

# Design and Functions of DUO: the First Italian Academic OPAC

Maristella Agosti\* and Maurizio Masotti†

\* Department of Electronics and Informatics, University of Padova

† Computer Centre, University of Padova

## Introduction

This paper presents the main features of an Online Public Access Catalogue (OPAC) [5] called DUO, which stands for Dialogue between University Users and Online SBN Database. DUO is the first Italian academic OPAC developed to provide direct access to the co-operative online catalogue database of the Universities of Padua, Venice, and Verona to final users of different university sites. The maximum distance between these universities is about a hundred kilometres, also these three universities are all in the same region: the Veneto Region that is situated in the North-East of Italy. At present, forty-three libraries of these universities are cataloguing and updating the co-operative catalogue database.

The online catalogue database has been created and is currently managed by one of the software systems which have been designed and developed under the Italian National Project for library automation: the SBN project (SBN stands for "Servizio Bibliotecario Nazionale") [7,8].

SBN software systems have been developed to implement the same specifications in order to give the same library automation functionalities to all librarians of the libraries participating in the project. These software systems have been primarily designed to interact with a professional user: the librarian. Due to this, the functionalities which have been made available are those usually carried out by the librarian himself, e.g. cataloguing, acquisition, serials control, and subject indexing.

In a university environment, as that previously depicted, it is not possible to envision all the different types of end-users who interact with the librarian, or to continue using only traditional cataloguing techniques (e.g.: card catalogues) which make localisation of books and serials available in different libraries within the same university or from different universities very difficult if not impossible. For this reason in early 1989 the necessity to launch a project for the development of a prototype of an OPAC to make the SBN catalogue more easily available to end-users was evaluated.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1992 ACM 0-89791-502-X/92/0002/0308...\$1.50

In mid-1989 a two-year project for planning and development of an experimental prototype of an OPAC for university users was set up. At the beginning of June 1991 the OPAC prototype was available to university users for experimental use. At present log and other types of data concerning the interaction between user and system are being collected in a systematic manner in order to monitor the prototype and to consolidate it into a tool satisfying end-user search requirements.

The paper is organised as follows: next section introduces the main project scheme, afterwards the innovative aspects of the architecture of the OPAC prototype and also of the database used by the system to interact with the users are presented. Querying and browsing facilities are detailed introduced, and at the end some considerations on future developments are given.

## Guidelines of the Project

The OPAC design and development project has been divided up into three main stages:

1. identification of end-user categories and analysis and design of end-user requirements [11];
2. design of functional specifications of an OPAC for university library end-users;
3. design and development of the OPAC prototype.

The first step to be addressed was dedicated to the identification of end-user categories and to analysis and planning of end-user requirements [11].

In order to identify the different end-user categories of university library facilities, an initial analysis had been carried out to schematically distinguish these categories: students, PhD students, researchers, and professors. After completion of this first target, it was decided to undertake the analysis and design of the end-users' requirements by means of interviews. A representative sample of the identified end-user categories was designated, and the corresponding interviews were carried out.

During selection of the sample, the end-users' discipline of interest was taken into consideration because it was necessary to find out the different possible end-users' search attitudes according to the specific subject of study: humanitarian studies, law, science or engineering. It was decided to add to the sample some librarians who work in Faculty libraries, due to the fact that generally this type of librarian often supports students in

their document searches, therefore it was necessary to include their experience of students' information search attitudes also.

The results of this work have been fully reported in [11]. We report here those results which have directly influenced the architecture of the prototype.

Results have shown that in order to identify the end-users' search attitudes it was important to find out:

- the state of development and settlement of the subject of study;
- whether the end-user is engaged in applied or theoretical research work;
- whether the subject of study is of interdisciplinary nature or not;
- the necessity to use documents coming from one or more countries (i.e. for some ancient historical studies the references are almost all written in one language and come from one single country, whereas for computing science studies the documents come from various different countries);
- the speed of document obsolescence.

End-user requirements that have emerged from the interviews are as follows:

- the end-users' expectations in terms of system interaction are similar to those of present-day on-line database query languages typically used for second-generation OPAC document attributes and for subject searching [4];
- the use of a unique, controlled vocabulary for subject indexing seems to be quite difficult, due to the past usage of specific indexing language at library level;
- the interest in searching through university library databases seems more interesting for interdisciplinary and applied study subjects;
- all end-users expect data concerning the presence and availability of documents in the library system to be accurate.

