Beyond Boolean: Designing the Next Generation of Online Catalogs

CHARLES R. HILDRETH

Introduction

THE ONLINE CATALOG will never be a finished, perfected product. Nor will the library catalog in its many online manifestations ever achieve the universal, familiar uniformity experienced by users of the twentieth-century card catalog. These twin realities threaten many librarians, represent a myriad of problems for catalog users, and challenge designers and developers of online catalogs to improve their systems specifically for the untrained occasional user of the library catalog.

Some writers view the online catalog as a new *form* of the library catalog, succeeding the earlier book, card, and COM catalogs. This perspective, although too narrow and unimaginative, has served as a useful point of departure for identifying the unique characteristics of the online catalog. Five years of examination and reflection have led this author to conclude that the online, interactive catalog has the potential to overcome all the major limitations of earlier forms of the library catalog (book, card, and COM). When its unique characteristics are fully understood, it becomes clear that the online catalog is far more than the traditional (read "card") library catalog executed in a new medium. Stated in somewhat general terms, the online catalog stands apart from earlier catalogs because it is *interactive*, infinitely *expandable*, and *public*.

As an interactive system, the online catalog can dynamically communicate with its user; it can be responsive and informative at a given time to a given need. The online catalog is "fence resistant." Its form

Charles R. Hildreth is Chief Consulting Scientist, READ Ltd., Worthington, Ohio.

does not constrain its development and expansion. Access points and pathways to the stored data can be continually added and redefined. Search, display, and support functions can be added or modified easily. Related data files (e.g., abstracts and book reviews) can be brought within the online library catalog. Linkages can be established between online catalog systems and other online information systems. Finally, the online catalog is public and very revealing of its use. Many of the mysteries of what actually takes place when a user is searching the catalog can now be solved. The searching activity can be logged (in its entirety, if desired) for examination and analysis. What users of the catalog actually do in the search process, if not *why* they do it, can be objectively ascertained. Patterns of search behavior, including encounters with problem situations, can be discovered for an entire population of users of a given online catalog.

The unique potential of the online catalog, together with the ever changing technologies that support online catalogs, leads to the inescapable conclusion that "we may have to adapt to a continuing state of mutability. The online catalog is not only an instrument of change in today's libraries, it is also everchangeable."[1] Automated library systems in general, and specifically online catalogs, will continue to be produced and enhanced from a variety of sources: in-house development, library consortia, and commercial firms. This will result in a diversity of online catalogs for some time to come.

Although dozens of different online catalog systems can be found in hundreds of libraries in North America and Europe, a determined observer can produce a "snapshot" (somewhat blurred and fuzzy around the edges) of today's online catalog scene. This article presents a brief overview of the state of the art of online catalogs. It discusses recent progress in the design and development of operational online catalogs, why the current generation of online catalogs falls far short of their potential, and what new directions for online catalog design should be expected.

Second-Generation Online Catalogs

In an earlier paper, this author introduced a classification scheme of three generations of online catalog developments to chart recent history and to cast some light on the likely course of future catalog design.[2] This approach assumed we could identify qualitative stages of evolution in the design and production of online catalogs. Each of the three generations was defined by a characteristic set of features. No

attempt was made to assign a fixed span of dates to any of the three generations. Conceivably, first-generation systems could be in operation in a world of second- and third-generation systems. The aim of the classification scheme was to generate informed criticism of the state of the art, especially those online catalog systems available in the library marketplace. This author did not expect that several commercial suppliers of online catalogs would subsequently refer to their new or updated products as "second-generation" catalogs.

The three-generation classification of online catalogs is useful once again, because it provides a framework for explaining precisely where online catalog development stands today. Almost without exception, we have moved beyond first-generation online catalogs. That is the good news. However, online catalog development has slowed to a snail's pace. Many of the commercial suppliers of second-generation online catalogs believe they have "finished" the job by adding online *public* access catalogs to their product lines. The danger exists that these commercial suppliers of online catalog systems will become stuck on the plateau of second-generation developments.

Several factors contribute to this apparent complacency: the vendors of turnkey library systems more and more have to assign scarce development resources to the support of existing installations. A vendor's choice of hardware architecture, software, and lack of imagination (read: "insensitive to the real needs of public access") may make it extremely expensive (or impossible) to provide enhancements that address more than the housekeeping tasks of a library. Kenneth Dowlin, director of the Pikes Peak Library District, suggests that most of the "integrated library systems" available in the marketplace "freeze the library into the housekeeping tasks phase and allow for little expansion" into later phases directed toward improving public access and public services.[3] However, the commercial suppliers have demonstrated that they will respond to competitive pressures and the demands of librarians for additional functions and features. Witness the rapid development of subject access (however rudimentary) and some measure of authority control once these appeared as standard "requirements" in Requests for Proposals (RFPs) during the years 1982-1985.

