PAPER • OPEN ACCESS

Studies on Experimental Ontology and Knowledge Service Development in Bio-Environmental Engineering

To cite this article: Yunliang Zhang 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **301** 012146

IOP Conf. Series: Materials Science and Engineering **301** (2018) 012146 doi:10.1088/1757-899X/301/1/012146

Studies on Experimental Ontology and Knowledge Service Development in Bio-Environmental Engineering

Yunliang Zhang

Institute of Scientific & Technical Information of China, Beijing 100038; Key Laboratory of Rich-media Knowledge Organization and Service of Digital Publishing Content, SAPPRFT, China
Email:zhangyl@istic.ac.cn

Abstract. The existing domain-related ontology and information service patterns are analyzed, and the main problems faced by the experimental scheme knowledge service were clarified. The ontology framework model for knowledge service of Bio-environmental Engineering was proposed from the aspects of experimental materials, experimental conditions and experimental instruments, and this ontology will be combined with existing knowledge organization systems to organize scientific and technological literatures, data and experimental schemes. With the similarity and priority calculation, it can improve the related domain research.

1. Introduction

In the fields where experiments are the main research methods, a good experiment scheme is very important. In this paper, taking the interdisciplinary field of bio-environmental engineering as an example, conducts research on the development of knowledge service model for experiment schemes.

Bio-environmental Engineering [1] started with the relevant departments set up by the U.S. Air Force in 1947 and divided the bio-environmental projects into four parts: radiation, industrial hygiene, environmental protection, and emergency response, emphasizing human health concerns, providing a good working environment. Universities have also started to focus on this area since the 1960s. They are both based on the biological process methods to solve the environmental problems and provide a good living and living environment for human beings.

At present, there are some information services and knowledge services for biological environment engineering, but the intervention of these services on the research process is very shallow. Mainly in this area is based on experimental research, which is usually a long period. So it is important to choose a good experiment scheme.In this paper, we hope to reduce the corresponding risks and improve the research results by constructing the knowledge service ontology and integrating with other existing knowledge organization systems to drive the knowledge service for the experiment scheme formulation and adjustment.

2. Related Ontology Construction and Service

2.1. Experiments Related Ontologies

2.1.1. EXPO. The scientific experiment ontology (EXPO) [2] is to construct a unified ontology between the SUMO (the Common Upper Ontology) and the experimental ontology of a specific topic proposed. The goal of this ontology is to promote the standardization and facilitation to use the description of the relevant documents in a scientific experiment. EXPO demonstrated its utility and its

IOP Conf. Series: Materials Science and Engineering **301** (2018) 012146 doi:10.1088/1757-899X/301/1/012146

ability to describe different experimental areas through two experiments, high-energy physics and phylogenetics. The EXPO made the goals and structure of these experiment documents clearer and demonstrated similarities in different fields. From the ontology file [3], we can find that the structure of the EXPO ontology is very clear. Under the Entity class, there are two subclasses, Abstract and Physical. The ontology has a positive effect on the standardization of experimental documents and the effective resolution of the natural language ambiguity of documents.

2.1.2. PEAO. Plant Experimental Assay Ontology (PEAO) [4] was jointly developed by INESC-ID in Portugal and other agencies. The purpose of this ontology construction is to explain the experimental results and the integration of the experimental data. It is not concerned with the specific scientific problems associated with the experimental analysis, but focusing on the development of ontology for these plant experiments. The ontology includes entities from three different fields, including the experimental products, the relationships between them, and the specifications that describe their operation. Although the entity is divided into three different areas, there is no obvious distinction in ontology. According to the ontology file [5], no clear clues are found in the organization of the class and there is no known practical application. The authors also point out that the ontology is related to the experimental analysis of the plant field so that there are likely to be different in other areas.

2.1.3. EFO. Experimental Factor Ontology (EFO) [6] provides a systematic description of lots of the experimental factors in the European Bioinformatics Institute (EBI) database as well as other similar external projects such as those associated with the National Human Genome National Human Genome Research Institute, Genome-Wide Association Studies (GWAS). EFO includes multiple ontologies including anatomy, disease and compounds. The purpose of EFO is to support EBI's multiple study groups in indexing, analysing and visualizing data. Apart from some fragmented classes, most of the classes are grouped into five core categories: information entity, material entity, material properties, process, and site. The ontology caters to bioinformatics and has some classes of bioinformatics such as information and location. But the services provided based on this ontology are still relatively simple.

2.1.4. BiOnto. BiOnto [7] is about the biological refining process, and it is an integration of two parts: the biomass types and biological treatment processes. The biomass part includes five sub-categories: Products, PlatformMaterial, MaterialByPhysicalForm, FeedstockMaterials and OtherMaterials. The biological treatment process includes three sub-categories of ByProcessType, ByProcessStage and ByProductHeatingValue.The properties used in BiOnto focus on the handling characteristics of the technology and the compositional (chemical) properties of the biomass. As reflected in the processing technology attributes such as hasConversionRate, hasOutput, hasInput and so on. The attributes that reflect the composition of biomass are hasHeatingValue, hasMoistureContent and composition of hasLinein , hasCelluloce, and hasHelicelluloce and so on.