Taken into consideration these results on user requirements, these have been set out as guidelines for the planning and implementation of the OPAC database and system [1]:

- when the project will be fully operative the end-user will only have to interact with one single software interface in order to make use of different information sources, therefore the OPAC tool needs to be implemented for operation through the same interface used to access other tools such as on-line databases, local CD-ROMs, e-mail sources, etc.;
- the OPAC tool needs to be modular, in order to permit expansion to be carried out in the future to introduce new capabilities;

- the design of the database and of the software tool needs to support different subject indexing procedures [15];
- end-users need to have a user-friendly and "intelligent" interface [2];
- the end-user must be provided with the possibility of printing out search results employing the standard approach used for producing bibliographies or lists of references, as well as to retain his or her own search strategies;
- mechanisms for storage of useful data coming from search sessions have to be included for subsequent evaluation [6,10,12];
- it will have to be possible to use the OPAC tool from any terminal of the computing network in the universities, that is the use must not be restricted to graphics or other specific terminals.

## System Architecture

This section presents the architecture and characteristics of the OPAC database designed to implement the project guidelines introduced in the previous section; it is important to bear in mind that one important target of the OPAC database design was to avoid data duplication logically integrating SBN and OPAC databases. This is an innovative aspect of this project, because at present it is quite common to see OPAC implementations operating on specific databases which duplicate that data of the cataloguing database which is considered informative for the end-users.

It is imperative to recall here that the main feature of the SBN database is to contain data for administrative and cataloguing purposes. Because of this, all query access operations that have been implemented by the SBN software permit a pre-defined access only. This means that it is not possible to obtain textual attributes through single words or of single word expressions connected by Boolean operators (post-coordinate access [9]).

Among the items available in the SBN schema, only these have been deemed useful for construction of the OPAC database:

- the title;
- the "nature" (or type) of title;
- the terms of the controlled vocabulary, developed by the National Library of Florence, directly related to the document title;
- classification subject headings directly related to the title;
- name(s) of document author(s);
- document language;
- the country in which the document was published;

- date type and date of publication;
- the type of publication.

Document abstracts and indexes, or any other form of semantic representation of document contents are not included in the SBN database, so it was not possible to base access operations on a "by content" principle designed for the OPAC on this data.

The architecture of the OPAC database was focused on making OPAC information retrieval [13,14] capabilities as efficient as possible; for this reason it was necessary to program the database and the software to work efficiently for words and terms access operations on the textual data [3] of the SBN database which have been previously mentioned. Since the SBN database does not contain specific data for word searching in the textual document attributes, it was necessary to decide how to make the necessary data available in the OPAC database.

We decided to avoid a total duplication of the SBN files and we physically designed the OPAC database to logically include a controlled vocabulary in order to make efficient access to the textual attributes of the SBN (or cataloguing) database. The software tool which is used for management of library functions makes use of the ADABAS database management system and of the primitives of the NATURAL language; the primitives of textual access are made available by a TRS product. All these three software products, that is ADABAS, NATURAL, and TRS are property of the SAG Company.

One important engineering problem to solve was that concerning updating of the OPAC database which needs to be done in sequence with the SBN database. We adopted the solution of accessing the OPAC database for search purposes only; after a query has been formulated and solved, the OPAC prototype accesses the SBN database in order to make the bibliographic data available to the end-user; in this way the bibliographic data displayed to the end-user is always consistent with the most recent SBN database updates.

The most difficult task in the development of the OPAC prototype was the construction of a controlled vocabulary [9] in the multi-disciplinary and multi-language environment of university library systems in order to provide for the advanced and sophisticated searching capabilities reported in the next section.

The controlled vocabulary is predominant to the functioning of the prototype and is generated and maintained by means of the TRS primitives. The data managed for each word includes:

- the word itself;
- the word written backwards (i.e., for the word "vocabulary" for example, the sequence of letters which is maintained is "yalubacov");
- word root(s);

- word aspect(s), where by the term aspect we mean the role the word plays in the dictionary; one aspect used in vocabulary construction was: "stopword", that is a word not considered useful in the generation of the OPAC database and which cannot be used in query formulation (e.g., some words often considered as stopwords in English are: about, as, but, for, of, the, with); another aspect which has been used in the construction of the database is the language of origin of the word: to each word we associate its language; with this information it is possible to design the partitions of the complete vocabulary for the different document languages in order to manage a multi-lingual document collection; and
- word synonym(s).

