THÜNEN
Digitalization sponsored
by Thünen-Institut

TRAWLBASE, a configurable database for trawl survey data

G. Garofalo - L. Cannizzaro - D. Levi

Istituto di Tecnologia della Pesca e del Pescato
Consiglio Nazionale delle Ricerche
Mazara del Vallo (Tp) - Italy

Abstract

This work presents an advanced software environment which is aimed at defining and managing databases for data from trawl surveys. In particular, it describes the conceptual model adopted for organizing the data, and the basic characteristics of the management system.

Introduction

This work describes the architecture of an advanced software environment, called TRAWLBASE, which is based on a conceptual database model for data collected during trawl surveys. The system was designed with the aim of helping the user to define, create and use customized database applications in the above mentioned context.

Our project derives from the central and critical role which the data storage, processing and retrieval phases play in a trawl survey programme. These programmes engage considerable human and financial resources and produce huge quantities of data. The value of this data is greater the more effectively it is made available for suitable analysis and processing.

A proper project for a database management system must therefore take into account the need to keep the dimensions of the database within limits which ensure reasonable performances, optimize data storage and correction operations, and provide the most flexible possible output interface so as not to limit the potential use of the program.

TRAWLBASE adopts a particular approach towards solving the above mentioned problems and integrates some functions which make it particularly flexible and adaptable to the needs of different users.

1

Conceptual model of the database

TRAWLBASE can be considered a "shell" for developing a database application for trawl surveys. As such, it only contains the description of a conceptual model of the database; i.e., it gives the organizational outline of the database files but not their content. The user has to fill this empty structure according to his needs.

The model proposed for data organization is a relational type; it includes two types of structure which are different in their characteristics and information content:
1) data files
2) reference files

The data files are the structures which carry the actual datum where "datum" means the basic information taken directly from the surveys. They cannot contain aggregate data. TRAWLBASE allows the use of three types of data files for each trawl survey, which are different to each other not only in their content but also in the type of handling they require:
1) haul file
2) catch file
3) species files

The haul file is made to contain all technical and operational information regarding the performance of the haul, such as, for example, the geographical coordinates, the duration, the speed, the depth, etc. It consists of one record for each haul in every survey. Only one haul file can be created for each survey.

The catch file is structured to contain the total catch data for each haul. It therefore has one record for each haul and each species and only one file is created for each survey.

Finally, the species files contain the definition of all the biometrical parameters relating to the target species. Each file has one record for each specimen in a given haul. The number of files of this type is equal to the number of species caught and processed in each trawl survey. Their structure, however; does not necessarily have to be the same. In fact, the choice of defining groups of one or more species characterized by different sets of parameters is left up to the user. A different species file structure corresponds to each of these groups.

The references files are, on the other hand, structure which help in arranging and codifying the information related to some of the parameters which make up the data files. They cannot contain data for processing but carry out some important tasks; they provide an automatic check system on data input regarding the codified parameters; they allow global or selective data file record searches; they ensure efficient use of mass storage without sacrificing the clarity and integrity of the information.

A further class of files is also automatically created when the program is installed. These files are used to store information relating to the configuration of the system, the structure of the database; the state of the data files and all the support information for handling the database in general. Some of these files are automatically updated with every database work session. Their creation and handling take place in a totally user-transparent way.

System architecture

The creation of an automatic data management system takes place in the following three main phases:
1) analysis of the requirements, definition of the database structure, and definition of the content of the data files;
2) creation of programs for storing and correcting data;
3) creation of programs for accessing the data.

Each of these activities corresponds to one of the three modules which the TRAWLBASE architecture can be visualized as consisting of: the Application Generator, the Application Administrator and the Query Interface.

Normally a database management system does not contain routines which implement functions relating to the first phase of the project. As was already mentioned, one innovative aspect of TRAWLBASE is that it gives the user partial but significant access to these functions. The Application Generator module allows the user to select the set of parameters which is relevant to his aims and guides him in configuring the record format for each type of file. This involves defining the name, type length and admissible range of values for each record field. During this phase, all fields which are associated with codes are established by defining reference files which can also be easily called during the data storage phase.
The advantage of this approach is that the user can configure a multi-file relational database which meets his needs without having to go into specific implementation details, since these are automatically handled by the program.

The second module, the Application Administrator, consist of input and data correction routines. The main characteristic of this module is the possibility of dinamically creating screen formats for dispalying the data based on the set of parameters which make up the structure of the data files. In practice, by using a screen format definition function, the user can change the number and the order of the parameters any time he wishes to exclude non measured parameters or selectively edit data already stored. The advantage of this approach is the possibility of later changing the number and type of parameters which are to be studied without having to alter the structure of the files.

