Title
Why Are Online Catalogs Still Hard to Use?

Journal
Journal of the American Society for Information Science, 47(7)

ISSN
0002-8231

Author
Borgman, Christine L.

Publication Date
1996-07-01

Peer reviewed

Why Are Online Catalogs *Still* Hard to Use?

Christine L. Borgman
Department of Library and Information Science, 2320A Moore Hall, Box 951521, University of California, Los Angeles, Los Angeles, CA 90095-1521. E-mail: cborgman@ucla.edu

We return to arguments made 10 years ago (Borgman, 1986a) that online catalogs are difficult to use because their design does not incorporate sufficient understanding of searching behavior. The earlier article examined studies of information retrieval system searching for their implications for online catalog design; this article examines the implications of card catalog design for online catalogs. With this analysis, we hope to contribute to a better understanding of user behavior and to lay to rest the card catalog design model for online catalogs. We discuss the problems with query matching systems, which were designed for skilled search intermediaries rather than end-users, and the knowledge and skills they require in the information-seeking process, illustrated with examples of searching card and online catalogs. Searching requires conceptual knowledge of the information retrieval process—translating an information need into a searchable query; semantic knowledge of how to implement a query in a given system—the how and when to use system features; and technical skills in executing the query—basic computing skills and the syntax of entering queries as specific search statements. In the short term, we can help make online catalogs easier to use through improved training and documentation that is based on information-seeking behavior, with the caveat that good training is not a substitute for good system design. Our long term goal should be to design intuitive systems that require a minimum of instruction. Given the complexity of the information retrieval problem and the limited capabilities of today's systems, we are far from achieving that goal. If libraries are to provide primary information services for the networked world, they need to put research results on the information-seeking process into practice in designing the next generation of online public access information retrieval systems.

Introduction

This Special Issue of the *Journal of the American Society for Information Science* on online catalog research is an appropriate time to revisit the questions asked in an article that appeared a decade ago: "Why are online catalogs hard to use? Lessons learned from information-retrieval studies" (Borgman, 1986a). We called for substantial changes in online catalog design based on our increasing knowledge of user behavior. Despite the attention that and other articles on online catalogs received (Efthimiadis, 1990), little seems to have changed. Subsequent research studies continue to report that users have great difficulty searching online catalogs. Thus we ask the question, "Why are online catalogs *still* hard to use?"

We argue that online catalogs continue to be difficult to use because their design does not incorporate sufficient understanding of searching behavior. Research in information seeking indicates that users formulate questions in stages, gradually coming to the point where they can begin to articulate a query. Even then the search process may be iterative and searching may serve to refine the question rather than to build a set of documents that matches an explicit query. A "search" may be conducted over a number of sessions with different information technologies and sources, both online and offline, picking and choosing from multiple options to answer a question or explore an issue. Yet the design of most operational online catalogs assumes that users formulate a query that represents a fixed goal for the search and that each search session is independent.

The first generation of online catalogs followed either of two query-oriented design models: Online "card" catalog models, emulating the familiar card catalog, or Boolean searching models, emulating information retrieval systems such as DIALOG or Medline. Second-generation online catalogs merged these two design models and improved access points, search capabilities, and display options (Hildreth, 1987, 1993). Most online catalogs currently in use provide second-generation functionality.

The record structure, content, and primary searchable fields are drawn from card catalog design models, while the searching functions and many of the interface design characteristics are drawn from retrieval system models. While user input is simpler and screen displays are much clearer and more attractive, the basic functionality of online catalogs has changed little since the late 1980s. We limit the scope of our discussion to the automated catalog

per se and do not address the full range of information systems and services in which it now may be embedded.