## Querying and Browsing Capabilities

The capabilities of the prototype can be grouped into three main classes:

1. **query formulation and searching**, with presentation of available data relating to retrieved documents,
2. **retrieval aids**,
3. **help facilities**.

### 1. Query formulation and searching

DUO implements two different techniques for end-user interaction: the first way is intended for inexperienced end-users who have never used the tool before or who will use the system only occasionally; the second type of interaction has been designed for end-users who use the system regularly or already have experience in on-line database query operations.

In the first interaction method the end-user is guided through his or her query formulation and searching by a sequence of self-explaining menus with fill-in masks; the second method lets the user employ Boolean operators for query formulation and searching; in both ways the end-users have direct availability to innovative retrieval aids and help facilities of DUO.

At present the retrieval process consists of two different steps: the initial search and subsequent search operations. While the initial search asks the end-user to make a "blind" query formulation, the subsequent search operations start after the end-user has examined the documents retrieved in response to the previous query.

The initial query formulation reveals three possible searching techniques, depending on the kind of information which is needed by the end-user and also on the data he or she has concerning the documents he/she wants to find:

- a) searching through a specific document attribute;
- b) searching for a set of documents which all have the same value of a specific document attribute;

c) subject searching.

One of the scopes of our prototype is to investigate new solutions for subject searching in Italian University libraries. With this aim in mind, in the prototype's current version, though it is possible to perform search operations on a specific document attribute, the particular structure of the OPAC database allows search operations using all the document attributes of the cataloguing description concurrently, and the cataloguing description can contain any semantic data.

The end-user can ask the system to find one or more words anywhere in the textual segments within the OPAC database: this particular retrieval behaviour is typical of someone who has a subject in mind on which he or she needs information, but does not know the specific title of an actual book or publication dealing with that subject.

The logical textual segments which have been created for subject searching using all the document attributes of the cataloguing description concurrently are:

- the document's author's name;
- title;
- the controlled vocabulary developed by the National Library of Florence which is used in subject indexing; in the foregoing the expression: "subject headings" will be used to refer to this controlled vocabulary;
- classification subject headings, and
- the publisher's name.

The *document author's name* allows free text searching. The option for word searching on the author's name is particularly important for author's corporate names. The cataloguing rules defined in the SBN database consider the pseudonyms and the acronyms of an author's name in a special way: they are linked to the author's name in the authority file but they are not linked with the documents related to the author. The OPAC database puts together in the same textual segment all the words which form the original author's name, the pseudonyms and the acronyms, maintaining the Internal Sequence Number (ISN) of the record in the SBN database which contains the full, original author's name. The prototype is able to find all the author's works even if a query formulation contains a pseudonym or an acronym.

In the SBN database the *titles* are related to each other in a hierarchical structure or at an equivalent level (i.e. titles which are related to each other), depending on the nature of the title. Let us give an example: a monograph can be related at a higher level due to the monograph title's relation to the title of the collection it belongs to; the same monograph can also be related to a title in its original language version. Among all these titles only the monograph title is linked to the physical book

it represents and to the circulation data (data concerning its availability, etc.). The prototype again gathers all the words which form the titles related to the one linked to circulation system and puts them in the same textual segment. In this manner the hierarchical order is lost but the set of words which could directly lead to data in relation to availability of the actual book is enriched. This operation is justified also by the fact that these hierarchical links which have some meaning for the library staff are generally unknown to the end-users of a library.

The *subject headings* also have hierarchical links: one subject heading can be linked to a title as well as to other subject headings. As for titles, these links set on different levels are smoothed down onto same root level of the tree of relationships. The same operation has also been done for the *classification subject headings*. These links differ from those of the titles, because they mainly depend on the subject matter of the document and the descriptions relating to contents in the hierarchy become more and more specific as we move down to the lower levels. It can be useful for end-users to be able to view the hierarchy of subject headings or navigate along the classification subject headings tree: this is still possible on the original file of the SBN database. In fact, the files created within in the OPAC database are complementary to the original ones of the SBN database: while the former improve the possibilities of retrieving a set of relevant documents, the latter assure navigation throughout the linking network involving titles, subject headings and classification subject headings.

