

An Interface for Self-Service Searching

A STUMBLING BLOCK to end user searching is the lack of a truly easy-to-use and helpful user system interface. It is a difficult problem because a really good interface should incorporate knowledge of the online host, the databases, the search process, and the needs of users. Some librarians are beginning to solve this problem themselves by creating friendly interface programs that allow searching on the major online systems. One of the best I've seen is the self-service database searching interface that is being developed under a Council on Library Resources grant by William Mischo, Mitsuko Williams, and Linda Smith, all of the University of Illinois at Urbana-Champaign.

The Illinois system helps users develop search strategy offline; selects a BRS, BRS/After Dark, or DIALOG database; automatically inserts appropriate commands and operators; logs on to the host system; executes the search; downloads up to 20 references; logs off from the host; and links to the University of Illinois online public access catalog (OPAC) to provide local call numbers for journals cited in the articles retrieved.

A typical search session

Since development is funded by a grant, all searches are currently recorded for later analysis and all are free of charge to the user. (Regular intermediary searches are billed back to clients.) Users are asked to fill out a questionnaire when the search is completed. There are now sites at Illinois's Engineering Library and the Biology Library. In fall 1988, the Agriculture Library will test the interface and plans are to include the undergraduate library and Beckman Institute library in the near future. Each

site's interface is custom-tailored to the subjects and databases most useful to their specialized clientele. Such individualized development is more time-consuming, but allows features to be developed that make the interface truly impressive.

The interface is tied into the Illinois OPAC so a user can search the online catalog or use the remote online searching interface through the same microcomputer. If the menu option "Search for Current Articles in Journals and Magazines" is selected, the user enters the interface program.

Once into the interface, the first screen to appear is a "subject areas" screen. This screen is different in each site because it is designed to reflect the departments and subjects of interest to the primary users of the branch library. Only the undergraduate library and test programs have a generalist subject screen that includes choices, e.g., biology, engineering, education, social sciences, popular magazines, newspaper articles, etc. Asking a user to select a predefined subject area allows the system to select a search database.

For each of the subject areas, a key database was selected by the librarian/programmer. The databases tend to be the major and most obvious choices—e.g., BIOSIS for biology, Compendex or Inspec for engineering, ERIC for education, etc. For subjects without one major database or for multidisciplinary research topics the forced choice of one database may not always be so clear or successful. Institutions without subject-specific branch libraries may need to have two subject screens to narrow down a subject area sufficiently to allow the system to select a database, or develop software enhancements that allow more than one database to be searched for each query.

After choosing a subject area, the user is asked to read two instruction screens. These two screens are the total search instruction and, in most cases, are all that is necessary to conduct a successful search using this system. Two instruction screens are, Mischo believes, the maximum

that users will put up with, and the minimum needed to explain the most important searching concepts. As the user progresses through the program the screens provide context-specific information and help. The interface program takes over as much of the mechanics of searching as possible (commands, proximity operators, modification of strategy) so these important parts of searching don't have to be explained to users.

And we're off

After these few preliminaries, users are ready to be led through their search. They are asked to enter a "search title" to help them focus their topic, followed by words representing the most important concept in that topic. For example, for "Microcomputers used in CAD/CAM," the user could enter CAD/CAM as concept one. They are then asked to enter synonyms, alternate spellings, abbreviations, or alternative terms for the first concept. Meaningful examples are given in each version of the interface.

The interface is programmed to supply singulars, plurals, and common alternate endings automatically. A dictionary table of substitution terms allows common synonyms to be searched even if a user does not input them. For example, if a person puts in "corn" in the agriculture interface, the system will automatically search "zea" or "maize" as well. Some thesaurus terms for the databases searched most often can also be incorporated.

For some topics commonly searched at the library site, the program includes prestored term hedges consisting of thesaurus terms. When a user enters one of the hedge "go" words, such as Artificial Intelligence, they are shown a list of common synonyms or related terms for AI from which to select. The terms selected are OR'd together by the system.

Control by the interface

The Boolean OR operator and proximity operators are inserted by the system as needed. BRS stop words and some other misleading



Carol Tenopir is Assistant Professor at the Graduate School of Library Studies, University of Hawaii, Manoa

words (such as "techniques" or "effects") are ignored. Users can build up to three concept groups that will be ANDed together. When the last concept group is built, the interface logs on to BRS and runs the search. All operations from this point on are controlled by the interface. Users are told what is happening at each stage of the search process, but they never actually see the direct interaction with the online host.

