From Authority Control to Informed Retrieval: Framing the Expanded Domain of Subject Access

Prudence W. Dalrymple and Jennifer A. Younger

The expanding domain within which subject searching takes place creates new challenges in providing effective subject access. Three changes in the domain of subject access are described. The purpose of this article is to propose a broader framework within which to analyze barriers to effective subject access and to propose directions for research and action. Methods of facilitating subject access are categorized as end points on a continuum: authority control, controlling the results of indexing, and informed retrieval—informing the process of searching. Identification of these two categories signals a paradigmatic shift toward a reliance on both authority control and informed retrieval to facilitate subject access.

This article reexamines some of the methods and approaches taken to provide subject access, especially in light of the many technological changes that have occurred in academic libraries over the past several years. The approaches are not necessarily new; some of them were proposed several decades ago. However, just as Vannevar Bush’s dream of a MEMEX seems to be embodied in today’s sophisticated workstations, the growing availability of integrated information systems, networks, and artificial intelligence may indicate that the ideal solutions which were once posed only as hypotheses are being developed today and will be adopted tomorrow. Solutions to problems previously limited by technical constraints can now be realized and brought to bear in today’s environment.

A central problem in academic libraries today is subject authority control. Subject authority control has generally been considered the domain of catalogers rather than reference librarians, of technical services departments rather

Prudence W. Dalrymple is an assistant professor at the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, and Jennifer A. Younger is an assistant director of Central Technical Services for the General Library System of the University of Wisconsin-Madison, Madison, Wisconsin 53706. The authors wish to thank our colleagues who provided thoughtful comments on earlier drafts of this paper; we particularly wish to acknowledge F. W. Lancaster and Pauline Atherton Cochrane for their valuable assistance in helping us place our ideas in a historical context.
than public services departments. In an online searching environment, the capabilities of keyword searching and Boolean operators have enhanced subject retrieval by freeing searchers from the necessity of knowing the exact form of a subject heading and giving them the power to combine the terms that bring together the concepts of interest. Before the widespread availability of online information retrieval, indexes, not librarians and library users, used thesauri. With the vision of the electronic "library without walls" rapidly becoming a reality, however, it has become possible, even necessary, to consider issues of information representation and storage together with issues of information search and retrieval. One of the most challenging issues in electronic information retrieval (CD-ROM, online catalogs, and online information retrieval systems) is providing effective subject access to users. This article examines three dimensions along which changes in the domain of subject access may be considered and describes current methods of providing subject access by controlling the results of indexing as well as by informing the search process. This study also suggests an integrated framework through which subject access may be facilitated and proposes an agenda for practice, education, and research. The major purposes of this article are twofold. First, we propose that methods intended to assist in the searching process be grouped together and the group subsequently be called "informed retrieval" methods. Informed or guided retrieval is defined as using feedback from an information system in order to improve retrieval results. A simple example of using such feedback is scanning the subject headings assigned to a book about a topic of interest, and then incorporating those terms in a subsequent search. This concept of informed retrieval parallels that of authority control, the establishment and use of controlled vocabularies in indexing. Second, we propose that an integrated framework be developed, one that includes both authority control and informed retrieval as equally important methods in providing effective subject access. Whereas practical and research efforts are currently under way, these efforts should be coordinated within a framework that considers authority control and informed retrieval as partners, rather than distant relatives, in improving subject access.

DEFINITIONS

Subject headings have been most closely associated with catalogs, and descriptors with indexes. In their electronic forms, indexes are commonly referred to as databases, while catalogs retain the name catalog, modified by the term online. In this article, we use the term "database" to refer to the online form of an index, and refer to catalogs, regardless of form, as "catalogs." Of course, both online catalogs and online indexes are databases in that they gather information units in a structured form to facilitate access.