The library staff of the library system at the University of Padua is not forced to adopt a common, sole, subject-headings and classification scheme which would be unique for all the libraries, but there is sufficient freedom of choice of a method which is semantically closer to the specialised documents of each library. This can be an advantage for the end-user who is searching in a specific library collection, but it becomes a disadvantage when an end-user wants to run a subject search throughout the entire collections of one or all of the universities in the Veneto Region. The solution which has been implemented in our prototype involves use of the *classification subject heading terms* just for their very meaning, but providing the OPAC database architecture with multiple logical segments in order to distinguish different classification systems.

Also for the *publisher's name* a specific segment is prepared following the rules used for creation of the author's name segment. It is therefore possible to conduct a free, text-search operation also on the publisher's name.

The *initial search* can be constructed specifying in the query some precise values of document attributes, such as data concerning publication, language of publication, or the particular library in which the end-user wants to look for the documents.

The initial query can also be constructed in an entirely different way in order to make it possible to set up the database for subject searching operations combined with a deterministic query on one or more specific document attributes. This retrieval method is not normally used by end-users in initial search operations, owing to the fact that it supposes that the end-user knows exactly what he or she is looking for. This possibility can however become very useful in *subsequent search* operations: in fact, the structured attributes of cataloguing data, language and library can be used by the end-user as "filters" to reduce the number of documents retrieved by the initial or previous query.

If the *initial query* retrieves one or more books, the prototype lists the results in two steps:

1. concise descriptions of all retrieved titles;
2. extended description of each title in the set.

The end-user can select one or more titles from the concise list (step 1) and obtain the screen display of all pertaining data (step 2): information concerning the complete bibliographic description and the description of related titles, the subject headings and the classification subject headings, the names of all the libraries which have the book or that have just ordered it can therefore be obtained.

The prototype gives three different print options for printing out data concerning the documents retrieved:

- a) a concise data printout of all retrieved documents;
- b) a concise data printout of some documents chosen by the user;
- c) a concise data printout of all or some of the retrieved documents which is made following a few specific bibliography construction rules.

## 2. Retrieval aids

The retrieval aids are: on-line dictionaries, and retrieval path history.

The *on-line dictionary* feature has been provided for in order to support the end-user in the formulation of the query and during browsing, because one *on-line dictionary* gives a further possibility of finding exact access points within the vocabulary of words and in the list of pre-defined codes.

Specific on-line dictionaries of words are available for words contained in the: title, document author's name, subject headings, and classification subject headings.

When an end-user selects a word in one specific dictionary, the word is first searched for within the selected segment: if it is not found in that dictionary, DUO automatically checks if it exists in any other textual segment, if so, the relevant dictionary is then displayed to the end-user.

When the word is entered by the user of the system, a stemming algorithm performs an automatic truncation

of the word which is sought and lists the set of words beginning with that stem, each one with the number of occurrences in the titles. If the end-user selects one or more words from the dictionary, the system formulates the equivalent query in the syntax accepted by the system.

The *retrieval path history* feature provides information on the end-user's individual interaction history with the library catalogue database. This feature is available at any moment during the interaction: it gives a text display of previous queries and the number of corresponding titles which have been retrieved. As for the other features of the prototype, the retrieval path history is currently under further development and will give the end-user the possibility to restart a new search operation from a particular set of retrieved documents in response to a previous specific query.

## 3. Help facilities

The prototype has been designed to be user-friendly; due to this feature, except for the difficulty of understanding the meaning of Boolean expressions in the query formulation, all the options are easy to use and they are always followed by help messages.

As has been previously underlined, the DUO system must be made directly usable by untrained and inexperienced users. Help facilities are therefore very important, because they have to teach end-users how to use the system while they are using it.

Two types of on-line help options have been provided for in the prototype:

- a) specific contextual help;
- b) general help on prototype use and features.

Many help messages also give examples of use of the prototype's facilities. Also some sort of monitoring on incorrect attempts of the end-user has been implemented in the prototype; in fact, after a series of subsequent errors made by the end-user during interaction with the system, the help facility gives advice for an alternative search strategy.

