INFORMATION TECHNOLOGY

DESIGN AND APPLICATIONS

EDITED BY NANCY D. LANE & MARGARET E. CHISHOLM

G.K. Hall & Co.
Boston

First published 1991
by G.K. Hall & Co.
70 Lincoln Street
Boston, Massachusetts 02111

10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Information Technology: Design and Applications /
edited by Nancy D. Lane and Margaret E. Chisholm
p. cm.
Includes index.
ISBN 0-8161-1908-2. – ISBN 0-8161-1909-0 (pbk.)
1. Library science – Technological innovations.
2. Information technology. 3. Libraries – Automation.
I. Lane, Nancy D. II. Chisholm, Margaret E.
Z678.9.I59 1990
020'.285 – dc20 90-43577
CIP

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials. ANSI Z39.48-1984 ♾™
MANUFACTURED IN THE UNITED STATES OF AMERICA

Contents

Foreword

We are in the midst of John Naisbitt's proclaimed change from an industrial society to an information society, popularized by his 1982 book, *Megatrends*. Each day we face a multitude of information options and decisions, and we have made tremendous advances over the last few decades in how we manage the increasing amounts of information available. By gaining a better understanding of information management, we have been able to effect global change. For the first time, local health officials in many Third World countries have access to relatively up-to-date medical information. Satellite transmission allows for the rapid, global "sharing" of research and the commmunication of news. Within hours of the earthquakes in Iran in June 1990, for instance, virtually every country in the world was aware of the catastrophe and began delivering humanitarian aid. We are living proof of Marshall McLuhan's "global village" theory.

In spite of the advances in information management, we are experiencing many of the symptoms of "information overload." For instance:

- Between 6,000 and 7,000 scientific articles are written each day.
- The quantity of scientific and technical information now increases by more that 13 percent annually.
- The *Wall Street Journal* reported in 1988 that less than 50 percent of all information received from satellites in the last 10 years has been processed.

Our challenge is to develop and incorporate faster and more efficient means to filter, process, store, and retrieve information. Today's information professionals are hard-pressed to acquire and maintain all the print materials they need for their collections to meet the needs of their users. At the same

■ 11 ■

Research on Information Access

Raya Fidel

The new developments in information technology provide faster and better access to information. These developments are gradually introducing major and significant changes in our lives and, quite strikingly, in libraries and information systems. These qualitative changes are mostly caused by an important social development: with the new information technology, direct access to information is no longer reserved for a few trained professionals; instead, various information systems are accessed daily by most members of society, at work and at home.

Technology, however, caught information science unprepared. As a result, the new technological and social developments have had a qualitatively marked effect on research in information science. New areas of research in information science are emerging, and the focus of research in the traditional areas has shifted significantly.

The new areas of research are well represented in this book, with each chapter addressing the application of a new technology or issues that are a direct result of these applications. Although not always apparent, the need for these new areas of research is strongly motivated by the fact that the use of information technology is now widespread.

Consider expert systems, for example. The technological developments that are associated with artificial intelligence made these systems possible. But the task of an expert system is specific: to act like an expert in a certain field. Clearly, the need for such systems was not urgent when the use of information systems was reserved for experts only. Today, however, when information systems are made available to the public, this subject is of supreme importance.

The shift in the focus of research in traditional areas is more subtle. To trace it, one needs to examine research about one of the oldest and most established branches of information science, retrieval from bibliographic databases. This chapter summarizes the trends in information retrieval

research since it first encountered computerized systems and online bibliographic databases, through online catalogs, to CD-ROM and expert systems. It shows that the focus of research is gradually shifting from the study of retrieval performance to the study of users' searching behavior.

INFORMATION NEEDS AND RETRIEVAL TOOLS

First, a delineation of the topic is required. The process of retrieving information typically consists of two stages: users first identify the information that will answer their information needs, and then they retrieve the relevant information.

Reality, however, is much more complex than this description. More important, information need is an elusive concept: even if a real and precise need for information exists in an objective sense, it is difficult to determine it accurately. Asking users to define their information needs requires them to describe in exact terms what they do not know – a situation that is most often contradictory in nature.[1]

For the purpose of this discussion, however, we assume that what is expressed by users of libraries when they want to retrieve information are information needs that *are* clearly defined. We also assume that most of these needs are subject related: users want to find information about a certain subject. Further, users' requests include an additional aspect: the purpose of the request. For example, a request for information about CD-ROM systems by a library patron who wishes to learn how to use such systems is essentially different from the one raised by this author who needed a literature review on the subject.

Users can employ two types of retrieval tools when they search subject-related requests: the library catalog, or abstracting and indexing (A&I) services. Traditionally, users have been expected to use the card catalog and the printed volumes of the A&I services on their own with the librarian available on call. Today, a growing number of libraries provide access to their collections through online catalogs and most A&I services can be accessed online via search systems such as ORBIT, Dialog, and BRS, which were discussed in chapter 3. In addition, there is a steady and rapid growth in the number of databases that are available for searching through CD-ROM systems, discussed in chapter 7.

The first search systems that provided access to A&I services (now called bibliographic databases) were geared to the professional librarian. Online catalogs and most CD-ROM systems, on the other hand, were always designed for direct user searches.

Library catalogs and A&I services, whether manual or computerized, organize information in a manner that is useful for retrieval. The construction of retrieval tools involves the creation of subject indexes. For

printed tools, subject indexes include index terms – descriptors or subject headings – taken from an authority list of controlled vocabulary. A thesaurus is such a list of controlled vocabulary terms that are used both for searching and for indexing (assigning descriptors or subject headings to bibliographic items). Indexing can be performed either by humans or by computers (automated indexing).

