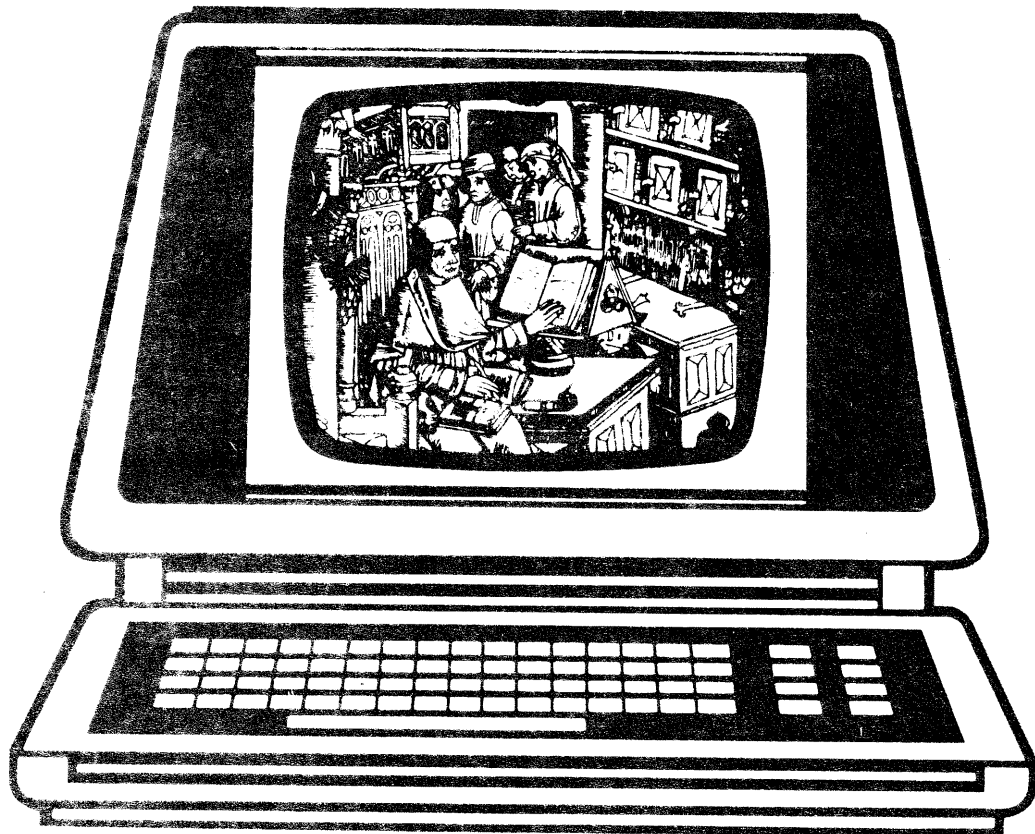


RETHINKING THE LIBRARY



in the information age

VOLUME II

ISSUES IN LIBRARY RESEARCH:
PROPOSALS FOR THE 1990s

U.S. Department of Education
William J. Bennett
Secretary

Office of Educational Research and Improvement
Chester E. Finn, Jr.
Assistant Secretary

Library Programs
Anne J. Mathews
Director

Information Services
Ray Fields
Director

Prepared for the Office of Library Programs under contract 300-86-0089 with the U.S. Department of Education. Contractors undertaking such projects are encouraged to express freely their professional judgment. This report, therefore, does not necessarily represent positions or policies of the Government, and no official endorsement should be inferred. This report is released as received from the contractor.

CONTENTS

	Page
Introduction	1
Anne J. Mathews	
Information Policy Issues: Putting Library Policy in Context	9
Louis Vagianos and Barry Lesser	
Education and Training of Librarians	43
Robert M. Hayes	
Education and Training of Librarians	75
Beverly P. Lynch	
Research Issues in Information Access	93
Pat Molholt	
Selecting Information of Enduring Value for Preservation	115
Richard J. Cox and Lynn W. Cox	
Organizing, Indexing, and Retrieving Information	131
Raya Fidel	
Information Needs: Old Song, New Tune	159
Joan C. Durrance	
The Role of the Public Services Librarian: The New Revolution	179
Brian Nielsen	
Library Funding and Economics	201
Yale M. Braunstein	
Libraries in the 1990s: The German Perspective	215
List of Participants	219

ORGANIZING, INDEXING, AND RETRIEVING INFORMATION

Raya Fidel

Graduate School of Library and Information Science
University of Washington
Seattle, Washington 98195

Abstract

Technological developments have brought significant changes in the organization and retrieval of information, and recent research has begun to investigate the use of new technologies in libraries. A review of the literature shows that new capabilities as well as new problems are associated with online catalogs, online searching of bibliographic databases, indexing (whether manual or automated), full-text retrieval, and expert systems. Future research should focus on individual needs and at the same time adopt an integrated approach that is relevant to all online bibliographic systems, with indexing and retrieving viewed as interrelated processes. The widespread use of automated systems in libraries provides an excellent environment for research because it facilitates the observation and analysis of actual searching behavior of individuals, provides a large body of data, and presents an ideal setting for studying the applications of new technological developments.

Introduction

Technological developments are no longer limited to easing our life at work and at home; they are gradually introducing qualitative changes in libraries and information systems. Technology has become a central theme in library and information science, and many articles in its literature begin with the phrase "with the advent of computers."

The vision of a paperless society, where "traditional" libraries no longer exist and all information is electronically stored and manipulated, has motivated much recent research in librarianship. Two of the basic assumptions that guide this vision are:

- Computers can be programmed to organize information for retrieval; and
- With the help of computers, users will become more and more independent in searching for information.

In other words, this vision of the future implies that the role of librarians in organizing and retrieving information is diminishing and may eventually disappear.

The paperless society is, however, only a vision. At present, users still benefit from services provided by librarians, even though much of the information

they now obtain has been processed electronically. Nevertheless, a growing number of administrators subscribe to the notion that the need for professional librarians to organize and retrieve information is diminishing. For example, some public and academic libraries assign computerized literature searching to technicians, and some special libraries have been asked to delegate searching to the users, or library patrons, themselves.

This paper reviews and summarizes the contribution of recent major works about the organization and retrieval of information. It examines the relevance of the findings to library decisionmaking, and focuses on whether research in librarianship has provided any evidence to substantiate the assumption that the processes involved in organizing and retrieving information no longer require the expertise of a professional librarian.

Definition of the Topic Area

The process of retrieving information typically consists of two stages: users first identify the information that will answer their information needs, and then they actually retrieve the relevant information. Reality, however, is much more complex than this description. More important, the concept of "information needs" is an elusive one: even if a real and precise need for information exists in an objective sense, it is difficult to define that need 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 paper, however, we assume that the information needs expressed by library patrons are clearly defined. We also assume that most of those needs are subject-related: that is, users want to find information about a certain subject. Further, users' requests include an additional facet: the purpose of the request. For example, a request for information about online catalogs by a library patron whose purpose is to learn how to use such catalogs is essentially different from the request that was made by the author of this paper, whose purpose was a literature review on the subject.

Thus, to identify information that will answer their needs, users express these needs in subject-related requests. Once they are set to retrieve the information they have already identified -- with the help of the library or any other source of information -- they look for known items.

