patterns and Frameworks. France: Project Sardes - INRIA Grenoble Rhône-Alpes, 2009.

[28] C. Reynoso y N. Kicillof. Estilos y Patrones en la Estrategia de Arquitectura de Microsoft. Universidad de Buenos Aires, Buenos Aires, Argentina, 2004.

[29] D. Wagner. Handheld Augmented Reality. PhD Thesis, Graz University of Technology Institute for Computer Graphics and Vision, Austria, 2007.

[30] M. R. Endsley. Measurement of Situation Awareness in Dynamic Systems. Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1), 1995, pp. 65- 84.

[31] M. R. Endsley. Situation Awareness Global Assessment Technique (SAGAT). Aerospace and Electronics Conference – Proceedings of the IEEE 1988 National, 1988, pp. 789-795.

[32] B. Bloom. Taxonomy of educational objectives: the classification of educational goals. Handbook I: Cognitive Domain, by Benjamin S. Bloom (ed.). New York: Longmans, Green and Company, 1956, p. 207

K. Finstad. The Usability Metric for User Experience. Interacting with Computers, 2010, pp. 323-327

MODELOS DE MADUREZ PARA LA MEJORA DE CALIDAD DE LOS DATOS DE LOS INDICADORES DE DESARROLLO SOSTENIBLE

Ignacio Marcovecchio[1,2], Elsa Estevez[1,3], Pablo Fillottrani[1,4]

[1] Depto. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Argentina
[2] United Nations University Institute on Computing and Society, Macao SAR, China
[3] Instituto de Ciencias e Ingeniería de la Computación, (UNS-CONICET), Argentina
[4] Comisión de Investigaciones Científicas Provincia de Buenos Aires, Argentina

ignaciomarcovecchio@gmail.com, {ece,prf}@cs.uns.edu.ar

Tesis de Doctorado expuesta en la ciudad de Bahía Blanca el día 24 de Octubre de 2019

RESUMEN

Disponer de datos útiles y confiables para poder tomar decisiones informadas es uno de cimientos necesarios para alcanzar los objetivos de la Agenda de Desarrollo de 2030. Sólo mediante datos confiables que describan claramente la realidad se pueden definir estrategias y asignar recursos que contribuyan a alcanzar los Objetivos de Desarrollo Sostenible. El trabajo de investigación que se presenta en este documento contribuye con el cumplimiento de Agenda 2030 a través de intervenciones que buscan fortalecer las capacidades de las entidades responsables de producir datos que describan la realidad de los distintos países. En particular, formula un modelo de madurez de la capacidad prescriptivo y multidimensional para evaluar y mejorar la capacidad de los procesos que las Oficinas Nacionales de Estadísticas utilizan para obtener e informar datos estadísticos oficiales a los indicadores de los Objetivos de Desarrollo Sostenible. Las principales contribuciones de la tesis, además del modelo mencionado, incluyen una matriz de alineación de indicadores que permite determinar la similitud entre indicadores de desarrollo, un conjunto de indicadores de gobierno digital que aportan datos y metodologías para los indicadores de desarrollo sostenible, y un conjunto de recomendaciones para utilizar modelos de madurez de la capacidad como instrumentos para estandarizar la producción de datos estadísticos.

Palabras clave: Objetivos de Desarrollo Sostenible, Modelos de Madurez de la Capacidad, Capacidad Institucional

CONTEXTO

El presente documento exhibe un resumen extendido de la tesis de doctorado realizada por el Dr. Ignacio Marcovecchio en el marco de la colaboración entre la Universidad Nacional del Sur (UNS) y el Instituto de Computación y Sociedad de la Universidad de Naciones Unidas (UNU-CS) en Macau SAR, China. El trabajo de investigación fue realizado bajo la dirección de los Profesores del Departamento de Ciencias e Ingeniería de la Computación de la UNS Dra. Elsa Estevez y Dr. Pablo Fillottrani, y con la colaboración del Dr. Mamello Thinyane, Investigador Principal de UNU-CS y Director del Laboratorio de Datos para Desarrollo Sostenible de UNU-CS [1]. En la UNS, este trabajo se enmarcó dentro de las líneas de investigación en Modelos y Aplicaciones de Interoperabilidad Semántica en Gobernabilidad Electrónica del Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información [2] y en Ingeniería de Software y Gobierno Digital del Instituto de Ciencias e Ingeniería de la Computación [3].

1. INTRODUCCIÓN

En septiembre de 2015 los líderes de 193 países definieron 17 objetivos para el desarrollo sostenible que el mundo se propone alcanzar para el año 2030. Estos objetivos, que se conocen como los Objetivos para el Desarrollo Sostenible (Sustainable Development Goals – SDGs), definen la agenda global de desarrollo y plantean un conjunto ambicioso de objetivos que deben mantener el balance entre los tres pilares

fundamentales del desarrollo sostenible: inclusión social, desarrollo económico y sostenibilidad del medio ambiente. Los 17 objetivos buscan alcanzar 169 metas que van a ser controladas y evaluadas a través de 232 indicadores. La Comisión de Estadísticas de Naciones Unidas [4] desarrolló un sistema global de indicadores que permite controlar el progreso de los países para cumplir con los SDGs [5].

Se prevé que el control de los indicadores de los SDGs va a demandar grandes esfuerzos para que se produzcan datos confiables y de calidad, manteniendo la premisa de que "nadie sea dejado afuera" [6]. Sin embargo, estos objetivos representan un gran desafío a la capacidad de muchos países para medir el progreso para alcanzar las metas de los SDGs [7] ya que la capacidad de los actores principales del ecosistema debe ser potenciada para que se pueda utilizar y sacar provecho de los datos. Por este motivo, resulta indispensable asegurar que todos los países cuenten con un sistema nacional de estadísticas capaz de producir y controlar datos estadísticos confiables que cumplan con los estándares y expectativas globales [6].

Disponer de datos confiables es un factor crítico para poder transformar los SDGs en herramientas útiles para la toma de decisiones y la solución de los problemas. Sin datos actualizados y confiables, el diseño y la implementación de políticas adecuadas resulta muy difícil. Por estas razones, los datos juegan un rol fundamental en el sistema de control de los SDGs.

Para supervisar el progreso, responsabilizar a los gobiernos y promover el desarrollo sostenible es indispensable contar con instituciones fuertes, capaces de satisfacer de manera rápida y eficiente las cambiantes demandas de calidad de datos e información. Este trabajo busca fortalecer la capacidad de las instituciones responsables de liderar y coordinar los Sistemas Nacionales Estadísticos (National Statisticals Systems, NSS) para que puedan ser más eficientes en la obtención e integración de datos para la generación de información que permita la toma de decisiones informadas con el fin de lograr los objetivos de desarrollo sostenible.

Para poder realizar intervenciones que resulten útiles y novedosas, se intentarán contestar las siguientes preguntas de investigación:

P1. ¿Qué prácticas estadísticas e instrumentos garantizan la calidad de los datos para los indicadores de desarrollo sostenible?

P2. ¿Qué elementos del ecosistema de los indicadores sociales contribuyen a la producción de valor y cómo se relacionan?

P3. ¿Qué instrumentos de soporte y mejora de la calidad de datos son adecuados para asegurar la calidad de los datos estadísticos que producen las oficinas nacionales de estadísticas?

Construyendo sobre la premisa de que cuanto más maduras son las organizaciones dentro del ecosistema nacional de datos, más confiables serán los datos que pueden producir, esta tesis propone la utilización de modelos de madurez de la capacidad (Capability Maturity Model, CMM) como herramientas para mejorar la capacidad de las entidades responsables de obtener e informar datos sobre el progreso en la implementación de la agenda de desarrollo. Más específicamente, propone la formulación de un CMM prescriptivo y multidimensional para evaluar y mejorar la capacidad de los procesos que las Oficinas Nacionales de Estadísticas (National Statistical Office, NSO) utilizan para producir datos oficiales para los indicadores de los SDG.

El resto de este documento está organizado como se detalla a continuación. En la sección 2 se presenta la metodología que sirvió de guía para el desarrollo de este trabajo de investigación. A continuación, en la sección 3, se posiciona y enmarca el presente trabajo en la base de conocimientos a partir del relevamiento y estudio de trabajos relacionados, identificando los faltantes e identificando las similitudes y diferencias. En la sección 4 se presenta el modelo propuesto como herramienta para la mejora de la calidad de los datos de los indicadores de desarrollo sostenible, mientras que en la sección 5 se describen las actividades realizadas para la validación del modelo propuesto. Finalmente, en las últimas dos secciones se presentan una discusión de los descubrimientos realizados durante el desarrollo de este trabajo (sección 6) y las conclusiones alcanzadas (sección 7).

2. METODOLOGÍA

Con la motivación de contribuir al ecosistema mediante la introducción de nuevos artefactos y procesos para la construcción de dichos artefactos [8], este trabajo adopta una filosofía Pragmática y sigue el enfoque de la Ciencia del Diseño (Design Science, DS). En particular, adopta la vista de tres ciclos de la DS propuesta por Hevner y que entiende a la DS como tres ciclos de actividades fuertemente relacionados (*relevancia, diseño y rigor*) que diferencian a la DS de cualquier otro tipo de paradigmas [9].

Por la naturaleza del problema que se quiere resolver, que implica influir cambios en la forma de producir datos estadísticos con propósitos sociales a través del diseño y la apropiación de nuevas herramientas computacionales, la estrategia de investigación seleccionada es Investigación-Acción (*Action Research*) [10]. Asimismo, debido a que el ecosistema de datos es estudiado y evaluado de acuerdo con su evolución en el tiempo, esta investigación se enmarca en un horizonte temporal longitudinal.

Todas las decisiones de diseño de la investigación enunciadas en esta sección están fundamentadas en la visión del autor y en función de su comprensión del dominio y de los objetivos perseguidos, y están sustentadas por una extensa revisión de la bibliografía del dominio. Por ejemplo, un estudio sistemático de mapeo de la literatura en modelos de madurez que analizó más de 200 artículos resalta que la DS es el enfoque más utilizado cuando se desarrollan modelos de madurez [11]. La Figura 1 ilustra y resume el diseño de esta investigación.

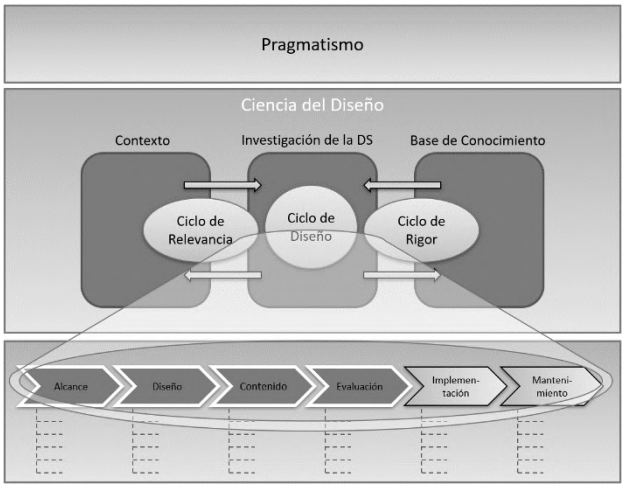


Figura 1: Diseño de la Investigación

3. TRABAJOS RELACIONADOS

A partir de una revisión extensiva de la literatura se identificaron numerosos esfuerzos que buscan mejorar el funcionamiento y los resultados generados por las entidades estadísticas nacionales. Debido a la gran cantidad y diversidad de trabajos existentes, y con el fin de facilitar su comprensión, los trabajos relevantes identificados fueron categorizados de acuerdo con el tipo de la solución. Las seis categorías definidas fueron marcos de trabajo (10), arquitecturas empresariales (2), estudios de nivel de preparación (3), procesos y estándares (2), y modelos (4). La Figura 2 resume el universo de instrumentos relevantes para la gestión de la calidad de datos de organizaciones estadísticas.

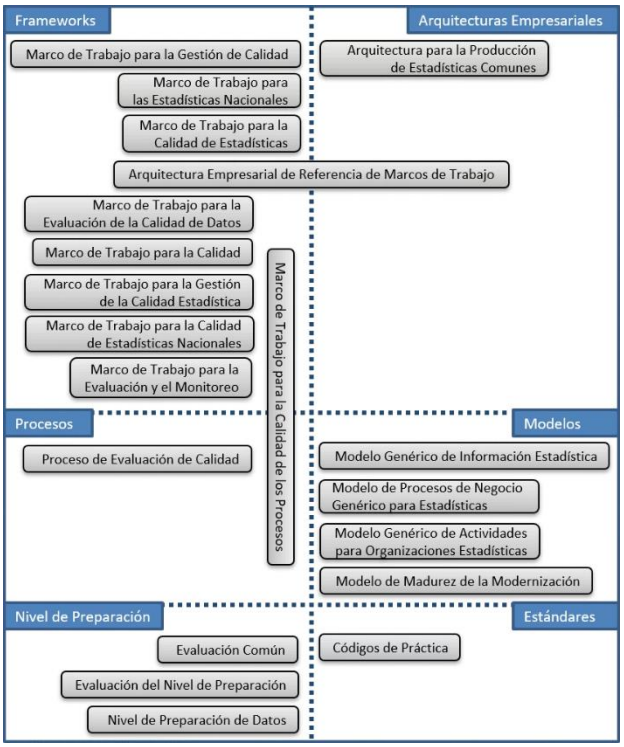


Figura 2: Trabajos Relacionados

La primer observación es que existe una demanda manifiesta de contar con datos e información confiable dentro de la comunidad estadística internacional y que existen numerosas iniciativas para mejorar la calidad y la precisión de los datos. Se observaron además dos grandes limitaciones; la primera es que gran parte de los trabajos se enfocan en evaluar y mejorar la calidad de los datos producidos, prescindiendo de cómo fueron producidos; la segunda es que ninguno de los trabajos se concentra en las demandas específicas de la Agenda 2030.

Mientras que ninguno de los trabajos relevados persigue los mismos objetivos que esta tesis, existen algunas similitudes y superposiciones. Desde el punto de vista de la utilización de modelos de madurez en el contexto de las agencias estadísticas, el trabajo que presenta más similitudes es el Modelo de Madurez de la Modernización (Modernization Maturity Model, MMM) ya que puede ser utilizado no sólo para identificar la madurez de las organizaciones estadísticas, sino que también las ayuda a modernizar la forma en la que operan [12]. Una de las diferencias fundamentales con el modelo propuesto en este trabajo radica en su enfoque: mientras que el CMM se centra en los procesos que producen datos para los indicadores de los SDGs, el MMM promueve la modernización de las formas de producir estadísticas oficiales en general. La evolución de ambos modelos también es distinta: mientras que el CMM es prescriptivo, el MMM es sólo descriptivo. Desde el punto de vista de especializarse en la producción de datos para la Agenda 2030, sólo dos de los trabajos satisfacen esta característica: la Evaluación Integrada Rápida (Rapid Integrated Assessment, RIA) que se enfoca en los planes nacionales de desarrollo [13], y la Evaluación Común de Países (Common Country Assessment, CCA) que se centra en políticas y programas nacionales [14]. Trabajos como el CCA pueden complementar al CMM que se propone en este trabajo ya que ofrecen el potencial de asegurar que el soporte otorgado por las agencias de Naciones Unidas en su conjunto, dentro de un país, sea coherente y complementario, sacando provecho de la experiencia, los recursos y el mandato de cada agencia.

Por lo tanto, luego de una extensiva revisión de la literatura – y de acuerdo con el conocimiento del autor – no existen herramientas diseñadas específicamente para mejorar los principios de los datos generados por las entidades estadísticas para el control de los SDGs y para evaluar la madurez de la capacidad de los procesos que emplean dichas entidades para producir datos estadísticos para los indicadores de los SDGs.

4. MODELO DE MADUREZ

El ciclo de diseño es el núcleo de todo proyecto de investigación de DS y es donde los artefactos se construyen y evalúan. Para guiar el desarrollo dentro del ciclo de diseño se ha seleccionado un marco de trabajo que incluye un proceso de seis fases para el desarrollo de modelos de madurez [15]. Las fases de desarrollo incluyen la definición del alcance, diseño, contenido, evaluación, implementación y mantenimiento. El orden de las fases es importante porque las decisiones que se realizan en una fase pueden tener influencia en las fases siguientes. En esta sección se describe instanciación de las fases de diseño y contenido.

4.1 Diseño

El diseño de la arquitectura del modelo representa las bases para su desarrollo y aplicación [15]. La arquitectura de los modelos de madurez se define a través de los componentes y su interacción [16] y se compone, mínimamente, por los objetos o capacidades a ser medidos (las dimensiones sobre las cuales se realizará la evaluación de la madurez) y por una medida que permita determinar su estado (los niveles de madurez) [11].

De los ejercicios de diseño se obtuvieron cuatro dimensiones que cubren todos los aspectos relevantes para la producción de datos estadísticos para los indicadores de los SDGs. Las dimensiones resultantes son:

- *Entorno* – representa el ecosistema en el cual las NSOs están inmersas y se compone de los factores institucionales y organizacionales que conforman el NSS y que tienen influencia en la efectividad y credibilidad de las NSOs para producir datos confiables y de buena calidad para los indicadores de los SDGs.
- *Recursos* – para poder desarrollar sus actividades, las NSOs deben contar con los recursos suficientes que les permitan contar con la última tecnología y personal bien preparado, y que asimismo les permita desempeñarse de manera independiente de cualquier presión económica. Los recursos no sólo deben ser suficientes, sino que deben ser administrados de manera eficiente.

- *Datos* – son tanto la materia prima como los productos finales que utilizan las NSOs para cumplir con sus responsabilidades.
- *Procesos* – si los datos son la materia prima con la que trabaja una NSO, los procesos son las herramientas que determinan cómo dichos datos se procesan para convertirse en productos estadísticos confiables y de buena calidad que permitan la medición, el control y la toma de decisiones basadas en evidencia..

De la evolución en el desarrollo del modelo se obtuvieron cuatro niveles para categorizar la madurez:

- *Básico* – la madurez de la capacidad de la organización es baja y la confianza en sus resultados es escasa debido a que los datos producidos pueden proporcionar una descripción de la realidad inexacta, parcial o incompleta.
- *Respaldado* – la madurez de la capacidad de la organización es intermedia y los resultados prometen un grado aceptable de calidad y confiabilidad, aunque podrían no ilustrar de manera precisa y completa la realidad.
- *Gestionado* – la madurez de la capacidad de la organización es alta y ofrece una descripción completa y precisa del estado nacional.
- *Experto* – la madurez de la capacidad de la organización es muy alta, lo que proporciona un reflejo preciso y altamente confiable de la realidad a nivel nacional.

El modelo en desarrollo, además de los niveles de madurez y las dimensiones de estudio, incluye los procesos a evaluar dentro de la organización, que son los que prescriben la secuencia de actividades que las NSOs realizan para producir datos estadísticos. Los procesos fueron agrupados en fases y representan el ciclo de vida de los datos. Las fases y los procesos no prescriben el modelo de procesos de negocio que las NSOs deben implementar para la producción de datos, sino que representan un conjunto de procesos genéricos que puede mapearse con los procesos específicos que cada NSO ejecuta para producir datos en el contexto de la Agenda 2030. El ciclo de vida diseñado para este modelo se compone de las siguientes cuatro fases:

- *Recolección* – comprende los procesos que se ejecutan para adquirir los datos.
- *Procesamiento* – agrupa los procesos que procesan los datos recolectados y validan los resultados hasta alcanzar un estado en el que puedan publicarse.
- *Utilización* – incluye los procesos relacionados con la difusión de los resultados además de la vinculación de los productos producidos con los consumidores.
- Impacto – abarca los procesos llevados a cabo para fomentar la utilización de los datos para generar impacto y crear cambios.

El modelo resultante es el resultado de la intersección de las fases con las dimensiones, segmentadas por los niveles de madurez, como se ilustra en la Figura 3.

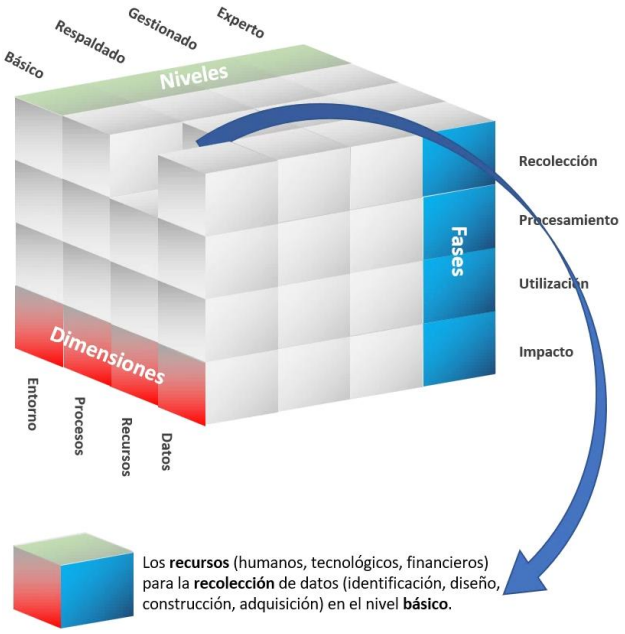


Figura 3: Arquitectura del Modelo

4.2 Contenido

En esta fase de desarrollo se identifica qué se debe medir en la evaluación de madurez y cómo se va a medir. El objetivo que se persigue en la definición del contenido es lograr componentes y subcomponentes del dominio que sean mutuamente excluyentes y colectivamente exhaustivos. Por tratarse de un dominio conocido, se realizó la identificación inicial de los componentes a través de una revisión de la literatura, la cual se mejoró a través de discusiones entre pares y entrevistas con expertos en el dominio.

De los ejercicios de definición de contenido, las cuatro dimensiones de análisis se subdividieron en 20 subdimensiones:

- *Entorno* – los factores que determinan la madurez del entorno de una NSO son su independencia profesional, un mandato claro y respaldado por el marco regulatorio, su ubicación dentro del NSS y una estructura interna que favorezcan su independencia y el acceso a los recursos necesarios, y un contexto que favorezca la colaboración dentro del ecosistema.
- *Recursos* – entre los recursos fundamentales para la producción de datos estadísticos confiables se encuentran los recursos humanos, tecnológicos, económicos y financieros, y físicos.
- *Datos* – los atributos que definen la calidad de los productos estadísticos y que determinan la madurez de una NSO para producir datos para los indicadores de los SDGs son su capacidad para gestionar la calidad de los datos; su confidencialidad; relevancia y vigencia; precisión y confiabilidad; puntualidad y oportunidad; coherencia, comparabilidad y consistencia; accesibilidad y claridad; así como su origen y nivel de desagregación.
- *Procesos* – esta dimensión evalúa la madurez de la capacidad de las técnicas y metodologías que emplea la NSO para la producción de datos estadísticos para los indicadores de los SDGs; su transparencia, imparcialidad y objetividad; la forma en que una NSO comparte e intercambia datos para, entre otras cosas, reducir la carga que se genera sobre los productores; y sus prácticas para mejorar de manera continua.

Similarmente, el contenido de las cuatro fases del ciclo de vida de la producción de datos estadísticos fue detallado a través de 14 procesos que se describen a continuación y se grafican en la Figura 4:

- *Recolección* – incluye los procesos de identificación de la necesidad de nuevos datos estadísticos, el diseño de cómo se conseguirán, la construcción de los artefactos y las herramientas necesarias para la recolección, y la adquisición de los datos necesarios.
- *Procesamiento* – contiene los procesos para el procesamiento de los datos recolectados en la fase anterior, su análisis, y la evaluación de los resultados para garantizar que sean adecuados para su propósito antes de ser difundidos.
- *Utilización* – consiste de la publicación, difusión y conexión que cubren el espectro de actividades que van desde la preparación de la difusión hasta su lanzamiento.
- *Impacto* – los procesos en esta fase incluyen actividades que promueven la utilización de datos para la toma de decisiones, la influencia en las decisiones basadas en datos y la reutilización y combinación de datos para la creación de nuevo conocimiento. Los procesos en esta fase suelen no existir entre los procesos tradicionales de las NSOs, pero son de gran importancia en el contexto de la Agenda 2030 y forman parte de la transformación y modernización que se persigue en las NSOs.