In addition to thesauri, computerized tools provide subject indexes that list all the meaningful words in the bibliographic citations and in their abstracts. Such indexes are generated by the computer and do not require manual indexing or the construction of a thesaurus, and users search them with free-text terms. In recent years, databases have been created that provide the complete text of documents in a machine-readable form. Such full-text documents as journal articles, textbooks, or chapters in an encyclopedia can usually be searched with both free-text terms and descriptors.

ONLINE BIBLIOGRAPHIC DATABASES

Research about information retrieval from computerized systems had begun in the late 1950s, before such systems were commonly used. Anticipating the new capabilities that computerized systems offer, in particular free-text searching, a few investigators examined the effect that the then-new technology would have on information retrieval. Most noted are the Cranfield studies which tested retrieval performance of a variety of index languages – including free-text searching.[2] While the results are still controversial, the Cranfield studies established the procedure for retrieval experiments and the measurements to use: precision (what proportion of the citations retrieved are relevant) and recall (what proportion of relevant citations have been retrieved).

Searching bibliographic databases was first possible in the batch mode only. During that period, the use of computerized databases was limited, and various researchers replicated the Cranfield studies with some variations. When in the late 1960s databases became accessible online, however, their use spread rapidly and with this change came a shift in research interest.

The results of these early studies were not considered when the first online bibliographic databases were designed. These databases were simply computerized versions of the printed A&I services. In fact, even today, many A&I services are publicly available in both printed and computerized versions which are often identical in content and structure. More important, the first search systems for bibliographic databases were designed for professional searchers, mostly librarians. Their use required expert knowledge, training, and experience.

Characteristics of Searchers

Research in online searching began in the 1970s with two large-scale studies that focused on users' attitudes, satisfaction, and success.[3] Most of the research that followed, however, concentrated on the attributes of a "good searcher": those personal characteristics a librarian should possess to become successful online. Experience in online searching,[4] type of training,[5] personal traits such as creativity,[6] and cognitive abilities[7] are among the characteristics examined. No conclusive results have been found, and most investigators observed that the large individual variability among searchers impeded their studies.[8]

The method used in online searching studies is well established today. Briefly, an experiment is set up in which the searchers, known as subjects in experimental studies, are assigned to a group, depending on their score on the tested characteristics (e.g., experience, cognitive style) or variables. All subjects are asked to search a given set of requests, and their search processes and results are analyzed and compared. These analyses employ a well-established set of measurements. The search process is evaluated by the number of commands used, number of search terms entered, length of search, and similar measurements. Search results are evaluated by precision, recall, and unit cost (the cost of each citation).

Why did these studies fail to produce conclusive results? There are several possible reasons, but what immediately comes to mind is that the personal characteristics (the variables) that have been tested do not affect the quality of online searching. Unfortunately, such a generalization contradicts common sense: it is common knowledge among librarians that experienced searchers perform better than novices and that training is important for online searching. Therefore, one conclusion is that the measurements that have been used are inadequate. This conclusion was substantiated when 10 librarians with almost identical subject background, training, and experience scored very differently from one another on search process measurements.[9]

Another impediment to obtaining conclusive results is the experimental setting in which variables that are not studied are assumed to be controlled. In reality, however, a large number of variables affect online searching behavior,[10] and no single experiment can control all of them. Variables that were ignored in these experiments, such as the ability of a subject to perform a search without interviewing the user, may have stronger effects than the ones tested.

But it is more important to ask: why did researchers focus on the attributes of a good searcher? The answer is apparent if we consider the patterns in which information retrieval technology was used at that time. When the access to information is provided primarily by experts it is important to investigate and identify the characteristics of these experts – the training, knowledge, and experience that is required of them.

With the new developments, however, we gain a different perspective on the subject matter. While identifying the characteristics of a good searcher may at times help administrators to decide which employee should perform online searching in their libraries, the issue is of very little significance when online bibliographic databases are searched by library users themselves. The important issue is to discover the characteristics of a good search. Identifying strategies and moves that can enhance the success of online searches is beneficial to all users of all online retrieval systems.

The Search Process

Online searchers have long known the importance of the search process. Journals such as *Online, Online Review,* and *Medical Library Reference Quarterly* include many articles that describe useful strategies in certain databases and search systems.

Research in this area began in the 1980s, but it is still sparse and requires different research methods from those that attempted to identify attributes of good searchers. Based on the experience of librarians, Bates proposed a number of information search tactics that could be employed in online searches.[11] Tactics to use early in the search include an examination of information already found in the search ("trace"), and those to use later in the search include the rejection of items indexed by certain terms ("block").

Fidel analyzed search protocols and verbal protocols of seven experienced searchers performing approximately 90 searches as part of their regular workload. As a result, she listed the moves (changes in search strategies) that searchers made to increase or decrease the number of items retrieved, or to improve the search in other ways.[12]

Studying the search process is a complex task requiring the probing of phenomena that are not easy to observe or analyze. Nevertheless, user training and the design of useful online retrieval systems cannot be successful before this process is thoroughly understood.

User Searching

The idea that users can search bibliographic databases for their own requests is rapidly gaining popularity, particularly with the proliferation of CD-ROM systems. The interest in this idea, however, emerged earlier, when search systems and other commercial vendors started to provide gateway (front-end, or intermediary) systems such as Search Helper or Knowledge Index, which are supposed to mediate between the end user and online bibliographic systems. Such gateway systems do indeed provide a simple interface with search systems. They free users from having to learn a command language or

deal with the idiosyncrasies of specific databases. This is especially important since most users are not likely to perform online searches frequently. Most of the current gateway systems, however, achieve their simplicity by providing users with a limited range of capabilities, thereby simplifying the search process itself.