Users can employ two types of retrieval tools for subject-related requests: the library catalog and abstracting and indexing (A&I) services. Traditionally, users have been expected to use card catalogs and printed volumes from 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.

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

Organizing, Indexing, and Retrieving Information

These two types of retrieval tools, 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). In addition, 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; users search them with free-text terms. In recent years, databases have been created that provide the complete text of documents in machine-readable form. Such full-text documents as journal articles, textbooks, and encyclopedia chapters can be searched with both free-text terms and descriptors.

It is beyond the scope of this article to provide a comprehensive review of research projects in this area. Instead, research projects that set new trends or those that reflect new approaches are reported. In addition, to keep the bibliography manageable, most often only one citation is provided for a research project, the one that is most accessible to the public.

Retrieving Information

Online Catalogs

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

Unlike research in the searching of online bibliographic databases, studies of online catalogs have focused on user requirements. Major studies have been set up to discover users' attitudes to and acceptance of the new catalogs by examining characteristics of both users and catalogs. Most studies, however, have concentrated on the human-computer interface -- that is, on how easy it is to "converse" with the computer. Only a few have addressed the retrieval problem: whether users are satisfied with the results of their searches, and what could be done to improve the results.

Further, studies of online catalogs are guided by an administrative approach that is essentially different from the approach taken by researchers in the area 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 should be examined in order eventually to design an online catalog that is most useful. In contrast, recent research on retrieval from online databases assumes that the databases and the search systems are a given, focusing

instead on the characteristics of users (i.e., professional searchers) that make for successful online searching.

Surely, the difference in research approaches stems from the reality of library decisionmaking. 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 among a number of candidates for the persons who will best perform online searches.

Nevertheless, the searching of bibliographic databases and of online catalogs is the same process: 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 provides only 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 catalogs are a relatively new phenomenon in libraries. In 1981-1983, the Council on Library Resources (CLR) funded 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 (e.g., Walton et al.² and Kranich et al.³), this first large-scale study provides most of our knowledge about the online public access catalog (OPAC). The study was conducted by various agencies, and although attempts to summarize and synthesize its results have been made (e.g., Matthews and Lawrence⁴), further interpretation and integration of the vast amount of data are required before specific conclusions can be drawn.

Typical of a large-scale study (although a novelty in research about online searching behavior), the CLR study applied a variety of methods. A survey of 8,094 users and 3,981 non-users of online catalogs in 31 libraries collected data through a questionnaire administered by various agencies. A summary of the data collected is provided by Matthews and his colleagues.⁵ 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 the users.

Focused-group interviews with library staff and patrons -- both users and non-users of online catalogs -- were carried out in six libraries,⁶ and were conducted with a group through an open, in-depth discussion led by a moderator. While such 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.⁷ These interviews supported the analysis of the questionnaires

and addressed issues such as problems in using the catalogs, possible system improvements, and the impact of online catalogs on library staff and patrons.

In addition, transaction log analyses, in which the protocols of individual searches are analyzed, were performed in seven libraries.^{8,9} It was found, for example, that while there was great variability in the length of searches and in their types (author/title or subject searches), users tended 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 (non-zero hits) in subject and author searches, frequency of commands used (in particular, "sophisticated" commands), and patterns of searches.

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

The CLR study stimulated further analyses of the data, additional explorations, and considerations of possible implications. For example, Cochrane and Markey showed how the results from focused-group interviews were useful in interpreting questionnaire findings.¹¹ Borgman, who compared the study results with research findings in other areas of online bibliographic retrieval, concluded that more similarity existed in conceptual than in mechanical problems.¹² 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 their concept of the card catalog.¹³ Last, a collection of articles assessed the impact of online catalogs on technical services, reference services, subject access, and library administration.¹⁴

In summarizing the CLR study, Matthews and Lawrence⁴ have outlined the principal findings:

- Experience with the library and its catalog is the most important factor in determining success and satisfaction in using the catalog.
- Online catalogs should provide a variety of interfaces, depending on the type of search and the level of user experience.
- User attitudes adapt to the capabilities and limitations of the online catalogs being used.
- The form and nature of training and user assistance are important.

These findings only substantiate common knowledge among librarians; they do not provide insight as to how online catalogs should be designed.

Subject Access Through Online Catalogs. The importance of subject access through online catalogs was the most significant finding of the CLR study.^{4,15} This is not surprising, since early studies of catalog use (Lipetz¹⁶ and Bates¹⁷) showed that subject access through card catalogs was inadequate. With online catalogs -- which are actually automated card catalogs -- the issue assumes even greater importance because library users who are not satisfied with subject searching

in a card catalog expect to be more successful with online catalogs.⁵ 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.¹⁸ It is not clear, however, which classification scheme is most suitable for this purpose: Dewey Decimal Classification¹⁹ or Library of Congress Classification.²⁰ Each scheme raises both conceptual and technical problems.^{21,22}

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 in titles (i.e., free-text searching),²³ or even in indexes and tables of contents of books, can improve subject access.¹⁵

Research on online catalogs has 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.²⁴ With the development of more powerful online catalogs, subject searching becomes more similar to subject searching with bibliographic databases. (For example, CITE -- the online catalog at the National Library of Medicine (NLM) -- can also access the NLM databases.²⁵) All in all, research in subject access to online bibliographic databases is more developed, and studies of online catalogs can draw on that research experience.

Online Searching of Bibliographic Databases

The first search systems for bibliographic databases were designed for professional searchers, mostly librarians. Research in online searching began in the 1970s with two large-scale studies that focused on user attitudes, satisfaction, and success.^{26,27}

Most of the research that followed, however, concentrated on the attributes of a "good searcher": the personal characteristics a librarian should possess to become a successful online searcher. Experience in online searching,²⁸ type of training,²⁹ and personal traits such as creativity³⁰ and cognitive abilities³¹ are among the characteristics examined. No conclusive results have been found, and most investigators have observed that the large individual variability among searchers impedes their studies.³²

The method used in online searching studies is well established today. Briefly, an experiment is set up in which each subject is assigned to a group, depending on his or her score on the tested characteristics (e.g., experience, cognitive style). 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 (what proportion of the citations retrieved are relevant), recall (what proportion of relevant citations have been retrieved), and unit cost (the cost of each citation).

Why have these studies failed 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 which have been used are inadequate. This conclusion was substantiated when ten librarians with almost identical subject background, training, and experience scored very differently on search process measurements.³³

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,³⁴ and no single experiment can control for all of them. Variables ignored in these experiments, such as the ability of a subject to perform a search without interviewing the patron, may have more significant effects than the ones tested.

While identifying the characteristics of a good searcher may at times help administrators to decide which employees should perform online searching in their libraries, the issue is of very little significance. This is particularly so when both online catalogs and bibliographic databases are searched by library patrons 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 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 for certain databases and search systems.

Research in this area is sparse, however, and requires a different research method. Based on the experience of librarians, Bates has proposed a number of information search tactics that could be employed in online searches.³⁵ Tactics to be used early in the search include the examination of information already found in the search, and those to be used later in the search include the rejection of items indexed by certain terms.

