CoBRA/RUG: Expert System for User Queries

Fred Bosman, Frank den Hollander, University Library, Groningen
Leo van Maanen, COWOG, University Groningen, Netherlands

Abstract: The University Library Groningen (the Netherlands), COWOG (centre for research on higher education) and PICA (the Dutch organisation for library automation), developed a computer assisted bibliographic reference and advisory system (CoBRA). It is an expert system that advises users of the University Library when they want to execute a search for literature on a certain subject. In fact, it produces custom made guides to the literature in the library. Explained is why the system has been built, the architecture of the system is outlined and some comments are made on aspects of implementing such a system, and the need for library cooperation to build expert systems.

1. Introduction

CoBRA, an acronym which stands for Computer Assisted Bibliographic Reference and Advisory system, is an expert system that advises users in the University Library of the State University Groningen in the Netherlands. The system was developed in three years by a small team of experts in the field of library and information science, in cooperation with the library’s subject specialists. It was a joint venture in which PICA, the Dutch organisation for library automation, COWOG, the centre for research on higher education, participated together with the University Library. In this paper we will first explain why the system is built and comment on the possible use of expert systems for subject queries. After that the architecture of the system will be outlined. Attention will also be given to some problems that were encountered while building and implementing the system in the library. Finally some remarks will be made about the possibilities for a further development of the system, about the necessity for library cooperation in this respect and about the role of the subject specialist in libraries which use expert systems.

2. Research questions: an outline of the problem

A patron in an (academic) library wanting to perform a literature search on a certain subject is confronted with a rather complex world. He has a large number of instruments at his disposal to help him find what he is looking for: reference books, collections in open access, catalogues, CD-ROM’s, online databases etc. A search for literature must be executed in a sound and efficient manner, with sources that have their own structure and rules for consultation. Libraries provide an assortment of facilities to assist these users, ranging from printed guides, instruction books to help desks and reference departments. Automated facilities which offer guidance for users have not yet been developed on a large scale. The dominant trends in library automation can be found in other areas, like electronic publishing, network building, delivering of documents etc. Reference work in libraries...
still has a traditional character, a patron formulating a question and a librarian formu­
la-ting an answer to his best knowledge.
Reference work has often been described as the essence of librarianship. Even if this is
not true, it can contribute greatly to an efficient use of collections and other facilities.
A lot of questions asked in libraries are not complex. It is sufficient to refer to another
department, to help with the OPAC, to tell about opening hours and facilities. One type
of question is usually difficult to answer: the question for literature on a subject. We
mentioned it already: there exist many, potentially useful sources, and often the search
involves several stages which need to be worked through in a sound and theoretically
good manner.

How good is the quality of reference work, especially for the questions for literature on a
certain subject? Hernon and McClure (1987) tried to find the answer using a rather
controversial method: volunteers posing as library patrons submitted enquiries to
reference librarians and noted the answers. The (shocking) results of their experiments in
American libraries, largely confirmed by many other studies, revealed that:
- factual and bibliographic enquiries received a complete correct answer only 50-60 % of the time;
- the average time spent on a reference question was no more than five minutes;
- conducting a proper reference interview and formulating an effective search
  strategy proved a problem;
- if the answer was not found, the patron was not referred elsewhere.

It can be doubted whether these results are representative for the reference work uni­
versity libraries perform in response to subject queries. The questions used by Hernon and
McClure, and most other researchers, differ in at least one major aspect from the kind of
questions most often asked in university libraries. The test questions were requests for
factual information or simple bibliographical searches and the correctness of the answer
was almost always easy to measure. In university libraries, however, requests for
complex subject searches dominate among the subject questions. That is by itself no
reason to expect a better performance, on the contrary.