While several in-house studies have been performed on gateway systems, studies of users searching their requests directly would have far more impact. The only example of such a large-scale study is a research project that examined 11 years of searching the databases of the National Library of Medicine (NLM) in the United States by pathologists and pharmacists.[13] The results of this study are many, but as yet not all have been published. Of particular interest here are the findings that users did not encounter many problems with the technique of searching but rather with the vocabulary and content of the system and that most performed simple searches. In addition, the major problems encountered were with the more sophisticated capabilities of the databases – problems that sometimes caused a substantive loss of citations.

Although this study examined a specific population, its results substantiate the findings of online catalog use studies, as described in the next section: the weakest point in patrons' doing their own searching is their inability to formulate successful strategies. Gateway systems do not provide help because they do not employ any of the sophisticated capabilities of databases – in fact, some even eliminate simple ones. But gateway systems do not employ sophisticated capabilities because research that can guide the design of such systems is still sparse. At present, the only source on which we can rely for search strategies is the community of experienced online searchers.

ONLINE CATALOGS

Online catalogs are designed for library users. Though libraries had prior experience with online retrieval through bibliographic databases, the introduction of online catalogs in the late 1970s and early 1980s marks the beginning of a new era in which users have begun to do their own searching. This new scenario has had an important effect on research in information science.

Unlike research in searching online bibliographic databases, studies in online catalogs focus on users' requirements. Major studies were set up to discover users' attitudes to and acceptance of the new catalogs by examining characteristics of both users and catalogs. Most studies, however, concentrated on the human-computer interface, that is, on how easy it is to "converse" with the computer. Only a few addressed the retrieval problem:

whether or not users are satisfied with the results of their searches, and what could be done to improve these results.

Further, studies in online catalogs are guided by an administrative approach that is essentially different from the approach taken by researchers of online searching in bibliographic databases. In studying online catalogs, researchers assume that the user population is a given and that features of the catalogs themselves need to be examined in order to eventually design an online catalog that is most useful. In contrast, research about retrieval from online databases assumes that the databases and the search systems are a given and focuses instead on the characteristics of the users (i.e., professional searchers) that are most suitable for online searching.

Surely, the difference in these research approaches stems from the reality of library decision making, which is determined by the use of technology. Librarians do not elect which patrons to serve, but they may want to design an online catalog and should be able to select the system that provides the best catalog for their users. In contrast, library administrators cannot change databases and search systems, but they can select from among a number of candidates those persons who would best perform online searches.

Nevertheless, the searching of bibliographic databases and of online catalogs is one and the same: searching for bibliographic information. There is a diminishing distinction between the database as a store of citations to journal articles, which can be manipulated with sophisticated techniques, and the online catalog as a store of information about monographs, which only provides for simple manipulation. The difference in research approaches is, therefore, an impediment to both areas of research: studies of online catalogs cannot rely on research that already has been performed in online searching, and vice versa.

User Interface with Online Catalogs

Online public access catalogs (OPACs) are a relatively new phenomenon in libraries. In the United States, the Council on Library Resources (CLR) funded in 1981-1983 a nationwide study of 17 online catalog systems (both in-house and commercial) in 29 libraries. While many articles about experience with online catalogs in specific libraries have been published, this first large-scale study provides most of our knowledge about OPACs.

The study was conducted by various agencies, and although numerous attempts to summarize and synthesize its results have been made (see, for example, Matthews and Lawrence's article[14]), more meaningful interpretation and integration of the vast amount of data are required before specific conclusions can be drawn.

CLR Study Methods

Typical of a large-scale study (although a novelty in research about online searching behavior), the study applied a variety of methods. A questionnaire survey of 8,094 users and 3,981 nonusers of online catalogs in 31 libraries was administered by various agencies. A summary of the data collected is provided by Matthews and his colleagues.[15] They point, for example, to demographic characteristics of users, the manner in which most users are informed about the availability of an online catalog, the persons to whom they address their requests for help, problems with the interface, and the rate of success and satisfaction perceived by users.

Focused-group interviews with library staff and patrons – both users and nonusers of online catalogs – were carried out in six libraries and were conducted through an open, in-depth discussion led by a moderator.[16] While such group interviews do not supply quantitative data, they can explore the degree of satisfaction and expectation of both patrons and staff. Among other things, these particular interviews revealed that while users of online catalogs are happy to use them, they have problems with subject access to information, and they expect online catalogs to provide more services than are currently available.

The results from the focused-group interviews were complemented by the results of individual and group interviews conducted among library staff at three research libraries.[17] These interviews supported the analysis of the questionnaires and addressed issues such as problems that users have in using the catalogs, possible system improvement, and the impact of online catalogs on the library staff and its patrons.

In addition, transaction log analysis, in which the protocols of individual searches are analyzed, were performed in seven libraries.[18] It was found, for example, that while there is a great variability in the length of searches and in their types (author, title, or subject searches), users tend to remain in the type of search that was initiated and to repeat their mistakes. These analyses also provided statistics about issues such as the rate of success (nonzero hits) in subject and author searches, frequency of commands used (in particular, "sophisticated" commands), and patterns of searches.

Feature analysis of 10 existing online catalog systems was also performed for the CLR study.[19] This was a functional analysis, documenting the functions and commands of each system, its interface capabilities, and the documentation available to users.