Fidel analyzed search protocols and verbal protocols of seven experienced searchers performing approximately 90 searches as part of their regular workload.³⁶ As a result, she listed the moves (changes in search formulations) that searchers made to increase the size of the retrieval, to decrease it, or to improve it altogether.

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 until this process is thoroughly understood.

Patron Searching. The idea that patrons can search bibliographic databases for their own requests is rapidly gaining popularity. Search systems and other

commercial outfits are providing an increasing number of gateway (front-end, or intermediary) systems such as Search Helper or Knowledge Index that are supposed to mediate between the end user (the patron) and online bibliographic systems. Such gateway systems do indeed provide a simple interface with search systems that frees end users from having to learn a command language or deal with the idiosyncrasies of specific databases (this is especially important since most end users are not likely to perform online searches frequently). Most of the current gateway systems, however, achieve their simplicity by providing end 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 end users searching their requests directly would have far more impact. The only example of such a study is a research project that examined 11 years of searching the NLM databases by pathologists and pharmacists.³⁷ The results of this study are many, but as yet not all have been published. Of particular interest here are the findings that end 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 finds of online catalog use studies: the weakest point in patrons' searching their own requests 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. At present, the only source on which we can rely for search strategies is the community of experienced online searchers.

The Role of the Librarian

Online catalogs and gateway systems are likely to increase patron use of catalogs and bibliographic databases. Moreover, as indicated by the CLR study, patrons expect more from the new catalogs than from card catalogs. While we would like to think that online systems of the future will be friendly and helpful enough to be self-explanatory, much research and development is still needed before this ideal can materialize. Although current online systems provide increasingly friendly interfaces, patrons still harbor many misconceptions about the coverage of these systems and about successful search strategies.

For this reason, the role of the librarian in assisting patrons to perform their searches has increased: training patrons in the use of online systems is now a major responsibility for librarians. The significant issue is which method of patron training is most effective.

Having online bibliographic systems for public access in their libraries has forced librarians to train their patrons. The literature on user training for both patrons and librarians abounds with reports on experiences in individual libraries, their successes and problems. Most libraries have applied instructional methods that seemed useful to them: group and one-to-one instruction, handouts, help screens, workbooks, charts, and manuals.³⁸

Some libraries have relied on examining patrons' searching behavior, either by analyzing transaction logs (e.g., Dickson¹³) or by the verbal protocol method, which consists of analyzing recorded accounts of patrons describing their work and thinking during problem solving (Sullivan and Seiden³⁹).

While such reports are anecdotal, the conclusions that are shared by all studies should be seriously considered. A summary of this literature reveals two significant findings:

- Library instruction, reference assistance, and staff experience are of prime importance, because patrons need help in forming a correct mental model of the online catalog.
- There is a need for a variety of formal and informal methods of instruction, tailored to patrons' individual needs.³⁸

Despite the findings that have emerged from the literature, as yet no evaluation study of user training has been performed.

Research Needs for the Future

Research needs for the future are greatly affected by the qualitative change in libraries and information systems that has been introduced by technological developments. Two major themes are apparent.

First, the widespread and ever-increasing use of automated systems for information retrieval enables researchers to investigate phenomena under actual, rather than experimental, conditions. The initial large-scale studies of online catalogs and bibliographic databases provided general and descriptive data about use, satisfaction, and problems. While important in pointing to research needs and in providing evidence that substantiates common knowledge among librarians, most of these data cannot be used directly in library decisionmaking or in system design. At present, however, librarians and library patrons are routinely searching online catalogs and bibliographic databases, and researchers can observe and analyze their searching behavior. Such analyses, which are based on actual searching, provide a deeper understanding of the search process because they can answer the "why" questions. They also facilitate investigations into individual variability in online searching, and they will produce findings that are relevant to library decision-making and to systems design.

Second, recent developments in libraries are removing the traditional barriers among specialties in librarianship. For instance, searches of online catalogs and bibliographic databases are essentially the same. In fact, in the future, libraries probably will provide access to both online catalogs and bibliographic databases through a single interface. Realizing this, one sees immediately that online catalog studies would be more useful if they examined the quality of retrieved sets rather than defining success as a non-zero hit; and that research on bibliographic databases should look at system features. Similarly, investigating subject access through online catalogs is no longer limited to the study of Library of Congress Subject Headings, but must consider other controlled vocabularies, indexing (manual and, possibly, automated), full-text retrieval, and expert systems.

Thus, the future calls for integrated, rather than fragmented, research. Studies should be designed to make their results relevant to the searching of both online catalogs and bibliographic databases. Moreover, with the descriptive data already collected, research in retrieval systems for subject access can now focus on its central issue, the search process. Examination of the search strategies that are available, and the conditions under which each of them should be used to achieve satisfactory retrieval, is necessary for the research to be relevant to library decisionmaking.

Human-Machine Interface. Ideally, human-computer interface should make human-computer communication as fluent as person-to-person conversation. Indeed, natural language interface, in which users converse freely with the machine, is a subject of much research in linguistics and computer science. Unlike automated bank teller machines, which are used to solve well-defined problems that are also narrow in scope, systems that provide subject access to bibliographic information are used to solve problems that are ill-defined and that range over a variety of subject matters. It seems that interfaces developed for solving simple problems are not likely to be adequate for library patrons, and that personal communication with computers in libraries is only a vision in the foreseeable future.

At present, there is some evidence that users have technical problems when communicating with the current automated systems in libraries,¹² but the nature of these problems is not clear as yet. Studies of users performing actual searches can identify the nature of these problems and guide systems designers in improving the human-computer interface. Such studies should examine problems in using commands and in understanding the computer's responses and displays.

Special attention should be given to factors that are important to the search process, such as experience or level of motivation. It is likely, for instance, that inexperienced users and experienced ones would encounter different problems, and consequently might require different sets of commands or computer displays.

Subject Access. Research on subject access through automated systems should cover two main themes: user aids and models of the search process. As described earlier, a variety of existing schemes (e.g., Dewey Decimal Classification) and arrangements (e.g., alphabetic display of the Library of Congress Subject Headings) can be made available to users to enhance subject searching. These aids should be tested with actual users, and their effectiveness and suitability for specific user needs and characteristics should be compared. Technical difficulties may preclude large-scale studies that employ all possible aids. Nevertheless, the combined results of studies examining a limited number of aids would be useful.

Most of the aids to subject access suggested in the literature are automated versions of aids that already exist in print form. Effort should be put into discovering new aids that might be applicable only in the online environment. One example is Bates' end-user thesaurus, which has a variety of special features, such as listing all terms in use in a catalog or database at any given time.⁴⁰

Models of the search process are useful for a variety of purposes. When they are based on searching behavior of "successful" searchers, they represent successful strategies. Such "success" models can be used either as examples to

be followed, in user training or as a source of knowledge for intermediary expert systems. Both uses are likely to enhance subject access.

Models of tactics and moves used to improve search results as well as a model of the selection of search keys -- whether free-text terms or descriptors -- already exist.^{35,36,41} They have been developed for bibliographic database searching, and their applicability to online catalogs should be tested. Further, other decision-making processes in the search process need to be described by such models: when to use automated tools and when printed tools are better; the selection of databases to search; and when to stop searching.