In the 1986 article (Borgman, 1986a), we examined studies of information retrieval system searching for their implications for online catalog design. In this article, we examine the implications of card catalog design for online catalogs. Although comparing card and online catalogs may seem like revisiting old territory, we find this analysis to be of current importance for several reasons. The library community is considering radical changes in cataloging codes to address the needs of networked, distributed computing environments and the description of materials in a vast array of new media. Today's online catalogs hold many millions of records that were constructed for card catalogs and retain the same underlying structure. Changes to cataloging rules and relevant standards must support the migration of old data to new formats, for libraries have created far too much data already to start over. Online catalogs are an established technology in major research libraries in developed Western countries, but in much of the world, card catalogs are current technology for all types of libraries, and they continue to be maintained in many smaller libraries in developed countries. Libraries in Central and Eastern Europe and in many other parts of the world are making the transition directly from card catalogs to the most modern online catalogs, having skipped the interim generations of technology (e.g., Borgman, 1995), and are faced with an array of research, training, and assistance issues in introducing these systems. Online catalog technology has not kept pace with expectations for ease of use or functionality, resulting in calls for a return to the trusted card catalog (Baker, 1994). We claim that many of the problems that remain in online catalogs are due to the remnants of the card catalog in its structure and a failure to design user interfaces based on the knowledge and skills of online catalog users. We hope that the arguments presented here will contribute to a better understanding of user behavior and lay to rest the card catalog design model for online catalogs once and for all.

We first discuss card and online catalogs as query design models, then we analyze the process of formulating and executing queries in each of these types of systems. We argue that most current online catalogs are based on card catalog design models, that this model does not map well onto online systems, and that the model is not based on information-seeking behavior. We conclude with proposals for new design models for online catalogs.

Query Design Models

Information retrieval is a difficult problem because it requires describing information that you do not yet have. Searchers must translate their information needs into a description of the information sought, relying on their own knowledge of the problem, their understanding of the tools the system provides to assist in describing the problem, and if available, the services of skilled reference librarians or search intermediaries—the original "intelligent agents." At one extreme, information retrieval system interfaces require searchers to specify the search completely in a single set of statements, or a "query"; such systems return a set of records that match the query as a final results set. At the other extreme, information retrieval system interfaces allow searchers to enter whatever fragments of the idea are available as a starting point, providing various tools to assist in exploring the information need; a query may never be stated explicitly nor a final results set retrieved. Most online catalogs are based on query design models that allow some degree of search modification but are far from being exploratory systems.

Query systems were designed for highly skilled searchers, usually librarians, who used them frequently, not for novices or for end-users doing their own searching. Query matching is effective only when the search is specific, the searcher knows precisely what he or she wants, and the request can be expressed adequately in the language of the system (e.g., author, title, subject headings, descriptors, dates). Even with graphical user interfaces, the searcher must enter terms and specify relationships that match those in the database. Many online catalogs allow users to browse authority files and indexes, but usually only within the constraints of a specified query. Few systems allow searchers to retain fragments of prior search strategies and recombine them in other ways, to pursue non-linear links in the database, or to explore by other means.

Online catalogs must serve a population of information seekers that is heterogeneous in terms of age, language, culture, subject knowledge, and computing expertise, most of whom will be perpetual novices at information retrieval. To design systems that can support question answering rather than simply query matching, we need to learn more about the search process. One way to identify the features and functions to incorporate into the next generation of systems is to examine why systems based on query design models are hard to use.

The "Online Card Catalog" Design Model

Current online catalogs continue to be criticized as being more difficult to use and less serviceable than card catalogs. Baker (1994) popularized these criticisms in a widely-discussed article that appeared in *The New Yorker,* provoking support and disdain throughout the library community and well beyond it. Baker's argument is based on the use of catalogs to find known books rather than to seek information or to solve information-based problems. Indeed, he claims (p. 78) that "the function of a great library is to store obscure books." We take the perspective that the function of a great library is to assist people in finding the information they need, whether

held in that library's collection or elsewhere. The catalog, whether online or hard copy, is one of several classes of tools for finding information, and is the primary tool for finding information held in a library's collection.

Online catalogs grew out of card catalogs, automating records that were designed for a manual environment, with interfaces that were intended to bring a generation of library users familiar with card catalogs into the online world. Card catalogs remain largely structured as they were in the 19th century (Buckland, 1992). The card catalog design model is best explained by Cutter's "objects" of the catalog (Cutter, 1904, p. 12):

1. To enable a person to find a book of which either
   (a) the author
   (b) the title
   (c) the subject
   is known.
2. To show what the library has
   (d) by a given author
   (e) on a given subject
   (f) in a given kind of literature (poetry, drama, fiction).
3. To assist in the choice of a work
   (g) as to its edition (bibliographically)
   (h) as to its character (literary or topical).