CLR Study Findings

The CLR study stimulated further analyses of the data, additional explorations, and considerations of possible implications. Cochrane and Markey showed how the results from focused-group interviews were useful in

interpreting questionnaire findings,[20] and Borgman compared the study results with research findings in other areas of online bibliographic retrieval.[21] She concluded that more similarity existed in conceptual than in mechanical problems. Dickson, on the other hand, analyzed a sample of zero-hit author and title searches to discover reasons for failure and concluded that users have a conceptual model of the online catalog that is different from the one they have of a card catalog.[22] Last, a collection of articles assessed the impact of online catalogs on technical services, reference services, subject access, and library administration.[23]

In a summary of the CLR OPAC study, Matthews and Lawrence outlined the principal findings:

1. Experience with the library and its catalog is the most important factor in determining success and satisfaction in using the catalog.
2. Online catalogs should provide a variety of interfaces, depending on the type of search and the level of user experience.
3. Users adapt their attitudes to the capabilities and limitations of the online catalogs they use.
4. The form and nature of training and user assistance are important.[24]

These findings only substantiate common knowledge among librarians; they do not provide insight as to *how* to design online catalogs that are most useful.

Studies from Great Britain

The spread of online catalogs in Great Britain has been slower than in the United States. Nevertheless, much research has been conducted there, most of it sponsored by the British Library Research and Development Department. While only a few large-scale studies have been completed, researchers take a multifaceted approach: they study retrieval techniques and interface design, as well as the quality of the database. Controlled experiments are common, and most are based on theoretical developments in the area of bibliographic databases. In addition to direct use studies, researchers have conducted comparative studies – either before-and-after surveys or examinations of the performance of different systems – and created experimental systems to study the user-catalog interface.[25] Findings from these various projects (some of which are still incomplete) have yet to be integrated.

Subject Access in Online Catalogs

The importance of subject access to online catalogs was the most significant finding of the CLR study.[26] This is not surprising because early studies of catalog use (e.g., by Lipetz and by Bates[27]) showed that subject access in card catalogs was inadequate. With online catalogs – which are actually automated card catalogs – the issue assumes an even greater importance because library users who were not satisfied with subject searching in a card catalog *expect* to be more successful with online catalogs.[28] Here again, the CLR study only substantiated previous findings but did not provide guidelines for improved subject access.

Present online catalogs are more sophisticated than the early ones and stand ready to facilitate improved subject access. However, the various ways that users can actually be helped in subject searches are not yet known. A display of a classification scheme, for example, can help users to "browse" in a subject area.[29] It is not clear, however, which classification scheme is most suitable for this purpose: the Dewey Decimal Classification[30] or Library of Congress Classification.[31] Each scheme raises both conceptual and technical problems.[32]

Another approach to aid users is to provide online help in the use of the Library of Congress Subject Headings. Such assistance could include an alphabetical display of the headings with cross references or a display of related subject headings. In addition, searching keywords (i.e., free-text searching) in the titles of books[33] or even in their indexes and tables of contents[34] can improve subject access.

Researchers of online catalogs have only begun to consider subject access, and most of the literature in this area is limited to expert opinion about the direction online subject access should take.[35] An exception is provided by the European research community. Several projects have begun to experiment with methods to improve subject access: the use of a thesaurus to suggest narrower terms if the user retrieves many references or broader terms if too little is retrieved; the development of automatic stemming techniques such as removing word suffixes to improve recall; and the preliminary design of an expert system that derives classification numbers from natural language terms, to name a few.[36] It remains the task of future research to integrate the results of these projects, most of which are ongoing.

With the development of more powerful online catalogs, subject searching becomes more similar to subject searching of bibliographic databases. For example, CITE – the online catalog at the NLM – can also access the NLM databases.[37] All in all, research in subject access to online bibliographic databases is more developed, and studies of online catalogs can draw on its experience. It is clear, however, that while current technological capabilities provide for improved subject access, research in this area is lagging behind.

CD-ROM SYSTEMS

The use of CD-ROM technology in libraries is spreading fast, and any statistics about this use are bound to be outdated by the time they are published. At present, libraries acquire CD-ROM software and hardware primarily for searching bibliographic databases, even though this technology enhances access to various types of databases, such as image or voice databanks.

Typical of an innovative technology, CD-ROM systems immediately introduced a qualitative change in libraries: because searching time is not a factor in the cost of online retrieval, users can search their own requests for free or at a low cost. Thus, while most CD-ROM products for bibliographic databases are designed for both users and professional searchers, most interest and research is focused on user searching.

As a result, the introduction of CD-ROM systems to libraries accelerated the emergence of users' searching behavior as a focus of research in information science. But while the technology is already in use and the importance of such research is widely recognized, actual research in this area has barely begun.

One problem typical of the new pattern in which this technology is used is the simple, unsophisticated manner in which users conduct their searches. A number of studies of end-user searching revealed that users carried out very simple searches, ignoring the sophisticated capabilities that are available to them. Moreover, there is an indication that, given a choice, simple systems (those that are commonly not powerful) would become more popular than sophisticated ones.[39]

Nonetheless, studies of users' searching show that CD-ROM users indicate a high rate of satisfaction (e.g., see Karp and Kleiner[38]): users believe most of the time that their searches are successful. Moreover, having retrieved information from the computer, they often assume that the search is complete and that there is no need to search other databases – whether printed or online. This tendency was even spotted among advanced students in a library school.[40]

The trade-off between sophistication and ease of use is not surprising, but it is alarming when users are not aware of this trade-off and when they are generally satisfied with the results of simple searches. This problem was clearly demonstrated in the CLR studies of online catalogs. These studies found that user satisfaction levels were surprisingly high given the relatively high error rates.[41]

There are two approaches to resolving the situation: to provide adequate training to users before–or even during–their searches; and to design systems and interfaces that are powerful and sophisticated, yet easy to use.