2.2. Bioenvironment Related Ontology

There are a lot of ontologies related to bioenvironmental engineering. In this paper, several important ontologies such as Plant ontology (PO), gene ontology (GO), environmental ontology (EnvO), and Chemical Entities of Biological Interest (ChEBI) are investigated. The related information is shown in Table 1. These ontologies objectively describe the composition of genes and chemical entities with important to biology and involves knowledge of some plants and biomes. It provides a basis for the organization of resources such as research literature.

Table 1. Information of several bioenvironment related ontologies

Ontology name	Main content	Main properties	Characters	Axiom*	URL
Plant ontology(PO)	Plant Anatomical Entity and Plant Structure ,Development Stage	Is_a(or Instance_of),Part_of , Develops_from,synonym (or synonyms) etc.	With a simple structure of Directed Acyclic Graph(DAG)	22732	http://archive.plantontology.org/
The Gene Ontology(GO)	cellular component, molecular function and biological process	is_a,part_of, regulates etc.	Provide a lot of tools, such as Ami GO 2,GO Enrichment Analysis etc.	579983	http://www.geneontology.org/
Environmental Ontology(EnvO)	Biome, environmental feature and environmental material	Many object propertie, such as contained, created_by, depends on, hasSynonyms,ocated in etc.	Closely related to the GAZ ontology, providing placename indexing	49113	http://environmentontology.org/home/about-envo
Chemical Entities of Biological Interest(ChEBI)	Molecular entities of "small" chemical compounds such as atom, molecule, ion, ion pair, radical, radical ion, complex, onformer	hasDataSource,targets, localized, part_of ,hasPhysicalProperty etc.	Use IUPAC and NC-IUBMB-approved naming, symbol, and terminology, etc.search	1973330	https://www.ebi.ac.uk/chebi/aboutChebiForward.do

* Until 2017-06-12

3. Relevant Knowledge Service

3.1. Common Information Service Providers and Their Service Patterns

At present, all the major information service providers, such as Web of Sciences, Engineering Village, Scopus, and Google Scholar and so on, can support the R & D information needs of bioenvironmental engineering to some extent and the service patterns and service features of several information services as shown in table 2.

Table 2. Service Patterns of Some Common Information Services Providers

provider name	service pattern	service features	URL
Web of sciences	Retrieval, cluster analysis, citation analysis, visualization	Multi-dimensional screening refinement, molecular search	http://apps.webofknowledge.com/
Engineering Village	Retrieval, analysis, visualization	Thesaurus Retrieval, Tags & Groups	https://www.engineeringvillage.com/
Scopus	Retrieval, analysis, track, visualization	Use ORCiD as the core of user information management	https://www.scopus.com/
Google Scholar	Retrieval, citation analysis, resource link	Large amount of data, high-performance sorting	https://scholar.google.com/

3.2. Professional Information Service Providers and Their Service Patterns

Professional information service providers are more targeted. Of course, these services tend to cover a larger area than bio-environmental engineering. The main providers are the NAL Catalog of the National Agricultural Library, the Technical Library of American Society of Agricultural and Biological Engineers (ASABE), the BIOSIS CITATION INDEX and BIOSIS PREVIEWS in the CAB Direct from Creative Analytics, the ProQuest Agricultural and Environmental Science Database, the EBSCO Academic Search Complete and PubMed and so on. The main service patterns and service features of each service provider are shown in Table 3.

Table 3. Service Patterns of Some Professional Information Services Providers

provider name	service pattern	service features	URL
NAL Catalog(Agricola)	Retrieval	Provide separate agricultural data services and research services based on VIVO	https://agricola.nal.usda.gov/
ASABE Technical Library	Retrieval	Industry conference information service	http://elibrary.asabe.org/
CAB Direct	Retrieval	Advanced Search can provide Descriptor search and Organism descriptor, etc.	https://www.cabdirect.org/
BIOSIS series of professional services	Retrieval,Citation	Refinement of search results based on Major Concepts, Concept Codes, Super Taxa and Assignees and so on	http://wokinfo.com/products_tools/specialized/
ProQuest Agricultural and Environmental Science Database	Retrieval	Supports vocabularies retrieval of 7 agricultural and environmental science databases such as EIS Controlled terms, Life Sciences Thesaurus, etc. It is also support Classification Search	http://search.proquest.com/agricenvironm/advanced
EBSCO Academic Search Complete	Retrieval	Full-text data, research-related Associated Press thematic video content and search.	https://www.ebscohost.com/academic/academic-search-complete
PubMed	Retrieval	Support MeSH Major Topics MeSH terms , MeSH Subheadings [sh]), etc.	https://www.ncbi.nlm.nih.gov/pubmed

3.3. Other Referential Service Patterns

*3.3.1. Distributed Infrastructure to Support Scientific Experiments.*Tadeu Classe et al. proposed a distributed infrastructure to support scientific experiments for E-Science [8], which can utilize distributed web services and computing resources to handle distributed experimental data. In order to effectively describe and assemble the workflow, the authors also put forward a set of ontology system. The research takes bioinformatics experiments as an example to illustrate the working principle and the effect. However, it should be noted that this framework is still aimed at data and information, and only provides the processing and flow of data and information. In fact, the management, intervention and influence of entities related to the science experiment itself are still relatively small.