Based on Cutter's objects, a catalog is arranged on the assumption that searchers arrive at the catalog knowing at least one of the three access points (author, title, or subject). However, studies of information-seeking behavior in both manual and automated environments show that people arrive at a catalog with incomplete information for any of the access points (Borgman & Siegfried, 1992; Chen & Dhar, 1990; Tagliacozzo, Kochen, & Rosenberg, 1970; Taylor, 1984). They must use information external to the catalog (e.g., bibliographies, lists of subject headings) to obtain sufficient data to express their search within the scope of Cutter's objects. Online catalogs based on a card catalog model provide more searching options and may be searched with less complete information about what is sought. However, Cutter's objects do not adequately represent the way people seek information. While Cutter certainly claimed to have the user in mind, his perspective and others of his time was based on a rational, positivistic approach, and not on direct study of how people formulate questions and seek information; rather, they made assumptions about the knowledge people brought to the information-seeking process.

Because card catalogs and first- and second-generation online catalogs utilize the same data and same principles as do their predecessors, the underlying record structure and access points remain the same. Accordingly, much of the search process is the same in card and online catalogs. Online catalogs add a layer of functionality, providing more techniques for searching the same data, but also add a layer of complexity to the process.

Conceptual Knowledge, Semantic Knowledge, and Technical Skills

In our earlier analysis of online catalog searching behavior (Borgman, 1986a), we claimed that users apply two types of knowledge to the information retrieval task (p. 388):

- Knowledge of the mechanical aspects of searching (syntax and semantics of entering search terms, structuring a search, and negotiating through the system), and
- Knowledge of the conceptual aspects (the "why" of searching—when to use which access point, ways to narrow and broaden search results, alternative search paths, distinguishing between no matches due to a search error and no matches because the item is not in the database, and so on).

Within this framework, a searcher usually can achieve some results from the system, once the mechanical aspects of searching are conquered. Only when the conceptual aspects are understood can the user exploit the system fully and effectively.

Upon reflection, and after 10 more years of studying online catalog searching behavior, we see three layers of knowledge required for online catalog searching:

- Conceptual knowledge of the information retrieval process—translating an information need into a searchable query;
- Semantic knowledge of how to implement a query in a given system—the how and when to use system features;
- Technical skills in executing the query—basic computing skills and the syntax of entering queries as specific search statements.

Our 1986 analysis merged conceptual and semantic knowledge into one category (conceptual knowledge) and regarded technical skills as mechanical knowledge. While a useful dichotomy, this approach failed to make some important distinctions between users' ability to understand the overall process of information seeking that is transferrable across systems and to understand how to conduct a search in a given system. It was not explicit in including the user's knowledge of the search topic in conceptual knowledge. Studies of searching behavior tend to focus on semantic knowledge and technical skills, and system evaluation often considers only technical skills, measuring users' success and errors in executing search statements. Instruction in the use of online catalogs too often focuses on technical skills alone, failing to set those skills in the context of semantic knowledge, much less conceptual knowledge.

We analyze the knowledge and skills required to search online catalogs in terms of these three layers, beginning with the most general. At each layer, we compare the requirements of card and online catalogs, discussing

the knowledge and skills that transfer from one to the other and those that do not.

Conceptual knowledge. Information retrieval is a difficult task, whether conducted in a card or online catalog. The process is not always a single act of formulating a query; rather, it often begins with some vaguely-felt need of wanting to know something and gradually evolves to the point where one can describe some attributes of the documents that might contain the information. Once the need can be phrased sufficiently to begin searching, the question itself may change through multiple iterations of finding and using information resources. Thus people usually approach an information retrieval system with a partially-formed query to be negotiated. The most complex part of the process often occurs offline, thinking through the problem to reach a point where it can be articulated.

When searching for information, a person is seeking *knowledge* or *meaning* (e.g., what? why? how?), but must formulate a query in terms of the *content* (e.g., words, numbers, symbols) of information *entities* (e.g., documents, objects). Information retrieval systems can deal directly with information only as "things" (Buckland, 1991), "containers" (Barlow, 1994), or *entities* that carry some information *content* that may impart *meaning* to its creators and users. In online catalogs, those entities are bibliographic records.