Training

Training is an important factor. Although CD-ROM products are tailored to end-user searching, and some even claim that no training is needed for their use, experience shows that no matter how simple a system is, some training is always needed.[42]

Furthermore, training seems to be a promising approach because there is evidence that users' training has a great effect on their searching. A study at Drexel University, for example, found that students followed closely the instruction they received in a short training session when they selected databases and search strategies.[43]

To complicate matters further, training sessions are usually short because most library patrons are infrequent users and are not willing to spend long hours on training. As a result, these brief sessions tend to emphasize the mechanics of a search or the technical aspects of searching because these are the basics necessary to access and search databases. In contrast, a number of studies revealed that end users encounter very few problems in learning the technical skills, but found it difficult to formulate search strategies. This difficulty was found across a variety of user groups, from highly trained pharmacists[44] and physicians[45] to high school students[46] searching in an academic library.[47]

Clearly, preparing users for searching CD-ROM databases is a complex task: the trainer must explain the basic technical features, as well as the "art" of constructing effective search strategies, and all in a short period of time. Training for online searching is performed today on an ad hoc basis, depending on the resources available. Furthermore, there is no agreed-upon understanding of what should be included in such training. This is not surprising. User searching is a new phenomenon and librarians have not yet acquired enough experience to know what problems users face in searching. As with general bibliographic instruction, training of users for online searching will become more effective in time. However, conducting research to uncover the problems and difficulties that users encounter when searching is necessary to achieve this goal.

Such research should reveal the features–both technical and conceptual–that are most important to effective user searching. At present, only a few studies of actual CD-ROM searching have been reported in the literature. In addition, most of these studies are surveys in which users are asked to rate their satisfaction and provide some kind of demographic data. A number of studies examined the effect of CD-ROM searching on searching

activities in libraries. For instance, a study in a medical school's library found that while users searched CD-ROM systems more frequently than online systems, the introduction of CD-ROM systems to their library did not change the number of mediated searches, i.e., those performed by a librarian.[48]

While such surveys are important to the general management of online services, they rarely illuminate the problems users encounter in searching. A few attempts to discover these problems have already been made, however. One example is the study of students at Drexel University that examined actual search strategies,[49] and another is a study of searching behavior of elementary school children while searching a full-text electronic encyclopedia on CD-ROM.[50] As first studies of this kind, however, they provide only tentative observations about problems encountered in searching. Much research in searching behavior is still needed before training effectiveness can be improved.

Systems Design

Another approach to help users to improve their searches is to design CD-ROM systems that are powerful and sophisticated, yet easy to use. At present, the development of these systems is guided by a "seat of the pants" approach, rather than by drawing from systematic studies of user searching.[51] Moreover, a number of fundamental issues are not yet resolved.

A viable example here is the issue of standardization: the development and use of a common standard for all bibliographic retrieval systems. There are two opposing views about whether or not standardization would be useful for searching CD-ROM systems. On one hand, some software designers claim that their creativity would be inhibited by the enforcement of a standard. On the other hand, such a standard would be extremely helpful because it would enable users to learn how to use only one system and to move freely and easily from one database to another.

In reality the National Information Standards Organization (NISO) issued a proposed standard for common command languages (CCL),[52] but CD-ROM producers exercise their creativity, each designing their own command language. In fact, a study of 20 databases provided by 10 different CD-ROM system producers showed that the system varied greatly in the search capabilities they offer, in their ease of use, and even in their use of function keys.[53]

Only research into the searching behavior of CD-ROM system users can determine whether or not a standard is needed. Moreover, such research could determine which standard would be most effective. By extension, using the results of research about searching behavior, software producers will be able to spot the problems in searching–and learn how to resolve at least some of them–when they have both a description and an evaluation of

searching behavior with current systems. They could then design more useful interfaces and more powerful systems than are available today.

Thus, as is the case with bibliographic databases and online catalogs, CD-ROM systems are already in use, but the research that can facilitate both effective training and the design of powerful and useful interfaces has just begun.

INTERMEDIARY EXPERT SYSTEMS

It is believed that an increasing number of users prefer to interact directly with online bibliographic retrieval systems. Although no statistics exist as yet to support this assumption, a large amount of effort is being invested by software producers and search system vendors in developing systems, such as SciMate or Colleague, that facilitate online bibliographic retrieval from users' offices or homes. It is also believed that users will very likely search their own requests online when search processes are simplified or made friendlier. The prevailing approach to providing such user-system communication is to develop intermediary systems designed to mediate between users and complex information retrieval systems – online bibliographic databases, online catalogs, or CD-ROM systems.

With intermediary expert systems, users should be able to present their requests to a system, which would then make expert decisions about the search process (that is, make decisions about the databases to search and the search strategy to employ). Such systems should interrogate users to elicit request characteristics, but would use their own expertise to make decisions about matters that are beyond the knowledge of users. For example, an intermediary expert system should ask the user whether high recall or high precision is required, and then use this information to decide which search strategy would provide the best results.

Various intermediary systems are already available for public access, such as CITE,[54] while others are prototype systems being tested in experimental settings. Examples of the latter are CANSEARCH, PLEXUS, EP-X, and CoalSORT, each covering a limited subject domain and searching a single database.[55] Through such systems, users are freed from encounters with the numerous peculiarities of databases and search systems – such as ORBIT, Dialog, or BRS – and yet can benefit from a large range of capabilities.

In particular, an intermediary system allows users to enter a request in a loosely structured format, preferably in natural language, using a sentencelike expression. The system then processes the request terms, displays information to users, and asks for feedback. The information displayed may be in the form of a list of subject areas, databases, search keys, or actual citations from which users are asked to make a selection, possibly in ranked

order. Interaction of this nature usually proceeds until the user terminates the session.