Figura 4: Fases y Procesos del Ciclo de Vida de los Datos

5. VALIDACIÓN

Para servir a sus objetivos, los modelos de madurez deben estar suficientemente validados; en caso contrario, la idoneidad y utilidad de un modelo de madurez sin ninguna aplicación ni validación resulta dudosa.

Informada por los objetivos de investigación, el diseño de la investigación y el cuerpo de conocimiento (*body of knowledge*), los ejercicios de validación realizados tuvieron como objetivo determinar si el modelo propuesto resulta adecuado para su uso (*fit-for-use*) y adecuado y útil para su propósito (*fit-for-purpose*). Mientras que el cuerpo de conocimiento informó los criterios generales para la validación, los objetivos particulares de la investigación y su diseño prescribieron los criterios específicos para la validación. La validación se realizó utilizando una combinación de múltiples métodos para evaluar los distintos criterios de validación. Para determinar la robustez del modelo se realizó una prueba de concepto (*proof of concept*) donde se evaluaron tanto su estructura como su contenido, estudiando no sólo el modelo resultante sino también cómo fue desarrollado. Los criterios para determinar la robustez del modelo fueron:

- *completitud* – se evaluó si contempla los principios que dictan la producción de estadísticas oficiales y los principios que rigen la calidad de los datos en el contexto de la Agenda 2030, y si incluye las áreas de evaluación relevantes consideradas por otros instrumentos similares;
- *correctitud* – fue evaluada a través de principios de los diseño que prescriben pautas para determinar la calidad de los modelos de madurez y de otros principios generales que evalúan la calidad de los artefactos desarrollados en el marco de la DS;
- *rigurosidad* – se analizó la instanciación del proceso de desarrollo y si fue desarrollado siguiendo estándares y buenas prácticas.

Para validar si el modelo satisface el propósito para el cual fue construido se realizó una prueba de valor (*proof of value*) donde se evaluó si cumple con los objetivos planteados y con los requerimientos preestablecidos, y si agrega valor tanto al dominio para el que fue desarrollado como a la base de conocimiento en el área.

Los criterios utilizados para estudiar la utilidad del modelo desarrollado fueron:

- *aptitud* – se evaluó si tiene la capacidad de cumplir adecuadamente la función para la que fue construido;
- *idoneidad* – se investigó si satisface los requerimientos planteados durante su concepción en las etapas de alcance y diseño;
- *originalidad* – se estudió si aporta nuevo conocimiento o nuevos instrumentos a la producción de datos para los indicadores de los SDGs y si agrega valor a la base de conocimiento en modelos de madurez y a la confiabilidad y calidad de la producción de datos estadísticos.

Los métodos de validación se resumen en la Tabla 1 e incluyeron tanto trabajo de escritorio como la participación de expertos de dominio y expertos en modelos.

Tabla 1: Criterios y Métodos de Validación

Prueba	Criterios	Métodos
Concepto	Completitud	- Comparaciones/Mapeos - Entrevistas a expertos
	Correctitud	- Principios de diseño DS - Principios de diseño CMMs
	Rigurosidad	- Trabajo de escritorio (rigor de aplicación del proceso de desarrollo)
Valor	Aptitud	- Entrevistas a expertos
	Idoneidad	- Trabajo de escritorio (validación de conformidad de requerimientos)
	Originalidad	- Literatura

Los ejercicios de validación realizados fueron satisfactorios, determinando tanto que el modelo propuesto es confiable como que puede ser utilizado para el objetivo para el que fue desarrollado. Adicionalmente, se observó que muchas de las limitaciones identificadas no son propias del modelo, sino que tienen origen en el ecosistema donde el modelo va a utilizarse. El modelo se enfoca específicamente en las NSOs y, si bien considera entre sus dimensiones el entorno en el cual se encuentran inmersas, los NSS son muy complejos y tienen un gran impacto en los resultados que se pueden lograr dentro de ellos. Una NSO madura dentro de un NSS inmaduro probablemente no pueda cumplir con los niveles de calidad y confiabilidad esperados.

Adicionalmente, muchos de los desafíos identificados por los expertos y especialistas (marco legal y regulatorio, financiamiento, interferencia política, contratación y retención de personal calificado) no son exclusivos de las NSOs sino que son desafíos comunes a muchas de las organizaciones del sector público.

Se observó también que, si bien varios de los expertos consultados coincidieron en que la introducción del modelo puede ser beneficiosa para la mejora de la calidad y la confiabilidad de los datos producidos para la Agenda 2030, algunos también resaltaron que las NSOs se encuentran bajo mucha presión de tiempo y reciben muchas demandas (muchas veces de los mismos datos por parte de distintos organismos) y que, por lo tanto, la introducción de más procesos puede no ser bien recibida o puede no cumplir con los beneficios esperados.

Otra de las observaciones fue que, por sí sólo, el modelo no puede solucionar todos los desafíos de datos que presenta la Agenda 2030 pero que, combinado con otras herramientas y otros esfuerzos, puede contribuir a lograr confianza en los datos lo que a su vez permita la toma de decisiones basada en evidencia confiable.

6. DISCUSIÓN

De la realización de este trabajo de investigación se lograron muchos descubrimientos que pueden aportar valor a otros trabajos que se realicen en el dominio. A continuación se presenta una pequeña discusión los más relevantes.

Se observa que existe una necesidad manifiesta de información confiable dentro de la comunidad estadística internacional y se han realizado numerosos esfuerzos con el fin de garantizar la calidad y precisión de los datos. Sin embargo, el proceso de la Agenda 2030 es largo y la tecnología cambia rápidamente, afectando directa e indirectamente las vidas de los seres humanos y, a su vez, los datos que se producen. Por lo tanto, la confiabilidad debe ser salvaguardada por organizaciones sólidas y maduras que sean independientes de sus empleados y de las administraciones actuales y futuras. Para este fin, los NSSs deben ser empoderados para adaptarse rápida y fácilmente a las nuevas realidades de los datos.

Todos los países, independientemente de su grado de avance y su nivel de desarrollo, pueden beneficiarse de soluciones que ayuden a desarrollar sus capacidades, como el CMM propuesto en esta tesis. Mientras que los países más desarrollados suelen disponer de mayores recursos para innovación y mejoras y, por lo tanto, suelen liderar el camino y marcar tendencias, los países en vías de desarrollo se puede beneficiar enormemente de los esfuerzos y la experiencia de aquellos que ya recorrieron ese camino. Cumplir los objetivos de la Agenda 2030 no implica una competencia entre países y sólo puede lograrse si cada país logra alcanzar sus metas y cumplir sus objetivos. Uno de los principios y creencias de los SDGs es que “los Estados Miembros de las Naciones Unidas trabajan juntos con un alto nivel de cooperación para mejorar las circunstancias de todas las personas en el mundo y los ubican en el centro del desarrollo futuro” [17, p. 1].

Las prácticas y soluciones provenientes del sector privado deben ser analizadas y adaptadas cuidadosamente ya que sus prioridades y objetivos pueden ser distintos. A manera de ejemplo, mientras que los indicadores de desarrollo prestan especial atención a la inclusión y el respeto por la privacidad de los individuos y sus comunidades, las soluciones del sector privado muchas veces se centran en otras prioridades. El modelo propuesto en este trabajo está definido para los SDG en particular, y para los indicadores sociales para el bien público en general.

Existen numerosos esfuerzos para el control de indicadores sociales y se puede (y se debería) sacar ventaja de ellos. Entre ellos se destacan las grandes inversiones que se han realizado para mejorar los datos para el control y reporte de los Objetivos de Desarrollo del Milenio (Millennium Development Goals, MDGs) [18]. De manera similar, algunos Estados Miembros de las Naciones Unidas reportan datos sobre derechos humanos desde hace más de una década en conformidad con las Revisiones Universales Periódicas (Universal Periodic Reviews, UPR) [19]. Todos estos esfuerzos, y en particular sus resultados, deberían ser estandarizados y considerados para desarrollar las sinergias que puedan facilitar.

Los datos deben incluir a todos y deben ser útiles para todos. La tendencia muestra que las empresas y los gobiernos confían cada vez más en los Grandes Datos (*Big Data*) y los datos analíticos asociados. Mientras que las empresas utilizan grandes datos para informar las decisiones y la estrategia empresarial, los gobiernos los utilizan para proporcionar una mejor prestación de servicios y participación ciudadana [20]. Enfoques complementarios como el de Datos Pequeños (*Small Data*) – en el que los datos, en lugar de ser agregados, se procesan en la misma unidad en la que se tomaron las muestras [21] – son importantes para asegurar que nadie sea excluido. El modelo propuesto en esta tesis integra ambos enfoques para promover la inclusión.

7. CONCLUSIONES

Este documento presentó el trabajo de investigación que se desarrolló con el fin de mejorar la calidad de los datos estadísticos producidos por los ecosistemas nacionales de datos para el seguimiento y control del progreso en el cumplimiento de la Agenda 2030.

Para responder la primera pregunta de investigación (P1) se identificaron, estudiaron y clasificaron los instrumentos utilizados para la gestionar la calidad de los datos oficiales para desarrollo. Para poder realizar tal estudio, primeramente se definió un marco conceptual de desarrollo sostenible que brinda los fundamentos conceptuales para todo el trabajo.

Con el fin de responder a la pregunta P2 se estudiaron y describieron los ecosistemas de producción de estadísticas oficiales para el seguimiento y control del progreso en el cumplimiento de la Agenda 2030. Se definió un modelo conceptual que describe los ecosistemas nacionales y globales de producción y reporte de datos que informan el estado de cumplimiento de las metas y los objetivos de los SDGs. Además, para enriquecer la comprensión del ecosistema global de control y reporte se realizó un mapeo a nivel de los indicadores que estudió cómo los indicadores de los instrumentos de medición de otros dominios (en este caso, Gobierno Digital) pueden proveer datos a los indicadores de los SDGs.

Para responder la tercera y última pregunta (P3) se plantearon dos objetivos de investigación que consistían en diseñar y validar un modelo para la producción de datos que permita a las NSOs avanzar gradualmente en la producción confiable de datos de calidad que sirvan para medir el cumplimiento de las metas de los SDGs. El cumplimiento de ambos objetivos se logró a través del desarrollo de tres actividades que incluyeron el diseño y desarrollo del modelo, la validación de su calidad y la utilidad, y el desarrollo un marco conceptual de modelos de madurez que consistió en un relevamiento extensivo y exhaustivo de la literatura existente sobre modelos de madurez en general y modelos de madurez de la capacidad en particular.

Las principales aportes de esta tesis son el CMM para la producción de datos para los indicadores de los SDGs y su respectiva validación. Pero además, este trabajo de investigación realizó las siguientes contribuciones:

- Una clasificación de los instrumentos relevantes para la calidad en la producción datos estadísticos.
- Una matriz de alineación de indicadores que permite determinar la similitud entre indicadores de desarrollo.
- Un conjunto de indicadores de Gobierno Digital que aportan datos y metodologías para 10 indicadores de los SDGs.
- Un conjunto de recomendaciones para reutilizar trabajos de medición en áreas relacionadas con el desarrollo sostenible para la medición de los SDGs.
- Un conjunto de recomendaciones para utilizar modelos de madurez de la capacidad como instrumentos para estandarizar la producción de datos estadísticos.

Todas estas contribuciones fueron compartidas mediante publicaciones y otras actividades académicas y de difusión. Los resultados de esta investigación sirven además para establecer un marco de referencia para investigaciones posteriores y como plataforma para continuar con las mejoras en la calidad y confiabilidad de los datos que permiten la toma de decisiones de manera informada y sustentados por evidencias. Posibles líneas de investigación futuras incluyen el estudio de la adaptabilidad del modelo para ser utilizado en otros procesos y dominios.

8. BIBLIOGRAFÍA

[1] United Nations University Insitute on Computing and Society, "Last Mile Data Enablement and Collaboration, and Building Trust in Indicators Data." http://cs.unu.edu/research/sdgs/ (accessed Mar. 13, 2018).

[2] Universidad Nacional del Sur, "Líneas de investigación – LISSI DCIC CIC." https://lissi.cs.uns.edu.ar/?page_id=26 (accessed Mar. 18, 2018).

[3] Universidad Nacional del Sur, "Departamento de Ciencias e Ingeniería de la Computación." https://cs.uns.edu.ar/home/ (accessed Mar. 18, 2018).

[4] United Nations Statistics Division, "United Nations Statistical Commission," 2018. http://unstats.un.org/unsd/statcom (accessed Jan. 31, 2017).

[5] United Nations, "Global Indicator Framework for the Sustainable Development Goals and Targets of the 2030 Agenda for Sustainable Development," 2017. [Online]. Available: https://unstats.un.org/sdgs/indicators/Global Indicator Framework_A.RES.71.313 Annex.pdf.

[6] Independent Expert Advisory Group on a Data Revolution for Sustainable Development, "A World that Counts: Mobilising the Data Revolution for Sustainable Development," 2014. [Online]. Available: http://econpapers.repec.org/RePEc:ecr:col013:40319.

[7] Business for 2030, "Metrics & Indicators." http://www.businessfor2030.org/metrics-indicators/ (accessed Jan. 31, 2017).

[8] H. A. Simon, *The Sciences of the Artificial (3rd Ed.)*. Cambridge, MA, USA: MIT Press, 1996.

[9] A. R. Hevner, "A Three Cycle View of Design Science Research," *Scand. J. Inf. Syst.*, vol. 19, no. 2, pp. 87–92, 2007, doi: http://aisel.aisnet.org/sjis/vol19/iss2/4.

[10] E. Ferrace, *Action Research*. 2000.

[11] R. Wendler, "The maturity of maturity model research: A systematic mapping study," *Inf. Softw. Technol.*, vol. 54, no. 12, pp. 1317–1339, 2012, doi: 10.1016/j.infsof.2012.07.007.

[12] "Modernisation Maturity Model (MMM) - Roadmap for Implementing Modernstats Standards - UNECE Statistics Wikis." http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=129172266 (accessed Feb. 28, 2017).

[13] UNDP, "Rapid Integrated Assessment (RIA)," 2017.

[14] United Nations Development Group, "Common Country Assessment," 2002. [Online]. Available: http://www.sprep.org/att/IRC/eCOPIES/Countries/Vanuatu/3.pdf.

[15] T. De Bruin, R. Freeze, U. Kaulkarni, and M. Rosemann, "Understanding the Main Phases of Developing a Maturity Assessment Model," in *Australasian Conference on Information Systems (ACIS)*, 2005, pp. 8–19, doi: 10.1108/14637151211225225.

[16] D. L. Moody and G. G. Shanks, "What Makes a Good Data Model? Evaluating the Quality of Entity Relationship Models," in *Proceedings of the13th International Conference on the Entity-Relationship Approach*, 1994, pp. 94–111, [Online]. Available: http://dl.acm.org/citation.cfm?id=647516.727205.

[17] "Sustainable Development Goals, Beliefs, and Principles." https://www.agora-parl.org/resources/aoe/sustainable-development-goals-beliefs-and-principles (accessed Jun. 05, 2017).

[18] United Nations, "The Millennium Development Goals Report," 2015. doi: 978-92-1-101320-7.

[19] United Nations Office of the High Commissioner for Human Rights (OHCHR), "Universal Periodic Review - A Practical Guide for Civil Society," 2014.

[20] M. Thinyane, "Investigating an Architectural Framework for Small Data Platforms," in *Proceedings of the 17th European Conference on Digital Government (ECDG 2017), Lisbon, Portugal, 12-13 June*, 2017, pp. 220–227.

[21] M. Best, "Small Data and Sustainable Development," in *International Conference on Communication/Culture and Sustainable Development Goals: Challenges for a new generation.*, 2015, pp. 1–6.

Generación automática inteligente de resúmenes de textos con técnicas de Soft Computing

Tesis doctoral realizada en cotutela por:

Augusto Villa Monte

Directores:





Laura Cristina Lanzarini
Universidad Nacional de La Plata
Buenos Aires, Argentina

José Angel Olivas
Universidad de Castilla-La Mancha
Ciudad Real, España

Fecha de exposición:
La Plata, 18 de Marzo de 2019

---

1. Introducción

Muchos años después de que el escritor y científico Toffler [1970] pronosticara que se produciría más información de la que sería posible procesar, la explotación de datos y el descubrimiento de conocimiento se volvieron fundamentalmente necesarios en todos los ámbitos. Así fue como en los últimos tiempos, los avances tecnológicos favorecieron la generación y almacenamiento de grandes volúmenes de datos y, como resultado, se volvió esencial el desarrollo de métodos inteligentes capaces de representar la información disponible [Fayyad et al., 1996].

Si bien los sistemas y aplicaciones actuales generan datos en diferentes formatos, la mayoría produce y almacena texto. Este formato resulta menos atractivo que otros como el sonido, las imágenes y el video, pero es, sin lugar a duda en la actualidad, el principal medio de comunicación entre seres humanos [Schreibman et al., 2016].

Desde la invención de la escritura, el ser humano ha almacenado el conocimiento en textos. Desde entonces, la cantidad de documentos disponibles ha aumentado exponencialmente mientras que el costo de generar, almacenar, duplicar y compartir dicha información fue disminuyendo [Saracco, 2017]. Al mismo tiempo, el consumo de información aumentó notablemente [Johnson, 2011].

Generalmente, los documentos de texto digitales no están estructurados ni mucho menos organizados en bases de datos tradicionales. El texto por sí mismo no tiene ningún tipo de estándar ni restricción al crearse y almacenarse y, por lo tanto, procesarlo se ha vuelto una tarea extremadamente difícil. El desarrollo de soluciones computacionales que permitan resumir texto pretende reducir los problemas generados por el crecimiento desmedido de información textual.

2. Motivación

A partir de la evidente explosión de información, sería ideal que el ser humano pudiera recordar toda la información disponible, pero en ese caso el cerebro colapsaría. El ser humano recuerda muchas cosas al mismo tiempo que olvida muchas otras. Inconscientemente capta la información esencial. Esta tarea, tratándose de texto, se conoce como "resumir".

Durante los últimos 60 años se han logrado grandes avances en lo que refiere a resúmenes automáticos. Sin embargo, desarrollar un programa por computadora que resuma requiere instrucciones precisas [Borko and Berniers, 1975]. Existen dos grandes enfoques para construir un resumen de forma automática, uno de los cuales tiene que ver con la interpretación del significado del contenido del documento (enfoque abstractivo) y el otro con el análisis más bien de su estructura (enfoque extractivo) [Hahn and Mani, 2000]. Esta tesis trata ambos enfoques a través de la propuesta de dos soluciones diferentes aplicadas a documentos médicos.

Contar con estrategias capaces de identificar automáticamente lo principal de un documento facilita su procesamiento, no sólo en lo que se refiere a lectores humanos, quienes son beneficiados agilizándoles la lectura, sino también en lo que se refiere a técnicas de aprendizaje automático y de recuperación de información. Por ejemplo, cuando se hace una búsqueda web resulta más eficiente trabajar sobre algunas partes de los documentos. Esto implica un menor esfuerzo en el proceso de búsqueda al operar únicamente con lo más importante de cada página web. Además, la búsqueda se concreta en menos tiempo y se resuelve más rápido al buscar el documento completo solo en caso de corresponder el criterio de búsqueda con el resumen de los documentos.

3. Objetivos y contribuciones

Esta tesis tiene por objetivo principal contribuir al área conformada por el *Procesamiento de Lenguaje Natural* y la *Minería de Texto* con dos estrategias distintas capaces de identificar a partir de un conjunto de documentos lo relevante y construir con eso un resumen en forma automática. Por un lado, identificando el criterio del usuario al seleccionar las partes principales de un documento y, por otro, extrayendo de los documentos patrones textuales específicos sumamente útiles en la toma de decisiones.

La primera de las soluciones, para documentos con cierta estructura, permite crear resúmenes extractivos utilizando una técnica de *Optimización mediante Cúmulo de Partículas* (PSO por sus siglas en inglés) a partir de la representación vectorial de los mismos basada en un conjunto amplio de métricas de puntuación de sentencias. Dicha técnica identifica el criterio del usuario al seleccionar las partes del documento que considera importante. En lugar de utilizar las distintas métricas en forma independiente para construir el resumen, la respuesta del método sugiere la combinación que mejor se ajusta a la valoración que el usuario realizó de cada parte de un documento en forma previa. El método propuesto, que combina una representación binaria y una continua utilizando una variante original de la técnica mencionada, no sólo permite identificar los coeficientes asociados a cada una de las métricas sino también cuáles son las métricas que permiten resumir lo más parecido al criterio del usuario.

La segunda solución extrae las sentencias causales con restricciones temporales existentes en un conjunto de documentos médicos y luego convierte dichas oraciones en un grafo causal equivalente. Para ello, se identifican y extraen patrones textuales específicos del texto. El modelo proporciona las relaciones que describen el contenido de los documentos originales mostrando únicamente todos los vínculos "causa-efecto" junto con las restricciones temporales que afecten su interpretación. Identificar estos patrones específicos en documentos médicos resulta sumamente útil para la toma de decisiones en el área de salud. El grafo, además, tiene la información necesaria para generar nuevas frases a partir del recorrido entre sus nodos y arcos.
Es importante aclarar que los términos de las afirmaciones que se obtienen del grafo están en los documentos originales pero las afirmaciones que se consiguen no necesariamente son extracciones textuales de los mismos, a diferencia de la primera solución en la cual se transcriben partes textuales de los documentos y que al conformar el resumen final no necesariamente tienen una coherencia narrativa. En las próximas dos secciones se describirán brevemente ambas soluciones.

4. Resumen utilizando una técnica de optimización mediante cúmulo de partículas

Un resumen extractivo está formado por un conjunto de porciones de texto (desde palabras sueltas hasta párrafos enteros) literalmente copiadas de la entrada de datos [Mani, 2001a]. Estas porciones de texto generalmente son llamadas "sentencias" haciendo referencia a las oraciones del mismo. Para extraer las partes del documento que formarán el resumen, este enfoque requiere asignar una puntuación a cada sentencia. Esto permite clasificar el contenido del documento ordenando las sentencias en una lista cuyas primeras posiciones son ocupadas las sentencias más relevantes que recibieron el puntaje más alto [Edmundson and Wyllys, 1961]. Hay muchas maneras de puntuar las sentencias. Cada una de ellas permite seleccionar las "mejores" sentencias para producir el resumen pudiendo variar el resumen resultante en cada caso. En las Secciones 2.10 y 2.11 de la tesis se describe detalladamente la representación vectorial de los documentos de texto y las métricas más utilizadas para puntuar las partes de un documento.
Es así como, los documentos se modelan como vectores n-dimensionales obtenidos de calcular n métricas. Por lo tanto, cada documento está representado por una matriz S de p filas (sentencias) y n columnas (métricas). Dicha vectorización transforma el documento en un conjunto de p vectores "sentencia" de la forma $S_i = [s_{i1}, s_{i2}, \ldots, s_{in}]$. Luego, estos vectores son utilizados para obtener resúmenes automáticos aplicando algoritmos más sofisticados [Nenkova and McKeown, 2012].
En general, obtener un resumen extractivo puede considerarse un problema de clasificación de dos clases en el que a cada parte del documento se la etiqueta como "correcta" si forma parte del resumen, o "incorrecta" caso contrario [Neto et al., 2002]. Trabajos recientes consideran la obtención de resúmenes extractivos un problema de optimización, donde una o más funciones objetivo se formulan para seleccionar las "mejores" oraciones del documento que forman el resumen [Vázquez et al., 2018]. Sin embargo, en este tipo de trabajos los documentos se representan a través de un conjunto de métricas establecido a priori, y su selección no forma parte del proceso de optimización, como sucede en [Meena and Gopalani, 2015] por ejemplo. Esta es la característica clave del método desarrollado en este capítulo de la tesis.
Cuando la solución exacta a un problema es difícil de obtener, las estrategias de búsqueda aproximadas mediante técnicas de optimización han demostrado ser sumamente efectivas. Este tipo

de técnicas mejoran, a través de procesos biológicos, un grupo de soluciones denominado población de individuos. Cada individuo representa una posible solución al problema y se desplaza por el espacio de soluciones observando su propio comportamiento y el de su entorno. En este principio se basa el primero de los métodos propuestos en esta tesis, el cual construye el resumen automático utilizando PSO. Dicha técnica fue propuesta por Kennedy and Eberhart [1995] y desde entonces se han desarrollado diferentes versiones. Originalmente se definió para operar sobre un espacio de búsqueda continuo pero años más tarde Kennedy and Eberhart [1997] definieron una versión discreta. Sin embargo, uno de los principales problemas que tuvo dicha versión fue la dificultad para cambiar de 0 a 1 y de 1 a 0 una vez estabilizado. Esto impulsó el desarrollo de diferentes versiones binarias del PSO que buscaron mejorar su capacidad exploratoria.