Librarians must continue to play the role of change agent for the online catalog. But this will require that they make efforts to learn about the potential of online retrieval, catalog access issues that cannot be couched in the familiar terms of card catalog use, and user-system interface problems and promises. More importantly, a fundamental shift in priorities is needed. In her recent review of library automation

and networking developments over the past two decades, Markuson reminds us that most of our efforts have been aimed at automating the library and the functions of the librarians, not at automating access and retrieval systems for our users.[4] The concentration has been "on control rather than access," according to Markuson, and she sees much "evidence of the continuing priority of control over access."[5] Successful efforts to bring more needed information (beyond that contained in MARC catalog records) to the users of online catalogs, and efforts to make this new access instrument both easier to use and more effective, must be based on a radical shift in our priorities from bibliographic control to information access. This requires a shift in our demands from better automated systems for serving librarians to systems designed to more effectively provide direct service to library patrons, the "primary" users of our libraries.

This period of developmental slowdown or complacency on the part of the commercial suppliers of online catalogs has its positive side. For librarians who will be involved in the evaluation and selection of online catalogs in the future, it provides time for learning and "catching up" on the state of the art, online access issues, and users' needs. It is necessary to understand how today's online catalogs have moved beyond the first-generation systems. First-generation online public access catalogs were characterized as being "known item" finding tools, which provided few access points (typically only author, title, and control number) to short, nonstandard bibliographic records. They were either crude attempts to replicate the card catalog online, or automated circulation database query systems masquerading as public access library catalogs. Many agree with Malinconico's astute analysis of circulation control systems as falling far short of any system deserving to be called a library catalog.[6] In first-generation catalogs, searching was initiated by derived-key input or by exact term or phrase matching on at least the first part of the term or phrase (as with heading searches in the card catalog). In addition to lacking subject access, including any keyword access to titles and subject headings, first-generation online catalogs provided only a single display format, a single mode of interaction with the system, and little or nothing in the way of online user assistance. Refining and improving a search in progress, based on an evaluation of intermediate results, was out of the question. Without subject access, authority-based searching with cross references, and meaningful browsing facilities, first-generation online catalogs were understandably criticized as inferior to traditional library catalogs.

Today's second-generation online catalogs represent a marriage of the library catalog and conventional online information retrieval (IR)

systems familiar to librarians who search online abstracting and indexing databases via DIALOG, BRS, ORBIT, MEDLINE, etc. Improved card catalog-like searching and browsing (via headings and cross references) capabilities have been joined with the conventional IR keyword and Boolean searching approaches. Many online catalogs support the ability to restrict searches to specified record fields, to perform character-masking and/or right-hand truncation, and to limit the results by date, language, place of publication, etc. Also, bibliographic records may be viewed and printed in a number of different display formats.

Second-generation online catalogs should be viewed as bibliographic information retrieval systems. But when compared to their conventional IR forebears, these key differences should be kept in mind:

—the online public access catalog must be usable directly by untrained and inexperienced users (online assistance is usually provided to help with the mechanics of searching);
—records in the catalog database lack abstracts, the subject indexing is sparse and uses broad terms not representative of current terminology; and
—the catalog database, in covering a library's collection, includes information on a wide variety of knowledge fields and subject areas.

Designers of second-generation online catalogs have addressed these differences in two ways: by providing card catalog-like precoordinated phrase searching and browsing options (along with keyword/Boolean capabilities), and by providing more and more online user assistance in the form of menus, help displays, suggestive prompts, and informative error messages. On the other hand, post-coordinated keyword searching on subject-rich fields (e.g., titles, corporate names, series entries, notes, and subject headings) serves to alleviate the twin problems associated with the sparse subject indexing of most library materials by the Library of Congress (using its list of subject headings—"LCSH") and the users' unfamiliarity with the controlled indexing vocabulary.

A library catalog that fulfills Cutter's classic objectives for the catalog in the online environment is a significant accomplishment. It succeeds in at least two ways: users prefer the online catalog to either the card or the COM catalog, and the online catalog is easier to maintain and update than earlier forms. Designing a keyword/Boolean information retrieval system as an online catalog that is easier to learn and easier to use than the conventional, commercial IR systems is also a significant accomplishment. The traditional, well-structured library catalog has been joined with the power and flexibility of conventional IR systems.

The prevailing temptation to be satisfied and to rest on our laurels is easily understood. We have come far, and the journey has been costly.

The Need for Further Improvements

Second-generation online catalogs can be used effectively by library staff and by library patrons trained to use and understand their particular indexing and search idiosyncracies. Most of these online catalogs are not yet effective, usable "self-service" information retrieval systems for a wide variety of their users. These conclusions are based on a number of factors: personal experience with the use of dozens of online catalogs, numerous discussions with librarians who have monitored the introduction and use of online catalogs in their libraries, discussions with system designers with expertise in human factors engineering, and review of the findings of research studies on the use and users of online catalogs.