Helper vs. Expert Systems

Some intermediary systems are actually helper systems: they provide menu-driven interaction that frees users from learning the command language while still requiring them to make most of the decisions during a search process, or they drastically simplify searching by reducing the number of options to a minimum. CITE, for example, leaves the selection of search keys to the user. It displays a list of search keys that can be used for a request concept – both descriptors and free-text terms – and asks the user to select the terms. In contrast, CONIT – which provides an interface with a number of databases covering a variety of subjects – simplifies the selection of search keys because it searches each search key as a free-text term and, under certain circumstances that depend on the search system rather than the request, also searches each as a descriptor.[56]

Intermediary expert systems, on the other hand, attempt a more powerful form of user assistance: they replicate the performance of an expert in online bibliographic retrieval by incorporating the knowledge of an expert with rules for making inferences on the basis of this knowledge. As such, they have attracted attention and controversy.

Most researchers agree about the nature of intermediary expert systems, even though a variety of definitions for expert systems currently exist. Studies examining users searching their own requests with no intermediary assistance show repeatedly that users need intermediary expertise mostly for formulating search strategies, while they seem to master the command language with no difficulties. Therefore, every intermediary expert system that is being developed today must include a component that supports decisions about search strategies.

Requirements for Intermediary Systems

Daniels, Brooks, and Croft, among others, delineate the requirements for such intermediary systems.[57] One such requirement is that an intermediary expert system should be able to take into account request (and user) characteristics that are beyond the topical description of the search. Existing intermediary expert systems, however, are unable to account for such characteristics because their knowledge is derived primarily from the information stored in bibliographic databases.

For example, to users who ask for information about online catalogs, such a system may suggest that they search under the term "OPAC," because this new term appears in the titles and abstracts of many items indexed with the descriptor "Online Catalogs." Deriving its knowledge only from the

stored text, however, prevents the system from considering aspects that are not directly related to the subject of a request. For instance, the system would not "know" that it is useful to find out what level of material is required: introductory and instructional material, or data about recent research.

Although the debate about whether knowledge for expert systems should be derived from human experts or from other sources is not yet settled, some attempts have been made to extract knowledge from librarians. Among the first steps in this direction is a project supported by the British Library that produced a model of the search process[58] and a list of moves to improve search results[59] with a decision tree that guides decisions about whether to enter free-text terms or descriptors.[60] In fact, a few prototype systems such as PLEXUS and EP-X are already based on such knowledge. In addition, in the area of indexing, two institutions in the United States, the American Petroleum Institute and the NLM, have used indexers' knowledge and practice to develop expert systems to aid in-house indexing.[61]

Research on intermediary expert systems is in its early years, even though various attempts have already been made to design such systems. Our short experience indicates, however, that the most promising approach to the creation of knowledge bases is to model searching behaviors. Thus, the technological developments that made expert systems possible also direct research to focus on users' searching behavior.

CONCLUSION

The application of new technology has a significant effect on the retrieval of information; access to information and the capabilities to manipulate information for retrieval have undergone revolutionary changes. Research to support these new possibilities, however, has just begun.

In a report about a recent study on information seeking and retrieving, Saracevic and his colleagues conclude their summary of the review literature with the observation:

> It is most indicative that an identical conclusion appears in every one of these reviews despite different orientation of the review and different backgrounds of the reviewers. They all conclude that research has been inadequate and that more research is needed. In the words of Belkin and Vickery: "... research has not yet provided a satisfactory solution to the problem of interfacing between end-user and large scale databases." Despite a relatively large amount of literature about the subject, the research in information seeking and retrieving is in its infancy. It is still in an exploratory stage.
>
> Yet, the future success or failure of the evolving next generation of information systems (expert systems, intelligent front-ends, etc.) based on built-in intelligence in human-system interactions depends on greatly increasing our knowledge and understanding of what is really going on in human information seeking and retrieving. The key to the future of information systems and searching processes (and by extension, of information science and artificial intelligence from where the systems and processes are emerging) lies not in increased sophistication of technology, but in increased understanding of human involvement with information.[62]

Technology is racing ahead and is being applied at a rapid pace, but research that is needed to guarantee the useful exploitation of the new applications is lagging behind. This disadvantage, however, can be turned into a significant advantage. Because new technological developments are already applied – indeed, some are becoming common – it is possible to study the use of new information systems as it actually happens in real life. For instance, researchers do not have to simulate the use of CD-ROM systems; they can go to the many libraries that provide CD-ROM searching and investigate searching behavior in reality.

Clearly, exploring information seeking and retrieving as it regularly occurs is a powerful approach to improving our understanding of users' searching behavior. In effect, the applications of information technology have opened a new and exciting avenue for research in information science.

NOTES

1. N. J.Belkin and A. Vickery, *Interaction in Information Systems: A Review of Research from Document Retrieval to Knowledge-Based Systems,* Library and Information Research Report 35 (London: The British Library, 1985).

2. C. W. Cleverdon, *Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems* (Cranfield, England: College of Aeronautics, ASLIB Cranfield Research Project, 1962).

3. F. W. Lancaster, *Evaluation of Online Searching in MEDLARS (AIM-TWX) by Biomedical Practitioners,* Occasional Paper no. 101 (Urbana: University of Illinois, Graduate School of Library Science, 1983); and J. Wanger, C. A. Cuadra, and M. Fishburn, *Impact of Online Retrieval Services: A Survey of Users, 1974-1975* (Santa Monica: Systems Development Corp., 1976).

4. C. H. Fenichel, "Online Searching: Measures that Discriminate among Users with Different Types of Experience," *Journal of the American Society for Information Science* 32, no. 1 (1981): 23-32.

5. J. Wanger, D. McDonald, and M. C. Berger, *Evaluation of the Online Process* (Santa Monica: Cuadra Associates, 1980).

6. T. Bellardo, "An Investigation of Online Searcher Traits and Their Relationship to Search Outcome," *Journal of the American Society for Information Science* 36, no. 4 (1985): 241-50.