Patron Training. Theoretically, patron training can be performed either by computers or by librarians. Realistically, most patron training is performed by librarians. Although studies in individual libraries, as well as the CLR study, show that some training is needed, two basic questions require rigorous testing:

- When is training needed?
- Which methods are most effective for each user group?

These questions are not simple to answer, and straight comparative studies would not provide the necessary results. Additional exploratory research is needed before studies to answer these questions can be designed. For example, observation of users' searching behavior and an in-depth analysis of their thought processes (e.g., Janosky⁴²) will point to problems that might be remedied with training. Once these problems are identified, the contribution of training to their solution can be tested. Further, the effectiveness of each training method is likely to depend on the characteristics of individual users. While experience in online searching immediately comes to mind as an example of such a characteristic, observations of actual training sessions and the resultant searching behavior suggest other factors that should be considered in testing instructional methods.

The Need for Librarians. The complex nature of subject access and the high degree of expertise required indicate that patrons are likely to require the assistance of librarians -- whether as mediators or as consultants -- for some time to come. It is not clear, however, what makes librarians uniquely suitable for such tasks: their education, their experience, or both. It is not uncommon for administrators to assume that any person with training in online searching who performs searches frequently is as good a searcher as a librarian with the same experience. This assumption has not been tested in a rigorous manner.

A project to test the need for librarians should be more sensitive than the common experiment. It should compare retrieval performance (precision, recall, and unit cost) of librarians' searches with those of non-librarians, as well as examining the search process itself. Here, the definition of the search process should not be limited to counting occurrences of activities but should be open-ended. One may observe, for instance, that librarians are more methodical -- following a plan or a pattern -- in their searching behavior than non-librarians. Once the differences in the search processes are discovered, the effects of these differences on search results and costs should be examined. Further, the examination should include the nature of the material retrieved in addition to the numeric values of recall and precision.

Similarly, studies of patrons searching their own requests should be conducted. Such studies can draw on the experience gathered in analyzing the searching behavior of pathologists and pharmacists.³⁷ Data that describe patron behavior, including errors and problems, should be collected and analyzed, as well as data that compare the results of patron searches and those performed by librarians. Although each study of actual patrons is limited to one user group, findings common to a number of studies can be considered as valid on a general level.

Indexing

The crucial role of indexing in facilitating subject access was never questioned when only printed retrieval tools were available. Indexing was, and to some degree still is, considered to be the "heart" of organizing information for retrieval. Indexing theory was relatively rich during the 1970s;⁴³ it drew on a variety of models, with the probabilistic model⁴⁴ attracting the most attention.

Although theoretically sound and attractive, most of these models were never developed enough to show their applicability to library decisionmaking, and some were indeed abandoned. One exception is the Five-Axiom Theory of Indexing developed by Fugman.⁴⁵ Though general in nature, the theory draws on practical experience in database retrieval and could be applied to specific situations.⁴⁶

The Quality of Indexing

Indexing quality is a major concern for investigators who focus on the applications aspect of research in indexing. One measurement developed for that purpose is indexer consistency, defined as the degree to which index terms assigned to a set of documents by one indexer agree with those assigned by another, or by the same indexer at a later time.

Many studies of indexer consistency were carried out during the 1960s and 1970s.⁴⁷ They examined an array of indexer characteristics, such as subject knowledge, and a variety of indexing methods, such as the degree of vocabulary control in the index language. Here again, no conclusive results were obtained except that indexing was most consistent when the index language used was highly controlled. Also, investigators were frequently disappointed with the low consistency scores obtained.

Although indexing is performed to make subject retrieval possible, its quality has never been tested against actual retrieval performance. In other words, we do not know which indexers' work guarantees the best retrieval results. One attempt to correlate indexer consistency (as opposed to indexer characteristics) with retrieval performance revealed that indexing with controlled vocabulary -- the method that maximized consistency -- resulted in the best retrieval.⁴⁸

A more sensitive approach was taken in a study of 760 articles that were mistakenly indexed twice for the MEDLINE database.⁴⁹ Results showed that consistency depended on the subject field (e.g., anatomy vs. public health), and on the degree to which a concept was central to an article.

Organizing, Indexing, and Retrieving Information

Some opposition to indexer consistency as the sole measurement of indexing quality has been voiced, most notably by Cooper, who showed that indexer consistency is not a measure of quality because bad indexers can achieve high consistency among themselves.⁵⁰ The failure to measure indexing quality was not lamented for long, however. With the introduction of online retrieval that provides free-text searching, the question of whether indexing is needed at all has become central to research in indexing.

The Need for Indexing

Indexing is a labor-intensive and costly process, thesauri are expensive to construct, and several theoretical issues relating to their construction remain unresolved. Therefore, the capability of retrieving documents without the aid of indexing and a thesaurus is very appealing. Beginning with the Cranfield studies,⁵¹ tests have been carried out to compare the success of retrieval by free-text searching with the success of retrieval using index languages of various types.

Results are inconclusive, and the debate continues as to whether future systems should restrict themselves to free-text searching capability. In fact, one is almost forced to belong either to the "free-text camp" or to the "controlled vocabulary camp." On the basis of practical experience with database searching, however, librarians have already recognized that both access methods are needed for successful retrieval. Further, a summary of a number of case studies and of several research papers indicates that free-text and descriptor searching complement each other.⁵² This evidence is significant for both database producers and librarians. While database producers may decide to avoid the cost involved in providing for descriptor searching, they should make efforts to include abstracts because they are essential to free-text searching. A survey of 123 A&I services showed that only 46% included abstracting guidelines that were pertinent to this mode of retrieval.⁵³

Librarians, on the other hand, need to become knowledgeable about the trade-offs between these modes of searching. Their training should include clear demonstrations of the circumstances under which it is best to search with free-text terms and the conditions under which descriptor searching is more useful.

Automated Indexing

Because of the "imperfection" of human indexers, much attention has been directed toward automated indexing, whereby a computer "reads" the text of a document and assigns index terms to it. The first experiments in which machines were programmed to index were carried out in the mid-1950s. Since then, automated indexing has undergone much research by a number of talented researchers, led by the work of Gerard Salton.

Automated indexing provides several powerful capabilities: retrieval systems can be designed for natural language interface in which users express their requests in sentence-like phrases; index terms can be assigned weights according to their "importance" to a document; internal networks can lead to additional search terms; and indexing can be modified with the help of feedback from users. Indeed, several experiments have shown that automated indexing performs as well as, or even better than, the conventional methods of indexing.⁵⁴

Because of the growing number of databases that provide full-text retrieval, automated indexing seems more realistic today than in the past. While automated systems continue to improve, however, they have been tested only in experimental settings and with simplistic assumptions about users' information needs and the variety that exists in the stored documents. Therefore, such systems cannot be considered for libraries until they have been tested on databases of realistic size and in situations that reflect the complexity of organizing and retrieving diverse collections of materials.

Research Needs for the Future

Future investigations in indexing should focus on determining when indexing is necessary, rather than on whether indexing is necessary at all. Such investigations could be conducted by examining how indexing is actually used (i.e., examining actual searches).