The success of a query is a function of the ability to translate the intended meaning into a set of search terms that are contained in the bibliographic records in the catalog and that convey the intended meaning. We have yet to design methods of organizing information (e.g., cataloging, classifying, indexing) that adequately bridge the gap between the way a question is asked and ways it might be answered, despite the application of sophisticated techniques to bring together terms with similar meanings and to distinguish among multiple meanings of a given term. Formulating author and title queries is difficult because the searcher must have accurate information that matches the entry in the catalog or know how to identify alternative forms that might exist. Subject searching is even more difficult because the searcher must find ways to articulate his or her intended meaning using terms that match those in the catalog, whether assigned by an author, indexer, or subject cataloger. The difficulties in query formulation, particularly with regard to subject, are well-studied, and a full analysis is beyond the scope of this article; we refer the reader to Bates (1986, 1989); Crawford, Thom, and Powles (1993); Efthimiadis (1992, 1993); Hildreth (1993); Lancaster, Connell, Bishop, and McCowan (1991); Larson (1991a, b,c); Markey (1986); Matthews, Lawrence, and Ferguson, 1983; O'Brien (1994); Taylor (1984); Walker (1988); and Walker and Hancock-Beaulieu (1991).

Searchers also need the conceptual knowledge of how search terms can be combined if they are to construct search statements with multiple concepts. Knowledge of Boolean logic is specific to automated systems and is not transferrable from card catalog usage. It is a difficult concept to grasp, as we have learned from studies of online catalogs and other online retrieval systems (Borgman, 1986a,b). Even scientists and engineers who have expertise in logic for other applications often use "and" and "or" backward in the searching process (Borgman, Case, & Meadow, 1989). Research by cognitive psychologists (Tversky & Kahneman, 1974) on heuristics and judgment finds that people do not apply normative logical models such as Boolean logic in their everyday reasoning. Rather, they follow intuitive judgement, which includes applying "and" and "or" in their linguistic sense—*and* is inclusive, making things bigger; *or* is exclusive, as in either/or, making things smaller—the reverse of the way these words act as Boolean operators.

Semantic knowledge. Given some conceptual knowledge of the search process, the information seeker needs the semantic knowledge of how to implement a query in a given system. Computer scientists distinguish between *semantic knowledge,* i.e., the meaning of computing and task concepts, and *syntactic knowledge* about a system, i.e., the way in which commands or actions are arranged (Shneiderman, 1992). We regard technical skills as syntactic knowledge in this framework.

Using card catalogs requires the semantic knowledge of how a particular catalog is structured, such as whether it has multiple filing sequences by time period or access point. The structure of online catalogs is less physically apparent than in card catalogs. Rather than opening a card catalog drawer, online catalog searchers must enter a search statement. In most online catalogs, a search statement consists of three parameters in sequence: *Action* (e.g., find, select, scan, browse), an *access point* or *field tag,* (e.g., author, title, subject), and *search terms,* either alone (e.g., "Shakespeare," "biology") or in combination with Boolean operators (e.g., "computers and behavior," "heat or thermal").

Action. The action in searching a card catalog is to scan or flip through cards in sequence, after determining the access point and the set of drawers in which the appropriate cards are filed. The action in online catalogs is the first parameter to specify, although not necessarily the first that the searcher needs to determine. Online catalogs typically offer at least two actions: A "find" or "keyword" search that matches individual words or word combinations in the specified field and retrieves a set of records; and a "browse" search that matches words in either the name or subject authority file and retrieves a set of headings with a set of records linked to each heading. The distinction in function between these two actions can be difficult to explain to the "perpetual novice" searcher who has limited knowledge of bibliographic organization. O'Brien (1994, p. 224) notes that while li-

brarians see a vital difference between keyword and subject searching, they "are as yet unsure of the role of each."

Access points. Searchers must determine which access point to use based on their assessment of the information they have available. They need to decide what fragments of each access point are most complete, distinctive, or reliable. Card catalogs in the United States, following Cutter's objects, usually offer author, title, and subject access. The fact of three access points is more apparent in divided catalogs than in dictionary catalogs, in which the two or three sequences are interfiled. Some card catalogs complicate the matter further by combining two sequences (typically author and title) into one catalog and keeping the third sequence in another. In Europe and elsewhere, card catalogs may be searchable by main entry only (usually author) and perhaps by classification. Providing comprehensive subject access (e.g., Library of Congress Subject Headings) to catalog records is largely a U.S. practice, although other subject thesauri or systems of organization may be applied.