The term "subject access" refers to the ability to conduct a search for information on a topic by various means: subject headings or descriptors, keywords, or the name of something or someone thought to be associated with the topic. (Citation indexes facilitate the latter form of access.) Closely tied to this issue of subject access is the issue of authority control. Authority control has been regarded as a service to library users, ensuring consistency in the choice and form of words or groups of words that are used to represent a specific concept in a bibliographic record. From the indexer's perspective, authority control means the process of matching the intended term with the terms prescribed by a thesaurus. When no match occurs, the process of adding a new term to the thesaurus can be initiated or the next best term may be selected.

THE DOMAIN OF SUBJECT ACCESS

Librarians are accustomed to thinking of subject searching as taking place in a number of locations and using multiple sources. Users routinely search sources such as library catalogs, printed indexes, and online databases in their efforts to find bibliographic citations for items pertaining to the topics of their interest. Traditionally, these sources not only are physically separate, but also are searched in distinct ways. An examination of the dimensions of the environment, or the domain as we have termed it, of subject access reveals accelerating changes that are transforming it, offering new options to users and creating new challenges for librarians. The three dimensions to be examined are: (1) new online searching capabilities that differ from those of manual searching; (2) decreasing visibility of the boundaries between multiple subject indexes and catalogs, as access is provided through a single computer terminal; and (3) the diversity of approaches to providing subject access.

NEW ONLINE SEARCHING CAPABILITIES

The introduction of online public access catalogs during the decade of the 1980s has extended online searching capabilities to catalogs. This development is presented here to users the options of keyword and Boolean searching, truncation, and cross-field access (title, author, subject) to enhance their searching. In all of these approaches, the user now has the choice of more than one method of searching, and the user has the power to choose the method that best suits his or her needs.

DECREASING VISIBILITY OF INDEX CATALOG BOUNDARIES

On more and more campuses throughout the country, databases produced by external suppliers are being purchased and mounted on campus computer networks along with the local online catalog. The "library without walls," in which faculty and students can search the library's catalog and selected databases such as MEDLINE and ERIC from a single terminal in home or office, has all but erased the distinction between bibliographic records produced by catalogers and indexes. These new search capabilities significantly change the kind of subject access available to searchers, giving them powers previously available only to indexers and catalogers.

These new search capabilities significantly change the kind of subject access available to searchers, giving them powers previously available only to indexers and catalogers.

(The proliferation of optical disc systems that resemble online systems has further clouded these distinctions and may contribute to users' difficulty in knowing exactly what they are searching through a particular terminal.) The once clear division between subject headings and descriptors—the former occurring on cataloging records, the latter appearing in indexing records—has also...
THE IMPACT OF THE CHANGING ENVIRONMENT ON USERS

In the face of these environmental changes, the problem becomes one of continuing to provide subject access in a shifting environment. Although effective subject retrieval has been understood as an optimal balance between recall and precision, there are currently no standards which can be used to determine when this optimal balance has been reached. Rather, effective retrieval from specific search strategies carried out in individual situations and evaluated by individual users whose judgments of relevance may be influenced by a variety of factors such as time, money, information need, and previous knowledge. To examine the problem of providing effective subject access, the authors describe current methods of providing subject access by controlling indexing results as well as by informing the search process, thus defining a continuum on which methods established to assist in subject retrieval can be identified and placed.

CURRENT METHODS OF CONTROLLING INDEXING RESULTS

A thesaurus or subject heading list is the traditional approach to defining concepts and mapping the relationships among concepts in a field in order that material on those subjects can be represented in a consistent way. From the vocabulary "at large," some terms are selected for use in the controlled vocabulary and references are made from terms not used, synonyms or near synonyms, to the selected terms. Many choices are made in the construction of a thesaurus; among them is the choice of whether a "top-down" or "bottom-up" approach will be employed in assembling terms; that is, whether the terms represent all knowledge in the field, regardless of the frequency with which they appear in the literature, or whether terms shall be determined by literary weight, i.e., their appearance in the literature of the field. Choices about the number of cross-references from popular usage or new terms must also be made. Overall, considerable time and intellectual energy are brought to bear in the construction of a well-designed subject authority list or thesaurus. The expenditure of these resources is justified in the belief that controlled vocabularies, because of their classifying functions, are the primary means of facilitating recall and, therefore, provide a needed service to users.