Researchers should be sensitive to a variety of factors that may determine the need for indexing. For example, the use of controlled vocabulary may not be required in scientific databases that are searched to answer information requests from engineers in industry, while searches of the social science literature for graduate students may be most successful when descriptors are used.

Indexing quality can also be assessed anew by examining its usefulness to retrieval. Indexer-requester consistency⁵⁰ can be measured when the index terms entered by a user are compared with the index terms that retrieve the best results. Although such tests may not produce general conclusions, their results could be used to improve the indexing of the tested databases.

Most important, online systems of the future will probably provide one interface with the same mechanism for subject access to all information, whether it is stored in library catalogs or in bibliographic databases -- a practice already in use at NLM. When searching systems of the future, users will be able to search a variety of databases with one search strategy. In addition to providing one retrieval mechanism, online systems should store information in a variety of "packages," that is, employ a variety of indexing methods and practices each corresponding to a type of information need.

Thus, research should examine the usefulness of existing index languages for subject access through online systems. Although some index languages were developed specifically for online databases and others are being constantly updated to facilitate online searches, some index languages were created for manual tools and then simply transferred for use in online systems. The Library of Congress Subject Headings are the most relevant example: they are used to index information in most online catalogs and in databases with a general subject coverage (e.g., newspapers and magazines indexes). Tests should be conducted to compare the usefulness of the Library of Congress Subject Headings -- in relation to individual user groups and needs -- with index languages derived from classification schemes such as the Dewey Decimal Classification or the Universal Decimal Classification; string languages⁵⁵ such as PRECIS,⁵⁶ and possibly a general thesaurus constructed by integrating a number of index languages, each covering a specific subject.

Similarly, indexing practices should be examined. Certainly, indexing practices employed in the Library of Congress cannot satisfy the requirements of all users all the time. Moreover, it is possible to identify which indexing practices are most suitable for specific types of information needs. The level of indexing (e.g., whether to index a monograph as a whole or to index each of its chapters), the degree of exhaustivity (whether to express in indexing only the most important aspects, or every aspect that is conceivably relevant), the indexing technique (e.g., whether to check each document against a predetermined list of descriptors), and many other practices should be tested to determine the methods that are most effective in meeting each type of information need.

Full-Text Retrieval

Central to the vision of a paperless society is the prediction that information will be available in electronic form. Texts of newspapers, magazines, scientific journals, legal literature, and encyclopedias are already available online through bibliographic search systems and other systems. The provision of the full text of documents online is supposed to serve two purposes: to eliminate the need for the printed materials, and to improve subject access to those materials.

Replacing printed materials with their electronic counterparts would seem to be beneficial to libraries: instead of purchasing materials that may or may not be used, access to a central store of all materials would be possible when specific needs arise. But this scenario raises a host of problems, ranging from technical difficulties and the discomfort of reading text on a screen to cost considerations that may limit such services to a privileged few.⁵⁷ Thus, the benefit of such replacement is still questionable.

In terms of improving retrieval performance, however, the benefits of full-text systems may be more apparent. Although bibliographic search systems currently provide only a few additional search capabilities that are especially suited to full-text searching, the ability to perform free-text searches on the full text of documents is regarded as promising.⁵⁸ On the other hand, there is very little evidence to substantiate this assumption. A test of the Harvard Business Review Online showed that to achieve maximum recall, one must search both the bibliographic citation with its indexing and the full text.⁵⁹ In contrast, a full-text retrieval test of the legal literature produced very low scores for recall.⁶⁰

Full-text retrieval is a new phenomenon in librarianship. It offers solutions to some existing problems, but it also raises a number of new questions. Much research and improvement are needed before full-text retrieval can be fully implemented in libraries. The use of full-text and other non-bibliographic databases in libraries is so recent that there is still very little information available to provide a basis for research. For this group of databases, exploratory surveys are needed to uncover the range and magnitude of their use, the rate of user satisfaction, and any problems associated with having full-text databases as both a substitute for the printed source and an enhancement to retrieval.

An assembly of research methods similar to the CLR study should be employed for an initial exploration. Some of the issues to be examined are: Can readers

retrieve information by themselves? What equipment and knowledge are required for the use of these databases? What are the useful features of the display of non-bibliographic and full-text information? Do users read the text at the terminal screen or do they print a hard copy to read from? When and how can subject access be improved for full-text databases?

Expert Systems

An expert system acts as an expert in a specific field. It is projected that expert systems could be used in library management, cataloging, and reference,⁶¹ and that as a result, librarians could become more service managers and user advisors than intermediaries between patrons and the desired information service.⁶²

A crucial component of an expert system is its knowledgebase -- a database of facts, which is an expression of accumulated information in a specific field, and a set of rules. To create a knowledgebase, one must decide what knowledge to include in the database and how to represent it. While knowledge can be derived from human experts or from other sources, its representation requires techniques that have been developed in the field of artificial intelligence (AI).

Current attempts to design expert systems for libraries focus most notably on intermediary systems for online searching: IIDA provides help to novice searchers and thus can be used for computer-assisted instruction,⁶³ and CITE (the NLM online catalog) offers natural-language query input and automatic subject headings display, among other capabilities.²⁵ Still experimental is CONIT, which provides an interface with a multitude of databases.⁶⁴ In-house, microcomputer-based systems to aid reference work have also been developed at the National Agricultural Library⁶⁵ and in one library's government documents department.⁶⁶ Most of these systems are not "truly" expert systems, however, either because their operations are not based on knowledge derived from experts, or because they do not utilize AI techniques.

Existing intermediary expert systems derive their knowledge from the information stored in bibliographic databases. 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 information, 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 are a project supported by the British Library that produced a model of the search process;⁶⁷ and a list of moves to improve search results³⁶ with a decision tree that guides decisions about whether to enter free-text terms or descriptors.⁴¹ In addition, two institutions -- the American Petroleum Institute⁶⁸ and NLM⁶⁹ -- use indexers' knowledge and practice to develop expert systems to aid in-house indexing.

Organizing, Indexing, and Retrieving Information

Further, some systems employ AI techniques. CANSEARCH, for example, which has been developed as an intermediary for searching the cancer literature,⁷⁰ is based on "frames" -- a technique that has attracted much attention.⁷¹ Also, the Indexing Aid Project at NLM is experimenting with the use of a variety of AI techniques.⁶⁹

Research on expert systems for libraries is in its early years. Many basic issues, such as how to express knowledge accumulated in librarianship, remain to be investigated. Our short experience does indicate that systems developed for searching or indexing in a specific subject area and within a certain setting are the most powerful among both the experimental and operational systems. It is sound to assume that the future will see a variety of expert systems, each tailored to a specific group of users, as opposed to "universal" expert systems for searching, indexing, or reference work.

Costs

Future studies may reveal the best methods for training patrons, or the best searching aids. The value of such discoveries should be weighed against their costs. Cost-benefit analysis is difficult if not impossible to perform for subject access, but weighing costs against the frequency of use can help library administrators select a system or a method for patron training. Data to support such decisions should be made available by cost studies.

Of further help would be a detailed cost analysis of all the components and factors involved in systems for subject access. Such an analysis should examine the tradeoffs between quality control and costs, and between the adequacy of the support given to users and costs.