Most online catalogs force the searcher to choose among author, title, and subject access points for the initial search statement. Other fields in the records, such as classification numbers, language, series, date or date range, or type of material, can be used to limit a search statement, but usually must be entered either as an additional parameter or as a subsequent command. The forms of subject access to card and online catalogs vary widely due to local, national, and regional variations in cataloging practice, as well as to differences in online catalog software capabilities.

Search terms. The search terms may be the first parameter determined, although not the first to be specified in card or online catalog searching. Card catalogs and first-generation online catalogs offer limited choices of search terms, typically as follows:

- Author catalog: Surname of personal author or the first significant word in a corporate name;
- Title catalog: First significant word in the title;
- Subject catalog: First significant word in subject headings or index terms;
- Shelf list: Classification number or class mark.

Classification access to card catalogs rarely is offered to the public in the U.S. Shelf list catalogs are essential for catalog maintenance but usually are kept in staff areas. Classification access is more common in Europe and may be offered in place of the subject catalogs that are typical in the U.S.

Search term options increased greatly with the introduction of second-generation online catalogs, which allow keyword searching of the access points. Current systems usually offer the following options:

- Author search: Surname and given name of personal author or any significant word in a corporate name;
- Title search: Any significant word in the title;
- Subject search: Any significant word in subject headings or index terms;
- Shelf list: Classification number or class mark, or a range of numbers.

Searchers need to know whether they can enter the catalog only under the search terms corresponding to the first significant word in the access point or whether all words are searchable. If only the first word is searchable, searchers must be able to determine which word that might be. Variations in cataloging practice result in inconsistent entries, both among different catalogs and within any given catalog. For example, the author catalog may include surnames of all authors or only first authors of a work; determining the appropriate search terms for corporate authors is even more difficult. Searchers may need to know a title precisely enough to know the first significant word and understand the notion of stop words, i.e., words deemed insignificant or too common to be useful for filing, typically articles. Identifying search terms for the subject catalog is hardest of all, since people often do not recognize that the subject entries are drawn from a controlled list or thesaurus that is separately searchable itself. Instead, they enter the catalog using the free-text keywords they know best, often on a trial-and-error basis.

Keyword searching of subject headings provides a powerful mechanism to locate all headings containing a given word and all records containing a given word in the subject fields. However, it does so at the expense of the structure of the subject heading system, which was designed to provide context with headings and subheadings. Many researchers have addressed the problem of subject access searching and display mechanisms in online catalogs, including Bates (1989), Borgman, Gallagher, Hirsh and Walter (1995), Leazer (1994), Markey and Demeyer (1986), McGarry and Svenonius (1991), Micco (1991), Rosenberg and Borgman (1992), and Tillett (1991).

Boolean logic. In card catalogs, terms must be searched individually. In online catalogs, search terms may be combined with Boolean operators, whether the operators are stated explicitly by the searcher or are implied by the system. Specifying queries that contain Boolean logic requires conceptual knowledge of how search terms can be combined and semantic knowledge of how a given system executes Boolean operators. Even if a searcher understands the concepts of combining terms with operators, anomalies in system features may interfere with employing that knowledge. Online catalogs vary widely in how they treat operators, particularly in the use of implicit operators and in the order they execute them. Implicit Boolean and execution order are

difficult features to explain and few systems document them clearly. We provide several examples to illustrate the complexity of these concepts and the significance of their effect on search results.

Early online catalogs allowed single-word search terms only; multiple-word search terms required explicit Boolean operators to specify how the words should be treated. Second-generation systems typically allow the searcher to enter multiple-word phrases, which simplifies the input process but leaves to the system the decision on how to combine the words. Some systems apply an implicit OR between words, others apply an implicit AND, and yet others treat multiple-word search terms as phrases bound together in that sequence. Thus, a search for *research design* would yield a large set of all records containing either of these general words, on a system that employs implicit OR; a small set of records containing both of these words, in a system that employs implicit AND; and a still smaller set of records in a system that treats the words as a phrase in the specified sequence.