The third module implements the data file inquiry functions. Although it is last in the development process, this module carries out a fundamentally important task in the database management system since it determines the real possibilities of using the information contained therein. Its basic functions must ensure the possibility of extracting data in aggregated forms and obtaining an output in the required format (chosen from a set of options). The TRAWLBASE Query Interface satisfies these requirements by using utilities which allow the user to create output files whose fields can be defined as operation fields: i.e, they can be the result of elementary arithmetical operations carried out on parameters in the data files. In this way, the user can extract information from database in a personalized "private" structure which meets the specific aims of the analysis to be carried out. It should be noted that it was decided not to allow aggregate data to be defined in the data files, but to calculate it dynamically when the database inquiry is made.

The increased complexity of the data access program slows down
the calculation time, but drastically reduces the mass storage
occupied by the data, which can be a critical factor in this type
of application.

Hardware and software requirements

The system has been created for an IBM-compatible personal
computer with MS-DOS.
The language used for its development is Clipper 5.0. It is a
DBMS of tha Dbase family and therefore the files have a DBf
format. Since it is also a compiler, the programs developed in
this environment do not require particular software in order to
run. No hardware interface is required either.
The minimum hardware configuration required is 1 Mb RAM and a
hard disk whise dimensions depend on the characteristics of the
trawl surveys. It is estimated that, on average, a trawl survey
of 100 hauls requires 5 Mb of mass storage.

Development state of the system

From the implementation point of view, the state of progress
of the system is as follow:
    - the Application Generator subsystem is complete as far as
the functions described in this work are concerned; however, the
possibility of extending the structure of the database cannot be
excluded;
    - the Application Administrator module has been completely
created, and its operation does not depend on any changes in the
other two modules;
    - the Query Interface module at the moment ensures only the
minimum data access functions. It is in the process of being
defined as regards its more sophisticated functions.

Conclusions

TRAWLBASE is both an advanced environment for trawl survey
database management, and a system which guides the user in
defining the database and creating customized application
software at the same time.
    As a database management system, it has the following
characteristics:
- it is complete and very flexible regarding the data flow
management functions;
- it is simple to use;
- it does not require the user to know programming
- it is designed in a modular way so that it is open to future
development;
- it is compatible with network architectures which allow
centralized data files to be created with several peripheral

accesses.

As an environment for the creation of database applications, it is designed for research laboratories which undertake trawl survey programmes. Thanks to its open configurable structure the system is widely adaptable to different needs.

The present system is in use at the Istituto di Tecnologia della Pesca e del Pescato del Consiglio Nazionale delle Ricerche which is based at Mazara del Vallo (Italy). The institute has been carrying out trawl surveys since 1985 in the Italian part of the Sicilian Channel. In particular, from 1985 to 1992, it completed 16 trawl survey for a total of about 1500 hours of trawling. TRAWLBASE is the result of a progressive development of an initial database model, which has taken place during the years the institute has been active in this field.

Also since 1985, analogous trawl surveys have been carried out in all Italian seas by a further 14 units which operate within the Italian national fishing plan. Furthermore, the European Economic Community has promoted the setting up of international trawl surveys in all parts of the Mediterranean sea which belong to EEC countries as part of the EEC fishing policy.

We believe that, with the prospect of creating common databases, TRAWLBASE could provide a data representation standard and a common working environment, and at the same time ensure complete compatibility with autonomous research programmes.

Bibliography

Cannizzaro L., (1988) - Struttura dei Databases TRAPES e TRALAB - Nota tecnica ID/TN/LC/3/11/88, 1988

Ceri S., (1980) - Progettazione di basi di dati - Clup Milano, 1980

Date C. J., (1985) - An introduction to Database Systems - Addison-Wesley - pub., 1985

Garofalo G., L. Cannizzaro, D. Levi, (1991) - A first note on the database organization of the program T.R.A.W.L. - Technical report, NTR-ITPP n° 36, 1991, pp. 33

Levi D., (1989) - Programma TRAWL: Procedure e modelli raccolta dati a partire dal 9° survey Nota tecnica ID/NT/DL/3/0189, 1989

Martin D., (1987) - Banca Dati - Tecniche nuove, 1987

Teory T. J., and J. P. Fry, - Design of Database Structures - Prentice Hall, 1982

Straley S. J., (1990) - Programmare in Clipper - Addison-Wesley pub., 1990

Ullman J. D., (1982) - Principles of Database Systems - (2nd ed.), Computer Science Press, 1982