Conclusion

Library automation has paved the way for new approaches to organizing, indexing, and retrieving information. The ramifications of the widespread use of computers to retrieve information are far-reaching. Users today expect more from computer-assisted retrieval than from printed sources. As a result, the demands placed on librarians and researchers are more immediate than ever in the past. Librarians must acquire expertise in a large variety of subjects, ranging from knowledge of computers to principles of bibliographic control and methods of training. While users will become more independent as computer technology advances, their need for librarians' assistance to solve complex problems will become more apparent than ever before.

Research in organizing and retrieving information faces a host of new issues and an urgent requirement to address old issues that have not yet been resolved. Because actual searching of online systems can be easily monitored and observed, and because of the large volume of such searches, researchers can conduct in-depth analyses of a large body of data. Some of the tasks that should be accomplished are listed below:

Issues in Library Research: Proposals for the 1990s

- To investigate information needs expressed by users as well as the purpose for which information is retrieved, in order to develop a typology of needs. Such a typology, linked to our knowledge about cognitive information processing, is essential to the development of information systems tailored to respond to individual information needs.
- To examine methods that can help users express their search requests and formulate their search strategies. New as well as existing methods should be investigated to determine their effectiveness and suitability for different user groups (e.g., professionals, students) and for specific information needs.
- To develop a research method to study online searching behavior that will produce results relevant to user training, systems design, and library decisionmaking. The method should take into consideration the nature of information needs, should develop adequate measurements for both the search process and search results, and should establish a test procedure that is as close as possible to actual (nonexperimental) searching.
- To identify the variables that have the greatest effect on online searching behavior and their interaction, so that individual variability in online searching can be controlled for studies of searching behavior.
- To discover the characteristics of a "good search" by identifying strategies and moves that enhance the success of online searches, and the conditions under which each of them should be used to achieve satisfactory retrieval. By this means, user training, system design, and actual searching will be based on a set of well-established search strategies for specific situations in which they are likely to be successful. A model of a "good search" is also important for the design of intermediary expert systems.
- To investigate the difficulties encountered by users when they attempt to interact with a computer in order to search online bibliographic systems. Examples include using command language and understanding computer responses and displays. Resolution of these difficulties will lead to human-computer interfaces that are tailored to individual needs and requirements.
- To examine the searching behavior of patrons searching their own requests in order to identify cognitive patterns and to determine the conceptual and technical problems they encounter, as well as the quality of the answer sets they retrieve -- in particular, as compared with searching performed by professional librarians.
- To review the new roles that librarians should assume (and the modifications introduced into existing roles) with the introduction of online systems for public access, in order to examine the performance of these roles and their effectiveness.

Organizing, Indexing, and Retrieving Information

- To investigate the methods that should be used by librarians in training patrons to perform their own searches, and to examine the effectiveness of each training method with relation to individual characteristics of patrons, so that instructional methods can be tailored to patrons' individual requirements.
- To compare the performance of professional librarians with that of technical staff in searching subject requests online. Both search processes and search results should be analyzed, so that the contribution of professional training can be identified explicitly. The results of such comparisons will also point to specific skills needed by professional librarians and thus should enhance curriculum design in library schools.
- To explore the effects of integrated online searching systems that provide a single interface for both catalogs and bibliographic databases on users' searching behavior, users' need for assistance, and database and catalog design.
- To develop measurements of indexing quality. These measurements should focus on the contribution of indexing to the success of retrieval rather than on indexer characteristics or indexing conditions and situations. These measurements are necessary to begin uncovering the effects of indexing methods and practices on retrieval.
- To investigate the relationship between indexing quality and the characteristics of the information being indexed (e.g., how well subject terminology is defined in the area), so that indexing practices and methods can be made sensitive to specific characteristics of the information being indexed.
- To discover the relationship between indexing methods or practices and the success of different user groups in retrieving various kinds of information, so that indexing can be tailored to the needs of individual users.
- To investigate the conditions under which descriptor searching is most appropriate and those under which free-text searching is likely to produce better results, so that users can make informed decisions when selecting search terms, and database and catalog designers can decide which indexing method will be more effective for their users.
- To examine whether the modification of indexing practices in existing systems as a result of user feedback has an effect on retrieval performance. Experiments with automated indexing systems have demonstrated that modifications guided by user feedback can improve retrieval performance significantly. Investigating whether such improvements could be achieved with conventional

Issues in Library Research: Proposals for the 1990s

systems would provide the basis for further investigations into feedback-based modifications, their effectiveness, and their value.

- To investigate the construction of internal semantic networks that will enrich either document indexing or the selection of search terms by users, so that the potential of these networks to enhance search results can be determined with relation to individual information needs.
- To explore the feasibility of constructing a general thesaurus (not limited to a subject area) by integrating a number of thesauri, each covering a specific subject. With such a tool, general databases could employ an index language that would be more suitable for online searching than the existing languages, and a standard index language could be created for use in cooperative projects among libraries regardless of their type.
- To evaluate the effectiveness of various general thesauri (not limited to a subject area) with relation to individual user groups and needs, so that index languages used for general purposes might be tailored to individual needs.
- To investigate the relationship between indexing practices (e.g., level of indexing, exhaustivity of indexing) and retrieval requirements of user groups and needs, in order to facilitate the development of indexing practices that are tailored to individual needs.
- To survey the use of full-text databases in order to determine the range and magnitude of their use, their rate of user satisfaction, problems associated with searching them, and problems associated with using them as a substitute for printed sources. Because full-text retrieval is a new phenomenon and its use is still limited, this survey is necessary to identify specific research issues in this area.
- To survey the use of non-bibliographic databases in libraries to determine the range and magnitude of their use, their rate of user satisfaction, the subject access methods used, and problems associated with searching them. Although various search systems provide access to a number of non-bibliographic databases, no statistics about their use are available. This first step will pave the way for a more in-depth understanding of the organization and retrieval methods required for non-bibliographic databases.
- To generate and investigate new display features that may be more suitable for full-text and other non-bibliographic databases. A variety of techniques (e.g., windowing) and approaches to text display (e.g., providing intermediary displays of text before the full text is displayed) should be tested. These examinations should be led by known patterns of cognitive information processing associated with problem-solving behavior. At the same time, they should

Organizing, Indexing, and Retrieving Information

enrich our understanding of the cognitive processes associated with searching non-bibliographic databases.

- To determine whether the information needs that are answered through searches of bibliographic databases and catalogs are different in nature from those that are answered through searches of full-text databases. Also, these two types of information needs should be compared with those that require searches in numeric databases, in order to determine whether searching non-bibliographic databases is essentially different from searching bibliographic databases and catalogs.
- To explore the feasibility and to assess the advantages and drawbacks of creating a set of standards that would guide the design of all library information systems and databases. Such standards may determine which index languages should be employed, what indexing practices should be followed, and what interactive features should be provided. Because the increased use of technology facilitates increased cooperation among libraries and other information-providing agencies, such standardization may be necessary.

Future research can no longer separate the organization of information from its retrieval. Research in the organization of information can and should consider retrieval performance, and investigations into retrieval performance are not useful if they ignore the organization methods used. Organizing and retrieving information are now one and the same subject.