Many commercial online retrieval systems execute Boolean operators in algebraic sequence, e.g., reading the whole string and first executing statements within parentheses, then NOT, AND, and OR. Most online catalogs do not allow parentheses and simply execute words and operators from left to right, ignoring the algebraic hierarchy of operators. Thus the same phrase will receive very different results depending on the sequence of operators and the order in which they are executed. To illustrate this difference, let us compare the queries *behavior or psychology and computers* and *computers and behavior or psychology*. We assume the searcher intends the same literal meaning for these phrases, which is to combine *computers* with any records containing either *behavior* or *psychology*. Most online catalogs would treat the first query as the user intended [(behavior OR psychology) AND computers], while any system that executes operators in algebraic order would not, instead retrieving [behavior OR (psychology AND computers)]. The second phrasing would retrieve unintended results whether the system executes left to right or in algebraic order [(computers AND behavior) OR psychology]; it would retrieve the intended results only in those systems that read the full phrase and then execute OR before AND.

Only the most sophisticated searchers are sufficiently knowledgeable of Boolean logic to seek documentation on how the system treats multi-word search terms and the order in which it executes operators. If they devote enough time and effort, they can experiment with multiple search statements that will reveal the system features. However, even expert searchers may achieve confusing, incomplete, or misrepresentative results by assuming that the system operates in a fashion similar to other systems with which they are familiar. People carry over knowledge from one automated system to the next, just as they carry knowledge of card catalogs to online catalogs (Borgman, 1986b; Gentner & Stevens, 1983).

File organization. Once some results are retrieved, the searcher must locate the records in the filing sequence, which requires some semantic knowledge of file organization. Without further instruction, people assume catalogs are arranged like a telephone book or some other well-known filing sequence. However, card catalogs typically are filed by an elaborate set of rules known only to catalogers and para-professional catalog filers—various editions of the *American Library Association Filing Rules*[1] range from 50 to 260 pages in length! Some of the features of the filing rules can be explained in simple terms, such as the difference between word-by-word filing and letter-by-letter filing, but the real difficulties lie in the fact that cards may be filed based on information that is not on the card, such as whether the entry is treating a person as an author or a subject, or the full spelling of a word abbreviated on the card (e.g., cards for English works starting with St. (for Saint) may file before cards for German works starting with St. (for Sankt)). Because the filing rules are so complex, cards often are misfiled, leading to clusters of cards correctly filed behind one misfiled card. Thus, even if the correct access points and search terms are known, it may be difficult to locate the cards, or to determine if the library indeed has any books that match the searched terms.

Just as cards are filed in a specified alphanumeric sequence, online catalog records are displayed in a specified alphanumeric sequence. However, the sorting sequence is programmed based on data in the record. This means that the *ALA Filing Rules* cannot be fully replicated because they rely on information external to the record that is supplied by the human filer (although librarians went to great effort to replicate these filing rules, in the early years of online catalogs, before succumbing to a machine filing sequence). As a result, most online catalogs sort catalog records in a sequence that is more similar to a phone book than a card catalog. Online catalogs based on MARC records have made some improvements over straight alphanumeric sorting, such as coding access points to identify the first significant word for filing, which allows fields beginning with a stop word (e.g., "The Story of O") to file under the first significant word ("story") but allows records to file under stop words when necessary (e.g., "A is for Apple" should file under "A," not "is" or "Apple"). In addition, most of the commercially-produced online catalogs presently in use were developed in the English-speaking world and incorporate only a few characters beyond the standard

[1] The first edition of the *ALA Filing Rules* (American Library Association, 1942) is 109 pages; the second edition (American Library Association, 1968a) is 260 pages; the second edition abridged (American Library Association, 1968b) is 94 pages; and the 1980 edition (American Library Association, 1980) that is the successor to the prior editions was reduced to 50 pages. The complexity of these rules also is apparent from the existence of substantial training manuals, such as Carothers (1981) at 120 pages and Hoffman (1976) at 176 pages.

English alphabet. The rest of the characters are lost in transliteration or in omitting diacritics. As the non-English-speaking world comes online and preserves their full character sets in their online catalogs and other retrieval systems, machine filing order, keyboard input, and display, will become ever more complex (e.g., Peruginelli, Bergamin, & Ammendola, 1992).