The potential of the online catalog to provide improved access to library materials and the information they contain is still largely untapped. Eventually, the forces of innovation and market competitiveness will boost online catalog development off the second-generation plateau. However, we should not expect a giant, discontinuous leap forward to the next generation of online catalogs. Rather, progress is likely to be made in small, incremental steps. Some of the new developments will almost certainly be technology driven. Combinations of new hardware, especially more intelligent workstations, and software techniques will be applied to new and improved library catalogs and retrieval systems. We will see more "WIMPs" (Windows, Icons, Menus, and Pointers) at the user interface. Already, the CD-ROM-based online catalog is being touted as yet another new form of the catalog. The danger is that future design and development efforts will neither be "user driven," nor incorporate the knowledge learned from information retrieval research and experimentation to improve conventional Boolean retrieval systems.[7]

Online catalog research studies have uncovered a number of common problems experienced by users of second-generation online catalogs. Solutions to these problems should constitute the design agenda for improved online catalogs. In general terms, the major problems include:

—too many failed searches (search attempts that are aborted, or that result in no matches—"0-hits"—or too many hits);[8]

—navigational confusion and frustration for the user during the search process (Where am I? What can I do now? How can I start over?);[9]
—unfamiliarity with or ignorance of the subject indexing vocabulary leading to the failure to match search terms with the system's subject vocabulary;[10]
—misunderstanding and confusion about the fundamentally different approaches to retrieval and search methods employed in today's online catalogs (e.g., precoordinate phrase searching and browsing, and postcoordinate keyword/Boolean searching);[11] and
—partially implemented search strategies and missed opportunities to retrieve relevant materials (e.g., searches in which large retrieval sets are not scanned or narrowed in size, and title keyword searches that are not followed by searches on the call numbers or subject headings of the found records).

Chan points out that online searching is a process of extracting a subfile of useful documents from a large file, a process where "in most cases, a sequence of search statements is required for even minimally satisfactory retrieval."[12] To optimize retrieval results in subject searching, more than one search approach may have to be employed in the overall search strategy: "Through combination, keywords and the [controlled] vocabularies of DDC, LCC, and LCSH should offer far greater possibilities in search strategies than any one of them can provide alone."[13] Markey has demonstrated, for example, that different records on a particular subject would be retrieved by using a classified approach from those retrieved using keyword or alphabetical subject heading browsing approaches.[14]

Conventional IR systems place the burden on the user to reformulate and reenter searches until satisfactory results are obtained. This is typically the case with second-generation online catalogs as well. This approach assumes, however, that the user knows what he wants and can describe it in the language of the catalog database being searched. Hjerppe quite correctly rephrases this problem as the *fundamental paradox of information retrieval*: "the need to describe that which you do not know in order to find it."[15] Even the best second-generation catalogs do little to help the user transform an information need to explicit descriptions of the information understandable by the system. Nor do these catalogs lead the user from "found" information to related, linked information that has not yet been discovered. It is unrealistic to expect our catalog users to know in advance the structure and language of our library databases. It is equally unrealistic to expect online catalog

users to be proficient in the various search approaches and techniques before they engage an interactive system in the retrieval process. Hjerppe reminds us that humans are much more adept at recognizing something than generating a description of it.[16] Online catalogs could take advantage of this human facility by permitting requests such as Give me more like this!

In summary, second-generation online catalogs fall short in that they:

—do not facilitate open-ended, exploratory searching, by following preestablished trails and linkages between records in the database in order to retrieve materials related to those already found;
—do not automatically assist the user with alternative formulations of the search statement or execute alternative search methods when the initial approach fails;
—do not lead the searcher from successful free-text search terms (e.g., title words) to the corresponding subject headings or class numbers assigned to a broader range of related materials;
—do not provide sufficient information in the retrieved bibliographic records (such as tables of contents, abstracts, and book reviews) to enable the user to judge the usefulness of the documents; and
—do not rank the citations in large retrieval sets in decreasing order of probable relevance or "closeness" to the user's search criteria.

Common Sense Enhancements

To move beyond second-generation online catalogs, we do not have to wait for the arrival of "fifth-generation" computers, or the "trickle-down" benefits of artificial intelligence (AI) research. Online catalogs can be made more intelligent, responsive, and usable employing already proven software methods. A measure of common sense, not AI, needs to be applied in our design efforts. The primary "common sense" assumption may be stated: All types of catalog users can benefit from additional interactive assistance and guidance, not only with the mechanics of searching and query formulation, but also with the selection and use of appropriate search strategies which may retrieve all materials of possible interest, or which may refine the search to produce precisely what the user is looking for.[17]

With the aforementioned problems and shortcomings of second-generation online catalogs in mind, we can focus on a short list of attainable, commonsense enhancements. Some of these enhancements have already been incorporated in a few advanced online catalogs;

others are undergoing testing in experimental public access retrieval systems. The list begins with recommendations that are relatively easy to implement, followed by several that require more sophisticated software methods and database techniques.

Enhancements to the User-System Dialogue

Online catalogs have the potential to communicate interactively with the user as the search progresses. Even the simplest, most constrained search dialogues require transition through three to four different screen displays. As search and display options increase in more powerful systems, the network of possible screen displays and sequences of displays expands considerably. The searcher may not be familiar with specific search sequences or the overall network of displays available and may not be proficient in the mechanics (the "how to") of transversing the network, giving appropriate requests and commands to the system. New and occasional users of today's online catalogs often express a sense of disorientation, of being lost, not knowing what to do next, and thus they often underutilize the capabilities of the system. Some walk away in anger and frustration.

A *usable* online catalog will display, along with data retrieved from the database (e.g., alphabetical browsing lists of headings or keywords, ordered lists of citations, full bibliographic records, etc.), information informing the user of the status of the search in progress. This information should include the query as the system has processed it, basic navigational prompts (e.g., "BACK," "FORWARD," "START OVER," etc.), and instructions for additional, required retrieval actions or optional search methods (where available). Frequent, experienced users may wish to "turn off" this structured on-screen guidance mode. This should be allowed. As we do with the many highway, road, and traffic signs in our daily transportation environment, the experienced user will probably just ignore the status messages and prompts until the need for their assistance arrives. The goal is to make the online catalog comfortable and effective for occasional users who are not trained search specialists.