Along with this trend toward consolidation, technological developments and actual activities in libraries facilitate the focus on individual searching behavior. Though these two trends seem contradictory, their combination is very powerful and most useful for library decisionmaking, bringing together specialties that have been somewhat artificially isolated, and providing guidelines for the provision of tailor-made services to library patrons.

Today's database producers and search systems vendors who compete among themselves to present a large array of features only confuse the issue of which features are useful and necessary. The future calls for system design based on established user needs rather than on bells-and-whistles products responsive to marketing-inspired needs. The future calls for general standards for the development of databases and computer interfaces that are also truly adaptable to individual needs.

Acknowledgment

The author wishes to thank Kathleen McCrory for her assistance in compiling the bibliography.

Bibliography

1. Belkin, N. J., and Vickery, A. 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. Walton, C., et al. "Resistance to online catalogs: A comparative study at Bryn Mawr and Swarthmore Colleges," Library Resources and Technical Services 30(4) 1986, 388-401.
3. Kranich, N. C., Spellman, C. M., Hecht, D., and Persky, G. "Evaluating the online catalog from a public services perspective: A case study at the New York University Libraries," in The Impact of Online Catalogs, ed. J. R. Matthews, 89-117. New York: Neal-Schuman, 1986.
4. Matthews, J. R., and Lawrence, G. S. "Further analysis of the CLR Online Catalog Project," Information Technology and Libraries 3(4) 1984, 354-376.
5. Matthews, J. R., Lawrence, G. S., and Ferguson, D. K., eds. Using Online Catalogs: A Nationwide Survey. Report of a Study Sponsored by the Council on Library Resources. New York: Neal-Schuman, 1983.
6. Markey, K. "Thus spake the OPAC user," Information Technology and Libraries 2(4) 1983, 381-387.
7. Ferguson, D. Public Online Catalogs and Research Libraries. Final Report to the Council on Library Resources. Stanford: Research Libraries Group, 1982.
8. Tolle, J. E. "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, 167-171. White Plains, New York: Knowledge Industry Publications, 1983.
9. Larson, R. R. "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.
10. Hildreth, C. R. Online Public Access Catalogs: The User Interface. Dublin, Ohio: OCLC, 1982.
11. Cochrane, P. A., and Markey, K. "Catalog use studies since the introduction of online interaction catalogs: Impact on design for subject access," Library and Information Science Research 5(4) 1983, 337-363.
12. Borgman, C. L. "Why are online catalogs hard to use? Lessons learned from information-retrieval studies," Journal of the American Society for Information Science 37(6) 1986, 387-400.

Organizing, Indexing, and Retrieving Information

13. Dickson, J. "An analysis of user errors in searching an online catalog," Cataloging and Classification Quarterly 4(3) 1984, 19-38.
14. Matthews, J. R., ed. The Impact of Online Catalogs. New York: Neal-Schuman, 1986.
15. Markey, K. "Subject-searching experiences and needs of online catalog users: Implications for library classification," Library Resources and Technical Services 29(1) 1985, 34-51.
16. Lipetz, B. "Catalog use in a large research library," Library Quarterly 42(1) 1972, 129-139.
17. Bates, M. J. "Factors affecting subject catalog search success," Journal of the American Society for Information Science 28(3) 1977, 161-169.
18. Svenonius, E. "Use of classification in online retrieval," Library Resources and Technical Services 27(Jan/Mar) 1983, 76-80.
19. Markey, K., and Demeyer, A. 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.
20. Chan, L. M. "Library of Congress classification as an online retrieval tool: Potentials and limitations," Information Technology and Libraries 5(3) 1986, 181-192.
21. Cochrane, P. A., and Markey, K. "Preparing for the use of classification in online cataloging systems and in online catalogs," Information Technology and Libraries 4(2) 1985, 91-111.
22. Hill, J. S. "Online classification number access: Some practical considerations," Journal of Academic Librarianship 10(1) 1984, 17-22.
23. Lawrence, G. S. "System features for subject access in the online catalog," Library Resources and Technical Services 29(1) 1985, 16-33.
24. Carson, E. "OPACS: The user and subject access," Canadian Library Journal 42(2) 1985, 65-70.
25. Doszkocs, T. E. "CITE NLM: Natural language searching in an online catalog," Information Technology and Libraries 2(4) 1983, 364-380.
26. Lancaster, F. W. Evaluation of Online Searching in MEDLARS (AIM-TWX) by Biomedical Practitioners. Urbana: University of Illinois, Graduate School of Library Science, Occasional Paper No. 101, 1983.
27. Wanger, J., Cuadra, C. A., and Fishburn, M. Impact of Online Retrieval Services: A Survey of Users, 1974-1975. Santa Monica: Systems Development Corp., 1976.

Issues in Library Research: Proposals for the 1990s

28. Fenichel, C. H. "Online searching: Measures that discriminate among users with different types of experience," Journal of the American Society for Information Science 32(1) 1981, 23-32.
29. Wanger, J., McDonald, D., and Berger, M. C. Evaluation of the Online Process. Santa Monica: Cuadra Associates, 1980.
30. Bellardo, T. "An investigation of online searcher traits and their relationship to search outcome," Journal of the American Society for Information Science 36(4) 1985, 241-250.
31. Woelfl, N. N. Individual Differences in Online Search Behavior: The Effect of Learning Styles and Cognitive Abilities on Process and Outcome. Ph.D. Dissertation, Case Western University, Cleveland, Ohio, 1984.
32. Fenichel, C. H. "The process of searching online bibliographic databases: A review of research," Library Research 2(2) 1980, 107-127.
33. Fidel, R. "Individual variability in online searching behavior," in ASIS '85, Proceedings of the 48th ASIS Annual Meeting, ed. C. A. Parkhurst, 69-72. White Plains, New York: Knowledge Industry Publications, 1985.
34. Fidel, R., and Soergel, D. "Factors affecting online bibliographic retrieval: A conceptual framework for research," Journal of the American Society for Information Science 34(3) 1983, 163-180.
35. Bates, M. J. "How to use information search tactics online," Online 11(3) 1987, 47-54.
36. Fidel, R. "Moves in online searching," Online Review 9(1) 1985, 61-74.
37. Sewell, W., and Teitelbaum, S. "Observations of end-user online searching behavior over eleven years," Journal of the American Society for Information Science 37(4) 1986, 234-245.
38. Fahey, B. J., and Campbell, D. G. New Catalog, Same Old User: Recommendations for Teaching the Online Catalog. Oshkosh: Wisconsin University, 1985. (ED264869.)
39. Sullivan, P., and Seiden, P. "Educating online catalog users: The protocol assessment of needs," Library Hi Tech 3(2) 1985, 11-19.
40. Bates, M. J. "Subject access in online catalogs: A design model," Journal of the American Society for Information Science 37(6) 1986, 357-376.
41. Fidel, R. "Towards expert systems for the selection of search keys," Journal of the American Society for Information Science 37(1) 1986, 37-44.
42. Janosky, B. "Online library catalog systems: An analysis of user errors," International Journal of Man-Machine Studies 25 1986, 573-592.