7. N. N. Woelfl, *Individual Differences in Online Search Behavior: The Effect of Learning Styles and Cognitive Abilities on Process and Outcome* (Ph.D. diss., Case Western University, Cleveland, Ohio, 1984).

8. C. H. Fenichel, "The Process of Searching Online Bibliographic Databases: A Review of Research," *Library Research* 2, no. 2 (1980): 107-27.

9. R. Fidel, "Moves in Online Searching," *Online Review* 9, no. 1 (1985): 61-74.

10. R. Fidel and Dagobert Soergel, "Factors Affecting Online Bibliographic Retrieval: A Conceptual Framework for Research," *Journal of the American Society for Information Science* 34 (May 1983): 163-80.

11. M. J. Bates, "How to Use Information Search Tactics Online," *Online* 11, no. 3 (1987): 47-54.

12. Fidel, "Moves in Online Searching."

13. W. Sewell and S. Teitelbaum, "Observations of End-User Online Searching Behavior over Eleven Years," *Journal of the American Society for Information Science* 37, no. 4 (1986): 234-45.

14. J. R. Matthews and G. S. Lawrence, "Further Analysis of the CLR Online Catalog Project," *Information Technology and Libraries* 3, no. 4 (1984): 354-76.

15. J. R. Matthews, G. S. Lawrence, and D. K. Ferguson, eds., *Using Online Catalogs: A Nationwide Survey,* Report of a study sponsored by the Council on Library Resources (New York: Neal-Schuman, 1983).

16. K. Markey, "Thus Spake the OPAC User," *Information Technology and Libraries* 2, no. 4 (1983): 381-87.

17. D. Ferguson, *Public Online Catalogs and Research Libraries,* Final report to the Council on Library Resources (Stanford: Research Libraries Group, 1982).

18. R. R. Larson, "Users Look at Online Catalogs: Results of a National Survey of Users and Non-users of Online Public Access Catalogs," in *Interacting with Online Catalogs,* Part 2 (Berkeley: University of California, Division of Library Automation and Library Research and Analysis Group, 1983); and J. E. Tolle, "Understanding Patrons' Use of Online Catalogs: Transaction Log Analysis of the Search Method," in *Productivity in the Information Age, Proceedings of the 46th ASIS Annual Meeting,* ed. R. F. Vondran, A. Caputo, C. Wasserman, and R. A. V. Diener (White Plains, N.Y.: Knowledge Industries, 1983), 167-71.

19. C. R. Hildreth, *Online Public Access Catalogs: The User Interface* (Dublin, Ohio: OCLC, 1982).

20. P. A. Cochrane and K. Markey, "Catalog Use Studies since the Introduction of Online Interactive Catalogs: Impact on Design for Subject Access," *Library and Information Science Research* 5, no. 4 (1983): 337-63.

21. C. L. Borgman, "Why Are Online Catalogs Hard to Use? Lessons Learned from Information-Retrieval Studies," *Journal of the American Society for Information Science* 37, no. 6 (1986): 387-400.

22. J. Dickson, "An Analysis of User Errors in Searching an Online Catalog," *Cataloging and Classification Quarterly* 4, no. 3 (1984): 19-38.

23. J. R. Matthews, ed., *The Impact of Online Catalogs.* (New York: Neal-Schuman, 1986).

24. Matthews and Lawrence, "Further Analysis of the CLR Online Catalog Project."

25. J. Kinsella and P. Bryant, "Online Public Access Catalog Research in the United Kingdom: An Overview," *Library Trends* 35, no. 4 (1987): 619-29.

26. K. Markey, "Subject-Searching Experiences and Needs of Online Catalog Users: Implications for Library Classification," *Library Resources and Technical Services* 29, no. 1 (1985): 34-51.

27. B. Lipetz, "Catalog Use in a Large Research Library," *Library Quarterly* 42, no. 1 (1972): 129-39; and M. J. Bates, "Factors Affecting Subject Catalog Search Success," *Journal of the American Society for Information Science* 28, no. 3 (1977): 161-69.

28. Matthews, Lawrence, and Ferguson, *Using Online Catalogs: A Nationwide Survey.*

29. E. Svenonius, "Use of Classification in Online Retrieval," *Library Resources and Technical Services* 27 (January/March 1983): 76-80.

30. K. Markey and A. Demeyer, *Dewey Decimal Classification Online Project: Evaluation of a Library Schedule and Index Integrated into the Subject Searching Capabilities of an Online Catalog,* Final Report to the Council on Library Resources (Dublin, Ohio: OCLC, 1986).

31. L. M. Chan, "Library of Congress Classification as an Online Retrieval Tool: Potentials and Limitations," *Information Technology and Libraries* 5, no. 3 (1986): 181-92.

32. P. A. Cochrane and K. Markey, "Preparing for the Use of Classification in Online Cataloging Systems and in Online Catalogs," *Information Technology and Libraries* 4, no. 2 (1985): 91-111; and J. S. Hill, "Online Classification Number Access: Some Practical Considerations," *Journal of Academic Librarianship* 10, no. 1 (1984): 17-22.

33. G. S. Lawrence, "System Features for Subject Access in the Online Catalog," *Library Resources and Technical Services* 29, no. 1 (1985): 16-33.

34. Markey, "Subject-Searching Experiences and Needs of Online Catalog Users: Implications for Library Classification."

35. E. Carson, "OPACS: The User and Subject Access," *Canadian Library Journal* 42, no. 2 (1985): 65-70.

36. R. M. Jones, "Online Catalogue Research in Europe," *Journal of the American Society for Information Science* 40, no. 3 (1989): 153-57.

37. T. E. Doszkocs, "CITE NLM: Natural Language Searching in an Online Catalog," *Information Technology and Libraries* 2, no. 4 (1983): 364-80.