Before the reference librarian can help with the selection of sources and the formulation
of a search strategy, he must understand what the patron really wants. That makes the
reference interview a key stage in the entire process. It may take up little time, still it is
an interaction involving a complex task for librarians, especially those trying to provide
the "liberal or maximum service" advocated by Katz (1982) and most other librarians
today. According to this standard, the librarian either provides the correct answer or
points out the sources where the answer can be found along with the necessary instruc­
tions for their use. The complexity of this task is due to the complexity of the sources:
extensive subject catalogues, sometimes thousands of bibliographies and other works,
printed or online, and many other relevant sources and facilities. The librarian must be
familiar with all of these and know the most important ones thoroughly.
The patron, especially a student in a university library, generally possesses a limited
acquaintance with the resources, facilities and possibilities of the library. As a result, he
tends to formulate his enquiry in such general or vague terms that the librarian is unable
to (immediately) give a specific answer (Yahoda and Braunagel, 1980, devoted their
whole book to the problem). Kahn pointed to the aforementioned unawareness of the
patron as the most important cause of the problem. Mount (1966) emphasized a series of
psychological elements in the interaction between the patron and the librarian. An additional problem for university libraries is the fact that a large number of the patrons are students working on assignments with little or no knowledge of the subject they are researching. Sometimes the definite object of their research has not yet been determined. The result is that the reference librarian must choose a search strategy as open-ended as the patron’s subject or the reference interview is used to limit and redefine the topic.

A number of authors have developed models for reference interviews with the emphasis on obtaining the information needed to choose the proper sources. Only subject questions, more specifically the research questions will be dealt with here. What must be determined, given the assumption that the patron in a university library usually performs the search himself?

1. The object of the enquiry - as precise as possible including the discipline-related context and, if relevant, limited by land and period.
2. The kind of information desired. The answers to most subject searches can be found in books, reports, journal articles and/or reference books, but the question remains which type is needed, how much information is required, how recent must it be, on what level and in which languages.

Furthermore, the librarian’s answer must take into account what the patron already knows. After all, many enquiries at the reference desk are made (long) after the searching has begun.
The help given must be complete and correct (Yahoda and Braunagel, 1980 p. 94 ff) which implies that locating the proper sources is not enough. Information about their use and the search strategy to be followed is also required.

This sketch of the reference interview is complete. It is clearly a complex task demanding expertise from the librarian not only in the subject area but also in interviewing skills. The conclusions drawn by Hernon and McClure in their study no longer seem surprising, nor is it likely that the performance would improve for the more complex subject questions.

3. Expert systems for user queries

Interest for using computers to solve problems that are normally reserved for experts originated in the 1950's. In the beginning the expectations were high, today they are more realistic, due to the fact that the performance of experts is mostly superior to that of expert systems because of the role played by intuition and similar mental processes. These processes contradict computer logic but play a large, sometimes crucial role in the way experts make decisions (Dreyfus and Dreyfus, 1986, pp. 101-122).

Under the following circumstances expert systems can be designed and put to good use (Dreyfus and Dreyfus, 1986, pp. 120-121. Alberico and Micco, 1990, p. 110):
- the problem must be of a kind that no algorithmic solution exists
- a performance inferior to that of an expert still results in a correct or otherwise satisfactory result
- nonexperts working in a traditional way usually make mistakes
- those mistakes have serious consequences
- the problem undergoes little or no change while the computer system is being consulted
- the knowledge domain is relatively static
- an expert is thoroughly involved in the design
- management and staff support the application of computer systems for this goal

A number of tasks in libraries, among them reference work, satisfy the above conditions. The following points are worth considering also:
The library patron's questions are not usually answered by an expert in his research topic but by an expert in another field or a semi-professional. The knowledge domain is sufficiently static but also very extensive, a fact which often places non-experts for large problems. The failure to solve these problems results in an inefficient use of expensive facilities.
The use of expert systems provides several definite advantages. A number of routine tasks are eliminated, expertise is available even when the expert is not and the library acquires a new training tool for its staff. In addition knowledge accumulated from several different experts is stored and problem solving approaches are tested and refined.
Information scientists are showing more and more interest in the development of expert systems and their application in reference work, indexing, abstracting, cataloguing and assisting the end-user with online information retrieval. A survey of these developments can be found in Morris (1992).