43. Borko, H. "Toward a theory of indexing," Information Processing and Management 13(6) 1977, 355-365.
44. Cooper, W. S., and Maron, M. E. "Foundations of probabilistic and utility-theoretic indexing," Journal of the Association for Computing Machinery 25(1) 1978, 67-80.
45. Fugman, R. "The five-axiom theory of indexing and information supply," Journal of the American Society for Information Science 36(2) 1985, 116-129.
46. Fugman, R. "On the practice of indexing and its theoretical foundations," International Classification (Germany) 7(1) 1980, 13-20.
47. Markey, K. "Interindexer consistency tests: A literature review and report of a test of consistency in indexing visual materials," Library and Information Science Research 6(2) 1984, 155-177.
48. Leonard, L. E. Inter-indexer Consistency and Retrieval Effectiveness: Measurement of Relationships. Ph.D. Dissertation, University of Illinois, Champaign, Illinois, 1975.
49. Funk, M. E., and Reid, C. A. "Indexing consistency in MEDLINE," Bulletin of the Medical Library Association 71(2) 1983, 176-183.
50. Cooper, W. S. "Is interindexer consistency a hobgoblin?" American Documentation 20(3) 1969, 268-278.
51. Cleverdon, C. W. 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.
52. Svenonius, E. "Unanswered questions in the design of controlled vocabularies," Journal of the American Society for Information Science 37(5) 1986, 331-340.
53. Fidel, R. "Writing abstracts for free-text searching," Journal of Documentation 42(1) 1986, 11-21.
54. Salton, G. "Another look at automatic text-retrieval systems," Communications of the ACM 29(7) 1985, 648-656.
55. Craven, T. C. String Indexing. Orlando, Florida: Academic Press, 1986.
56. Dykstra, M. PRECIS: Recent Applications. Halifax, Nova Scotia: School of Library Science, Dalhousie University, 1987.
57. Katz, B. "The full-text online magazine," Collection Building 6(4) 1985, 26-29.
58. Goldsmith, N. "An appraisal of factors affecting the performance of text retrieval systems," Information Technology: Research and Development 1(1) 1982, 41-53.

59. Tenopir, C. "Full text database retrieval performance," Online Review 9(2) 1985, 149-164.
60. Blair, D. C., and Maron, M. E. "An evaluation of retrieval effectiveness for a full-text document-retrieval system," Communications of the ACM 28(3) 1985, 289-299.
61. Borko, H. "Artificial intelligence and expert systems research and their possible impact on information science education," Education for Information 3 1985, 103-114.
62. Mason, R. M. "Office automation and information technology trends -- their impact on libraries and information center management," Microcomputers for Information Management 3(1) 1986, 1-14.
63. Meadow, C. T., Hewett, T. T., and Aversa, E. S. "A computer intermediary for interactive database searching. II. Evaluation," Journal of the American Society for Information Science 33(6) 1982, 357-364.
64. Marcus, R. S. "An experimental comparison of the effectiveness of computers and humans as search intermediaries," Journal of the American Society for Information Science 34(6) 1983, 381-404.
65. Waters, S. T. "Answerman, the expert information specialist: An expert system for retrieval of information from library reference books," Information Technology and Libraries 5(3) 1986, 204-212.
66. Smith, K. F. "Robot at the reference desk?" College and Research Libraries 47(5) 1986, 486-490.
67. Williams, P. W. "A model for an expert system for automated information retrieval," in Proceedings, 8th International Online Information Meeting, 139-149. London: Learned Information, 1984.
68. Brenner, E. H., Lucey, J. H., Martinez, C. L., and Meleka, A. "American Petroleum Institute's machine-aided indexing and searching project," Science and Technology Libraries 5(1) 1984, 49-62.
69. Humphrey, S. M., and Miller, N. E. "Knowledge-based indexing of the medical literature: The Indexing Aid Project," Journal of the American Society for Information Science 38(3) 1987, 184-196.
70. Pollitt, A. S. "A front-end system: An expert system as an online search intermediary," Aslib Proceedings 36(5) 1984, 229-234.
71. Minsky, M. "A framework for representing knowledge," in The Psychology of Computer Vision, ed. P. Winston. New York: McGraw-Hill, 1975.

Comments by Reviewers

The three reviewers who commented on this article addressed the issues of retrieving, indexing, and full-text retrieval. They also suggested alternative approaches to the topic area, additional courses of investigation, and additional research objectives.

Retrieval

Fidel encourages constructing a "general thesaurus by integrating a number of thesauri, each covering a specific subject." This research objective could be expanded to include the study of retrieval language compatibility in general. This topic, long of interest to Europeans, is now the focus of research projects sponsored by the National Library of Medicine and is the topic of Multiple Thesauri in Online Bibliographic Systems, recently published by the Library of Congress.

Fidel's concept of successful online searching by patrons warrants further research as well, to establish the likelihood of patrons helping themselves with online searching. Available data suggest that the novelty of patron searching wears off quickly and the task reverts to the professional searcher, both in the public library setting and in the corporate setting, where time is critical and patrons prefer having results delivered to their desk.

The Council on Library Resources (CLR) study cited by Fidel is important, but was conducted too early in the development of online catalogs. More current research on online catalog systems could be considered in drawing conclusions about user interfaces with online catalogs. Another aspect of retrieval to be addressed is the use of microcomputer-based interfaces to facilitate user access to bibliographic and non-bibliographic data systems and as a means to link those systems.

Works by Brian Nielsen and Betsy Baker are recommended as further references on the use of online systems.

Indexing

Fidel recommends that research in the field "investigate the conditions under which descriptor searching is most appropriate and those under which free-text searching is likely to produce better results." Such studies could also include designing controlled vocabularies to be used in conjunction with free-text searching as well as in selecting search terms.

She also suggests that research be conducted to "investigate the relationship between indexing quality and the characteristics of the information being indexed." Studies along this line must bear in mind that methods of indexing developed for the sciences are inappropriate for the social sciences and humanities. Furthermore, a study of different sublanguages in the humanities and social sciences is prerequisite to designing vocabulary control mechanisms for them.

The need also exists, on a general scale, for a uniform filing, classification, and indexing or cataloging system to make retrieval of the documents a smooth

and uniform activity. This is essential in university or regional research collections, as well as in commercial environments.

Full-Text Retrieval

Fidel recommends surveying "the use of full-text databases . . . to identify specific research issues in this area." Although full-text searching has been acclaimed by some as a breakthrough in online searching, more research is needed to assess its precision, recall, and rate of user satisfaction.

She also encourages investigating "new display features that may be more suitable for full-text and other non-bibliographic databases." This research could usefully be extended to bibliographic databases as well. Wajenberg would be a good source of information.

Additional Research Objectives

Descriptive cataloging is not emphasized in this paper. The following research objectives might be explored in this area:

- To investigate the principles underlying catalog code design with a view to their appropriateness for the online era, using the work of Gorman, Wilson, and Malinconico as possible resources.
- To investigate the impact of technology on networking and particularly on how technical services related to the organization of materials will be performed in the future. Specifically, does the advent of CD-ROM technology pose a threat to large bibliographic utilities such as OCLC? How successful will cooperative ventures like the linked systems project be?
- To investigate the possibilities of automatic cataloging and expert systems to assist with cataloging. Work is being done in this area at OCLC, UCLA, Texas and in Europe.