Markey has identified three major difficulties encountered by online catalog subject searchers:

—discovering the most appropriate subject heading to use in a search statement,
—increasing the results when no or too few records are retrieved, and
—reducing the results when a large number of records is retrieved.[18]

Following a detailed analysis of each problem area, Markey presents an insightful list of suggested improvements to online catalogs, improvements designed to assist searchers overcome one or more of the difficulties. Short of incorporating automatic search routines which go into effect to reprocess the query when certain predetermined criteria of "failed" searches are satisfied, online catalogs can include a message-response system that tells the user to try available search and display options that offer ways out of the current difficulty. Any such message should tell the user what to do, how to do it, and why it may improve the results. In the case of no or few retrievals, the online suggestions may include shortening the search phrase or word, substituting synonyms or more general terms for the initial search words, or retrying the search using a different search method which may produce broader results. When an excessive number of records are retrieved (more than fifty or more than one hundred?), users seldom scan through the long lists of citations. Online suggestions addressing this problem could include asking the user to enter additional search words (with the system executing an implicit Boolean "AND" operation), or recommending the entry of limiting criteria to narrow the search results (e.g., date of publication, language, precise data field specification, etc.).

Automatic Correction of Search Term Spelling and Format Errors

Search failures (especially no matches) commonly result from misspellings of words and names. Several spelling correction or word approximation software routines are available to help with this problem. With systems having limited processing capacity (this is relative of course to the demands placed on the system at any time), routines that attempt to correct spelling or to find words that are orthographically similar to the entered word could be invoked only after a "no match" has resulted. While extending this additional effort, the system could inform the user what it is trying to accomplish.

Arlene Taylor has discovered that a large percentage of errors made in entering name searches results from the user not knowing the system's rigid requirements for the form and order of the elements in the entry.[19] Typically, online catalogs require that personal names be entered last name first, followed by the first name and middle initial, if known. Frequently, users enter personal names in their natural, uninverted order. Flexible system software can easily handle this problem. The software could invert the word order where required, or conduct a keyword, Boolean "AND" search on the name's components, ignoring the order in which they were entered. Then, if no matches result, the

system could reprocess the search using n-1 components of the name, etc. (such as the last name only, if this fact is discernible).

Automatic Search Aids

Most online catalogs assume that the display of a full bibliographic record represents a closure point in the search process. Online catalogs are generally inactive and silent at this point; one online catalog even displays the cryptic message, "The END." These assumptions are, of course, false. The bibliographic record contains data elements that link it with associated and possibly relevant records (authors, call number, series title, and subject headings). These linkages are not exploited in second-generation online catalogs. A single, relevant bibliographic record may serve as the "jumping-off" point for browsing a selected portion of the shelves, or it may be the point of departure for finding related materials. The user may wish to say, after viewing a bibliographic record, Great, I want more like this. The system could then use any of several methods to retrieve related records, including gathering all records in the database assigned one or more of the same subject headings or the same class number as the initial record. With a bit more design sophistication, the system could ask the user which data element in the displayed record (e.g., personal name, series title, specific subject heading, etc.) should be used as a gateway to related records.

The traditional distinction between "known-item" and "subject" searches is useful for designing search dialogues. In practice, the distinction blurs as one type of search may lead naturally into another type. A search for a specific work often becomes a search for materials related in some way to the work first sought. Transaction logs have shown that users of online catalogs frequently change from one type of search to another type during the same search session.[20] When conducting a subject search, the user might discover a series of interest (and wish to see immediately all the titles in the series), or learn of an author who has been listed as an "added entry" (a useless concept in the online environment), then ask to see all of this author's works in the collection. The bibliographic record as *displayed* can serve many related retrieval purposes. It can also be a source of relevance feedback from the user to the system. Additional dialogue and automatic search routines can assist the searcher in tracking down related materials without requiring the user to continually reformulate precise, well-structured queries until satisfactory results are obtained.

Online catalog users display no desire to search in the disciplined, highly-structured, linear manner of trained search intermediaries who

aim to produce a well-defined list of citations for an end user. Miller and Tegler summarize the view held by many researchers that the scholarly process of seeking and identifying information to assist with a problem typically follows a more circuitous, cyclical, and unstructured path of browsing and discovery.[21] This author suspects that this is also true of information seeking by the general public. We often do not know exactly what we want when we begin looking for it. The fuzzy model of information seeking activities implicit in this view should be incorporated into future online catalogs designed for scholars and general users. Efficient "known-item" and "known-subject" search methods should be retained as options for those searchers who know exactly what they are looking for.