The help messages can always be requested by the end-user. At any rate, in certain circumstances the help messages are automatically displayed by the system; for example, if the end-user makes use of a specific prototype feature for the first time, the system prompts a contextual help message.

## Further Developments

The prototype has been made available to users starting from June 1991; general impressions and initial logging data are currently being collected for subsequent analysis. Evaluation tools of the suitability of the features provided in the prototype and efficiency of the chosen software tools are presently under way.

In parallel with the planning of the prototype evaluation

tools, we are at present designing new features to implement in the next version of the prototype, two of these new features will be: automatic selection of alternative retrieval paths when the set of retrieved documents is empty; in "subsequent search" operations it must be possible to get the system to find other documents similar to that (or those) just retrieved.

## Acknowledgements

The initial version of the DUO prototype was designed and developed with the kind collaboration of Anna Maria Moressa, who is now working with another association.

There are many people who are constantly participating in this vast project; we would like to thank all the end-users who have participated in the design of the tool through our interviews and useful discussions. We would like to thank also Franca Gambato, Anna Maria Fornea, Anita Zorzi and Alberta Panti, who all work at the Computer Centre of the University of Padua, for their support and cooperation during the development of the project. The librarian of the Department of History at the University of Padua, Loredana Menapace, deserves a special mention of thanks for her continuous advice concerning the bibliographic characteristics of the SBN library automation system.

This work has made possible thanks to the support of the Committees of the SBN project of the Universities of the Veneto Region and the Library Automation project of the University of Padua.

The work has been partly supported by the Veneto Region authorities.

## References

- [1] M. Agosti, M. Masotti, A.M. Moressa. *Linee di progetto*. Progetto di un prototipo di catalogo in linea per l'utente finale (OPAC) per l'università - Base dati del Polo SBN Universitario Veneto, novembre 1989, pp. 13.
- [2] N.J. Belkin, P.G. Marchetti. "Determining the functionality and features of an intelligent interface to an information retrieval system". *Proc. of ACM-SIGIR 1990 Conf.*, Brussels, Belgium, 1990, 151-177.
- [3] D.C. Blair. "The data-document distinction in information retrieval". *Communications of the ACM*, 1984, 27 (4), 369-374.
- [4] C.L. Borgman. "The user's mental model of an information retrieval system: an experiment on a prototype online catalog". *Int. J. Man-machine Studies*, 1986, 24, 47-64.
- [5] C.R. Hildreth. "Online public access catalogs". *Annual Review of Information Science and Technology (ARIST)*, 1985, 20, 233-285.
- [6] C.R. Hildreth. "General introduction; OPAC research: laying the groundwork for future OPAC design". In: C.R. Hildreth (Ed). *The online catalogue: developments and directions*. Library Association, London, 1989.
- [7] Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU). *Il Servizio Bibliotecario Nazionale*. Roma, 1985.
- [8] Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU). *Guida alla catalogazione nell'ambito del Servizio Bibliotecario Nazionale*, Roma, 1987, Vol. 1.
- [9] F.W.Lancaster. *Information Retrieval Systems: Characteristics, Testing and Evaluation* (2nd ed). New York, 1979.
- [10] C.A. Lynch. "Applications of performance and usage data for online catalogues". In: [6], 127-141.
- [11] A.M. Moressa. *La fase di analisi dei requisiti utente - Relazione sulle interviste raccolte*. Progetto di un prototipo di catalogo in linea per l'utente finale (OPAC) per l'università - Base dati del Polo SBN Universitario Veneto, settembre 1989, pp. 15.
- [12] D.N.J. Mostert, J.H.P. Eloff, S.H. Von Solms. "A methodology for measuring user satisfaction". *Information Processing & Management*, 1989, 25 (5), 545-556.
- [13] G. Salton, M.J. McGill. *Introduction to Modern Information Retrieval*, McGraw-Hill, 1983.
- [14] C.J. Van Rijsbergen. *Information Retrieval* (2nd ed). Butterworth, 1979.
- [15] S. Walker. "Improving subject access painlessly: recent work on the Okapi online catalogue projects". *Program*, 1988, 22 (1), 21-31.