The next thing the user sees is a postings message for each concept and for the final search. Up to 20 citations are downloaded automatically and can be printed locally with call numbers and an explanation of the search attached.

It is what happens unseen by users during the online search that makes the Illinois interface so good. Unlike most of the commercially available front-end programs, this one incorporates the expertise and ideas of experienced online searchers by providing automatic search modification if too few or too many citations are retrieved.

Mischo has attempted to build in "what an online searcher/intermediary would do" to control output. Because he has found that most users enter very specific multiword terms, the system does not insert an adjacency operator between words, but instead uses the BRS WITH (within the same sentence) operator. (On DI-ALOG this would be the N operator with five or six intervening words.) If fewer than 30 citations are retrieved with the first concept, the searching stops there and the first 20 citations are downloaded.

Limiting the search

Searching is first done on the complete free text. If more than 100 citations are retrieved in total, the search is limited to the title and descriptor fields. If there are still more than 100 citations the Boolean AND between concepts is replaced by the BRS SAME operator (concepts must be found in the same paragraph or same field).

If there are zero postings after concepts are linked, the system will begin to drop concepts. If three concepts are entered by the user, the third concept (and, if necessary, the second concept) will be dropped in an attempt to get something. For some of the databases or subjects, the interface will provide more search modification

options such as limiting to English or allowing users to specify unwanted terms or concepts so the system can NOT them out.

Interesting statistics

Transaction logs are kept for all searches, questionnaires are distributed to users, and site monitors record information about every search. Together these provide some interesting data about the Illinois project. Between October 12, 1987 and March 31, 1988, over 440 database searches in 214 search sessions were conducted in the Engineering Library. Six databases were available (Inspec, Compendex, Computer Database, Magazine Index, Chemical Abstracts, Current Contents) of which Inspec and Compendex were searched by far the most frequently.

Most of the users like the system and the process of online searching. Over 97 percent had a favorable or very favorable "overall attitude toward the system" and over 88 percent were enthusiastic or very enthusiastic about online searching. Nearly 90 percent were satisfied with their search and over 83 percent indicated that "compared to a printed index, this system is better."

The Engineering Library users are very familiar with keyboards or terminals with over 87 percent indicating that they were experienced or very experienced. It is difficult to determine how much of an effect level of experience had on the 82 percent "successful or very successful" rating on ability to use the system. Results in the undergraduate site may provide a comparison.

Incorporating the results

Mischo has gained insight into what works and what doesn't in the interface and has incorporated much of this into the system. He has found that a "truncation algorithm is necessary" because people either assume the system can supply singulars and plurals or don't think about them. Other word form variations should be truncated, but sometimes strange results will occur (stemming when someone is searching for cement or -ing from sing.)

The NOT operator is necessary in "a small, but important, number of searches" and should be asked for at the beginning of the search strategy process. Users will enter "very specific multiword phrases" (e.g., fiber-

glass reinforced epoxy resin) that may appear slightly different in the literature. This makes free-text searching and WITH or NEAR proximity operations necessary.

A limit of 234 characters per concept group "can be of positive benefit" because "users tend to supply nonsignificant or harmful synonyms or related terms if given the opportunity." Users don't seem to have a clear idea of how terms are OR'd together and will enter redundant phrases such as earth metals OR alkaline earth metals.

Keep on testing

The Illinois system is still under development and testing. The data gathered are helping the developers to refine the interface and to gain insight that will be useful for other librarians who wish to develop or select interfaces for end user searching. Mischo believes "the interface will play a critical role in searching periodical indexes" whether they are available on remote systems, locally mounted on the OPAC, or on local CD-ROMs. The end user should not have to know which of these heterogeneous systems they are accessing and libraries should provide access to all of them through a single central interface.

One important lesson learned is that an interface that can be customized for certain databases and can reflect the subject interests of a group of users can allow more assistance in search strategy, term choice, and search modification. This, of course, complicates the development process. This project is already in its third year and development and testing will continue for at least one more.

Another lesson learned is that many prefer to do their own searches, but computer and sometimes human help is necessary. The system is designed for noncomprehensive searches. The developers assumed that complex or comprehensive searches require a human intermediary. The interface is meant to extend the online catalog by providing access to the journal holdings, not to replace intermediaries.

Mischo is willing to share his programs and ideas with interested librarians. Data analysis is going on now; results will be published in the library literature. Contact: William Mischo, Engineering Lib., Univ. of Illinois, 221 Engineering Hall, 1308 W. Green St., Urbana, IL 61801; 217-333-3576.