Second-generation online catalogs offer both precoordinate phrase searching (by name, title, or headings as in the card catalog) and postcoordinate keyword searching using Boolean operators (and, in some cases, truncation, range, and proximity operators). The limitations and advantages of each approach in various types of searching have become well-known to those familiar with the online search literature. Searches on precoordinated subject headings, for example, can improve recall (the number of relevant records retrieved from a specific database) and in some instances, improve precision. But this places the burden on the user to enter (or to be guided to) the correct subject terminology. On the other hand, keyword subject searching on component words in titles or subject headings is a powerful, perhaps more natural, search method. However, this method frequently produces very large retrieval sets which include many nonrelevant records ("false drops"). When the exact title is known, a title-phrase search (matching words in exact order) is likely to result in a precise retrieval. A title-keyword search (matching words in any position and order) would have retrieved additional, nonrelevant records.

Chan recommends that a combination of these (and other) search methods be used in subject retrieval attempts, because in combination they "offer far greater possibilities in search strategies than any one of them can provide alone."[22] The assumption here is that different sets of relevant records will be retrieved from an online catalog when different search methods are employed. Experiments by Chan[23] and Markey[24] appear to support this assumption. A combination of these methods would seem to increase recall. Unfortunately, searching in this manner requires an expertise that users of online catalogs are not likely to possess.

Mention has been made of online catalogs that judiciously *suggest* either alternative search formulations or alternative search strategies for

the user to try when certain "failure" conditions arise. Another design approach, borrowing heavily from IR research and experimentation, involves programming the system to execute automatically a search in a variety of ways until satisfactory results are obtained. Usually some interaction with the user is required for eliciting judgments of relevancy and satisfaction. The selection of search formulations and search methods is carried out automatically by the system software. The user needs to know, or recognize, what he wants, but does not need to know how to search the database in an "expert" manner. For example, automatic stemming routines may be applied to search words to broaden a search. Truncating a search word to a common root may retrieve records that match any form of the search word—i.e., plural, singular, etc.

Based on a view of online searching as a multilevel, trial and error process of seeking, relevance judgment, selection, and discovery, some online catalog designers have incorporated a set of search sequencing rules which, using the results of the search at given stages, determine the course of the search. The "rules" determine which search method will be executed next based on one or more measures of success or failure including feedback from the user (other measures include similarity of words in the query to index terms, their frequency in the database, etc.). The LCS/WLN-based online catalog at the University of Illinois at Urbana-Champaign takes a no-match subject heading search and automatically reprocesses it as a title keyword search, assuming implicit Boolean ANDs between the search words. The retrieved citations are displayed a few at a time for relevance assessment by the user. If a citation is selected for "full" display, the user is guided to the subject headings assigned to the document and is encouraged to continue the search using the subject headings to retrieve related materials.[25]

The British experimental online catalog, OKAPI (Online Keyword Access to Public Information), uses a built-in "search decision tree" approach to establish a conditional search sequence for various types of searches the user may enter (author, subject, title, etc.).[26] The search path followed once a search begins is determined by the preestablished rule system and conditions encountered by the system in interaction with the user (e.g., user's choice of search type, user's actual query, results from a previous retrieval, the user's feedback on those results, etc.). For example, the user may enter a title phrase; the system would execute an exact phrase-match search and display any matching citations. If no exact match occurred, the system would then automatically execute a weighted, combinatorial search retrieving records which had any of the search words in their titles and displaying the citations in a ranked order with titles having the greater number of words listed first.

Other optional search sequence rules that could be used following a phrase-match search include executing a word search with Boolean ANDs, or automatically stemming some words to improve the chances of retrieving something relevant.

These online catalogs represent attempts to make searching easier and more effective for untrained users by automating some of the "intelligent" judgments and activities of experienced search intermediaries. These activities include relevance feedback, stemming or truncation, finding synonyms, applying Boolean operators, and ranking some index terms as more important than others. CITE, the online catalog in use at the National Library of Medicine, incorporates a number of these automatic functions, including term weighting, combinatorial searching, and ranked display output.[27] CITE's rich dialogue also suggests limiting measures that might be applied to a search in progress and always asks if the user would like to see items related to any displayed citation. Subject headings, call numbers, and free-text terms weighted heavily as important terms (based on an inverse document frequency measure) are used by the system at various stages of the search as "new" search words to retrieve related and potentially relevant citations.

Generally speaking, online catalogs can be enhanced in ease of use and retrieval effectiveness in three ways:

1. improving the user-system interaction with richer dialogue, online assistance, and online guidance;
2. enriching the catalog records and improving the structure of the bibliographic database; and
3. adding reference data files which supplement the catalog file.

Enriching the Subject Vocabulary of Catalog Records

Before concluding this section on commonsense enhancements, mention must be made of attempts to enrich and augment the bibliographic records in our online catalogs. The need for and value of enriching our bibliographic records with data extracted from tables of contents and back-of-the-book indexes has been established for some time.[28] When indexed judiciously for online searching, the current, more specific terminology obtained from contents pages and book indexes can lead to vastly improved retrieval effectiveness and user satisfaction. The display of tables of contents after retrieval enables the user to make more meaningful judgments about the potential usefulness of documents in the collection.

Until very recently, adding later to the bibliographic record obtained from contents pages and back-of-the-book indexes has required a fair amount of skilled, manual labor. Editing must be applied to each publication to ensure only meaningful and informative words and phrases are selected to be added to the bibliographic record created for each work. Additional time is needed to manually key in the additional terminology. The need to keep down cataloging costs has been offered to explain why enriching the record has not been adopted in our national cataloging activities. Also, some librarians have expressed fears about the inconsistency that could result from the selection of words and phrases for these additional index terms. The cost issue is a serious one; how well the job is done is less serious. Any enhancement which adds indexable and displayable subject vocabulary to our sparse bibliographic records is better than none.