Usar PSO para generar, a partir de la representación vectorial de los textos, un resumen extractivo que combine adecuadamente varias métricas de puntuación de sentencias, requiere utilizar ambos tipos de PSO. Por un lado, se requiere seleccionar el subconjunto de métricas a utilizar (parte discreta) y, por el otro, se necesita establecer la relevancia de cada una de estas métricas (parte continua).

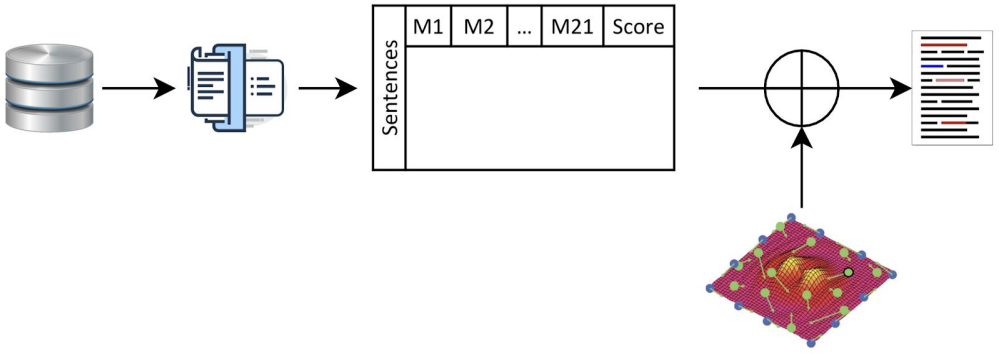


Fig. 1 - Construcción del resumen extractivo utilizando PSO

La Figura 1 muestra cómo se construye el resumen con la estrategia propuesta. En lugar de utilizar las distintas métricas en forma independiente para construir el resumen, la respuesta del método sugiere la combinación que mejor se ajusta a la valoración que el usuario realizó de cada una de las partes de un documento en forma previa. El método propuesto, que combina una representación binaria y una continua utilizando una variante original de la técnica mencionada, no sólo permite identificar los coeficientes asociados a cada una de las métricas sino también cuáles métricas permiten resumir lo más parecido al criterio del usuario.

El PSO inicia su población en forma aleatoria y mide iteración a iteración el desempeño de la solución de cada partícula para resolver el problema. Para eso se construye el resumen de los documentos de entrenamiento utilizando el criterio establecido por la partícula en cada iteración y el resumen obtenido se lo compara con el esperado midiendo cuán parecido ha resultado. Esto justifica la necesidad de preprocesar los documentos una única vez y tenerlos almacenados en forma de vectores, para que cada partícula pueda medir sobre los documentos su desempeño sin tener que volverlos a procesarlos. Una vez evaluadas todas las partículas, se mueve la parte continua de las partículas con un movimiento estándar y la parte binaria con algunas modificaciones propias para favorecer la estabilidad de las mismas una vez conseguido el óptimo. El detalle

completo del funcionamiento del método propuesto puede encontrarse en el Capítulo 3 de la tesis, así como el detalle de la representación y almacenamiento de los documentos en el Capítulo 2 y Anexo B respectivamente.

Para evaluar la calidad del resumen automático producido por el método propuesto, se utilizaron 3322 artículos publicados en PLOS Medicine entre 2004 y 2018. Se realizó el entrenamiento del modelo con los artículos de un mes y el testeo con los del mes siguiente. Se realizaron 30 ejecuciones independientes con 100 iteraciones como máximo y se resumió el 10% del documento. El PSO fue global y de población fija con 10 partículas inicializadas aleatoriamente. Además, para determinar el aporte del método propuesto se lo comparó con un método previo que no poseía la capacidad de seleccionar métricas.

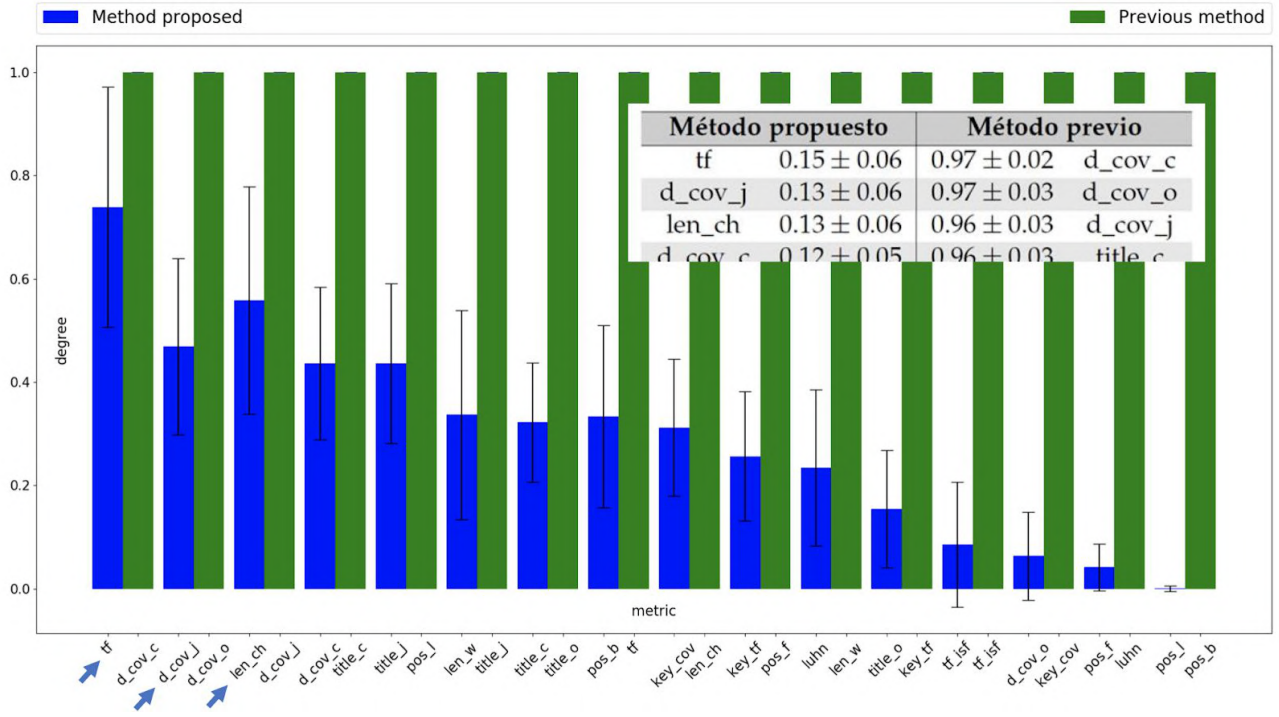


Fig. 2 - Participación de las métricas ordenadas descendente por valor de coeficiente

La Figura 2 muestra el nivel de participación de las métricas para los dos métodos evaluados, en orden decreciente según el valor de su coeficiente promedio indicado en la tabla que se encuentra superpuesta en la imagen. La tabla completa con la media y desviación de los coeficientes de todas las métricas utilizadas se encuentra en la página 86 de la tesis. Estos coeficientes son los que se usan para ponderar el valor de cada métrica al obtener el puntaje de cada sentencia. Por ejemplo, si se observan los tres primeros valores en la columna "Método propuesto" de dicha tabla, se puede observar que los coeficientes promedio de las métricas "tf", "d_cov_j" y "len_ch" son 0.15, 0.13 y 0.13, respectivamente. Por lo tanto, el criterio para resumir hace el mismo énfasis entre "d_cov_j" y "len_ch". Sin embargo, si se observa la Figura 2, se puede ver que el nivel de participación de "len_ch" es mayor que el de "d_cov_j". Esto se debe a que la primera ha sido seleccionada más veces por la técnica de optimización. En cambio, la métrica "tf" para el método propuesto tiene el coeficiente promedio más alto de la tabla y también el nivel más alto de participación. Por otro lado, la Figura 3 muestra cómo la precisión de cada uno de los métodos evoluciona a medida que se

incorporan métricas en la construcción final del resumen. Esto se hace en el orden indicado en la tabla de coeficientes promedio. Como se puede ver, el comportamiento del método propuesto aquí es más estable. Además, después de agregar la cuarta métrica, la precisión se vuelve notablemente mejor que la que se obtiene utilizando todas las métricas. Aunque el valor máximo se observa con la incorporación de la séptima métrica, cuatro métricas serían suficientes para obtener un buen rendimiento. También se debe tener en cuenta que, utilizando el método previo, incluso si la precisión resultante es mayor para las dos primeras métricas, las restantes arrojan un resultado inferior en comparación con el método propuesto, sin lograr superar el valor más alto.

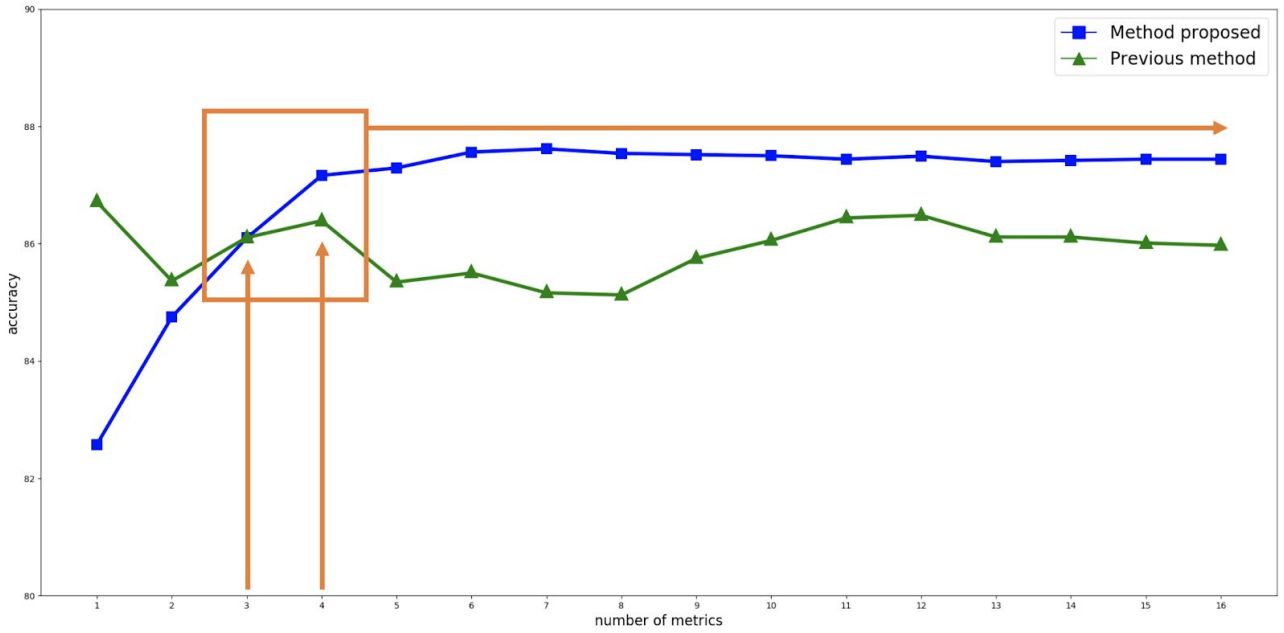


Fig. 3 - Evolución de la precisión a medida que se agregan métricas para obtener el resumen

5. Resumen mediante grafos causales y con componentes temporales

Así como existen métodos de construcción de resúmenes cuyo enfoque es extractivo, existen otros métodos que utilizan estructuras tipo grafo para representar el contenido de un conjunto de documentos a partir de la extracción de patrones textuales específicos. En esta dirección, se han realizado estudios utilizando relaciones conceptuales, centrándose principalmente en aspectos semánticos y no tanto en la causalidad presente en los textos. La causalidad cumple un rol importante en cualquier toma de decisiones ya que los efectos de una decisión pueden determinarse por sus causas. Desde ese punto de vista, la causalidad permite generar conocimiento científico objetivo [Kant et al., 1999; Mill, 1843] y describir fenómenos [Mach, 1976]. Si de Medicina se trata, este tipo de sentencias están presentes al detallar las causas de una enfermedad o al informar los efectos de un tratamiento. Con frecuencia, las Ciencias de la Salud muestran la causalidad como un proceso complejo que involucra la evolución de una causa anterior en causas intermedias a lo largo del tiempo antes de alcanzar el efecto final. La lógica causal ofrece un vocabulario y reglas específicas para explicar y predecir procesos complejos en términos de vínculos “causa-efecto”.
La causalidad es una relación en la cual la ocurrencia de una entidad B de cierta clase depende de la ocurrencia de una entidad A de otra clase. Generalmente por “entidad” se entiende un fenómeno, un

hecho, una característica, una situación o un evento, entre otras cosas. En este tipo de relación, "A" representa una causa y "B" un efecto. Entre ambos existe una clara relación de dependencia ya que la causa provoca un efecto, y el efecto se deriva de la causa considerándola su consecuencia. Se expresa no sólo utilizando el término "cause" sino también "produce", "bring about", "issue", "generate", "result", "effect" o "determine", entre otros [Kim, 1995].

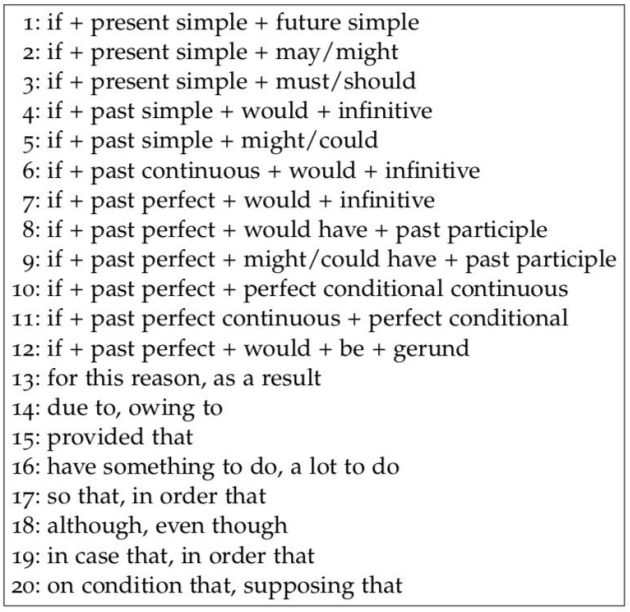


Fig. 4 - Estructuras condicionales y causales implementadas

La causalidad puede ser un proceso directo cuando A causa B y B es un efecto directo de A; o un proceso indirecto cuando A causa C a través de B, y C es un efecto indirecto de A. Tratándose de texto la fuente de información, para extraer las relaciones causales directas se necesita tener un amplio conocimiento del lenguaje que permita definir correctamente una serie de patrones de búsqueda específicos como los de la Figura 4. En cambio, no ocurre lo mismo con las relaciones causales indirectas. Este tipo de relaciones no se puede identificar a través de patrones. A pesar de ello, extrayéndose relaciones causales directas y representándolas adecuadamente en forma de grafo (como el de la Figura 5), a través del recorrido de sus caminos pueden establecerse relaciones indirectas que no eran obvias y que no podrían haberse obtenido de otra manera.

Como se dijo anteriormente, las explicaciones causales relacionan causas y efectos y, en Medicina, a menudo se las restringe temporalmente. El tiempo juega también un papel importante en diferentes escenarios médicos [Keravnou, 1996]:

- *Prevención* evaluando factores de riesgo en pacientes que se comportan de cierta manera ⇒ «Smoking for a long time causes lung cancer»
- *Pronóstico* prediciendo la probabilidad de supervivencia de una persona ⇒ «25 % of patients with septic shock will die within 15 days»
- *Tratamiento* administrando medicamentos durante ciertos períodos de tiempo o bajo ciertas restricciones temporales ⇒ «Omeprazole should be taken before aspirin»
- *Diagnóstico* identificando el origen o la causa de las enfermedades ⇒ «According to the symptoms that he presents, he has flu»

La toma de decisiones basada en el tiempo cubre áreas de la atención al paciente donde resulta fundamental la administración de medicamentos durante ciertos períodos de tiempo o bajo ciertas restricciones temporales.

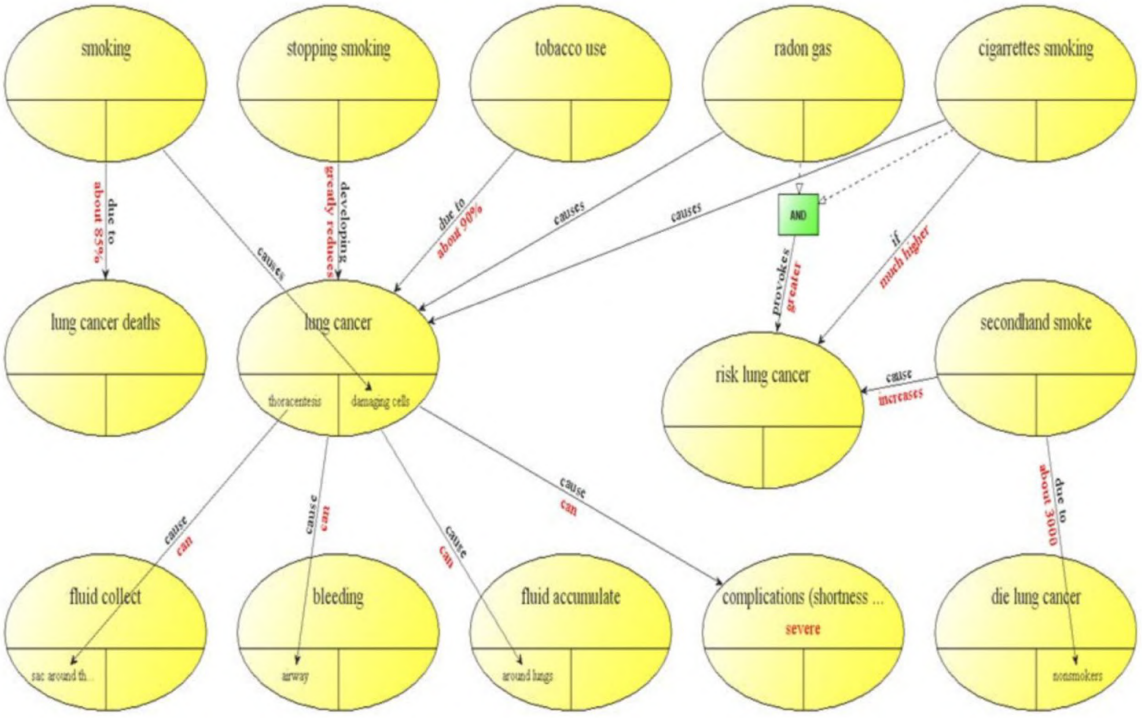


Fig. 5 - Grafo causal relacionado con «lung cancer» construido de forma automática

El tiempo utiliza un léxico específico. Las oraciones temporales incluyen con frecuencia referencias relacionadas con el calendario (año, mes, día) o momentos del día (mañana, tarde, noche). También utilizan conjunciones y preposiciones como "by", "until", "before", "since", "past", "next". Además, hay frases causales como "secondary to" o "because of" que también denotan influencia causal temporal. Para analizar el comportamiento de este tipo de sentencias en documentos de texto, se requiere contar con un proceso capaz de detectar, extraer y clasificar las sentencias con base en ciertos patrones estructurados. En el Capítulo 4 de la tesis podrá encontrarse la descripción completa del estudio realizado sobre las asociaciones causales que implican dependencias temporales y se podrá encontrar también más detalle sobre el proceso que permite detectarlas y extraerlas para construir el grafo causal.

Las sentencias finalmente extraídas por el proceso descrito en la tesis permiten luego crear una base de conocimiento causal sobre un determinado tema. Esto es posible siempre que se procesen documentos de una misma temática. Tratándose de páginas web, pueden conseguirse los documentos a través de una simple búsqueda en Internet y posterior recopilación del contenido de los sitios arrojados como resultado.

Una vez producida la base de conocimiento en un determinado ámbito y sobre un tema en particular, el usuario deberá introducir una pregunta. Dicha pregunta permitirá seleccionar aquellas oraciones que se encuentren directamente relacionadas con ella. Una vez que todas las frases hayan sido procesadas rastreando los conceptos asociados a causas o consecuencias, se construye el grafo

causal que las representará (como el de la Figura 5). En la Figura 6 se muestra el esquema del proceso completo que permite obtener el grafo.

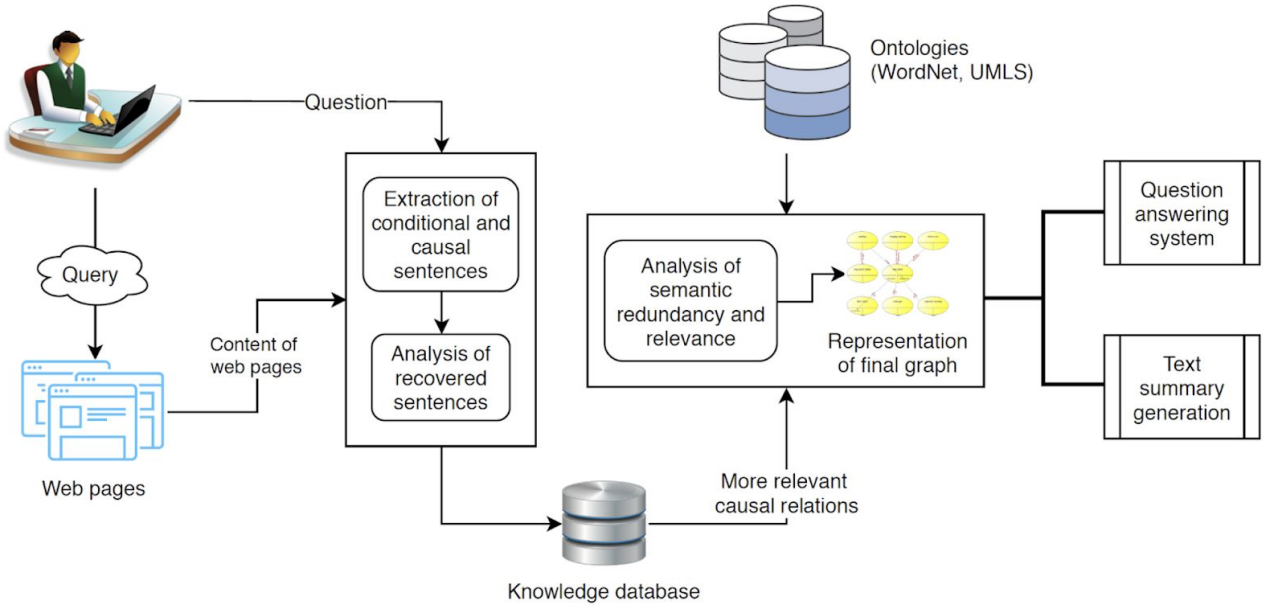


Fig. 6 - Proceso para obtener un grafo causal a partir de texto plano

Los grafos causales son una forma gráfica de mostrar las dependencias causales que posee la información no estructurada como lo es el texto. La causalidad implica una transferencia de la causa al efecto que en el grafo se denota mediante una flecha dirigida. Dicha flecha conecta el nodo "causa" con el nodo "efecto" en el grafo. Como podemos ver en la Figura 6, cada nodo representa un concepto mediante una elipse. Las relaciones entre nodos están representadas por arcos que llevan el tipo de conector causal. Entonces, de la lectura del grafo pueden obtenerse relaciones ocultas.