*3.3.2. Integrated Digitized Biocollections (iDigBio).*Integrated Digitized Biocollections(iDigBio) [9] is a project created by Florida State University and the University of Florida for the purpose of coordinating biological collections in 92 institutions in 45 states. Information in iDigBio includes field records, photos, 3-D images, and information on the geographic distribution of the associated organisms, habitat of the environment, and sample DNA samples. The site currently contains 105,505,060 Specimen Records, 21,983,112 Media Records and 1,662 Recordsets. For a specimen, it mainly includes the following six components: Specimen Record, corresponding Media Record, its own From Recordset, Data, Flags, Raw.

*3.3.3. Morphbank Biological Imaging Database.*The Morphbank Biological Imaging Database [10] was established in 1998 by entomologists from Sweden, Spain and the United States and it is a growing database of images used by scientists for international collaboration, research and education. The database includes records of various studies such as comparative sample-based research in anatomy, morphological phylogenetic studies, taxonomic studies and research in areas related to biodiversity awareness. There are now more than 216,000 published photos of more than 4,500 species. The project provides an image viewer; you can click the zoom button to browse the image in different formats.

*3.3.4. The Catalogue of Life (CoL).*The Catalog of Life (CoL) [11] Project started in June 2001 with Species 2000 and the Integrated Taxonomic Information System (ITIS). It is the most comprehensive and authoritative global index of species available today. It consists of a single integrated species checklist and a taxonomic hierarchy. The catalog contains basic information on the names, relationships and distribution of some 1.66 million species of existing species and nearly 50,000 species of extinct species. Although this number is very large, it is estimated that this is only one-fifth or less of all species. Of course, the number of species will continue to rise with the convergence of different sources of information from the entire world.

*3.3.5. N8 Biohub Information and Knowledge Management System.*N8 Biohub Information and Knowledge Management System project [12] was funded by Innovate UK, with Unilever UK Central Resources Limited, The University of Sheffield and other organizations to complete. The project established an information system (IS) to support the creation of a "Biohub". The system is used for the rapid identification of functional components of simple transformation of renewable plant and waste raw materials and to recommend the best raw materials.

4. Conclusion

Based on the investigation and research in the field of bioenvironmental engineering, we find that this field has the characteristics of being highly dependent on the experiment. Therefore, starting from the goal of experimental design and adjustment, this paper establishes a unified ontology framework and correlates relevant knowledge organization system to manage experimental materials, conditions, experimental apparatus and the final results of the information base of the entity. Based on the above information of the entity similarity and experiment scheme optimization, recommendations can be given in scheme design, and can direct the resources allocation and literature feedback. In the future, the ontology will be constructed firstly.

5. Acknowledgment

This work is partially supported by CKCEST-2017-1-12, 201605360003 and 2015BAH25F01.

6. References

[1] Bio-environmental Engineering [EB/OL]: https://en.wikipedia.org/wiki/Bio-environmental_Engineering, (2017-03-17) [2017-06-07].
[2] Soldatova L N, King R D. Ontology of scientific experiments. [J]. Journal of the Royal Society Interface, 2006, 3(11):795.
[3] EXPO-OWL [EB/OL]: https://sourceforge.net/projects/expo/files/?source=navbar, (2006-04-20) [2017-06-09].
[4] Mendes N D, Monteiro P T, Vaz C A, et al. Towards a Plant Experimental Assay Ontology: 10th International Conference on Data Integration in the Life Sciences (DILS), 2014[C].
[5] PlantExperimentalAssayOntologyImported.owl[EB/OL].https://bitbucket.org/PlantExpAssay/ontology/src, (2014-12-19)
[6] Experimental Factor Ontology [EB/OL]: http://www.ebi.ac.uk/ols/ontologies/efo, (2017-06-16) [2017-06-27].
[7] Trokanas N, Bussemaker M, Velliou E, et al. BiOnto: Ontology for Biomass and Biorefining Technologies [J]. Computer Aided Chemical Engineering, 2015, 37:959.
[8] Classe T, Braga R, David J M N, et al. A Distributed Infrastructure to Support Scientific Experiments [J]. Journal of Grid Computing, 2017:1-26.
[9] iDigBio[EB/OL].: https://www.idigbio.org/, [2017-07-10].
[10] Morphbank: Biological Imaging [EB/OL]. http://www.morphbank.net/About/AboutMb/, [2017-06-10].
[11] Catalogue of Life [EB/OL]: http://www.catalogueoflife.org/, (2017-06-30)[2017-07-12].

[12] N8 Biohub Information and Knowledge Management System [EB/OL]: http://gtr.rcuk.ac.uk/projects?ref=101508, [2017-06-02].