The online catalog, EIS (Engineering Information System), at Purdue University's Seigesmund Engineering Library has been enhanced through the addition of data from the edited tables of contents for most of the monographs in the library's collection.[29] This augmented monograph file is searchable by keywords, and the use of Boolean operators is supported. As new books arrive, library staff manually scan, edit, and subsequently input the tables of contents into the file. This file is very current, being updated (reloaded) weekly in the online catalog. The additional labor costs incurred by the library are believed to be more than justified by the increased search benefits to users. Both staff and patrons feel that subject searching of the monograph collection has been greatly enhanced. The number of search terms indexed per book has increased, thereby increasing the possibility of finding specific information in a particular book. Users of the Purdue online catalog are very happy with the expanded catalog because "the records augmented with tables of contents are searchable by terms in current use in engineering, obviating the necessity of mastering the intricacies of the LC (Library of Congress) subject classification."[30] Furthermore, when a bibliographic record is retrieved from the catalog, the user has the option of displaying its associated table of contents, permitting a more meaningful assessment of the potential utility of the book.

A related experimental project being conducted by the Bibliographic Services Division of the British Library bears close scrutiny, because it addresses the problem of reducing the time and costs associated with adding tables of contents data to MARC catalog records. The project aims to create a test file of United Kingdom MARC records augmented by words and phrases from tables of contents. Expanded

records for both monographs and conference proceedings may be included. The test file would be mounted on BLAISE-LINE (the British Library's online retrieval system), and perhaps other retrieval systems, for controlled evaluations by library staff and patrons. This project is unique because it employs a prototype digital page scanning system developed by OPTIRAM to automatically "read" selected tables of contents, editing and formatting the data according to criteria built into the scanning software. The process produces a machine-readable file of tables of contents for each publication that can be merged with the matching MARC records. The machine must be programmed to read a wide variety of printed tables of contents pages because no standard for layout, format, enumeration, and syntax, etc., is currently followed by publishers. If it succeeds, the project will significantly reduce the costs associated with the manual production of such augmented catalog records. Software may also be developed that will instruct the machine, after a scan of the title page, to identify the already-created catalog record stored in the database and to ascertain its unique control number. This would permit the entire process of scanning, editing, formatting, merging (contents and MARC records), and updating the catalog file to be accomplished automatically with very little manual effort. The machine scanning technique is proven and reliable. The challenge lies in developing the software to complete the editing and merging process.[31]

Integrating Periodical Indexes in the Online Catalog

Library catalogs do not index the periodical literature held by the library. Library patrons wishing search access to the articles and reports contained in periodicals have had to use a variety of separately published abstract and index sources in print, microform, or online media. These gateways to the periodical literature have not been integrated with the library catalog in any of its forms in this century. Furthermore, the periodical publications these "global" indexes cover do not represent the actual periodical holdings of any particular library. Searching them successfully in print, microform, or online brings no assurance that the library holds copies of the relevant documents. Although the twentieth-century library catalog has had a monographic orientation, nothing seems more natural than providing access to a library's periodical literature through its online catalog. A user entering an author search in an online catalog may understandably wish to retrieve all the author's publications held by the library: periodical articles, technical reports, papers in proceedings, as well as monographs. A searcher with a

subject information need probably does not care how the information that satisfies that need is packaged. A keyword search on titles or subject headings may represent a desire to retrieve documents of interest, regardless of the form in which they are published: book, magazine, newspaper, technical report, or scholarly periodical.

Adding "analytics" to the catalog record is costly and only goes so far in widening access to the library's periodical literature. The "fence resistant" online catalog provides us with an opportunity to break through barriers to the periodical literature inherent in earlier forms of the catalog. The necessary retrieval functions are already present in most second-generation online catalogs. All that is needed is to load from a global online abstract and index (A&I) database a subset of the bibliographic records that matches the issues of periodical titles actually held by the library. These A&I records would supplement the MARC catalog records in the bibliographic database. Indexes to the A&I records could be integrated with the indexes to the catalog records (for monographs, serials, etc.), or be maintained separately. In the latter case, the user could be offered a choice of bibliographic files to search: the "catalog" or the periodical indexes. A better approach would be to process an author, title, or subject search without regard initially to the form of publication which is indexed and which may result in a match.

Unlike library catalogers, most A&I database producers do not use the Library of Congress Subject Headings as their subject indexing vocabulary. But this does not present a very large problem for online catalog users. Researchers have shown that most (60 to 70 percent) subject searches in online catalogs do not use Library of Congress Subject Headings. Most catalog users do not seem to know what they are or what role they play in the catalog. They assume that the natural language terminology they use in their subject queries is also used in the catalog. Various kinds of keyword access are provided in today's online catalogs. Keyword queries on the A&I records in the expanded catalog database could be executed in the same manner as queries of the monograph file. Online catalog searchers could be guided to the no less familiar controlled subject vocabularies of the A&I file in the same ways they are now guided to Library of Congress Subject Headings, that is, via alphabetical displays of headings or descriptors (including broader and narrower terms from the thesaurus), or by special labeling of such subject descriptors when they appear in the displayed citation.