Para introducir restricciones de tiempo al grafo, se debe buscar por cada nodo aquellas oraciones que tengan modificadores temporales. El siguiente paso consiste en poder identificar si el indicador de tiempo viene asociado al nodo antecedente o al consecuente para introducirlo en el nodo correcto. A partir del grafo, se puede generar un resumen que incluya información causal con componentes temporales.

Aunque existen sistemas que recuperan sentencias causales de los textos, no hay muchos que organicen dichas sentencias para resolver un problema concreto y, mucho menos, haciendo uso de información temporal. Los diagnósticos de enfermedades o la prescripción de medicamentos frecuentemente implican dependencias temporales. En el caso de una enfermedad generalmente depende de cuándo haya aparecido el síntoma y, en el caso de los medicamentos, su uso está limitado a ciertas condiciones temporales.

Cada año la salud de los pacientes se ve perjudicada por la mala administración de medicamentos. Existen estudios que evidencian un alto porcentaje de medicamentos mal administrados en cuidados intensivos [Keers et al., 2013]. Dado que la correcta administración de medicamentos depende del tiempo y, además, se encuentra a cargo de seres humanos (enfermeras generalmente), desarrollar sistemas de alerta puede evitar errores, contribuir con la seguridad del paciente y mejorar la gestión

de los hospitales. Algunos medicamentos, por ejemplo, deben tomarse antes o después de las comidas, siendo “antes” y “después” restricciones temporales. Por eso en esta tesis, para contribuir al manejo del tiempo en tratamientos médicos que deben tener en cuenta el tiempo se propuso el desarrollo de una aplicación llamada “My Medicine”, que permite controlar las restricciones temporales indicadas por los médicos al prescribir tratamientos para ciertas enfermedades. La descripción detallada de la aplicación podrá encontrarla en el Capítulo 5.

6. Conclusiones y líneas de trabajo futuro

En esta tesis se han propuesto dos estrategias basadas en técnicas de Soft Computing para la generación automática de resúmenes de texto. Se las ha evaluado con procedimientos habituales en el área de estudio dando buenos resultados.

Por un lado se ha desarrollado un método para identificar el criterio del usuario al seleccionar las partes principales de un documento por medio de una variante original de PSO. Esta propuesta ha sido evaluada sobre una amplia colección de artículos científicos y los resultados reflejan que con pocas métricas es posible caracterizar el criterio del usuario para resumir. Además, la técnica utilizada no se limita a resúmenes, habiendo sido aplicada también en la obtención de reglas de clasificación.

Por el otro, se ha desarrollado una estrategia de extracción de relaciones causales en los textos a partir de las cuales se construye un grafo con anotaciones temporales que afectan su interpretación. Esto además dio lugar al desarrollo de una aplicación para dispositivos móviles que combina la causalidad y temporalidad para ayudar a controlar la administración de dosis de medicamentos.

También, se ha comparado la calidad de los resúmenes generados de forma extractiva a partir de métricas y aquellos creados a partir de las relaciones causales. El resultado ha demostrado que la calidad está vinculada al tipo de narrativa del documento. Los resúmenes extractivos basado en PSO son más adecuados para compactar el volumen de información textual y los resúmenes abstractivos basados en causales son mejores para generar resúmenes conceptualmente más ricos.

Como líneas de trabajo futuras se plantea ampliar el conjunto de métricas utilizado para caracterizar los documentos de entrada y así enriquecer su representación; incorporar conceptos de Lógica Borrosa o Difusa que permitan flexibilizar el criterio del usuario; continuar con el desarrollo de la aplicación My Medicine para utilizarla además en la gestión hospitalaria; e, incluir nuevas estrategias para verbalizar el grafo causal resultante.

Por último, cabe mencionar que la tesis desarrollada tiene asociadas varias publicaciones científicas en revistas y congresos internacionales, de las cuales el tesista es el autor principal. Por la tesis realizada, el autor fue reconocido como “Mejor Egresado de Postgrado” por la Universidad Nacional de La Plata y obtuvo la “Mención Cum Laude” en la Universidad de Castilla-La Mancha.

7. Referencias

Las referencias se encuentran en la tesis, publicada en http://sedici.unlp.edu.ar/handle/10915/74098.

Bocetado de Interacciones Enactivas

Resumen de Tesis presentada para obtener el Doctorado en Ciencias Informáticas
en la Facultad de Informática de UNLP

Andrés Rodríguez
arodrig@lifia.info.unlp.edu.ar
Centro de Investigación LIFIA Facultad de Informática UNLP
Director Dr. Gustavo Rossi
Centro de Investigación LIFIA Facultad de Informática UNLP
Codirector Dr. Pascual González López
LoUISE Esc. Sup. de Ingeniería Informática Universidad Castilla La Mancha
22 de Noviembre de 2019

Resumen

Esta tesis se constituye en el cruce de tres temas: el diseño de sistemas interactivos que combinan un pie en lo digital y uno en lo físico, las teorías de la cognición corporizada y enactiva y las prácticas creativas soportadas por el bocetado. Este trabajo incluye contribuciones de diferente carácter. Se realiza un estudio profundo de las teorías sobre cognición corporizada y enactiva, del diseño de interacción con dispositivos digitales y del bocetado como herramienta básica del diseño creativo. Sobre la base de este análisis de la bibliografía existente y con una caracterización de la práctica de bocetado de interacciones enactivas basada en estudios etnometodológicos se plantea un framework para organizar conceptualmente esa práctica y una herramienta de soporte a esa actividad concebida como una composición creativa. Se discuten las contribuciones y se plantean posibles líneas de trabajo futuro.

Palabras clave: HCI, diseño de interacciones, bocetado, cognición

1. Introducción

El desarrollo de nuevas tecnologías interactivas y una mejor comprensión de cómo participa cuerpo en los procesos cognitivos empuja al campo de Interacciones entre Personas y Computadoras (HCI) a la necesidad de resolver la relación del usuario con múltiples dispositivos que se extienden mucho más allá de los escritorios[1].

La tendencia creciente a integrar la forma física y el proceso digital afecta a todas las disciplinas involucradas en el diseño de esos sistemas. Desde la perspectiva de la informática significa crear interfaces *corporizadas*, que utilizan diversos objetos físicos para controlar la información digital e interfaces *contextuales* que dependen de las señales disponibles en el entorno y la situación.

El diseño se ha considerado central para HCI, comenzando con un enfoque en el diseño de software, de la interfaz y finalmente el proceso completo como diseño centrado en el usuario. Muchos trabajos han mostrado la complejidad de esta práctica de diseño que no puede reducirse a unas recetas de pasos a seguir. Los estudios en el marco de la *Tecnología como Experiencia* (cómo aplicar la fenomenología de la experiencia a la práctica del diseño) y los aspectos lúdicos que promueven la curiosidad, la exploración y el disfrute estético entre las personas cuando se encuentran con nuevas tecnologías agregan un factor más de complejidad al paisaje.

Las teorías de la Cognición Corporizada (CC) que entienden la cognición como un fenómeno *corporizado*, *situado* y *enactivo* (e.g. [2], [3], [4]) se expandieron notablemente en las últimas décadas y abrieron nuevos programas de investigación en HCI. A partir del éxito del libro de Paul Dourish *Where the action is*[1] se produjo un viraje hacia la corporización con la intención de comprender la

interacción en términos prácticos con el entorno social y físico. La CC concibe la cognición como un proceso dinámico donde el significado surge dentro de la acción en curso, en lugar de suponerlo como fijo o pre-configurado. La acción y la percepción están inherentemente acopladas[7]. Los nuevos enfoques de CC enactiva obligan a nuevos estudios acerca de su vinculación con la práctica del diseño de sistemas interactivos.

En este contexto, el foco de esta tesis se constituye en el cruce de tres temas de interés actual para la Interacción entre Personas y Computadoras (HCI):

- ***la actividad del diseño de sistemas interactivos***, con énfasis en la creación de los mecanismos de interacción que combinan un pie en lo digital y uno en lo físico. ¿Qué ocurre con el desafío de integrar la forma física con el software digital?
- las ***teorías de la cognición corporizada y enactiva***. ¿Cómo aproximar el diseño de sistemas interactivos tomando en cuenta de manera explícita que la cognición que los usuarios tienen de ellos es *corporizada y enactiva*?
- las ***prácticas creativas soportadas por el bocetado***, herramienta por antonomasia del diseño de interacciones, en particular los procesos de generación, evaluación y comunicación de propuestas de diseño. ¿Cuáles son las similitudes y diferencias entre el bocetado a mano alzada y los bocetos de interacciones enactivas? ¿Cómo deberían ser las herramientas para soportarlos?

A continuación se resumen los puntos más relevantes del trabajo realizado.

2. Background y trabajos relacionados

2.1. La cognición corporizada para el diseño de sistemas interactivos

La CC provee una base teórica relevante para el diseño de la nueva clase de formas artificiales que integran software e interacción tangible[8, 1]. La idea de CC surge como reacción al paradigma conocido como *Cognitivismo* que dominó el período inicial de investigaciones en HCI a partir de los trabajos de Newell, Card, Simon y Moran[9, 10]. En esa visión el conocimiento consiste esencialmente en *representaciones* almacenadas en el cerebro y todas las representaciones juntas forman un *modelo mental* del mundo exterior. El cerebro realiza *computaciones* sobre estas representaciones que permiten seleccionar una acción apropiada, dado el adecuado input perceptual. La CC rechaza esta división cartesiana entre la representación interna y el mundo externo y considera que el cerebro, el cuerpo y el entorno (y las relaciones entre ellos) forman parte del sistema cognitivo. La cognición tiene lugar **en la acción** como un aspecto del comportamiento real en situaciones concretas y que no se puede entender como un razonamiento abstracto separado de circunstancias concretas[2].

Se pueden identificar tres formas de abordar la idea de la CC: como un fenómeno distribuido de representación y computación, como una práctica socialmente situada o como un acoplamiento sensorio-motor (enacción). La primera, más cercana al cognitivismo, ha aportado conceptos como memoria externa, distinciones entre acciones epistémicas y pragmáticas y andamiaje cognitivo que dieron lugar a programas como las Interfaces de Usuario Tangibles de Ishii. La idea de cognición como práctica situada en contextos sociales donde los planes son adaptados continuamente, alimentó muchos casos de dispositivos interactivos para hacer música. Finalmente, la idea más radical de enacción con su enfoque de acoplamiento estructural entre la percepción y la acción, mediante conceptos como las affordances ha empezado a explicar, dar soporte y facilitar diseño de interacciones para generar nuevos sentidos y experiencias, interacciones transparentes o sin pantallas y enfoques sistémicos de la actividad.

2.2. Diseño de interacciones digitales y bocetado

El diseño es esencialmente una actividad intencional y compleja, que involucra incertidumbre y riesgo y que merece entidad propia en el proceso de creación de sistemas interactivos. Su complejidad reside, entre otras cosas, en las numerosas fuentes de información, restricciones y deseos, muchas veces

contradictorios, que es necesario compatibilizar hasta definir el problema a resolver. Sin embargo, la actividad de diseño no es un arte mágico e inescrutable sino que tiene su propia disciplina y rigor en la indagación, que involucra ciclos iterativos de enmarcado y re-definición de problemas, desarrollo de hipótesis, experimentación y actuación para cambiar progresivamente una situación incierta en una con mayor certidumbre. En esa práctica, el bocetado es esencial. Todas las externalizaciones, gráficas, tridimensionales o actuadas, que ayudan a pensar la situación, conversar sobre ella con otros o guardar ideas para más adelante, son centrales en el proceso de proponer una solución (hipótesis, *ver-como*) y analizar si se adapta al problema o lo redefine y abre un nuevo camino a explorar (testear, *ver-eso*).

En el caso de las relaciones entre personas y dispositivos digitales, donde lo temporal es relevante, el bocetado incorpora además la necesidad de experimentar la interactividad, la interacción-en-acción. Diseñar esa interacción es la actividad que intenta determinar los espacios de acciones (todo lo que la persona puede hacer) y operaciones (lo que el sistema es capaz de reconocer) en los que se vean involucrados el sistema o artefacto y sus usuarios.

2.3. Relevamiento de herramientas de soporte

Se realizó una revisión sistemática de la literatura sobre las herramientas propuestas para soportar la práctica del diseño de interacciones corporizadas. Tras un proceso de selección de dos ruedas, se recuperaron un total de 78 artículos que reportan *frameworks*, guías, toolkits y aplicaciones que se ofrecen para (a) abstraer y conceptualizar la actividad, (b) generar ideas de diseño y (c) construir modelos para evaluar la interacción. Estos trabajos se han centrado en las dimensiones de una idea de interacción, la corporización o *fisicalidad* en sí misma, la interactividad (*interacción-en-acción*) y las cualidades hedónicas de la experiencia de uso. Se reportan diversas herramientas para facilitar la creación, almacenamiento e intercambio de representaciones análogas a los bocetos en papel o las maquetas 3D enriquecidas con capacidades interactivas, aunque son pocos los trabajos que se apoyan sobre las observaciones de la práctica real de los diseñadores de interacciones corporizadas.

En el caso de las herramientas para generar, bocetar y evaluar ideas el principal foco ha sido puesto sólo en la disminución del umbral de acceso a la elaboración de modelos (básicamente mediante la facilitación de conexiones y circuitos mediante componentes pre-ensamblados). Esta revisión sistemática muestra que existe espacio para ampliar el soporte al proceso de la generación de ideas, no sólo bajando el piso, sino expandiendo el techo y las paredes del espacio de diseño, que es el enfoque que intentamos abordar en esta tesis.

3. Preguntas y método de investigación

El objetivo central de esta investigación es operacionalizar la Cognición Corporizada Enactiva para el bocetado y la generación de ideas en el diseño de interacciones y caracterizar y soportar esta práctica con herramientas computacionales. Para ello, se plantean dos preguntas que la tesis intenta responder:

1. **¿Cuáles son las características de los sistemas interactivos que se apoyan en la Cognición Corporizada y Enactiva?** Aquí buscamos comprender ¿de qué manera la CC provee información al diseño sobre la forma en que los sistemas interactivos se conectan con las prácticas corporizadas de la gente y situadas en el mundo real? ¿cuáles son las características de los sistemas interactivos digitales que facilitan una interacción Corporizada y Enactiva entre las personas y esos sistemas?
2. **¿Cómo podemos diseñar sistemas interactivos que se apoyen en la Cognición Corporizada y Enactiva?** En particular ¿qué características de la CC son relevantes para el proceso de diseño de los sistemas interactivos? ¿Qué características debe poseer el bocetado para dar soporte a la generación y evaluación de ideas para interacciones enactivas?

Para buscar respuestas a estas preguntas, esta tesis utiliza como método de investigación la aproximación triangular entre perspectivas múltiples que Mackay y Fayard [11] proponen para los problemas en HCI: a) una perspectiva teórica, en la cual las teorías, modelos y marcos de trabajo son estudiados y utilizados para entender un dominio del problema, b) una perspectiva de observación en la cual se observa cómo los usuarios interactúan con el mundo y c) una perspectiva de diseño en la cual el análisis de los datos de las observaciones y las teorías son trasladados en artefactos de diseño, los que a su vez pueden ser fuente de estudios de observación o teorización. Para nuestro trabajo, la triangulación se constituyó en:

- **Línea teórica: Entender la Cognición Corporizada y Enactiva**. Además de una revisión profunda del estado del arte en este tema, lo triangulamos para agregar en sus resultados en el los protocolos de análisis de observaciones y en los requerimientos de desarrollo de las toolkits.
- **Línea de diseño: diseñar herramientas para diseñar sistemas interactivos** El marco general dentro del cual se puede ubicar la aproximación en la línea central es el llamado Investigación a través del Diseño (RtD por su sigla en inglés)[12].
- **Línea de observación: Entender la práctica del bocetado** La observación en los talleres y las entrevistas situadas que se resumen más adelante integran en la investigación el uso de técnicas de etnometodología para el estudio del diseño de interacciones.

Durante la investigación, las diferentes actividades se realizaron de manera que se alternaran a lo largo del tiempo, informándose mutuamente. La investigación general que se lleva a cabo se construye de manera iterativa en el sentido que sugieren estos ciclos de estudio, análisis, diseño y evaluación. Dentro de cada actividad de diseñar o estudiar seres humanos en acción, también hay reflexiones que pueden aportar una visión parcial: los resultados de los talleres se interpretan a la luz de la teoría, y también lo son los problemas de diseño particulares y la forma en que los abordamos. En conjunto, esto dio lugar a una comprensión cada vez más enfocada de las relaciones entre la cognición corporizada y enactiva, el diseño de sistemas interactivos y el contexto particular de la práctica del bocetado.

4. Una caracterización del bocetado de interacciones enactivas

Para caracterizar la práctica del diseño de interacciones corporizadas y enactivas se realizaron estudios empíricos con técnicas de etnometodología. Se llevaron adelante cuatro instancias de talleres de 16 horas cada uno en las que participaron en total 69 profesionales de diferentes nacionalidades y disciplinas, todos con práctica activa en el diseño de sistemas interactivos. Organizados en grupos de 3 o 4 integrantes generaron 15 propuestas. En cada taller, el autor desempeñó el rol de observador participante alternando entre los diferentes grupos. Con las consignas *Ver con los oídos*, *Ver con la piel* y *Sentir en el cuerpo el entorno físico/digital* se pidió a los diseñadores que generaran ideas de interfaces enactivas basadas en la sustitución sensorial, estrategia de diseño que requiere un abordaje de cognición corporizada[13]. Todo el proceso fue registrado en audio y video y analizado con un protocolo basado en el modelo *Personas-Actividades-Tiempo-Espacio*[14].

El análisis arrojó luz sobre las concepciones de la cognición corporizada que emplean los diseñadores y sobre los procesos y herramientas que utilizan para bocetar, en particular:

Desde la computación distribuida hacia la enacción, ida y vuelta En los diferentes grupos hubo una recorrida permanente de un camino que va desde la posición representacional (casi cognitivista) al enfoque más radical del acoplamiento sensorio motor. Los datos obtenidos indican que cualquier caracterización del proceso de ideación de interacciones enactivas debería considerar esta traza como un camino normal y proveerle andamios para recorrer la vía hacia el acoplamiento sensorio motor sin forzar ni obligar un proceso que saque al diseñador de su propia racionalidad.



Figura 1: Un grupo realizando pruebas de experiencia con sus bocetos interactivos

Uso de huellas creativas Tomando prestados conceptos de estigmergia (el mecanismo de coordinación entre agentes a través del medio ambiente y las huellas dejadas en él), podemos decir que los artefactos persistentes producidos durante la ideación (bocetos, audios, maquetas, anotaciones, etc) pueden considerarse como *huellas* que ayudan a construir el camino hacia la producción de una idea. Podemos enfocarnos en el aspecto sensorio-motor de esos artefactos en lugar de su función representativa (externalizar una idea) o social (pensar con otros). Sin embargo, en línea con lo propuesto por van Dijk[15], en el caso de la ideación de interacciones se trata, además, de ***huellas creativas***, ya que son expresiones ***intencionadas***, actos que involucran objetos a través de los cuales una persona se expresa a si misma en un mundo social.

Durante la ideación los diseñadores gradualmente comienzan a construir un espacio compartido de huellas que perduran en los bocetos, las anotaciones en papel o pizarra, las configuraciones del espacio de trabajo, etc, con el objetivo de construir la idea de interacción que se expresa en la configuración final de los artefactos y los compromisos de cada participante. Ese camino de migas de pan no es lineal, como en cualquier exploración tiene curvas y contracurvas, pero puede ser recorrido en el sentido opuesto para abrir nuevas bifurcaciones. Todos los grupos hicieron este ida y vuelta en el camino de ideación. Avanzaron con una propuesta, la bocetaron y probaron. Sólo al repasar el *rationale* que los había llevado a esa idea (para lo cual fueron imprescindibles todas las huellas dejadas, tales como bocetos, anotaciones, videos, etc, que les permitieran **re-visitar** y **re-vivir** el momento previo de creación) se sintieron en la seguridad que la nueva idea que proponían realmente les abría un nuevo sendero de exploración.

Uso de interactividad como la dimensión del filtrado de los bocetos Todos los diseñadores trabajaron directamente sobre la *interacción-en-acción* o interactividad y dejaron para las etapas finales las pruebas sobre otras dimensiones de filtrado, como la apariencia, la funcionalidad o estructura espacial[16]. Mantener el foco sobre la interacción en sí misma involucra una postura de exploración y descubrimiento que favorece la emergencia de la forma del producto como una corporización de la interactividad buscada, tal como postula el abordaje enactivo. Esta emergencia impulsa además la toma oportunista de decisiones. De alguna forma el producto surge como una encarnación o corporización de la interactividad que se busca, en un ida y vuelta constante de definiciones sobre la extensión de campos de acción y operación[17].

Expresión de los atributos de interacción en cada tipo de boceto Los diseñadores registraron los atributos deseados de interacción en cada modelo que construyeron con recursos diferentes en cada caso: sobre los bocetos 2D o los storyboards lo hicieron con anotaciones textuales o un lenguaje de símbolos, pero también en las maquetas o bocetos en hardware trataron de registrar

cómo lograr la interactividad que imaginaron (post-Its en maquetas, comentarios en el código del software). En el contexto de acción del aquí y el ahora, los diseñadores hablaban, hacían gestos y modificaban representaciones gráficas o 3D. Los modelos crean un contexto objetivo en el que los diseñadores pueden realizar sus movidas. Lo observado en estos talleres respalda el trabajo de Robertson[18] que describe cómo los diseñadores utilizan las enacciones para crear y participar en una representación del proceso o actividad basada en el tiempo en la que otros pueden participar. Los bocetos de diferentes naturaleza (papel, hardware, experiencia) se pueden considerar como estados en un particular diagrama de estado; lo que falta en el diagrama son las transiciones. Para representar las transiciones entre estados, los diseñadores hacen uso de gestos y enacciones. Los diseñadores de interacción corporizada deben aprender a reconocer e incluso articular estas enacciones como formas importantes de expresarse a medida que desarrollan diseños.

Uso de bocetos como ideas de interacción *enactuadas* La concepción y creación de nuevas interfaces enactivas demanda la generación de una gran cantidad de externalizaciones y representaciones para complementar la generación de ideas que puede ser soportada por el papel con la acción sobre el mundo real, usando el cuerpo (propio o actuando en grupo) y modelos físicos tridimensionales. Es evidente que los diseñadores necesitan valerse de herramientas que les permitan explorar y confrontar la interacción que buscan, casi como una puesta en escena que implica herramientas de diseño de gestos, performances casi coreográficas y bocetado interactivo (con o sin hardware electrónico). Los diseñadores necesitan expresar la dinámica de la interacción. Jugar el rol de un usuario y explorar la tecnología potencial con la intención de usarla, ayuda a los diseñadores a explorar el diseño y sus ideas de diseño, y también a generar nuevas ideas. El uso de recorridos de interacción conducidos por gestos y enacciones también facilita ver que hay consecuencias en las otras dimensiones del boceto, fuera de la interactividad.

Uso de *tinkering* Todos los grupos emplearon un estilo de *tinkering* (una especie de "hágalo usted mismo" basado en la exploración de material existente) para construir modelos físicos e interactivos, con cualquier elemento a la mano o con sencillos circuitos sobre placas de prototipado, antes de intentar un modelo más refinado o con mayor fidelidad. El ajuste de los atributos de interactividad fue realizado con procesos iterativos de re-codificación, compilado y subida del programa a Arduino, lo que claramente interrumpía el flujo cognitivo de los diseñadores. Este aspecto fue el más destacado por los diseñadores: la necesidad de disponer de herramientas que agilizaran esta dinámica de cambio y les facilitaran mantener el flujo de pensamiento sobre las ideas de la interacción buscada. De acuerdo con las estrategias de *tinkering*, refuncionalización de partes desguazadas, elaboración de circuitos ad hoc, etc. observadas en los talleres es factible sugerir que los diseñadores se benefician con la disminución del umbral para la construcción con componentes electrónicos (por ejemplo, soporte para construir circuitos de manera correcta, herramientas de diagnóstico de error y correcciones apropiadas).