Most of the expert systems developed till now for use in libraries provide an extensive coverage within a limited domain, for example a discipline. They are often very sophisticated, but seem to have little practical value. Often they were designed for exploring the possibilities for using artificial intelligence in a limited domain. In those cases where a practical goal such as library reference work was pursued, the coverage was often limited to a certain type of sources.

In our opinion the development of expert systems designed to assist users doing subject searches and its employment in an academic library with a general collection is possible. The CoBRA system proves that possibility. Such a system must satisfy the conditions already mentioned, and also the following preconditions:

- the type of documents referred to must be limited. Otherwise the database will grow too large and become insufficiently static. The maintenance costs would then be prohibitive.
- it must respond to the concrete needs and wishes of librarians and end-users whose number and importance justify the development, maintenance and utilization of the system.
- the design and construction of the system must use those characteristics of reference work and literature study treated earlier.

Given the above conditions, it seems preferable to limit the type of documents included in the database to reference works in the broadest sense.
4. The CoBRA System

CoBRA is developed to solve the different kinds of problems that librarians encounter in answering user queries. The leading concept to solve these problems was developed by Yahoda and Braunagel (1980, pp. 52-61). They used the concept "lead-in-tools" for those instruments to be used when a librarian does not know which specific sources are needed to answer a patron's questions. Different kinds of such tools exist, catalogues, bibliographies and literature guides. However none satisfy the following requirements for an ideal instrument:

- to contain all potentially useful sources
- to evaluate the listed sources
- to provide the location of the sources in the library
- to be indexed for quick and efficient use.

Each of the above mentioned "lead-in-tools" satisfies one or two of the criteria, but a combination of them is in most cases necessary to find the sources which can satisfy the user. An ideal "lead-in-tool" for answering research questions is a composite combining the features of the three mentioned above. Such a system would designate all of the potentially useful sources, provide the information needed to make a selection among them, list the location of the sources and enable the user to find the primary documents quickly and efficiently.

This, in a nutshell, is the proper definition of CoBRA. More precisely defined: a custom made guide to the literature in a library.

CoBRA does not refer to primary documents, because the high tempo in which such publications appear would render the database unstable and no library could afford the high workload that is necessary for building such a database. The proper indexing level for publications is therefore that of reference works: catalogues, bibliographies, CD-ROM's, encyclopaedias, online databases etc. Another important quality worth mentioning is that CoBRA leaves the final selection of sources to the patron. The active role of the patron is necessary because a precise definition of the research question is so difficult, and often during the process the subject is redefined. The possibility of browsing is therefore necessary.

As Yahoda and Braunagel insist, the librarian's answer to research inquiries, as well as the expert system's search advice, should be complete and should anticipate future problems. In order to satisfy these requirements, an integrated user advice system for research questions must contain:

- references to collections which provide basic information on the topic researched;
- references to catalogues to find other documents available in the library;
- references to sources indexing materials not listed in the catalogues (for example articles).

In order to anticipate on future questions from the patron, the system should also include:

- the names and addresses of experts, especially the relevant subject librarians;
- information on collections located elsewhere.
The essential elements of the CoBRA-system, in use at the reference desk of the University Library Groningen, producing printed advices for patrons by the reference librarian are shown in figure 1. It consists of a number of modules with the product of each module in the right column. The nature of the product, modules shown and the information in a module, is determined by the subject of the question and the type, amount and topicality of documents desired. In some modules the information provided is a list of titles of documents while in other modules references are made to persons, areas in reading rooms etc.

The database contains the titles of approximately 7000 reference works including their classification codes, subject headings, form entries, country and period specifications when applicable and, for a core collection, also annotations. In addition the database contains a number of secondary bases, as listed in the diagram, the elements of which are equipped with the same index items as the title base (for example classification codes). These additional units together with the information stored in the system on the methodology of literature study are responsible for steering the system. The (printed) search advice contains the necessary comments and directions so that it also can be used as an abridged library instruction.