There are some practical problems associated with adding A&I citations to online catalog databases. Many different publishers produce online A&I files, and each file typically covers a specialized subject area. Many different A&I files would have to be acquired from various sources

to cover the literature in most libraries' collections and to support the wide variety of subject searches conducted in online library catalogs. But, once again, any improvement is better than none. Online messages can and should be used to tell the searcher what is included in the library's online catalog. Lack of consistency in headings such as names of persons and organizations across files may cause difficulties, but loading and retrieval software can do a great deal to resolve this problem. Related headings can be "normalized" or linked together during database loading, and software retrieval techniques can be used to bring potentially related items to the user for assessment and selection.

General coverage A&I databases that use Library of Congress Subject Headings present a natural target for acquisition and loading into the online catalog. One such database is the "Magazine Index" published by the Information Access Corporation (IAC). "Magazine Index" covers several hundred popular periodicals held by most libraries. IAC also publishes online indexes of the business and legal periodical literature. IAC is mentioned here because they have entered into an agreement with the Division of Library Automation at the University of California to load and index selected portions of their A&I periodical databases into the MELVYL online catalog. MELVYL serves as the public access catalog at the nine University of California campuses. During an upcoming test and evaluation period, users of MELVYL (which is primarily a keyword/Boolean search system) will have access at one terminal to indexed magazine and periodical articles held by the University of California libraries as well as to the monographs in their collections. This represents a giant step forward. Now that H.W. Wilson, Inc. has put their indexes online, perhaps both tables of contents and book reviews will be added to our online catalogs in the near future.

Conclusion

Reflecting on the vast potential of the online catalog, Malinconico writes:

> There is little doubt that we are standing on the threshold of changes that will alter the catalog and library service in ways that we can only dimly perceive. The library catalog will very likely change into something that bears little resemblance to the instrument we currently know.[32]

With a bow to tradition and the conventional wisdom, Malinconico goes on to claim that the catalog in its online form will "remain the principal means by which readers help themselves to use the resources of the library."[33] It is doubtful that this claim holds true for the present in

many libraries, and its truth is not guaranteed for the future. Today's online catalogs are not likely to serve and satisfy tomorrow's library patrons and other seekers of information.

Many suggestions for improving online catalogs, along with the motivation and rationale for each, have been discussed in preceding sections of this essay. The essay largely represents a compilation and reformulation of the ideas and efforts of many researchers and innovative system designers. It is hoped that this discussion, along with the others in this issue of *Library Trends*, will help librarians learn more about the problems of online catalog use and the promise the online public access catalog holds for vastly improved access to our libraries' collections.

A summary comment is offered to highlight the general aim of the many suggestions and recommendations that have been discussed. Enhancements to online catalogs should be guided by a principle which states: An online public access catalog should work intelligently with the user, engaging in meaningful dialogue, to elicit expressions of the user's information need (which may change during the course of the search), and to improve the results of the user's search activity.

Some corollaries of this principle can be stated as pleas to those responsible for the design and development of improved online catalogs:

1. Never assume the user can effectively navigate across an evermore complex database, presented with more and more sophisticated retrieval options, without generous assistance and guidance from the online system.
2. Never permit a search to fail and do nothing. The system should assume one or more records in the catalog will satisfy the user's need(s) and exhaust all approaches to finding those records until instructed by the user to stop.
3. Never assume the display of a bibliographic record represents the end of the user's search. Use the bibliographic record as a point of departure for related-item searching and browsing.
4. Never assume the user knows the "official," controlled vocabulary of the database, or understands the generative relationship between uncontrolled, free-text terms in a citation and the subject descriptors or classification numbers specially assigned to the document.
5. Never assume more useful information cannot be added to the online catalog. Patron access must be given priority over cataloger's control of the database. Especially never assume that the current MARC record contains enough displayable information to indicate the relevance and utility of a document to the user.

References

1. Hildreth, Charles R. "Pursuing the Ideal: Generations of Online Catalogs." In *Online Catalogs, Online Reference: Converging Trends* (Proceedings of a Library and Information Technology Association Preconference Institute, 23-24 June 1983, Los Angeles, Calif.), edited by Brian Aveney and Brett Butler, pp. 31-56. Chicago: ALA, 1984.

2. Ibid., p. 33.

3. Dowlin, Kenneth E. "The 'Integrated Library System.'" *The Electronic Library* 3(Dec. 1985):340-45.

4. Markuson, Barbara E. "Issues in National Library Network Development: An Overview." In *Key Issues in the Networking Field Today* (Proceedings of the Library of Congress Network Advisory Committee Meeting, 6-8 May 1985, Washington, D.C., Network Planning Paper, Number 12), pp. 9-32. Washington, D.C.: Library of Congress, 1985.

5. Ibid., p. 11.

6. Malinconico, S. Michael. "Circulation Control Systems as Online Catalogs." *Library Journal* 108(15 June 1983):1207-10.