While authority control exists as a service to users, in practice, users have had little or no input into the construction of thesauri. With the exception of consultative groups representing specific subject fields, it is usually librarians or indexers who determine controlled vocabularies. However, pointers provided involving users in the creation and maintenance of searching vocabularies nearly three decades ago. In the 1980s, Jean M. Tague proposed user-responsive subject headings and Marcia J. Bates described a "superthesaurus" in which virtually all terms would be available as a "front end" to assist users in finding their way through a series of cross references to the controlled vocabulary term. While these proposals have not been developed into working systems, other similar ideas have been implemented. Sara D. Knapp's early work in developing BRS/TERM, a vocabulary database for searchers, is a "home-grown" solution to the problems of rationalizing natural language to the numerous controlled vocabularies that exist in multifile systems such as BRS. In an experimental interface at the University of Illinois, one may propose additional terms for inclusion in the controlled vocabulary. Librarians review these for permanent inclusion. The personal HYPERCATalog project initiated at LILAB at Linkoping University in Sweden also has features whereby authors and users indicate indexing terms and leave a "usage trail" that subsequent searchers may follow. (A usage trail is analogous to a hiker through forests selecting "frontier" paths, they leave their footprints, eventually making a visible trail. Several trails may exist simultaneously, each having advantages in terms of access, destination, ease, efficiency, perspective, or resources. Subsequent searchers can select an already available trail or strike out on their own. Similarly, expert searchers can establish pathways through the library's catalog.)

The controlled vocabulary started within single disciplines and databases. Users can search multiple databases covering a number of distinct subject areas from a single station, vocabulary control becomes more complex. While some thesauri attempt nearly universal scope (LCSH), others, such as Medical Subject Headings (MeSH), are more closely related to a particular discipline. To facilitate working with concepts in one database to the same concept in another database, special tools such as the Cross Reference Index, which links already-existing terms in thesauri, have been developed. Through facilities such as DIALINDEX and BRS/TERM, major online database vendors such as Dialog and BRS enable searchers to ascertain the frequency of occurrence of terms across several databases to assist in term selection. Still another method of rationalizing multiple indexing languages is a prescriptive approach such as the Universal Medical Language System (UMLS). In this project, indexers are developing a specialized language to represent concepts in a variety of
CURRENT METHODS OF INFORMING THE SEARCHING PROCESS

As methods of controlling the results of indexing through authority control change to meet current challenges, users' efforts to design an effective search strategy must adjust accordingly. Librarians can assist users in developing improved search strategies through bibliographic instruction and system design. Because the bibliographic instruction approach is adequately described elsewhere in the literature, the authors do not discuss this approach here.8 We do suggest, however, that bibliographic instruction librarians possess a rich store of knowledge that could be incorporated into system design. Such systems would have as their objective effective subject access through "informed retrieval." As mentioned earlier, we define informed retrieval as using feedback from an information system in order to improve retrieval results. The system provides feedback and integrates it into system design. However, the user controls it and can invoke it as needed. Informed retrieval is not totally automatic and does not, therefore, perform exactly as a counterpart of an automatic indexing system. Instead, informed retrieval provides to the user greater control over the results of the indexing process. Therefore, informed retrieval constitutes a form of user-exercised "authority control."9

Almost twenty years ago, F.W. Lancaster described many of the premises on which these ideas are based. They are reflected in the design of such early systems as SUPARS, an online system for accessing Psychological Abstracts developed at Syracuse University.10 These ideas are closer to more widespread implementation because the technology is now available. However, librarians cannot expect users to search immediately with the same sophistication that indexers have, particularly when these searchers have not been educated in the principles of bibliographic control and access. Librarians must provide help, e.g., through vocabulary displays that make relationships among terms clear, leading users from unused terminology, allowing the user to conduct the search in natural language which the system then translates into the appropriate controlled vocabulary. Although early online systems were intended to be searched by end users, they failed to attract large numbers of users not because of the theoretical principles that underlie their design, but because the technology was too primitive to support them adequately. Despite the introduction of programs to assist users in devising suitable search strategies, truly effective information retrieval by end users remained an elusive ideal.