5. Herramientas para el bocetado de interacciones enactivas

Los toolkits se consideran en la investigación de HCI como contribuciones de tipo *artefactual* donde "el nuevo conocimiento está embebido en y manifestado por los artefactos y el material de soporte que los describen"[19]. Por lo tanto, a partir de los análisis de la Cognición corporizada y el diseño de interacciones y la caracterización de esta práctica en talleres, se abordó un proceso de diseño para embeber ese análisis un conjunto de herramientas de soporte.

Pluma 1.0 La primera contribución es Pluma, una pequeña aplicación para generar bocetos con el concepto de composición creativa basada. Pluma versión 1.0 permite componer y configurar rápidamente bocetos interactivos empleando gestos manuales como input y actuación háptica e imágenes visuales como output. Tiene dos módulos principales: uno para seleccionar los elementos de

sensado y actuación y otro para determinar la interactabilidad del artefacto en boceto. La aplicación permite la .[ej]ecución"del boceto y su modificación tantas veces como sea necesario sin requerir programación por parte del diseñador. Está implementada como una pequeña arquitectura web basada en Node.js[1] que se comunica mediante el framework Johnny-Five[2] con módulos de Arduino con Firmata[3] y el sensor LeapMotion para detectar gestos[4].

Para la evaluación de Pluma 1.0 se realizó un estudio de caso informal (en el sentido propuesto por Lazar[20]) con un grupo de 5 profesionales que participaban de un desarrollo para la Televisión Digital Interactiva (4 diseñadores y un ingeniero electrónico). La consigna fue generar ideas de control remoto de TV con interacción gestual y feedback háptico. El taller se extendió por 4 horas, se registró en audio y video y luego se utilizó un protocolo de análisis con codificación abierta.

Los diseñadores se enfocaron directamente en encontrar una buena interactividad y para eso filtraron todos sus bocetos con la interactabilidad que buscaban para el control. Sólo aparecieron algunos esbozos en papel sobre formas y geometrías, fundamentalmente analizando tipos de prensión y resolución de interactabilidad para ambidextros. Fue interesante comprobar que los 4 participantes con formación y experiencia de diseño utilizaron bocetos 2D, storyboards y maquetas de cartón con numerosas gesticulaciones además del bocetado en hardware elaborado con Pluma, pero sólo el ingeniero actuó sobre los circuitos de actuadores y sensores: la *cajanegrización* del código de software por sí sola no alcanza para bajar el umbral de acceso, es necesario incorporar en ese mecanismo de alguna forma también los circuitos y conexiones de componentes. La forma de trabajo en términos de composición creativa no pareció imponer restricciones fuera de la lógica propia de los diseñadores.

Un marco para el bocetado de interacciones enactivas Los estudios de CC y Diseño más la evaluación de Pluma 1.0, dieron pie a un framework conceptual para organizar y relacionar los diferentes elementos y variables que entran en juego en el proceso del bocetado para la ideación de interacciones enactivas, utilizando un esquema de dos dimensiones más una tercera que corresponde al desarrollo temporal del camino creativo:

- **La corporización de interactabilidad** Los diseñadores generan representaciones externas utilizando dos, tres o cuatro dimensiones (bocetos, maquetas de estudio con o sin interactividad, etc.).
- **La extensión y profundidad de la interactabilidad** con representaciones que van desde una expresión estrecha y profunda de la interactabilidad (un aspecto particular con todas sus variaciones) hasta una amplia y superficial (un poco de toda la interactabilidad que se busca).
- **La ubicación temporal del boceto** aunque es obvio que las representaciones se crean en distintos tiempos es necesario incluir esta temporalidad en la organización porque la secuencia que se produce de distintas externalizaciones es parte significativa del proceso de indagación del diseñador y, por otra parte, porque se podrán identificar trazas que vinculen una representación en un momento con otra de un momento posterior.

Tanto en el eje de la corporización como en el de la extensión-profundidad no es posible conocer de antemano el orden y formato de representación y modelado a utilizar por los diseñadores. Cada proceso, cada diseñador, cada problema es un universo en sí mismo en este aspecto, como bien pudimos observar en la práctica. Es necesario por lo tanto aceptar que el diseñador puede comenzar en cualquiera de los niveles y proveer mecanismos de traza que permitan relacionar la información entre aquellos cuando se trate de diferentes vistas de una misma idea. El hilo conductor que permite establecer esa conexión es la expresión de la *interacción-en-acción* que se busca conseguir (que es uno de los atributos a considerar en la dimensión de filtrado de la anatomía de prototipos[16]).

Esta forma de pensar la organización de la producción de externalizaciones puede servir como base para la definición de requerimientos de las herramientas que den soporte a la actividad. Para ello,

[1] http://nodejs.org
[2] http://johnny-five.io
[3] http://github.com/firmata/arduino
[4] http://www.ultraleap.com

las herramientas deberán combinar la posibilidad de integrar diferentes formas de externalización y además ser capaces de asegurar la trazabilidad de las características implícitas o emergentes en cada una de ellas hacia otras de distinto nivel de corporización y extensión-profundidad, ofreciendo por ejemplo el uso de lenguajes o vocabularios de interactividad.

Pluma 2.0 Finalmente, se propone una versión de Pluma completamente reformulada que implementa varias de las ideas y requerimientos explorados en secciones precedentes. Pluma 2.0 facilita el bocetado de interacciones enactivas conectando diferentes niveles de manifestaciones y baja el umbral para el bocetado interactivo con hardware.

Esta segunda versión de Pluma está diseñada como un Cuaderno de bocetos enactivos, donde el diseñador pueda desarrollar, incorporar, integrar y evaluar las diferentes manifestaciones del bocetado como quedaron planteadas en el framework de la sección anterior y la estrategia de composición creativa para la ideación. Presenta una biblioteca de componentes de sensado y actuación y otra de vocabularios de interactabilidad, que permiten componer bocetos en un Cuaderno que pueden ser enactuados en todo momento.

Para la evaluación de Pluma 2.0 se utiliza una demostración del espacio de diseño alcanzable, que sigue siendo el método de evaluación más utilizado en la investigación en HCI (como se muestra en un estudio reciente sobre más de 60 toolkits significativas en la historia de HCI[21]). Los espacios de diseño a menudo consisten en dimensiones con propiedades a las que se pueden alinear los ejemplos. Una exploración del espacio de diseño es, por lo tanto, una forma sistemática de tratar de trazar posibles límites de diseño. En el caso de Pluma, realizamos la exploración con una colección de tres ejemplos que replican casos presentados en la literatura: un reproductor de audio controlado con movimientos, un sistema de mensajería con feedback háptico y un browser controlado por gestos.

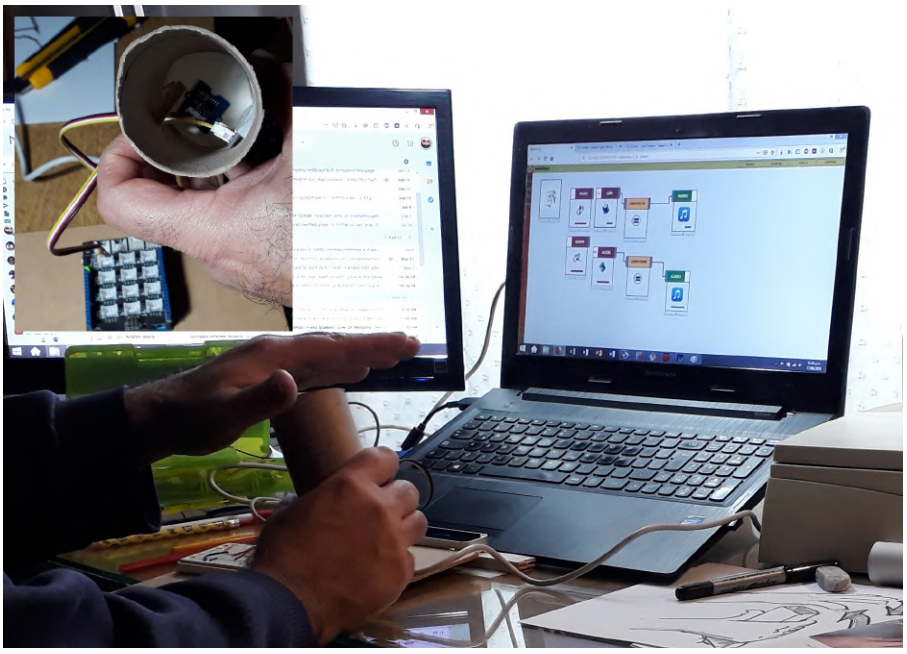


Figura 2: Demostración del caso Reproductor de Audio: construcción y prueba de boceto con hardware (con Arduino, sensor LDR y acelerómetro) y modelo de bloques en Cuaderno de Pluma

En particular, Pluma soporta la exploración de ideas y propuestas por parte del diseñador, permitiéndole trabajar con diferentes versiones de una misma idea, replicar su funcionamiento y re-visitar el camino de indagación que llevó a ellas tantas veces como el diseñador considere necesario.

Pluma baja el umbral al facilitar la modelización con unos pocos clics y *drag and drop* de proxies visuales que le permiten al diseñador disponer también de una guía sobre cómo construir su modelo en hardware. La adopción del sistema Grove[5] con conexiones y componentes estandarizados *cajanegriza* la

[5]http://wiki.seeedstudio.com/Grove_System/

elaboración de circuitos y permite que el diseñador pueda concentrarse en la idea de interactabilidad.

Hemos detectado a su vez un espacio para mejoras necesarias. Por una parte, el avance de ideas en paralelo con bocetos de hardware requiere tantas versiones de la composición como variantes se quiera desarrollar. Será necesario mayor trabajo para facilitar la conexión simultánea de al menos dos o tres variantes en hardware de una idea. Además, hemos intentado mantener el espacio de diseño con paredes amplias y techos altos y facilitar además la extensión mediante sub-clasificación de elementos. Sin embargo, esta sub-clasificación requiere en este momento de la intervención en el código de Pluma.

6. Conclusiones y trabajos futuros

Esta tesis involucra tres campos disciplinares (diseño de sistemas interactivos, cognición corporizada y enactiva, bocetado) cada uno con tradiciones, métodos y principios diferentes por lo que se ha utilizado una aproximación triangular de perspectivas múltiples[11]: a) una teórica, en la cual se estudiaron las teorías y modelos para comprender la Cognición Corporizada y Enactiva y dar forma su relación con el diseño de sistemas interactivos, b) una empírica, con la que se observa y analiza la práctica real de los diseñadores cuando idean dispositivos y sistemas para interactuar en y con el mundo y c) una de diseño, que permite trasladar el análisis de las observaciones y teorías a la construcción de herramientas para soportar el bocetado de interacciones enactivas, lo que a su vez retroalimentó nuevos estudios de observación y análisis teórico.

Se han presentado en detalle los conceptos de la Cognición Corporizada y Enactiva y las consecuencias particulares de cada una de sus vertientes (computación distribuida, práctica social situada y acoplamiento sensorio motor) para el diseño de sistemas interactivos.

Se ha revisado el conocimiento actual sobre la ideación en diseño de interacciones, los roles que el bocetado cumple en ese proceso y se ha efectuado una revisión sistemática de las herramientas propuestas para soportarlo.

Ese background permitió alimentar triangularmente la generación de tres tipos de contribuciones (según la clasificación de Wobborck[19]):

- **Una contribución empírica**: mediante observaciones etnometodológicas se ha obtenido un caracterización del bocetado de interacciones enactivas. Se ha mostrado tanto la forma en que los diseñadores utilizan las ideas de cognición corporizada y enactiva para fundamentar sus procesos de diseño, como los requerimientos a satisfacer por las herramientas de soporte a esos procesos.
- **Una contribución teórica**: un framework que se propone para dar estructura y organización conceptual al bocetado de interacciones enactivas. Clarifica las tres funciones del bocetado en este proceso de ideación y organiza la producción de externalizaciones..
- **Una contribución artefactual**: Pluma, una herramienta para el bocetado con un abordaje de composición creativa que facilita la producción de bocetos en diferentes niveles de corporización, permite una integración entre diversas s formas del boceto que sea manejable por el diseñador, facilita la definición y el diseño de la interactividad mediante el uso de vocabulario de la interacción, mantiene el umbral de acceso a la tecnología lo más bajo posible y permite el amojonamiento del proceso con huellas expresivas que sean re-visitables. La validación inicial de estas herramientas se ha presentado por estudios informales de casos y demostración del espacio de diseño alcanzable.

Para dar continuidad y profundidad a este trabajo será importante extender la serie de observaciones de la práctica del bocetado, incorporando la utilización de herramientas específicas como Pluma y otros abordajes de diseño por fuera de la sustitución sensorial. Tanto estudios de observaciones etnometodológica como el presentado en esta tesis como estudios de casos controlados podrán aportar mayor información sobre la práctica y contribuir a extender y profundizar el framework presentado.

Por otra parte, respecto de Pluma será necesario actualizar su arquitectura tecnológica, aumentar y complejizar formatos de creación y re-visitado de huellas expresivas y realizar experimentos controlados de validación.

Referencias

1. Paul Dourish. *Where the action is. The foundations of embodied interaction*. MIT Press, 2004.
2. Lucy Suchman. *Human-machine reconfigurations. Plans and situated Actions, 2nd edition*. Cambridge University Press, 2007.
3. F J Varela, E Thompson, and E Rosch. *The embodied mind*, volume 6. MIT Press, 1991.
4. Edwin Hutchins. *Cognition in the Wild*, volume 19. MIT Press, 1995.
5. Donald A. Schon. Designing as reflective conversation with the materials of a design situation. *Knowledge-Based Syst.*, 5(1):3–14, mar 1992.
6. Jonas Löwgren and Erik Stolterman. *Thoughtful Interaction Design: A Design Perspective on Information Technology*. MIT Press, 2004.
7. James Jerome Gibson. *The Ecological Approach to Visual Perception*. 1986.
8. Eva Hornecker and Jacob Buur. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, volume 1 of *CHI '06*, pages 437–446, New York, NY, USA, 2006. ACM.
9. Alan Newell and Herbert Simon. *Human problem solving*. Prentice Hall, Englewood Cliffs, 1972.
10. Stuart K. Card, Allen. Newell, Thomas P. Moran, and Allen. Newell. *The psychology of human-computer interaction*. Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1983.
11. W Mackay and A Fayard. HCI, Natural Science and Design: A Framework for Triangulation Across Disciplines. *DIS' 97 Amsterdam*, pages 223–233, 1997.
12. Ilpo Koskinen, John Zimmerman, Thomas Binder, Johan Redström, and Stephan Wensveen. *Constructive Design Research*. 2012.
13. Tom Froese and Shaun Gallagher. Getting interaction theory (IT) together: Integrating developmental, phenomenological, enactive, and dynamical approaches to social interaction. *Interact. Stud.*, 13(3):436–468, 2012.
14. Rosana Guber. *La etnografía. Método, campo y reflexividad*. Siglo XXI Editores, Buenos Aires, 2011.
15. Jelle van Dijk. *Creating Traces, Sharing Insight: Explorations in Embodied Cognition Design*. PhD thesis, 2013.
16. Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. The anatomy of prototypes. *ACM TOCHI*, 15(2):1–27, jul 2008.
17. Lars-Erik Janlert and Erik Stolterman. The Meaning of Interactivity—Some Proposals for Definitions and Measures. *Human–Computer Interact.*, 0024(September), 2016.
18. Toni Robertson. Cooperative Work and Lived Cognition: A Taxonomy of Embodied Actions. In *Proc. Fifth Conf. Eur. Conf. Comput. Coop. Work*, ECSCW'97, pages 205–220, Norwell, MA, USA, 1997. Kluwer Academic Publishers.
19. Jacob O Wobbrock and Julie A Kientz. Research Contributions in Human-computer Interaction. *interactions*, 23(3):38–44, 2016.
20. Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in Human-Computer Interaction*. Morgan Kaufmann Publishers, Cambridge, MA, USA, second edition, 2017.
21. David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. Evaluation Strategies for HCI Toolkit Research. *Proc. 2018 CHI Conf. Hum. Factors Comput. Syst. - CHI '18*, pages 1–17, 2018.

Resumen de Tesis

Medidas de Invarianza y Equivarianza a Transformaciones en Redes Neuronales Convolucionales

Aplicaciones al Reconocimiento de Formas de Mano

Facundo Manuel Quiroga

Laboratorio de Investigación en Informática III-LIDI,
Facultad de Informática, Universidad Nacional de La Plata

Directora	Dra. Laura Cristina Lanzarini
Carrera	Doctorado en Ciencias Informáticas Facultad de Informática, Universidad Nacional de La Plata
Fecha	Defendida el 13 de marzo de 2020
Contacto	{fquiroga}(at)lidi.info.unlp.edu.ar
Documento completo	`http://sedici.unlp.edu.ar/handle/10915/90903`

1. Introducción

Las Redes Neuronales son los modelos de aprendizaje automático con mejor desempeño en la actualidad en una gran variedad de problemas. Son modelos generales y aproximadores universales. Con algoritmos de optimización basados en descenso de gradiente, pueden optimizar miles o millones de parámetros en base a una función de error. Se distinguen de otros modelos en que no requieren un diseño manual de características de los datos para funcionar. Las mismas se aprenden automáticamente mediante el proceso de optimización, también llamado entrenamiento. Su diseño se organiza en capas que determinan su arquitectura. En los últimos años, se ha conseguido entrenar Redes Neuronales con múltiples capas mediante un conjunto de técnicas que suelen denominarse Aprendizaje Profundo (Deep Learning)[GBC16].

En particular, las Redes Convolucionales, es decir, Redes Neuronales que utilizan capas convolucionales, son el estado del arte en la mayoría de los problemas de visión por computadora, incluyendo la clasificación de imágenes. Las capas convolucionales permiten aplicar convoluciones con filtros aprendidos para un mejor desempeño y eficiencia[GBC16].

Muchos de los problemas para los cuales las Redes Convolucionales son el estado del arte requieren que los modelos se comporten de cierta manera ante transformaciones de su entrada. Existen tres propiedades fundamentales que capturan dicho requerimiento: la invarianza, la auto-equivarianza y la equivarianza.

Una red f es invariante a un conjunto de transformaciones si su salida $f(x)$ no cambia cuando una de esas transformaciones se aplica a su entrada x. La auto-equivarianza, un concepto relacionado, nos permite cuantificar que tan equivalente es transformar la entrada x o la salida $f(x)$. Finalmente, la equivarianza generaliza ambos conceptos, y nos permite establecer relaciones entre una transformación de la entrada x y su salida $f(x)$. La medición de estas tres propiedades nos permite comprender mejor la forma en que se comporta una red [GBC16; LV15].

Dotar con invarianza o equivarianza a los modelos tiene utilidades en varios dominios, como la clasificación de imágenes de galaxias, imágenes de microscopios o formas de mano. En particular,

el reconocimiento de formas de mano en imágenes es una de las etapas más importantes de los sistemas de reconocimiento de lenguas de señas o gestos mediante imágenes o video. En muchos casos, la rotación, traslación o escalado de la mano en la imagen no afectan a su forma, y por ende se requiere dotar de invarianza a la red para mejorar el desempeño del sistema [SG18].

La comprensión de la invarianza o equivarianza de una red o cualquier sistema puede ayudar a mejorar su desempeño y robustez. Dado el reciente éxito de las Redes Neuronales, y que son modelos tradicionalmente entendidos como de caja negra, resulta necesario avanzar en una mayor comprensión de su funcionamiento. Este requerimiento tiene especial importancia en el contexto del surgimiento de directivas gubernamentales como la de la Unión Europea que exigen que los modelos de inteligencia artificial puedan explicar sus decisiones [GBC16].

Existen dos opciones principales para otorgar invarianza o equivarianza a un modelo de red neuronal. La tradicional ha sido modificar el modelo para dotarlo de esas propiedades [CW16; Jad+15]. La otra opción es entrenarlo con aumentación de datos utilizando como transformaciones el mismo conjunto al que se desea la invarianza o equivarianza [SG18].

No está claro cómo se adquieren dichas propiedades con aumentación de datos, así como cuál es la mejor estrategia para aumentar los datos con este fin [SG18]. Tampoco está claro como las modificaciones de los modelos afectan la eficiencia y el poder de representación de los mismos.

Con este fin, pueden definirse métricas de invarianza y equivarianza para analizar a las redes neuronales desde una perspectiva transformacional [Goo+09; LV15]. Dichas métricas permiten, por ejemplo, entender como se comporta un sistema de reconocimiento de rostros ante transformaciones afines de una imagen del rostro. Para los sistemas que analizan el riesgo crediticio, dada una transformación del género del solicitante, permiten evaluar si dicho atributo influye en la decisión final. Para los sistemas que analizan texto permite evaluar si el significado predicho varía en base a una transformación que, por ejemplo, reemplace palabras por sus sinónimos.

La invarianza y equivarianza pueden estimarse midiendo los cambios en la red en base a las transformaciones realizadas a su entrada. Las metodologías actuales de evaluación y comprensión de la invarianza y equivarianza se enfocan solamente en los cambios detectados en la salida de la red. No obstante, para poder comprender como se codifican estas propiedades, el análisis debe realizarse en base a toda la red, es decir, considerando las representaciones intermedias. No obstante, las métricas de este tipo presentadas previamente han sido inadeacuadas para evaluar estas propiedades [Goo+09; LV15].

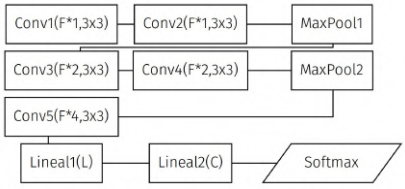
En resumen, si bien hay varios métodos propuestos para otorgar equivarianzas a las redes, y en consecuencia aumentar su robustez, no existe ninguna comparación exhaustiva entre los distintos métodos, y lo que es aún más acruciante, existen pocos análisis en profundidad sobre como dichos métodos codifican las equivarianzas. Comprender el funcionamiento transformacional de las redes ayuda a mejorar los sistemas existentes. No obstante, para ello se requieren herramientas para analizar las redes, como las métricas de invarianza y equivarianza. De estas interrogantes surgen los objetivos de investigación de esta tesis.

2. Objetivos, contribuciones y publicaciones

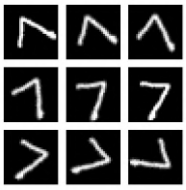
2.1. Objetivos

Nuestro objetivo general en esta tesis es contribuir al entendimiento y mejora de la equivarianza de los modelos de redes neuronales, en particular aplicados a la clasificación de formas de mano para la lengua de seña y otros tipos de gestos mediante modelos de redes convolucionales.

Para ello, establecimos los siguientes objetivos particulares:



(a)



(b)

Figura 1: (a) Modelo SimpleConv (b) Ejemplos rotados de la base de datos MNIST.

1. Analizar los modelos específicos para equivarianza en CNNs. Comparar los modelos específicos y la aumentación de datos para obtener equivarianza. Evaluar estrategias de transferencia de aprendizaje para obtener modelos equivariantes a partir de modelos que no lo son.

2. Desarrollar métricas de equivarianza para las activaciones o representaciones internas de las redes neuronales. Implementar las métricas en una librería de código abierto. Analizar el comportamiento de las métricas. Comparar con las métricas existentes.

3. Caracterizar modelos de CNN para la clasificación de imágenes en términos de su equivarianza con las métricas propuestas. Comparar modelos de CNN, con y sin equivarianza, para la clasificación de formas de mano.