Central in the system is the knowledge base in which one can express the knowledge of, among others, the librarian(s). A knowledge base consists of a collection of frames. In each frame one can select a database and set parameters for selection and/or for constraining information and formatting the output. This means that one frame can behave totally
different from another. Also it is possible that the behaviour of a frame is (partly) dependent on the user input. Frames can be merged to modules and modules can be merged to a so called profile. A knowledge base consists of at least one profile but mostly there will be more profiles. For the user each profile appears as a selection option in a menu. There is no limit to the number of knowledge bases.

The 'application' is an other distinguishing mark of the current system. In each application different kinds of information are combined, like knowledge bases, the query interface, special database definitions and text files. Each application appears to the user as a name which can be selected from a menu. The number of applications is also unlimited.

The prototype of the current expert system is an MS-DOS program. The software consists of three parts:
- One part for general services like selecting applications and printing advices.
- A second part handles the query definition according to the specifications of the current application.
- The third part translates the query by means of the knowledge base into an advice.

The software is linked with a database software library and a graphical user interface software library. Separate programs are written for defining and maintaining the knowledge and databases. There are also tools to keep data and knowledge bases downwardly compatible in case of upgrading the system.

5. Concluding remarks

CoBRA is developed by a small team of experts with the help of the library's subject specialists. Essential in the process was a sound organisation of the work to be done, with vital tasks for knowledge engineering, coordination, subject expertise and more general knowledge of reference work. During the process several prototypes were build. The project started with a rather explicit and thoroughly discussed specification of the systems architecture.

Essential is of course the acceptance of such a system by the librarians at the reference desk. Some remarks can be made in that respect. The reference librarians discovered that working with the system (which not everybody enjoyed at the beginning) meant structuring the reference interview according to the model built in the system. Without certain information from the patron the system does not work, just as an expert cannot give advice without knowing the vital elements of a query. That meant a new way of working for some of us. Certain facts had to be determined at the beginning of an interview before a CoBRA-answer can be formulated. Previously such data could be acquired while helping the patron.

The best way to help the user is to fill out the different screens together; especially the subject definition gives the opportunity to discuss the different ways of describing the subject. Working with a computer is of course a practice patrons did not expect when asking their questions.

The advice made can be inspected at the screen, but is often printed. An experienced reference librarian is able to refine the advice and delete some possibilities given on the print while recommending others. As we have discussed earlier that is an expected situation; real experts in reference work cannot be beaten by an expert system, although
CoBRA produced advices which reminded experienced reference librarians of sources they had overlooked, sources which where vital for those particular questions.

In more general terms, the system is developed to guarantee a more equal level of performance in the area of reference work. In the University Library Groningen, due to the liberal opening hours (seven days a week, 80 hours weekly), many questions are answered without the help of a subject specialist. Most questions asked at the reference desk are answered by an expert in another field of knowledge or by a semi-professional. For these situations the system was built and has been put to good use.

The role of the subject specialist is of course not threatened by such a system. Their role in the library is enlarged by building the system and maintaining it. In the University Library discussions about the role of the subject specialist have not been held, mainly because their experiences convinced our colleagues of the importance of their knowledge for building the system.

In the future the system will be developed in several directions. Due to the very large and differentiated database and the flexibility of the software it is possible to develop other applications for more specific user queries. An example can already be shown during this conference: in a short time we built an application for finding sources referring to book reviews.

Other applications will be built in the coming years. In most cases these applications will be built after considering the efficiency. More precisely, the question is whether or not the enquiries for that specific application are numerous enough to outweigh the investment in time and money. Two remarks are necessary in that respect. Often for these applications a user interface will be built that also allows end-users to use the system themselves. Other libraries have also shown great interest in cooperating with the University Library Groningen to implement the system in their own library. Such a cooperation should lead to the situation that libraries cooperate to develop new products.

Literature