Since the early 1970s, further research into end-user searching indicates that end users are not performing particularly effective searches. Comparisons of the search results of search intermediaries with those of end users reveal that trained intermediaries retrieve a greater number of relevant citations than do end users.11

A FRAMEWORK FOR FACILITATING SUBJECT ACCESS

The preceding discussion leads the authors to suggest that several trends are converging toward the shared goal of providing effective subject access to users. This convergence exercises control not only over the indexing process, but also over the searching process. Indexing and searching affect one another in many ways: depth, entry vocabulary, and the use of browse pages. The traditional controlled vocabularies supplied, maintained, and applied by librarians and indexers can be used to form the basis for user vocabularies. Additionally, informed retrieval provides feedback during the search process, notifying searchers of their location in the database relative to the controlled vocabulary's structure, as well as their progress in locating relevant materials. In other words, the system provides feedback about where the searcher is located in the search, whether more information is available, and what can be done to retrieve it.

We do not wish to diminish the importance of, or necessity for, controlling the indexing process. On the contrary, we believe that as long as humans or machines perform indexing, guidelines and control are essential. The concept of informed retrieval, when combined with controlled vocabularies, is powerful indeed. Two examples of this kind of combination are Knowledge Finder, a CD-ROM MEDLINE system for the Macintosh, and Autocat, an online catalog developed at Dickinson University. Both use the power of controlled vocabularies and provide the user with an array of tools that can be invoked as needed when monitoring the progress of the search. Other examples do not doubt exist.

The traditional controlled vocabularies supplied by librarians and indexers can be used to form the basis for user vocabularies.

Current research at OCLC exemplifies another line of investigation. A recent study examined the extent to which user-entered terms correspond to the catalog's controlled vocabulary and determined that up to two-thirds of user-entered subject terms do not match LCSH terms in exact, normalized, or keyword form. Further analysis of the searches will identify those subject terms that are close approximations of LCSH terms and will identify what system capabilities are needed to assist users in approaching LCSH terms more closely.12 To the degree that such research is successfully implemented and disseminated, subject authority control will be one means of achieving optimal recall and precision in subject searching. Subject authority control and informed retrieval will be seen as complementary and essential components of providing subject access.

While ideally we would like to see orderly progress toward effective subject access through directed efforts within a set of theoretical principles, in fact the practical steps are isolated and atheoretical. A concerted effort to recognize, coordinate, and promote work would encourage quicker and more efficient progress.

AN AGENDA FOR PRACTICE, EDUCATION, AND RESEARCH

Directed changes in practice, education, and research are needed to achieve
March 1991

We suggest that subject authority control and informed retrieval form a logical coalition for providing subject access.

For many years, curricula in library schools have reflected the traditional administrative division between public service and technical services in libraries. In so doing, the schools fail to establish leadership in designing an integrated approach to effective subject access. Subject indexing, bibliographic instruction, and information system design courses are taught with minimal attention given to the role that each function plays in enhancing subject access. A first step toward achieving this approach is integrating the teaching of the representation of information with the retrieval of information. Some library schools have already taken this step.

The theoretical foundations for a unified approach to research in subject access come from both within the field of library and information science and from other disciplines such as cognitive psychology, linguistics, and information storage and retrieval. One strength of this multifaceted foundation is that various insights can be applied to the broad problem of creating effective subject access. The three parts of this problem are controlled vocabulary and subject authority; informed retrieval applied during the search process; and the potential of natural language processing to either enhance or supplant traditional human subject cataloging.

While well-established research tradition on the performance of controlled and uncontrolled vocabularies in the retrieval process exists, many of these studies have taken place in a laboratory environment.