Debido a la existencia de múltiples métodos para lograr equivarianza, y a la falta de comparaciones rigurosas y profundas de los mismos, el alcance de esta tesis se limita al análisis y no propone nuevos modelos equivariantes de redes.

2.2. Publicaciones

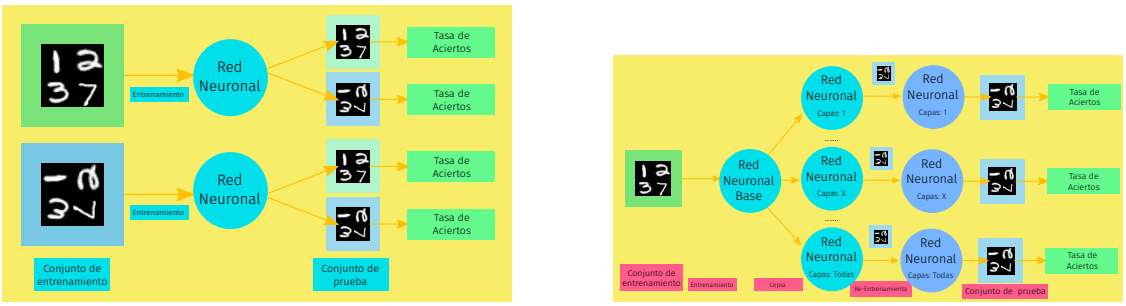
Las siguientes publicaciones están relacionadas directamente con las contribuciones que presentamos en esta tesis: [Qui+17; Qui+18; Qui+19]. Además, esta línea de investigación fue motivada por trabajos relacionados realizados en la primera etapa del doctorado [Ron+15; Qui+16a; Qui+16c; Qui+16b].

Adicionalmente, durante el transcurso de esta tesis creamos **Métricas transformacionales** [1], una librería de código abierto para computar las métricas propuestas. También creamos una librería de código abierto llamada **Handshape Datasets** [2] para facilitar la descarga y el preprocesamiento de conjuntos de datos de formas de mano.

3. Resumen de contribuciones

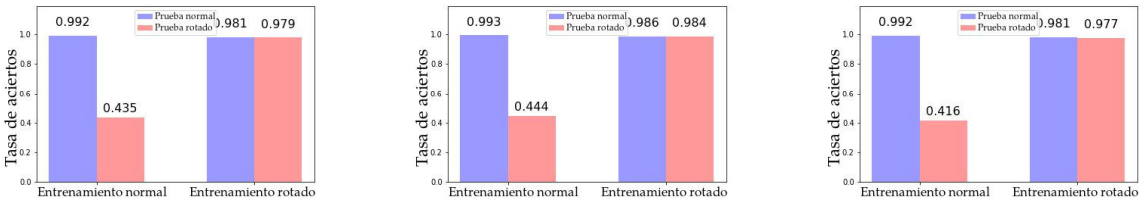
En esta sección, presentamos de forma resumida las contribuciones principales de la tesis. En términos de los experimentos, si bien hemos utilizado diversos modelos, transformaciones y bases de datos, **dadas las limitaciones de página** sólo mostraremos resultados para el modelo SimpleConv, la base de datos MNIST y rotaciones de 0° a 360° como conjunto de transformaciones (Figura 1). Por el mismo motivo, tampoco presentaremos el trabajo realizado sobre la clasificación de formas de mano, el tercer aporte de la tesis.

[1] https://github.com/facundoq/transformational_measures
[2] https://github.com/midusi/handshape_datasets



(a) (b)

Figura 2: (a) Metodología del experimento de comparación de modelos. (b) Metodología del experimento de reentrenamiento de capas específicas.



(a) Aumentación de datos (b) GCNN (c) STN

Figura 3: Resultados del experimento para el modelo base SimpleConv en la base de datos MNIST

3.1. Comparación de Modelos Invariantes con Aumentación de Datos

En el **primer aporte** de esta tesis, analizamos diversas estrategias para obtener invarianza en modelos de clasificación de imágenes con redes neuronales convolucionales.

3.1.1. Comparación de modelos invariantes

Comparamos los modelos tradicionales de CNN con aumentación de datos contra los modelos especializados Group CNN (GCNN) [CW16] y Spatial Transformer Networks (STN) [Jad+15] para determinar su desempeño, utilizando la base de datos MNIST.

Para ello, entrenamos un modelo con y sin aumentación de datos, y luego lo evaluamos también con y sin aumentación de datos, dando lugar a 4 combinaciones independientes (Figura 2 (a)).

Los resultados (Figura 3) indican que aún con ingeniosas modificaciones de las redes convolucionales, la aumentación de datos sigue siendo necesaria para obtener un desempeño similar al de los modelos no invariantes. Más aún, en varios casos la aumentación de datos por si sola puede proveer un desempeño similar al de los modelos especializados, siendo al mismo tiempo más simples de entrenar y comprender.

3.1.2. Análisis del re-entrenamiento de capas específicas para obtener invarianza

La utilización de redes pre-entrenadas es una técnica usual para mejorar el desempeño de los modelos. No obstante, al reentrenar un modelo para obtener invarianza no está claro a priori si toda la red puede o necesita ser re-entrenada, o si solo algunas capas requieren adaptarse.

Para determinarlo, entrenamos un modelo base con una BD sin rotar, y luego hacemos copias de la red. En cada copia re-entrenamos algún subconjunto de sus capas con el conjunto de entrenamiento y aumentación de datos. Luego, evaluamos el desempeño de la red con alguna de sus capas re-entrenadas (Figura 2 (b)).

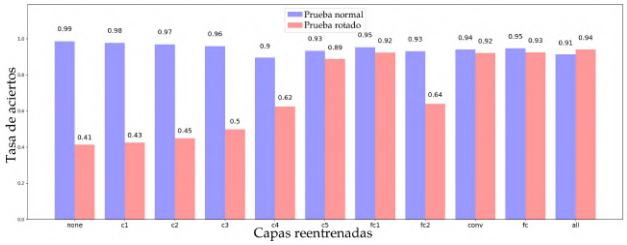


Figura 4: Tasa de acierto del modelo luego de re-entrenar una capa o conjunto de capas, para la versión rotada y la versión original de la BD. El eje x indica que capas re-entrenamos. Las etiqueta *conv* y *fc* denotan todas las capas convolucionales y lineales, respectivamente.

La Figura 4 muestra las tasas de aciertos obtenidas al re-entrenar algunos subconjuntos de capas. El re-entrenamiento para obtener invarianza muestra una dinámica similar a re-entrenamiento para hacer transferencia de aprendizaje: las capas finales, con características de alto nivel, son una mejor elección para re-entrenar de forma individual dado que están más cerca de la salida y pueden afectarla más. Esto indica que hay una redundancia de información en las capas anteriores, debido a que las versiones rotadas de las imágenes pueden codificarse con éxito a partir de la salida de dichas capas. Sorprendentemente, se puede también re-entrenar *solamente* las capas convolucionales para obtener invarianza a niveles originales. De esta forma, estos experimentos revelan que la codificación de la invarianza puede darse de distintas formas. Esto sugiere que es importante entender como se codifica la invarianza en las distintas capas, para lo cual desarrollamos las métricas de invarianza y equivarianza que describimos a continuación.

3.2. Métricas de Invarianza y Equivarianza

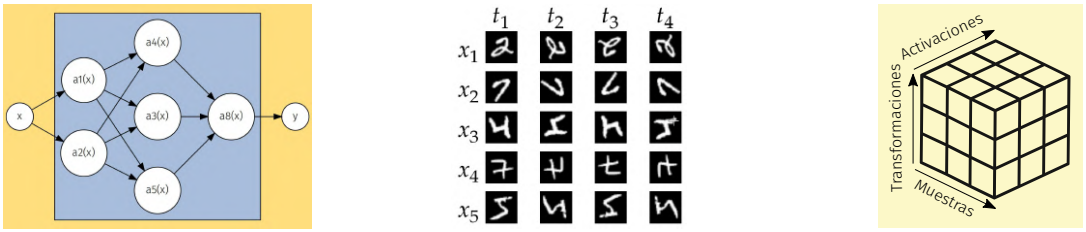
Como **segundo y principal aporte**, proponemos un nuevo conjunto de métricas para medir la invarianza y auto-equivarianza de las redes. Dichas métricas permiten cuantificar estas propiedades de forma empírica y no solo en la salida de la red sino también en sus activaciones o representaciones internas. De esta forma, podemos visualizar y cuantificar que tan invariante o auto-equivariante es una red, ya sea en su totalidad, por capas, o por activaciones individuales. Las métricas son aplicables a cualquier red neuronal, sin importar su diseño o arquitectura, así como a cualquier conjunto de transformaciones. Todas las métricas han sido implementadas y están disponibles públicamente en `https://github.com/facundoq/transformational_measures`. Para más detalles, consultar [Qui20].

3.2.1. Definiciones básicas

Dada una entrada x, un modelo de redes neuronales computa varios valores intermedios u ocultos, denotados como $a_1(x), \ldots, a_k(x)$. Por simplicidad, llamaremos a estos valores $a_i(x)$ *activaciones* de la red f (Figura 5 (a)). Para medir la equivarianza de un modelo f, medimos la equivarianza de las activaciones individuales $a_1(x) \ldots a_k(x)$. Dado que la métrica puede ser definida para una activación de forma independiente del resto, en lo siguiente nos referiremos a una activación simplemente como $a(x)$, sin su índice.

Dado un conjunto de n muestras $X = [\, x_1 \cdots x_n \,]$ y un conjunto de m transformaciones $T = [\, t_1 \cdots t_m \,]$ definidas sobre X, podemos calcular todas las posibles transformaciones de todas las muestras, y organizarlas en una matriz de muestras-transformaciones llamada $\mathbf{MT}(a, X, T)$ (Figura 5 (b)).

$$\mathbf{MT}(a, X, T)_{i,j} = a(t_j(x_i)) \quad [1]$$



(a) (b) (c)

Figura 5: (a) Activaciones de un modelo. (b) Muestras y transformaciones. (c) Tensor de activaciones, muestras y transformaciones.

3.2.2. Métricas de Invarianza basadas en Varianza

Podemos cuantificar la invarianza de una activación *a* midiendo su varianza. Debido a que existen 2 fuentes de variabilidad, transformaciones y muestras, medimos la varianza transformacional y muestral.

Varianza Transformacional Si una activación es invariante a t, entonces $f(t(x)) = t(x)$. Por ende, para un conjunto de transformaciones invariante $T = [t_1, \dots, t_m]$, tenemos $f(t_1(x)) = f(t_2(x)) = \cdots = f(t_m(x))$. Podemos calcular las desviaciones de la invarianza midiendo la varianza de los valores $f(t_1(x)), f(t_2(x)), \dots, f(t_m(x))$

La métrica de Varianza Transformacional (VT) se define entonces según la Ecuación [2], y su cómputo se ejemplifica con la Figura 6. Esta métrica mide la varianza en las activaciones con respecto a todas las transformaciones, para cada ejemplo, y luego promedia esas varianzas con respecto a todos los ejemplos. Si $VT(a) = 0$, entonces la activación *a* no varía con las transformaciones, y por ende es invariante.

$$VT(a) = Media\left(\begin{bmatrix} Var(\mathbf{MT}(a)[1,:]) \\ \cdots \\ Var(\mathbf{MT}(a)[n,:]) \end{bmatrix}\right) \quad [2]$$

Varianza Muestral Los valores de la métrica VT pueden observarse en la Figura 7 (a). Allí podemos observar que las unidades de la métrica VT dependen de la activación y capa. Por ende, utilizaremos la Varianza Muestral (VM) para normalizar la VT. La métrica VM es la transpuesta conceptual de la VT y se define de forma análoga (Ecuación [3]).

$$VM(A) = Media\left(\begin{bmatrix} Var(\mathbf{MT}[:,1]) & \cdots & Var(\mathbf{MT}(a)[:,m]) \end{bmatrix}\right) \quad [3]$$

Varianza Normalizada Combinando las Ecuaciones [2] y [3], podemos definir la métrica de Invarianza basada en Varianza Normalizada (VN), como la relación entre las otras dos métricas (Ecuación [4]).

$$VN(a) = \frac{VT(a)}{VM(a)} \quad [4]$$

La figura Figura 7 (c) muestra el resultado de la VN. Podemos ver que la escala difiere significativamente entre VT y VM, y por ende es necesaria la normalización que ofrece VN.

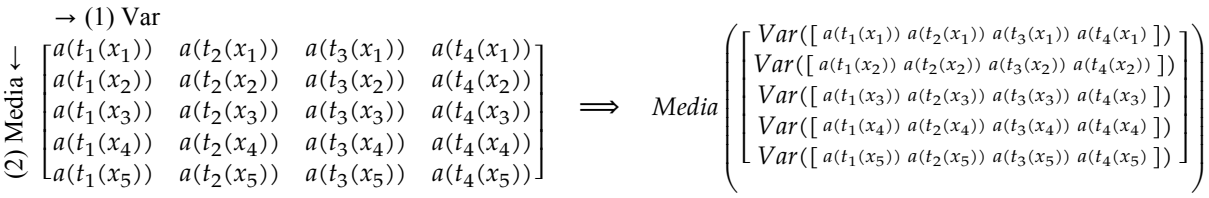


Figura 6: Cálculo de la Varianza Transformacional (VT).

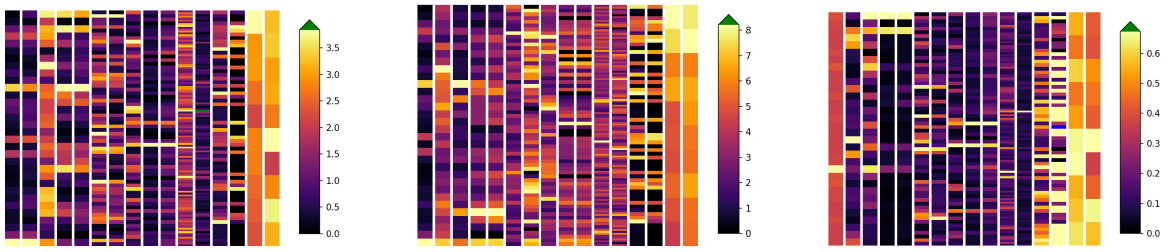


Figura 7: (a) VT, (b) VM y (c) VN de cada activación de un modelo SimpleConv, evaluado con rotaciones en MNIST.

Casos especiales En los casos donde tanto $VT(a) = 0$ como $VM(a) = 0$, definimos $VN(a) = 1$. Este caso corresponde a la definición de una activación *muerta*, que no responde a ningún patrón y por ende no tiene ningún uso en la red. En los casos donde $VT(a) > 0$ y $VM(a) = 0$, definimos $VN(a) = +\infty$. Este caso es similar al primero, pero ahora las transformaciones *si* causan variabilidad en la activación. Esto posiblemente se deba a que generan muestras muy por fuera de la distribución original de los datos.

Valores de la Varianza Normalizada Podemos analizar los valores posibles de VN y su interpretación. Si $VN(a) = 0$, entonces $VT(a) = 0$ y la activación es claramente invariante. Si $VN(a) < 1$, la varianza debido a las transformaciones es menor que la varianza debido a las muestras, y por eso podemos considerar aproximadamente invariante a la activación. Si $VN(a) > 1$ se aplica el mismo razonamiento pero con la conclusión opuesta. Si $VN(a) \simeq 1$, entonces ambas varianzas están en equilibrio.

3.2.3. Métricas de Auto-Equivarianza basada en Varianza

La auto-equivarianza es un caso especial de la equivarianza que nos permite analizar si al aplicar una transformación a las entradas, las activaciones se transforman de la misma manera: formalmente, $t(f(x)) = f(t(x))$ (Figura 8 (a)). Si asumimos que t tiene una inversa, entonces esto es equivalente a $f(x) = t^{-1}(f(t(x)))$ (Figura 8 (a)). Para un conjunto de transformaciones $T = [t_1, \ldots, t_m]$, esto implica que $t_1^{-1}(f(t_1(x))) = \cdots = t_m^{-1}(f(t_m(x)))$ (Figura 8 (c,d)).

Entonces, podemos definir las métricas de Auto-Equivarianza basadas en Varianza de manera similar a las métricas de Invarianza basadas en Varianza. Las métricas de Invarianza se basan en comparar los valores $a(t_1(x)), \ldots, a(t_m(x))$. En el caso de la Auto-Equivarianza, podemos en lugar de eso comparar los valores $t_1^{-1}(a(t_1(x))), \ldots, t_m^{-1}(a(t_m(x)))$. Para una definición completa de estas métricas, consultar [Qui20].

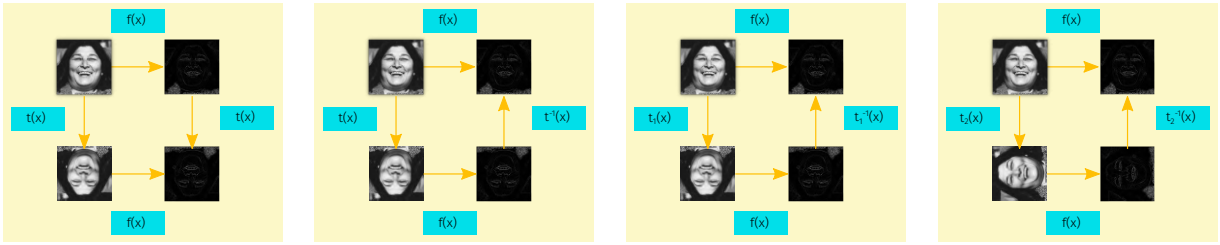


Figura 8: (a) Auto-equivarianza (b) Auto-equivarianza con t invertible. (c,d) Valores a comparar para distintas transformaciones.

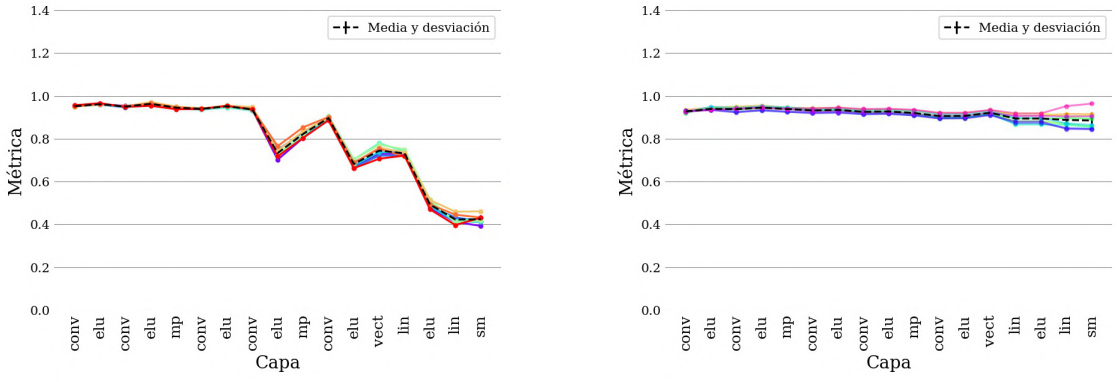


Figura 9: (a) Inicialización aleatoria (b) Pesos aleatorios.

3.3. Experimentos con Métricas de Invarianza

En esta sección, resaltamos 2 de los más de 30 experimentos[3] realizados con las métricas definidas y que nos permiten caracterizar el comportamiento de las redes neuronales en términos de su invarianza y equivarianza.

Nuevamente aclaramos que sólo mostraremos los resultados para SimpleConv, MNIST, rotaciones, y la métrica de Invarianza basada en Varianza Normalizada. Para una descripción completa de estos y otros experimentos, consultar [Qui20].

Invarianza de modelos inicializados aleatoriamente Para estudiar el efecto de la inicialización aleatoria del modelo en las invarianza, entrenamos 10 instancias de cada modelo utilizando la misma arquitectura, base de datos y conjunto de transformaciones, hasta que las mismas converjan. Cada instancia del mismo modelo fue inicializada con distintos pesos aleatorios. Luego, todas las instancias fueron evaluadas con la misma métrica.

La Figura 9 (a) muestra el valor promedio de la métrica para cada capa. Los resultados sugieren que la invarianza no varía con respecto a la inicialización de sus pesos. Esto indica que la invarianza es una propiedad emergente del modelo, sin importar la inicialización aleatoria de sus pesos. Por ende, la misma depende mayormente de la arquitectura del modelo, las muestras y las transformaciones, y no el proceso de entrenamiento.

Además, la estabilidad de las métricas con respecto a la inicialización del modelo es una propiedad deseable, ya que implica que no se requiere entrenar varias instancias de un modelo y medirlos para obtener un valor de invarianza promedio del mismo.

[3]El código de los experimentos está disponible en `https://github.com/facundoq/transformational_measures_experiments/`

Invarianza de modelos con pesos aleatorios Para comprender aún mejor el origen de la estructura del invarianza de las redes, medimos dicha propiedad en modelos sin entrenar, es decir, modelos con pesos completamente aleatorios. Dado que la inicialización es aleatoria, calculamos la métrica para ocho modelos aleatorios en cada caso (Figura 9 (b)).

Nuevamente, podemos ver que la invarianza no depende del conjunto de pesos aleatorio de la red. Esto sugiere que dada una inicialización controlada, la invarianza es mayormente una propiedad de la arquitectura de la red en lugar de los valores específicos de los pesos. Por otro lado, es interesante que la invarianza de las redes sin entrenar no tiene tendencia, es decir no varía en base a las capas, a diferencia de las ya entrenadas (Figura 9 (a)).

4. Conclusiones

La invarianza y la equivarianza son conceptos clave para comprender el funcionamiento de las redes neuronales.

Presentamos una serie de métricas flexibles, eficientes e interpretables para cuantificar la invarianza y auto-equivarianza de cualquier modelo basado en Redes Neuronales. Nuestras métricas son completamente adaptables en términos de la arquitectura de la red, tipos de capas, conjuntos de transformaciones y formatos de entrada. Las métricas tienen una granularidad alta, permitiendo medir éstas propiedades en cada activación de la red de forma individual, y luego combinar dichos resultados para evaluar estructuras de más alto nivel como capas o redes enteras. A diferencia de otras métricas, son fáciles de interpretar y eficientes para calcular en redes grandes con millones de activaciones. Además, desarrollamos una librería de código abierto que provee implementaciones de referencia de las métricas desarrolladas.

Los experimentos realizados nos permiten realizar un análisis de la invarianza y la equivarianza de los modelos y sus características. Las conclusiones pueden ser utilizadas para crear mejores modelos equivariantes. Además del análisis, las comparaciones con aumentación de datos y las métricas presentan en conjunto una metodología poderosa para evaluar la equivarianza de los modelos y así comprenderlos más completamente.

4.1. Trabajos futuros

Creemos que es posible aprender más acerca de las Redes Neuronales estudiando sus equivarianzas, y así mejorar los modelos existentes para hacer posible nuevas aplicaciones. Para avanzar en esta dirección, identificamos tres áreas de trabajo.

Primero, proponemos mejorar la facilidad de uso de la librería de métricas, implementando soporte para el framework TensorFlow, aumentando la eficiencia en el cálculo de las métricas y permitiendo su ejecución en GPU o CPU de forma indistinta. Segundo, buscaremos profundizar la caracterización teórica de las métricas. Por un lado, desarrollando pruebas de hipótesis específicos para cada métrica. Por el otro, generalizando las métricas para obtener una métrica de equivarianza general. Tercero y último, buscaremos ampliar el análisis de los modelos, incluyendo más modelos invariantes (GCNN, capsules, etc), distintos tipos de modelos (recurrentes, GANs, etc), y de otros dominios, como procesamiento de texto, audio, o ejemplos adversariales.

Referencias

[CW16] Taco S. Cohen y Max Welling. «Group Equivariant Convolutional Networks». En: *arXiv:1602.07576 [cs, stat]* (feb. de 2016). arXiv: 1602.07576. url: http://arxiv.org/abs/1602.07576 (visitado 08-05-2018).

[GBC16] Ian Goodfellow, Yoshua Bengio y Aaron Courville. *Deep Learning*. MIT Press, 2016. url: http://www.deeplearningbook.org.