7. Mitev, Nathalie Nadia, and Walker, Stephen. "Information Retrieval Aids in an Online Public Access Catalog: Automatic Search Sequencing." In *Advances in Intelligent Information Retrieval: INFORMATICS 8* (Proceedings of a conference organized by the Aslib Informatics Group and the BCS Information Retrieval Specialist Group, 16-18 April 1985, Wadham College, Oxford); Harper, D.J. "Relevance Feedback in Document Retrieval." Ph.D. diss., Cambridge University, 1980; and Oddy, R.N. "Information Retrieval Through Man-Machine Dialogue." *The Journal of Documentation* 33(March 1977):1-14.

8. Markey, Karen. "Users and the Online Catalog: Subject Access Problems." In *The Impact of Online Catalogs*, edited by Joseph R. Matthews, pp. 35-69. New York: Neal-Schuman, 1986; and Markey, Karen. *Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs*. Dublin, Ohio: OCLC, 1984.

9. Knipe, Nancy. "Hands-On: User-Directed System Evaluation." (A paper presented at the *Conference on Integrated Online Library Systems*, 23-24 Sept. 1986, St. Louis). To be published in the proceedings of the conference, edited by David C. Genaway.

10. Markey, "Users and the Online Catalog."

11. Kranich, Nancy C., et al. "Evaluating the Online Catalog from a Public Services Perspective: A Case Study at the New York University Libraries." In *The Impact of Online Catalogs*, edited by Joseph R. Matthews, pp. 89-140. New York: Neal-Schuman, 1986.

12. Chan, Lois Mai. "Library of Congress Classification as an Online Retrieval Tool: Potentials and Limitations." *Information Technology & Libraries* 5(Sept. 1986):181-92.

13. Ibid., p. 188.

14. Markey, Karen. *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, Report No. OCLC/OPR/RR-86/1, 1986.

15. Hjerppe, Roland. "Project HYPERCATalog: Visions and Preliminary Conceptions of an Extended and Enhanced Catalog" (Paper presented at IRFIS 6 Conference [International Research Forum in Information Science] in Frascati, Italy, 15-18 Sept. 1985). Available from Department of Computer and Information Science, Linköping University, Linköping, Sweden, p. 14. See also: Noerr, Peter L., and Bivins Noerr, Kathleen T. "Browse and Navigate: An Advance in Database Access Methods." *Information Processing & Management* 21(No. 3, 1985):205-13.

16. Ibid.

17. Excellent definitions of search mechanics, search formulation, and search strategy are provided by Marcia J. Bates, "Search Techniques." *In Annual Review of Information Science and Technology*, vol. 16, edited by Martha E. Williams. White Plains, N.Y.: Knowledge Industry Publications, 1981, pp. 139-69.

18. Markey, "Users and the Online Catalog," p. 35.

19. Taylor, Arlene. "Authority Files in Online Catalogs: An Investigation of Their Value." *Cataloging & Classification Quarterly* 4(Spring 1984):1-17.

20. Hildreth, Charles R. "Monitoring and Analyzing Online Catalog User Activity." *LS/2000 Communique* (an OCLC Local Systems publication) 4(Summer 1985):3-6.

21. Miller, Connie, and Tegler, Patricia. "Online Searching and the Research Process." *College & Research Libraries* 47(July 1986):370-73.

22. Chan, "Library of Congress Classification," p. 188.

23. Ibid., p. 189.

24. Markey, *Dewey Decimal Classification Online Project.*

25. Mischo, William, to whom I am indebted for many conversations and for sending me several samples of screen displays. Engineering Library, University of Illinois, Urbana-Champaign, Illinois, May 1986.

26. Mitev, Nathalie Nadia, et al. *Designing an Online Public Access Catalogue: OKAPI, a Catalogue on a Local Area Network* (Library and Information Science Report No. 39). London, England: British Library Publications Sales Unit, Boston Spa, Wetherby, West Yorkshire LS23 7BQ.

27. Doszkocs, Tamas E. "CITE NLM: Natural Language Searching in an Online Catalog." *Information Technology & Libraries* 2(Dec. 1983):364-80.

28. Atherton, Pauline A. *Books are for Use: Final Report of the Subject Access Project to the Council on Library Resources.* Syracuse, N.Y.: Syracuse University, School of Information Studies, 1978; and Cochrane, Pauline A. "Augmenting Subject Descriptions for Books in Online Catalogs." *Database* 5(Dec. 1982):29-37.

29. Posey, Edwin D., and Erdman, Charlotte A. "An Online UNIX-Based Engineering Library Catalog: Purdue University Engineering Library." *Science & Technology Libraries* 6(Summer 1986):31-43.

30. Ibid., p. 34.

31. Hoey, Michael, Head-Planning and Research Group, Bibliographic Services Division, The British Library, personal conversations with Hoey and members of the Research Group, May 1986.

32. Malinconico, S. Michael. "Catalogs & Cataloging: Innocent Pleasures and Enduring Controversies." *Library Journal* 109(15 June 1984):1213.

33. Ibid.

Boolean reasoning builds on the Boole-Schroder algebra of logic, which is based on Boolean equations, rather than on the predicate calculus. Al-though Boolean equations are predicates-statements that are either true or false for any values of their arguments-almost none of the apparatus of predicate logic is employed in Boolean reasoning. Neither the disjunction of two Boolean equations nor the negation of a Boolean equation is a Boolean equation; thus neither of these operations is generally allowed in Boolean reasoning (see Rudeanu [172, Chapt. 10] for results concerning disjunc-tion and neg