Earlier research concentrated primarily on the performance of indexing schemes and did not consider retrieval from catalogs. Only with the introduction of the online catalog have catalog-use studies looked more closely at the process of subject access. Furthermore, much of this research failed to consider the effect of either the user or the system interface on the retrieval process. Recent research has attempted to examine performance from many dimensions, taking into consideration the effect of the query, the retrieval technique, the searcher, and the nature of the search process. We call for greater attention to the potential in these areas and for discussion and research that involves researchers in other disciplines, librarians, and information scientists. By suggesting informed retrieval as a theoretical framework, we hope to enable librarians and information scientists working in both areas—indexing and retrieval—to coordinate efforts to produce truly exciting and informative systems.

FUNDING AND SUPPORT

A critical item on any agenda for action and research is the question of funding: Who will pay? In times of limited resources when new monies are scarce, funds must come from more efficient utilization of present programs, from cost shifting, or from changed funding priorities.

The bibliographic control community has borne a large part of the fiscal responsibility for subject access. Catalogers at the national level, through institutional or agency support, have maintained national libraries such as the Library of Congress, the National Library of Medicine, and the National Agricultural Library, or through cooperative cataloging efforts promoted by the bibliographic utilities such as OCLC and RLG, have carried the burden for current subject access. While users have indicated that enhanced subject access is a high priority, librarians have long stated their concerns for the restricted resources available to fund the development of subject headings. Yet, cooperation among decision-making agencies has brought a much-needed focus to addressing the problems of funding effective subject access through cooperative cataloging programs and shared research.

Recently, system designers and developers, both in the private and public sectors, have picked up some of these costs. In some cases, they have been recouped these development costs by marketing their systems directly to users, bypassing the traditional library channels. Perhaps the scarcity of available money to fund research and discussion between the research and practice communities further reflect this desire for a return on investment. With regard to the role users play in bearing the burden of subject access, Lancaster suggested that users also contribute to the cost of achieving precision in information retrieval because they spend time and effort to prepare a search strategy, run the search, and refine the results.

The implications of cost shifting suggest that the bibliographic control and system designer community, as well as users, have a substantial stake in decisions affecting subject access. If so, it is even more important to unite the knowledge and expertise of these communities in an integrated approach to creating effective subject access. As money continues to diminish, both cataloging and reference librarians are turning to other ways of managing their bibliographic control operations. Using paraprofessionally designed and implemented expert systems for both training and task performance, and simplifying cataloging are all attempts to cope with financial and staffing crises in both cataloging and reference areas.

SUMMARY

In the world of Charles Cutter, providing subject access was a manageable task. In retrospect, it appears that a
readily distinguishable topical nature characterized publications, mostly monographs. Catalogers and users alike identified topics similarly. The information explosion, together with the pluralism of today's society, produces a complex and diverse set of publications to which people must subject access.

If librarians agree that effective subject access is a realistic goal, and that authority control and informed retrieval are viable means for achieving that goal, they must begin to integrate presently fragmented efforts into a unified agenda. A great deal of time, talent, and money have been expended to achieve these goals. Now a clear path toward a unified framework is needed. Separate communities within librarianship—in indexers and catalogers, reference librarians, bibliographic instruction specialists, and system designers—have developed differing approaches to solving this universal problem. While positive effects may be derived from differing approaches, all approaches must contribute to a unified approach in recognition of the constraints of scarce resources.

The framework proposed embodies a comprehensive approach that includes authority control and informed retrieval as a holistic approach to subject access. Both representation and retrieval are integral parts of a holistic approach to subject access. Through the implementation of an agenda for action in practice, education, and research, librarians can ensure a future of effective subject access conceived within a unified framework of authority control and informed retrieval.

REFERENCES AND NOTES


15. Pauline A. Cochrane, "Improving the Quality of Information Retrieval—Online to a Library Catalog or Other Access Service... Or... Where Do We Go from Here?" Online 5:30–42 (1981).