[Goo+09] Ian Goodfellow, Honglak Lee, Quoc V Le, Andrew Saxe y Andrew Y Ng. «Measuring invariances in deep networks». En: *Advances in neural information processing systems*. 2009, págs. 646-654.

[Jad+15] Max Jaderberg, Karen Simonyan, Andrew Zisserman y Koray Kavukcuoglu. «Spatial Transformer Networks». En: *arXiv:1506.02025 [cs]* (jun. de 2015). arXiv: 1506.02025. url: http://arxiv.org/abs/1506.02025 (visitado 08-05-2018).

[LV15] Karel Lenc y Andrea Vedaldi. «Understanding image representations by measuring their equivariance and equivalence». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, págs. 991-999.

[Qui+16a] Facundo Quiroga, Franco Ronchetti, César Armando Estrebou y Laura Cristina Lanzarini. «Handshape recognition for argentinian sign language using probsom». En: *Journal of Computer Science & Technology* 16.1 (2016). issn: 1666-6038.

[Qui+16b] Facundo Quiroga, Franco Ronchetti, César Armando Estrebou, Laura Cristina Lanzarini y Alejandro Rosete. «LSA64: An Argentinian Sign Language Dataset». En: *XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016)*. Ed. por Springer. Red de Universidades con Carreras en Informática (RedUNCI). 2016, págs. 794-803.

[Qui+16c] Facundo Quiroga, Franco Ronchetti, César Estrebou, Laura Lanzarini y Alejandro Rosete. «Sign languague recognition without frame-sequencing constraints: A proof of concept on the argentinian sign language». En: *Ibero-American Conference on Artificial Intelligence*. Springer International Publishing. 2016, págs. 338-349.

[Qui+17] Facundo Quiroga, Ramiro Antonio, Franco Ronchetti, Laura Cristina Lanzarini y Alejandro Rosete. «A study of convolutional architectures for handshape recognition applied to sign language». En: *XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017)*. Springer International Publishing. 2017.

[Qui+18] Facundo Quiroga, Franco Ronchetti, Laura Lanzarini y Aurelio F Bariviera. «Revisiting Data Augmentation for Rotational Invariance in Convolutional Neural Networks». En: *International Conference on Modelling and Simulation in Management Sciences*. Springer. 2018, págs. 127-141.

[Qui+19] Facundo Quiroga, Jordina Torrents-Barrena, Laura Lanzarini y Domenec Puig. «Measuring (in) variances in Convolutional Networks». En: *Conference on Cloud Computing and Big Data*. Springer. 2019, págs. 98-109.

[Qui20] Facundo Quiroga. «Medidas de Invarianza y Equivarianza a Transformaciones en Redes Neuronales Convolucionales.» Tesis doct. La Plata, Buenos Aires, Argentina: Universidad Nacional de La Plata, mar. de 2020.

[Ron+15] Franco Ronchetti, Facundo Quiroga, Laura Lanzarini y Cesar Estrebou. «Distribution of Action Movements (DAM): a Descriptor for Human Action Recognition». En: *Frontiers of Computer Science* 9.6 (2015), págs. 956-965. issn: 2095-2236. doi: 10.1007/s11704-015-4320-x. url: http://dx.doi.org/10.1007/s11704-015-4320-x.

[SG18] Megha Srivastava y Kalanit Grill-Spector. «The Effect of Learning Strategy versus Inherent Architecture Properties on the Ability of Convolutional Neural Networks to Develop Transformation Invariance». En: *CoRR* abs/1810.13128 (2018). arXiv: 1810.13128. url: http://arxiv.org/abs/1810.13128.

SEDAR: Detección y Recuperación Automática de Fallos Transitorios en Sistemas de Cómputo de Altas Prestaciones

Diego Montezanti
Directores por UNLP: Armando De Giusti y Marcelo Naiouf
Directores por UAB: Dolores Rexachs del Rosario y Emilio Luque Fadón
Fecha de exposición: 18/03/2020
Facultad de Informática, Universidad Nacional de La Plata, Argentina
dmontezanti@lidi.info.unlp.edu.ar

1. Introducción

En primer lugar, se presenta la motivación de esta tesis. Luego se enuncian los objetivos y finalmente, se describen las contribuciones y limitaciones de esta investigación.

1.1 Motivación

La confiabilidad y, como consecuencia, la Tolerancia a Fallos (TF) se han vuelto aspectos críticos de relevancia creciente, en especial en el ámbito del Cómputo de Altas Prestaciones (HPC, por sus siglas en inglés), debido al considerable incremento en la probabilidad de que ocurran fallos de diferentes clases en estos sistemas [1]. Este incremento es consecuencia, fundamentalmente, de dos factores:

a) La creciente complejidad de los procesadores, en la búsqueda de mejorar la potencia computacional de los sistemas. Esta complejidad está dada por el aumento en la escala de integración y en la cantidad de sus componentes. Estos componentes, trabajando cerca de sus límites tecnológicos, son cada vez más propensos a la ocurrencia de fallos.

b) El rápido aumento del tamaño de los sistemas paralelos, entendido como la cantidad de nodos de procesamiento que lo componen, y la cantidad de cores por *socket*, de forma obtener mejores prestaciones en la resolución de problemas demandantes de potencia de cómputo.

En relación con estos aspectos de complejidad y tamaño, actualmente la escena de las supercomputadoras es dominada por los *clusters* de multiprocesadores, interconectados por redes de comunicaciones de alta velocidad, habiendo crecido en popularidad la utilización de dispositivos aceleradores, que proporcionan mejoras notables tanto en los tiempos de ejecución como en el consumo energético, si se los compara con los sistemas basados únicamente en las CPUs de propósito general [2].

A medida que las aplicaciones demandan mayores tiempos de cómputo ininterrumpido, la influencia de los fallos se vuelve más significativa, debido al costo que requiere relanzar una ejecución que fue abortada por la ocurrencia de un fallo, o, peor aún, que finaliza con resultados erróneos. Si las aplicaciones (y sus resultados) son relevantes, o incluso críticos, es necesario ejecutarlas sobre sistemas tanto de alta disponibilidad (es decir, que se mantienen funcionando un alto porcentaje de tiempo) como de alta fiabilidad (es decir, cuyo comportamiento no se aparta del especificado, proporcionando, por lo tanto, resultados correctos [3]). Las estrategias de tolerancia a fallos, capaces de proveer detección, protección y recuperación frente a fallos, se emplean en la búsqueda de obtener sistemas paralelos altamente disponibles y fiables.

En los próximos años está previsto alcanzar la era de la Exa-escala, en la que se construyan supercomputadoras formadas por millones de núcleos de procesamiento, y capaces de realizar en el orden de 10^{18} cálculos por segundo (1 Exaflop). Esto constituye una gran ventana de oportunidad para

las aplicaciones de HPC, aunque también se incrementa el peligro de que no completen sus ejecuciones. En el ámbito de HPC, el riesgo de fallos transitorios aumenta con la cantidad de procesadores que trabajan juntos [4]. Por lo tanto, el peligro de obtener datos y resultados corrompidos silenciosamente es directamente proporcional a la potencia de cómputo que se logra agregando más procesadores al sistema. Estudios recientes muestran que, a medida de que los sistemas de HPC continúan creciendo e incluyendo más componentes de hardware de distintos tipos, el Tiempo Medio Entre Averías (MTBF, por sus siglas en inglés) para una aplicación determinada disminuye, resultando en una tasa de fallos más alta en general [5]. Por lo tanto, parece claro que los sistemas futuros de la Exa-escala presentarán altas tasas de errores: está previsto que las grandes aplicaciones paralelas tengan que lidiar con fallos que ocurran cada pocos minutos, por lo que requerirán ayuda externa para progresar eficientemente [6]. Algunos reportes recientes han señalado a la Corrupción Silenciosa de Datos (SDC, por sus siglas en inglés) como el tipo de fallo más peligroso que se puede presentar durante la ejecución, ya que corrompen bits de la caché, de la memoria principal o de los registros de la CPU, produciendo que el programa finalice con resultados incorrectos, aunque en apariencia se ejecuta correctamente. Las aplicaciones científicas y las simulaciones a gran escala son las áreas más afectadas de la computación; en consecuencia, el tratamiento de los errores silenciosos es uno de los mayores desafíos en el camino hacia la resiliencia en los sistemas de HPC [21]. En aplicaciones de paso de mensajes, y sin mecanismos de tolerancia apropiados, un fallo silencioso (que afecte a una única tarea) puede producir profundos efectos de corrupción de datos, causando un patrón de propagación en cascada, a través de los mensajes, hacia todos los procesos que se comunican; en el peor escenario, los resultados finales erróneos no podrán ser detectados al finalizar la ejecución y serán tomados como correctos [20].

Dado que las aplicaciones científicas, intensivas en cómputo, presentan tiempos de ejecución prolongados, usualmente del orden de horas o días, resulta imprescindible encontrar estrategias de tolerancia a fallos que permitan que las aplicaciones se completen de forma satisfactoria, alcanzando una solución correcta, en un tiempo finito y de una forma eficiente a pesar de los fallos subyacentes del sistema. La utilización de estas estrategias tendrá además por efecto evitar que se dispare el consumo de energía, ya que de no utilizar ningún mecanismo de tolerancia a fallos, la ejecución tendría que volver a comenzar desde el principio. A pesar de esto, los modelos de programación paralela más populares que se utilizan para explotar el poder de cómputo proporcionado por supercomputadoras carecen de soporte de tolerancia a fallos [2].

1.2 Objetivos

En este contexto, que conjuga altas tasas de fallos, resultados no fiables y altos costos de verificación, el objetivo de esta tesis es ayudar a los programadores y a los científicos que ejecutan aplicaciones críticas, o al menos relevantes, a proporcionar fiabilidad a sus resultados, dentro de un tiempo predeciblemente acotado.

Con este objetivo, en los últimos años se ha diseñado y desarrollado la metodología **SEDAR** (*Soft Error Detection and Automatic Recovery* [7,8], a partir de una estrategia de detección llamada SMCV [9,10,11]), cuyo objetivo es proveer tolerancia a fallos transitorios en sistemas formados por aplicaciones paralelas que utilizan paso de mensajes y se ejecutan en *clusters* de *multicores*. **SEDAR** es una solución basada en replicación de procesos [12,22] y monitorización de los envíos de mensajes y el cómputo local, que aprovecha la redundancia de hardware que es intrínseca a los *multicores*.

SEDAR proporciona tres variantes para lograr la tolerancia a fallos: un mecanismo de detección y relanzamiento automático desde el comienzo de la aplicación; un mecanismo de recuperación automática, basada en el almacenamiento de múltiples *checkpoints* de nivel de sistema (ya sean

periódicos o disparados por eventos); y un mecanismo de recuperación automática, basado en un único *checkpoint* seguro de capa de aplicación. El objetivo principal de esta tesis es la descripción de la metodología y de sus pautas de diseño, y la validación de su eficacia para detectar los fallos transitorios y recuperar automáticamente las ejecuciones, mediante un modelo analítico de verificación. Si bien esta validación es esencialmente funcional, también se realiza una implementación práctica de un prototipo funcional (utilizando herramientas existentes y conocidas), y, a partir de las pruebas realizadas sobre ella, se caracteriza también el comportamiento temporal, es decir, la *performance* y el *overhead* introducido por cada una de las tres alternativas. Además se busca mostrar que existe la posibilidad de optar, incluso dinámicamente, por la alternativa que resulte más conveniente para adaptarse a los requerimientos de un sistema (por ejemplo, máximo *overhead* permitido o máximo tiempo de finalización), convirtiendo a **SEDAR** en una metodología eficaz, viable y flexible para la tolerancia a fallos transitorios en sistemas de HPC. A diferencia de las estrategias específicas, que proporcionan resiliencia para ciertas aplicaciones, a costa de modificarlas, y que no cubren todos los fallos [6,13], **SEDAR** es esencialmente transparente y agnóstico respecto del algoritmo al cual protege.

1.3 Contribuciones y limitaciones

Durante el desarrollo de la tesis hemos arribado a una metodología completa y funcionalmente válida, que contempla todos los escenarios posibles de fallos, y a un prototipo de un sistema automático capaz de recuperar sin intervención del usuario, garantizando así la fiabilidad de los resultados dentro de un tiempo factible de ser acotado.

Las principales contribuciones de este trabajo son:

- El desarrollo completo de una metodología de tolerancia a fallos, que integra la duplicación (efectiva para detectar los fallos transitorios), con el *checkpoint & restart* que se utiliza normalmente para garantizar disponibilidad frente a fallos permanentes, obteniendo, a partir de esta combinación, una estrategia que asegura tanto la finalización como la fiabilidad de los resultados.
- La descripción y verificación del comportamiento funcional en presencia de fallos, demostrando la eficacia de la estrategia de detección y la validez del mecanismo de recuperación basado en múltiples *checkpoints* coordinados de nivel de sistema.
- La comprobación empírica de las predicciones del modelo, por medio de inyección controlada de fallos, que evidencia la fiabilidad provista por las estrategias de **SEDAR**.
- La herramienta **SEDAR**, que implementa el mecanismo de detección y un algoritmo de recuperación basado en múltiples *checkpoints*, y el trabajo experimental realizado para incorporar **SEDAR** a aplicaciones paralelas.
- La caracterización temporal y la evaluación de los *overheads* introducidos para cada una de las tres estrategias alternativas de **SEDAR**, mostrando los beneficios que ofrecen las distintas variantes, tanto en el tiempo de ejecución como en confiabilidad de los resultados.
- La evidencia de la flexibilidad de **SEDAR** para adaptarse de forma de alcanzar un determinado compromiso entre costo y desempeño obtenido.
- La determinación de la relación entre la cantidad de recursos requeridos para implementar la estrategia y la ganancia de tiempo obtenida junto con la fiabilidad de la ejecución, mostrando la viabilidad y la eficacia de **SEDAR** para tolerar los fallos transitorios en sistemas de HPC.

Entre las principales limitaciones de **SEDAR**, su aplicabilidad está restringida a los sistemas de memoria distribuida, o, mejor dicho, a los sistemas que ejecutan aplicaciones de paso de mensajes, a causa de que la validación de los mensajes es el fundamento del mecanismo de detección. Otra

limitación es que, actualmente, **SEDAR** funciona únicamente para aplicaciones paralelas determinísticas, ya que sólo en esas aplicaciones se puede garantizar que el cómputo, sobre los mismos datos de entrada, por parte de dos réplicas, arroja los mismos resultados, que son los que se comparan para detectar los fallos. El modelo actual de caracterización temporal no permite predecir la respuesta cuando suceden dos o más errores distintos. El mecanismo de recuperación basado en múltiples *checkpoints*, en su implementación actual, no soporta de manera óptima la ocurrencia de varios fallos no relacionados, por lo que requiere ser refinado. Finalmente, el desarrollo y la implementación actual de **SEDAR** soportan únicamente *checkpoints* coordinados de nivel de sistema o no-coordinados (por proceso, en capa de aplicación); en este sentido, requiere ser extendida de forma de poder soportar diferentes tecnologías, como el *checkpointing* diferencial o incremental.

2. Metodología SEDAR

SEDAR consiste en tres estrategias alternativas y complementarias para alcanzar cobertura completa frente a errores silenciosos: **(1)** sólo detección con notificación al usuario y relanzamiento automático; **(2)** recuperación basada en múltiples *checkpoints* de nivel de sistema; y **(3)** recuperación utilizando un único *checkpoint* seguro de capa de aplicación. A continuación se revisan los principales aspectos de la metodología, los cuales han sido publicados en congresos internacionales [7,8], revistas [9,10] y capítulos de libros [11].

2.1 Detección y Relanzamiento Automático

Para detectar fallos transitorios durante la ejecución de aplicaciones paralelas determinísticas, **SEDAR** valida los contenidos de los mensajes que se van a enviar, acotando así la latencia de detección, aislando el fallo y evitando que se propague a los demás procesos. Para ello, duplica en un *thread* cada proceso de la aplicación. Al alcanzar el instante de envío de un mensaje, las réplicas utilizan un mecanismo de sincronización para garantizar que se encuentran en el mismo punto, y allí se comparan los contenidos de los mensajes computados por ambos hilos; si hay coincidencia, sólo una réplica envía el mensaje (bajo la premisa de que no hay fallos en las comunicaciones), no congestionando la red. El proceso receptor espera a que su réplica alcance el mismo punto y le copia los mensajes que ha recibido para luego reanudar la ejecución. Por otro lado, los fallos cuyos efectos no se transmiten a otros procesos, pero que se propagan localmente hasta el final, son detectados en una operación de validación de resultados finales. En [9] se estudia compromiso entre la latencia de detección y la sobrecarga computacional involucrada en realizar validaciones frecuentes.

Este mecanismo posibilita detectar todos los fallos que causan SDC, en sus dos variantes: Corrupciones en Datos Transmitidos (TDC, por sus siglas en inglés) y Corrupciones en el Estado Final (FSC, por sus siglas en inglés), asegurando la fiabilidad de la ejecución. Además, bajo la premisa de que, en un sistema homogéneo dedicado, los tiempos de ejecución de dos réplicas que realizan el mismo cómputo deben ser similares, es posible detectar los Errores de Temporización (TOE, por sus siglas en inglés) al notar una diferencia considerable entre los tiempos de procesamiento de ambas réplicas [11], a causa posiblemente de un fallo silencioso. Por lo tanto, **SEDAR** permite la configuración de retardos programables de sincronización entre réplicas, para evitar que las aplicaciones ingresen en esperas infinitas. La Figura 1 esquematiza el funcionamiento del mecanismo de detección frente a la ocurrencia de un fallo.

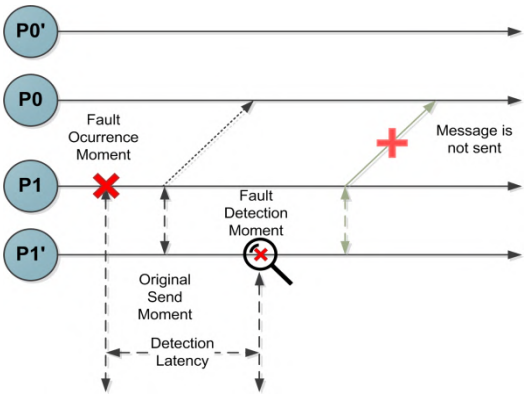


Fig. 1. Operación del mecanismo de detección frente a la ocurrencia de un fallo

El mecanismo de detección propuesto tiene situaciones de vulnerabilidad asociadas, que han sido descriptas previamente en [11]; sin embargo, las situaciones que lo afectan son extremadamente improbables, de modo que pueden ignorarse sin riesgos considerables.

En ausencia de un mecanismo de recuperación, la ocurrencia de un fallo de los tipos mencionados causa que la aplicación detenga su ejecución al momento de la detección, evitando así la costosa e innecesaria espera a que finalice con resultados incorrectos, y permitiendo su relanzamiento inmediato y automático desde el comienzo [7].

2.2 Recuperación Basada en Múltiples *Checkpoints* de Nivel de Sistema

El mecanismo de recuperación de **SEDAR** se basa en el almacenamiento de una cadena de múltiples *checkpoints* coordinados y distribuidos que se realizan por medio de una librería de *checkpointing* de capa de sistema. El hecho de guardar varios *checkpoints* es necesario debido a que, en el *checkpoint* de nivel de sistema, el usuario no tiene control sobre la información que se guarda. En consecuencia, frente a la perspectiva de los errores silenciosos, no hay garantía de que un *checkpoint* determinado sea un punto seguro de recuperación, debido a que sus datos pueden haber sido alterados por la ocurrencia de un fallo transitorio antes de ser almacenados. El *checkpoint* puede contener errores no detectados que permanecían latentes al momento de su almacenamiento. Por lo tanto, debe almacenarse una cadena de *checkpoints*, ya que previos al último no pueden ser descartados. En el caso de que se detecte un fallo, se debe determinar la factibilidad de recuperar la ejecución desde el último *checkpoint* (en el mejor caso) o desde alguno anterior [14]. Como casos particulares, si un fallo es detectado cuando aún no se ha almacenado ningún *checkpoint*, la aplicación debe ser relanzada desde el comienzo. En tanto, si en caso de detectar un fallo, se regresa a un *checkpoint* corrupto, en el intento de re-ejecución se vuelve a detectar el mismo fallo y es necesario retroceder hasta uno previo; esta situación puede ocurrir repetidas veces si la latencia de detección es muy alta [7]. En las Figuras 2 (a) y (b) se esquematizan estos comportamientos. El algoritmo de recuperación, en pseudocódigo, se muestra como Algoritmo 1.

Con esta estrategia, basada en múltiples *checkpoints*, es posible llegar al final de la ejecución, a costa de que a veces incluso se requiera regresar hasta el principio. Principalmente se intenta garantizar la seguridad de los datos del usuario.

La usabilidad de este método puede incrementarse si se automatiza el mecanismo. Esto se logra permitiendo que un proceso externo a la aplicación pueda acceder al valor de *extern_counter* y, de acuerdo a él, buscar el *script* de reinicio correspondiente (entre todos los generados por la librería de c*heckpointing*). Además, el *checkpoint* inválido, que ha causado el *restart* erróneo, debe ser eliminado, sobreescribiéndolo y volviendo a almacenarlo nuevamente durante la re-ejecución).

Algoritmo 1 Algoritmo de recuperación con múltiples *checkpoints* de capa de sistema

```
1: int extern_counter = 0;   ▷ un contador externo controla la cantidad de rollbacks (no incluido en el checkpoint)
2: boolean fault_detected = FALSE;   ▷ una variable boolean para reportar si se detectó un error en la última ejecución
3: SEDAR_run(app);   ▷ ejecuta la aplicación paralela app bajo el monitoreo de SEDAR
4: while fault_detected == TRUE do ▷ si la condición se cumple, se detectó un error en la última ejecución
5:     extern_counter++;                    ▷ extern_counter se incrementa en 1
6:     ckpt_count = get_ckpt_count();   ▷ obtiene el número de checkpoints guardados
7:     ckpt_no = ckpt_count - extern_counter + 1;   ▷ calcula el número del checkpoint para el restart
8:     fault_detected = FALSE;              ▷ resetea el flag de detección en el restart
9:     SEDAR_restart(app, ckpt_no);         ▷ relanza app desde el checkpoint ckpt_no
10: end while
```

Este método tiene como principal ventaja el ser transparente a la aplicación,

en el sentido de que no es necesario conocer su estructura interna, ni disponer de *checkpointing* a medida de la aplicación; además, el mismo mecanismo de C/R es el que ya se usa para los fallos permanentes [23]. Por lo tanto, la utilización de un mecanismo de detección basado en duplicación, y uno de recuperación basado en C/R, posibilita detectar y corregir fallos silenciosos sin la necesidad de recurrir a la triplicación con mecanismo de votación [15]. En tanto, sus limitaciones consisten, por un lado, en la cantidad de almacenamiento requerido, al no poder eliminar los *checkpoints* previos; por otro, en el gasto de tiempo involucrado en los múltiples intentos de reinicio posibles [14]; y, por último, en la baja escalabilidad de los *checkpoints* coordinados al crecer la cantidad de procesos, además del hecho de que almacenan gran cantidad de información relacionada al sistema [1].



(a) Baja latencia de detección: la recuperación se realiza desde el último *checkpoint*

(b) Alta latencia de detección: la recuperación se realiza desde el *checkpoint* anterior

Fig. 2. Posibles casos de recuperación utilizando múltiples *checkpoints* de nivel de sistema, dependiendo de la latencia de detección

2.3 Recuperación Basada en un Único *Checkpoint* de Capa de Aplicación

A pesar de requerir un conocimiento detallado de la estructura interna de la aplicación, los *checkpoints* de capa de aplicación constituyen una opción más adecuada, debido a que sólo guardan información relacionada con la aplicación. Además, son más pequeños (ocupando menor espacio de almacenamiento), más portables y escalan mejor que sus contrapartes de nivel de sistema. Por este motivo, en **SEDAR** se propone también un mecanismo de recuperación basado en un único *checkpoint* de capa de aplicación, en conjunto con una estrategia para garantizar la validez del último *checkpoint* guardado. Esto disminuye el tiempo y el trabajo asociados a la recuperación, a costa de requerir una validación de los datos contenidos en los *checkpoints* [7].