38. N. S. Karp, "ABI/Inform on CD-ROM: A First Look," *Laserdisk Professional* 1, no. 1 (1988): 28-34; and J. P. Kleiner, "InfoTrac: An Evaluation of System Use and Potential in Research Libraries," *RQ* 27, no. 2 (1987): 252-63.

39. M. D. Bonham and L. L. Nelson, "An Evaluation of Four End-User Systems for Searching MEDLINE," *Bulletin of the Medical Library Association* 76, no. 1 (1988): 22-31.

40. J. M. Day, "LISA on CD-ROM–A User Evaluation," in *Online Information 87; 11th International Online Information Meeting* (Medford, N.J.: Learned Information, 1987), 273-83.

41. C. L. Borgman, "Information Retrieval from CD-ROM: Status Quo or a Revolution in End-User Access?" *Canadian Journal of Information Science* 12, no. 3/4 (1988): 43-53.

42. Karp, "ABI/Inform on CD-ROM: A First Look;" J. A. Capodagli, J. Mardikian, and P. A. Uva, "MEDLINE on Compact Disc: End-user Searching on Compact Cambridge," *Bulletin of the Medical Library Association* 76, no. 2 (1988): 181-83; and Borgman, "Information Retrieval from CD-ROM: Status Quo or a Revolution in End-User Access?"

43. L. A. Wozny, "College Students as End User Searchers: One University's Experience," *RQ* 28, no. 1 (1988): 54-61.

44. Sewell and Teitelbaum, "Observations of End-User Online Searching Behavior over Eleven Years."

45. J. G. Marshall, "The Perceived Complexity of Database Searching among End-Users: A Multivariate Analysis," *Canadian Journal of Information Science* 12, no. 3/4 (1988): 89-97.

46. D. Barlow, B. Karnes, and G. Marchionini, "CD-ROM in a High School Library Media Center: A Research Project," *School Library Journal* 34, no. 3 (1987): 66-72.

47. Wozny, "College Students as End User Searchers: One University's Experience."

48. F. A. Brahmi, "The Effect of CD-ROM MEDLINE on Online End User and Mediated Searching," *Medical Reference Services Quarterly* 7, no. 4 (1988): 47-56.

49. Wozny, "College Students as End User Searchers: One University's Experience."

50. G. Marchionini, "Information-Seeking Strategies of Novices Using a Full-Text Electronic Encyclopedia," *Journal of the American Society for Information Science* 40, no. 1 (1989): 54-66.

51. J. H. Sweetland, "Beta Test and End-User Surveys: Are They Valid?" *Database* 11, no. 1 (1988): 27-32.

52. National Information Standards Organization (NISO), *Proposed American National Standard for Information Science–Common Command Language for Online Interactive Information Retrieval* (Gaithersburg, Md.: National Bureau of Standards, 1987).

53. Tian-Zhu Li, "Generic Approach to CD-ROM Systems–A Formal Analysis of Search Capabilities and Ease of Use," Paper presented at the Midyear Meeting of the American Society for Information Science, San Diego, May 1989.

54. Doszkocs, "CITE NLM: Natural Language Searching in an Online Catalog."

55. S. Pollitt, "CANSEARCH: An Expert Systems Approach to Document Retrieval," *Information Processing & Management* 23, no. 2 (1987): 119-38; A. Vickery and H. M. Brooks, "PLEXUS–The Expert System for Referral," *Information Processing & Management* 23, no. 2 (1987): 99-117; D. Krawczak, P. J. Smith, and S. J. Shute, "EP-X: A Demonstration

of Semantically Based Search of Bibliographic Databases," in *Proceedings of the 10th Annual International ACM SIGIR Conference in Research & Development in Information Retrieval,* ed. C. T. Yu and C. J. Van Rijsbergen (New York: ACM, 1987), 263-71; and I. Monarch and J. Carbonell, "CoalSORT: A Knowledge-Based Interface," *IEEE Expert* 2, no. 1 (1987): 39-53.

56. R. S. Marcus, "An Experimental Comparison of the Effectiveness of Computers and Humans as Search Intermediaries," *Journal of the American Society for Information Science* 34, no. 6 (1983): 381-404.

57. P. J. Daniels, "Cognitive Models in Information Retrieval–An Evaluative Review," *Journal of Documentation* 42, no. 4 (1986): 272-304; H. M. Brooks, "Expert Systems and Intelligent Information Retrieval," *Information Processing & Management* 23, no. 4 (1987): 367-82; and W. B. Croft, "Approaches to Intelligent Information Retrieval," *Information Processing & Management* 23, no. 4 (1987): 249-254.

58. P. W. Williams, "A Model for an Expert System for Automated Information Retrieval," in *Proceedings, 8th International Online Information Meeting* (London: Learned Information, 1984), 139-49.

59. Fidel, "Moves in Online Searching."

60. R. Fidel, "Towards Expert Systems for the Selection of Search Keys," *Journal of the American Society for Information Science* 37, no. 1 (1986): 37-44.

61. E. H. Brenner, J. H. Lucey, C. L. Martinez, and A. Meleka, "American Petroleum Institute's Machine-Aided Indexing and Searching Project," *Science and Technology Libraries* 5, no. 1 (1984): 49-62; and S. M. Humphrey and N. E. Miller, "Knowledge-Based Indexing of the Medical Literature: The Indexing Aid Project," *Journal of the American Society for Information Science* 38, no. 3 (1987): 184-96.

62. T. Saracevic, P. Kantor, A. Y. Chamis, and D. Trivison, "A Study of Information Seeking and Retrieving. I. Background and Methodology," *Journal of the American Society for Information Science* 39, no. 3 (1988): 161-76.