La solución propuesta se basa en almacenar *checkpoints* de aplicación (no coordinados, para mejorar la escalabilidad) a nivel de *thread* (cada *thread* guarda un *checkpoint* propio), aprovechando el mismo mecanismo de sincronización entre réplicas que se desarrolló en la fase de detección. Cuando uno de los *threads* llega a la instancia de almacenar un *checkpoint*, espera allí hasta que su réplica también haya guardado el suyo correspondiente. Estos *checkpoints* consisten en salvaguardar sólo el conjunto de variables que son relevantes para la aplicación en ese momento específico. Como se almacenan *checkpoints* de ambos hilos, es posible calcular un *hash* sobre cada uno de ellos y aplicar el mismo mecanismo utilizado para validar contenidos de mensajes, para comparar en este caso ambos *hashes*. Por lo tanto, un *checkpoint* se considera válido sólo si la verificación resulta exitosa, y en consecuencia sólo permanece almacenado el *checkpoint* cuyo contenido ha probado ser válido. En este escenario, el *checkpoint* previo puede descartarse sin peligro, reduciendo así el espacio de almacenamiento, ya que el *checkpoint* actual constituye un estado consistente y seguro para la recuperación de un error.

Algoritmo 2 Algoritmo de recuperación con *checkpoints* de capa de aplicación

```
1: function USR_CKPT(n)                          ▷ definición de la función usr_ckpt
2:     for (tid=0; tid < 2; tid++) do                          ▷ cada una de las réplicas
3:         store_all_significant_variables(tid);        ▷ almacenan su propio checkpoint
4:         hash_array[tid]=compute_hash(tid);
5:     end for
6:     synch_threads();                                          ▷ se esperan mutuamente
7:     if tid==0 then                          ▷ sólo una de las réplicas compara los hashes
8:         if hash_array[0]==hash_array[1] then                          ▷ si coinciden
9:             remove_all_significant_variables(tid);      ▷ elimina su propio checkpoint
10:            return TRUE;          ▷ este es un checkpoint válido, por lo que se puede
    descartar el anterior
11:        else
12:            return FALSE                              ▷ este es un checkpoint inválido
13:        end if
14:    end if
15: end function
16:
17: SEDAR_run(app)   ▷ ejecuta la aplicación paralela app bajo el monitoreo de SEDAR
18: if usr_ckpt(n)== TRUE then                          ▷ n representa el checkpoint actual
19:     remove_usr_ckpt(n-1); ▷ elimina el checkpoint previo, dado que el actual es válido
20: else
21:     remove_usr_ckpt(n);                                  ▷ elimina el checkpoint actual
22:     restart_from_usr_checkpoint(n-1);          ▷ restart desde el checkpoint anterior
23: end if
```

En tanto, si se detecta una diferencia en la comprobación, ésta necesariamente se debe a un fallo ocurrido dentro del último intervalo de *checkpoint*; por ende, este último *checkpoint* no puede ser utilizado como punto seguro para la recuperación, debiendo borrarse el *checkpoint* corrompido, y retomar la ejecución desde el anterior (que fue verificado). Como consecuencia, existe un único *checkpoint* válido almacenado en un momento determinado (excepto durante el intervalo de validación), independientemente del resultado de la comparación.

El mecanismo de recuperación propuesto se muestra como Algoritmo 2 (en pseudocódigo).

Cada una de estas alternativas tiene características particulares y provee un *trade-off* diferente entre costo y performance. La posibilidad de elegir entre estas alternativas brinda flexibilidad para adaptarse a la relación costo-beneficio requerida para el sistema particular.

3. Herramienta SEDAR y verificación funcional

En su estado actual, **SEDAR** permite 3 modos de utilización. El modo de **Sólo Detección y Relanzamiento Automático** permite evitar el coste asociado al almacenamiento de los *checkpoints*, posibilitando que la aplicación pueda ser relanzada desde el principio ni bien transcurre la latencia de detección. El modo **Recuperación Basada en *Checkpoints* de Nivel de Sistema Disparados por Eventos** es de utilidad cuando se tiene un conocimiento detallado del comportamiento de la aplicación a proteger. En este caso, **SEDAR** puede sintonizarse con la aplicación: tomando en cuenta que los datos a comunicar son validados previamente, almacenar *checkpoints* en los instantes de comunicación permite minimizar los riesgos de almacenar valores corruptos y de propagarlos. Por otra parte, seleccionar cuántos *checkpoints* se realizan (y en qué instantes) posibilita optimizar la relación costo/beneficio del mecanismo de protección, y mantener acotado el *overhead*. Este modo fue el utilizado principalmente en la etapa de validación funcional del mecanismo. Como desventaja, este mecanismo no es transparente (ni automático), en el sentido de que el código de los *checkpoints* debe insertarse dentro de la aplicación. Por último, el modo **Recuperación Basada en *Checkpoints* Periódicos de Nivel de Sistema** permite proteger a la aplicación "desde el exterior". Lanzando la ejecución bajo la órbita de un proceso coordinador de la librería de *checkpointing*, se establece un intervalo de almacenamiento periódico. En este caso, no es necesario conocer la aplicación en profundidad, a costa de no tener control sobre el instante preciso en que se realiza el *checkpoint*, ni de desde dónde podrá recuperarse la ejecución. El *overhead* tendrá un comportamiento asociado al costo de almacenar un *checkpoint* y a la cantidad de ellos que se almacenen dado el intervalo de periodicidad. Como ventaja, esta protección pude realizarse de forma completamente automática.

El mecanismo de detección se implementa en la forma de una librería que modifica las funciones y los tipos de datos definidos en MPI (sintácticamente, sólo cambia el prefijo MPI por **SEDAR**), y agrega las funciones **SEDAR_Call** y **SEDAR_Validate**. Por otro lado, utiliza ***Pthreads*** para la replicación y sincronización de hilos. Para incorporar la detección, el código de la aplicación debe ser levemente modificado y recompilado, en un procedimiento automatizable.

La recuperación automática se implementa mediante la utilización de la librería DMTCP [16] (versión 2.4.4) que provee *checkpointing* coordinado y distribuido de nivel de sistema. Durante la instalación, puede configurarse su comportamiento para almacenar varios *checkpoints* (con números correlativos) y sus correspondientes *scripts* de reinicio, posibilidad en la que se basa el mecanismo de recuperación de **SEDAR**. DMTCP genera un proceso coordinador, que monitorea la aplicación objetivo y es capaz de realizar los *checkpoints* periódicos desde el exterior, lo cual produce que el mecanismo de recuperación sea transparente para la aplicación [7].

Para validar el funcionamiento de la estrategia de recuperación automática, se ha desarrollado un modelo sobre un caso, basado en una aplicación de prueba bien conocida, combinada con una carga de fallos (*workfault*) completa y totalmente controlada. El caso modelado contempla todos los escenarios de fallos posibles que pueden ocurrir, a partir del profundo conocimiento sobre el comportamiento de la aplicación de prueba, por lo que el *workfault* está diseñado para para cubrir esos escenarios.

Cada fallo que se inyecta tiene un efecto predecible, un instante en el que se puede prever que será detectado, y un punto desde el cual podrá la recuperarse la ejecución, que puede determinarse con certeza. Obviamente, existen innumerables posibilidades de fallo (cualquier bit, de cualquier registro o unidad funcional, en cualquier momento de la ejecución); sin embargo, todas ellas están representadas en los escenarios previstos. Un escenario representa una clase de errores, por lo que engloba un amplio conjunto de casos que darán origen a un mismo comportamiento. Para cada experimento que se realizó en esta etapa, se inyectó un único fallo.

La aplicación sintética de prueba fue construida sobre la base de una multiplicación de matrices *Master/Worker* en MPI. Además de las modificaciones relativas a la replicación de procesos en *threads* para detección y validación final de la matriz resultante, se agrega el almacenamiento de un *checkpoint* de nivel de sistema cada vez que aplicación acaba de realizar una comunicación, debido a que los mensajes circulan sólo cuando los datos involucrados han sido validados (maximizando la probabilidad de datos confiables). El conocimiento detallado sobre el comportamiento de la aplicación de prueba permite identificar en qué instantes ocurren las comunicaciones entre procesos, y los datos que forman parte de cada comunicación. Como consecuencia, se puede predecir el efecto preciso de cada fallo inyectado, como también el estado de cada *checkpoint* almacenado ("seguro" o "corrompido"), y, por ende, qué *checkpoint* posibilita la correcta recuperación. Todos estos factores fueron combinados para dar origen a un conjunto de 64 experimentos de inyección de fallos que contemplan todas las situaciones que pueden ocurrir para la aplicación de prueba particular, obteniendo para cada experimento el comportamiento esperado, y verificando así el funcionamiento correcto de los mecanismos implementados en **SEDAR**.

4. Caracterización Temporal

Una vez realizada la validación funcional, el foco se puso en los aspectos prestacionales. Para ello, se realizó una evaluación cuantitativa del comportamiento temporal de las tres estrategias que componen la metodología. Por un lado, se desarrolló un modelo matemático que toma en cuenta los parámetros que influyen sobre el tiempo de ejecución de cada una de las alternativas. El modelo se materializa en la forma de ecuaciones que pueden utilizarse para estimar el tiempo de ejecución para cada caso, tanto en ausencia de fallos como cuando ocurre un único fallo durante la ejecución, si se pueden medir u obtener valores aproximados para los parámetros mencionados.

Para poner de manifiesto la forma en la cual el modelo presentado se puede utilizar para evaluar el comportamiento temporal de cada estrategia alternativa, se montaron experimentos en los cuales **SEDAR** fue utilizado en conjunto con tres *benchmarks* paralelos con diferentes patrones de comunicación y demandas de *workload*: la multiplicación de matrices; el método de Jacobi para la

solución de la ecuación de Laplace [17]; y el alineamiento de secuencias de ADN con el algoritmo de Smith-Waterman [18]. Estas aplicaciones fueron elegidas porque son bien conocidas, computacionalmente demandantes y representativas del cómputo científico. Además, presentan diferentes patrones de comunicación: *Master/Worker*, SPMD y *Pipeline*, respectivamente [10]. La comparación de los tiempos de ejecución de las tres aplicaciones de prueba seleccionadas, entre las versiones MPI puras y nuestra implementación de **SEDAR** basada en MPI, permitió obtener, en base a medidas y estimaciones, los parámetros que caracterizan la ejecución de cada aplicación. A su vez, la obtención de estos parámetros posibilitó evaluar cualitativamente la incidencia del patrón de comunicaciones, el intervalo de *checkpoint* y la latencia de detección sobre la *performance*.

Como resultado, se pudo observar que el comportamiento temporal de cada estrategia de **SEDAR** es dependiente del patrón de comunicaciones, el *ratio* de cómputo a comunicaciones, la cantidad y frecuencia de datos que se transmiten (que afecta al *overhead* de detección) y de la latencia de detección (cuanto más avanzada la ejecución al momento de la detección, más costoso es relanzar desde el comienzo sino hay ningún *checkpoint* válido cercano disponible). Las diferentes alternativas de **SEDAR** pueden alcanzar mejoras considerables cuando se enfrentan a la ocurrencia de un error silencioso, tanto en fiabilidad como en tiempo de ejecución, lo que resulta particularmente valioso en ejecuciones que pueden durar muchas horas. Más aún, cuanto más larga es una ejecución, más útil es la estrategia de tolerancia a fallos, porque la ocurrencia de los fallos se vuelve más probable, en tanto que el *overhead* de guardar *checkpoints* es demasiado alto para aplicaciones breves.

A pesar de que en un sistema que almacena una cadena de *checkpoints*, la recuperación es siempre posible luego de uno o más intentos de *rollback*, existen escenarios posibles en los cuales el tiempo que se gasta en esos intentos puede ser mayor al de simplemente detener la ejecución al momento de la detección y relanzar desde el comienzo, considerando el *overhead* introducido al hacer *checkpointing* y *rollback*. Por lo tanto, se evaluaron las situaciones en las que es conveniente almacenar múltiples *checkpoints* en lugar de activar sólo el mecanismo de detección, extrayendo información útil del modelo de comportamiento temporal desarrollado: si para un sistema particular existen estadísticas disponibles sobre la frecuencia y comportamiento típico de los fallos, es posible optar por la estrategia de protección más apropiada, basándose en el conocimiento de los parámetros del sistema. Si el fallo se detecta cerca de la finalización, incluso hacer varios reintentos de *rollback* puede representar una mejora, comparado con parar y relanzar. El conocimiento de la cantidad de *checkpoints* que conviene tener almacenados en un momento determinado permite también ahorrar espacio, ya que se pueden borrar los *checkpoints* a los que ya no valdrá la pena regresar a medida que avanza la ejecución, manteniendo constante el espacio de almacenamiento requerido. El *overhead* asociado con retroceder y volver a ejecutar es mucho más significativo que el costo temporal de almacenar *checkpoints*, por lo que el beneficio volver atrás un lapso acotado y rehacer sólo una parte del trabajo excede al *overhead* de guardar *checkpoints* más frecuentemente.

Por último, para cuantificar el efecto de aplicar las distintas alternativas de **SEDAR**, y su relación con las características de las aplicaciones, se realizaron mediciones del *overhead* de ejecución en escenarios realistas. Para ello, se han comparado los modos **Sólo Detección y Relanzamiento Automático** y **Recuperación Basada en Checkpoints Periódicos de Nivel de Sistema**, tanto en ausencia de fallos como con fallos inyectados en distintos puntos durante la ejecución, y con distintos intervalos de *checkpoint*, buscando ofrecer distintas alternativas de compromiso entre la protección brindada y el *overhead* introducido, frente a distintos valores posibles del Tiempo Medio Entre Errores (MTBE, por sus siglas en inglés, el equivalente a MTBF pero para errores silenciosos). De esta forma, se busca aportar información que permita alcanzar la configuración óptima de la herramienta según las características particulares del sistema, cuando se la utiliza, integrada de forma transparente, en entornos y problemas de mayor escala. Este análisis se realizó con el producto de matrices, que

demanda gran cantidad de cómputo local y requiere pocas comunicaciones al inicio y al final de cada tarea, y con la aplicación de Jacobi, que realiza una gran cantidad de comunicaciones regulares y permanentes pero con menor volumen de datos. Como conclusiones, para aplicaciones limitadas por cómputo la latencia de detección tiende a ser elevada, por lo que pueden requerirse varios intentos de recuperación; en esa situación, la estrategia de sólo detección y relanzamiento automático puede ser beneficiosa. En tanto, para las aplicaciones que realizan comunicaciones frecuentes, la latencia de detección tiende a ser baja y, por lo tanto, también la cantidad de intentos de recuperación; en ese escenario resulta conveniente el almacenamiento de múltiples *checkpoints*. Por lo tanto, se puede destacar la flexibilidad de **SEDAR** para ajustarse a las necesidades de un sistema particular, proveyendo diferentes posibilidades de cobertura que le permiten adaptarse para obtener un compromiso en la relación costo/beneficio. El intervalo de *checkpoint* óptimo y la frecuencia de validación de las comunicaciones pueden ser determinados a partir de las características de la aplicación a proteger.

5. Conclusiones y trabajos futuros

El camino hacia los sistemas de cómputo en la Exa-escala presenta varios desafíos para las próximas generaciones. La obtención de garantías respecto de la confiabilidad de las ejecuciones y el manejo eficaz de los fallos son algunos de los principales, constituyendo una de las preocupaciones crecientes en el ámbito del HPC. En el futuro, se esperan mayores variedades y tasas de errores, intervalos o latencias de detección más prolongados y errores silenciosos. Se proyecta que, en los sistemas de Exa-escala, los errores ocurran varias veces al día, y se propaguen de forma de generar desde caídas de procesos hasta corrupciones de resultados debidas a fallos no detectados. En este contexto, si se toman en cuenta los profundos efectos que un único fallo transitorio puede causar en todos los procesos que se comunican, la protección de las aplicaciones MPI a nivel de las comunicaciones es un método factible y efectivo para detectar y aislar la corrupción de datos, evitando su propagación.

En este trabajo se propone, diseña e implementa **SEDAR**, una metodología que permite detectar los fallos transitorios y recuperar automáticamente las ejecuciones, aumentando la fiabilidad y la robustez en sistemas en los que se ejecutan aplicaciones paralelas determinísticas de paso de mensajes, de una manera agnóstica a los algoritmos a los que protege. La metodología desarrollada está basada en la replicación de procesos para la detección, combinada con distintos niveles de *checkpointing* (de capa de sistema o de capa de aplicación) para la recuperación automática. Con el objetivo de ayudar a programadores y usuarios de aplicaciones científicas, hemos arribado a un prototipo de un sistema automático que brinda una calidad de servicio, siendo capaz de almacenar los *checkpoints* y de recuperar sin intervención del usuario, garantizando así la finalización de las ejecuciones con resultados correctos dentro de un tiempo factible de ser acotado.

Las principales líneas abiertas son:

- Ampliar la validación experimental, utilizando el algoritmo de recuperación basado en *checkpoints* no-coordinados con aplicaciones que incluyen sus propios *checkpoints* a medida, que será la tendencia más marcada en el futuro.
- Calcular el intervalo óptimo de *checkpoint*, de modo de minimizar el *overhead* de ejecución pero también la cantidad de trabajo que debe volverse a hacer en la re-ejecución, cuantificando la relación entre la latencia de detección y el patrón de comunicaciones.
- Refinar el mecanismo de recuperación basada en múltiples *checkpoints*, de forma de soportar de manera óptima la ocurrencia de varios fallos, y predecir la respuesta temporal cuando suceden dos o más errores no relacionados.

- Implementar una estrategia de adaptación dinámica del mecanismo de recuperación, y herramientas auxiliares que le brinden al usuario reportes de los fallos detectados, si han podido ser recuperados y cómo, de modo de servir como información estadística para realizar análisis posteriores.
- Integrar **SEDAR** con arquitecturas que utilizan C/R para tolerancia a fallos permanentes [19]. Debido a que **SEDAR** proporciona detección y la recuperación, se podrían soportar ambos tipos de fallos con una única herramienta funcional para las plataformas que se proyectan en la Exa-escala, y tomar en cuenta el impacto del consumo energético sobre la resiliencia.

Referencias

1. T. Martsinkevich, O. Subasi, O. Unsal, F. Cappello, and J. Labarta, “Fault-tolerant protocol for hybrid task-parallel message-passing applications," in 2015 IEEE International Conference on Cluster Computing , pp. 563-570.
2. Nuria Losada. Application-level fault tolerance and resilience in HPC applications. 2018.
3. Dolores Rexachs and Emilio Luque. High availability for parallel computers. Journal of Computer Science & Technology (JCS&T), 10(3), 2010.
4. Adam Oliner and Jon Stearley. What supercomputers say: A study of five system logs. In 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN’07), pages 575–584. IEEE, 2007.
5. Greg Bronevetsky and Bronis de Supinski. Soft error vulnerability of iterative linear algebra methods. In Proceedings of the 22nd annual international conference on Supercomputing, pages 155–164, 2008.
6. Anne Benoit, Aurélien Cavelan, Franck Cappello, Padma Raghavan, Yves Robert, and Hongyang Sun. Coping with silent and fail-stop errors at scale by combining replication and checkpointing. Journal of Parallel and Distributed Computing, 122:209–225, 2018.
7. Diego Montezanti, A De Giusti, Marcelo Naiouf, Jorge Villamayor, Dolores Rexachs, and Emilio Luque. A methodology for soft errors detection and automatic recovery. In 2017 International Conference on High Performance Computing & Simulation (HPCS), pages 434–441. IEEE, 2017.
8. Diego Miguel Montezanti, Dolores Rexachs del Rosario, Enzo Rucci, Emilio Luque, Marcelo Naiouf, and Armando Eduardo De Giusti. Sedar: Detectando y recuperando fallos transitorios en hpc. In XXV Congreso Argentino de Ciencias de la Computación (Río Cuarto, 2019), 2019.
9. Diego Montezanti, Emmanuel Frati, Dolores Rexachs, Emilio Luque, Marcelo Naiouf, and Armando De Giusti. SMCV: a methodology for detecting transient faults in multicore clusters. CLEI Electronic Journal, 15(3):1–11, 2012.
10. Diego Montezanti, Enzo Rucci, Dolores Rexachs, Emilio Luque, Marcelo Naiouf, and Armando De Giusti. A tool for detecting transient faults in execution of parallel scientific applications on multicore clusters. Journal of Computer Science & Technology, 14:32–38, 2014.
11. Diego Miguel Montezanti, Dolores Rexachs del Rosario, Enzo Rucci, Emilio Luque Fadón, Marcelo Naiouf, and Armando Eduardo De Giusti. Characterizing a detection strategy for transient faults in hpc. In Computer Science & Technology Series. XXI Argentine Congress of Computer Science. Selected papers, pages 77–90. Editorial de la Universidad Nacional de La Plata (EDULP), 2016.
12. Anne Benoit, Thomas Hérault, Valentin Le Fèvre, and Yves Robert. Replication is more efficient than you think. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, page 89. ACM, 2019.
13. George Bosilca, Rémi Delmas, Jack Dongarra, and Julien Langou. Algorithm-based fault tolerance applied to high performance computing. Journal of Parallel and Distributed Computing, 69(4):410–416, 2009.
14. G. Lu, Z. Zheng, and A. A. Chien, “When is multi-version checkpointing needed?" in Proceedings of the 3rd Workshop on Fault-tolerance for HPC at extreme scale. ACM, 2013, pp. 49-56.
15. F. Mathur and A. Avizienis, “Reliability analysis and architecture of a hybrid-redundant digital system: generalized triple modular redundancy with self-repair”, in Proc. Spring Joint Computer Conference (AFIPS '70), New York, USA, May. 1970, pp. 375 – 383.
16. Ansel, J., Arya, K., & Cooperman, G. (2009, May). DMTCP: Transparent checkpointing for cluster computations and the desktop. In 2009 IEEE International Symposium on Parallel & Distributed Processing (pp. 1-12). IEEE.
17. Gregory Andrews. Scientific computing. In Foundations of Multithreaded, Parallel and Distributed Computing, chapter 11, pages 527–585. Addison-Wesley, 2000.
18. Enzo Rucci, Armando De Giusti, and Franco Chichizola. Parallel smith-waterman algorithm for dna sequences comparison on different cluster architectures. In Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA’11), volume 1, pages 666–672. WorldComp, 2011.
19. Marcela Castro-León, Hugo Meyer, Dolores Rexachs, and Emilio Luque. Fault tolerance at system level based on RADIC architecture. Journal of Parallel and Distributed Computing, 86:98–111, 2015.
20. David Fiala, Frank Mueller, Christian Engelmann, Rolf Riesen, Kurt Ferreira, and Ron Brightwell. Detection and correction of silent data corruption for large-scale high-performance computing. In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, page 78. IEEE Computer Society Press, 2012.
21. Catello Di Martino, Zbigniew Kalbarczyk, and Ravishankar Iyer. Measuring the resiliency of extreme-scale computing environments. In Principles of Performance and Reliability Modeling and Evaluation, pages 609–655. Springer, 2016.
22. Kurt Ferreira, Rolf Riesen, Ron Oldfield, Jon Stearley, James Laros, Kevin Pedretti, and T Brightwell. rMPI: increasing fault resiliency in a message-passing environment. Sandia National Laboratories, Albuquerque, NM, Tech. Rep. SAND2011-2488, 2011.
23. Julien Adam, Maxime Kermarquer, Jean-Baptiste Besnard, Leonardo Bautista-Gomez, Marc Perache, Patrick Carribault, Julien Jaeger, Allen D Malony, and Sameer Shende. Checkpoint/restart approaches for a thread-based mpi runtime. Parallel Computing, 85:204–219, 2019.