Machine filing order probably is more intuitive than are the *ALA Filing Rules* because it relies on data in the record and because the sequence is easier to see when 10 to 20 items are displayed on a screen at once. Strictly speaking, records never are filed out of sequence in an online catalog, eliminating one of the problems of the card catalog. However, a number of other factors can make it difficult to scan a list of records for the item sought. For example, searchers have difficulty with the sort order of mixed letter and number sequences in call numbers, both on the screen and on the shelf. Which comes first, Z699.51 or Z699.6? Numerically, 6 is much less than 51; but alphabetically, 5 sorts before 6. In addition, typos (e.g., "Untied States"), variant punctuation (e.g., U.S. vs. US), and other variations in entry form (e.g., "addresses, essays, and lectures," "addresses, essays, & lectures," "AEL") introduce all the filing problems of card catalogs and a few new ones. Problems with sort order and with stop words presumably are less in online catalogs than in card catalogs because more alternatives exist for locating items. Variant forms, once identified, are more easily corrected in automated systems than in manual systems due to "search and replace" functions—in some systems one action can change all occurrences of "Untied States" to "United States," for example, where such a global change would be prohibitively expensive in a card catalog.

Technical skills. Given the semantic knowledge of how to implement a query in a given system, the searcher needs the technical skills to formulate and execute specific search statements. The primary technical skill required by card catalogs is the ability to manipulate cards in the drawers. The organization of the card catalog is physically apparent, particularly in a divided catalog, and a searcher selects a starting point by pulling out a drawer in the alphabetic vicinity of the desired known item or subject heading. Part of the appeal of card catalogs, that which Baker (1994) captures so well, is the tactile sense of flipping through cards, wandering through the library's collection. Online catalogs require technical skills that card catalogs do not, such as basic computing concepts and the syntax of commands or actions in the system.

Using computer terminals or work stations. We must not take for granted that all library users are computer users, or that their knowledge of computers is adequate or appropriate for online catalog use. A recent article in the *Wall Street Journal* (Carlton, 1994) describes some of the naive questions directed to the "help lines" operated by computer hardware and software companies. The anecdotes are instructive for librarians: A customer trying to operate a mouse as a foot pedal, others pointing the mouse at the screen rather than rolling it on a pad, some who have difficulty finding the power switch, and many customers who request the location of the "any" key in response to screen instructions to "press any key." And these are the people who *do* have a computer!

While searchers of online catalogs need not be computer experts, they must be familiar with the layout of the keyboard, both the keys in common with typewriters (letters, numbers, space bar, caps lock, etc.) and keys specific to computers (*control, return/enter, delete,* function keys, etc.), and with conventions such as pressing *return/enter* after typing a command and using arrow keys or a mouse to move about the screen. Often searchers must learn system-specific function keys, or special uses of general keys, such as *tab* for a new line, or differences between *enter* and *return* on systems that use these keys for different purposes. Searchers also must understand screen display conventions such as consistent placement of instructions, messages to the user, and user input, as well as the nature of menu selection.

These basic skills rarely are included in the instruction manuals for online catalogs, much less the reference sheet. For the skilled computer user, this information may seem trivial. For the searcher with no prior computer experience, the lack of these skills may prove to be an insurmountable barrier since it is difficult to know where to start.

Online catalogs were introduced in libraries long before computers were widely implemented in schools, work places, and homes. While several experimental online catalogs existed in the 1960s and early 1970s, it is generally acknowledged that the first large-scale implementations were at Ohio State University in 1975 (Miller, 1979) and the Dallas Public Library in 1978 (Borgman, 1978; Borgman & Kaske, 1980). By the early 1980s, a sufficient number of online catalogs were in place in the United States for the Council on Library Resources to commission a major study of online catalog usage (Matthews et al., 1983). Time-sharing systems appeared in the 1960s and hobbyist personal computers in the late 1970s (Press, 1993), but large-scale desktop computing is a 1980s phenomenon—the IBM personal computer was introduced in 1981 and the Apple Macintosh in 1984. Only in the last few years has it become reasonable to assume that many library users in the West are familiar with computers. In other parts of the world, where online catalogs are being introduced in societies with far less information technology in place, we cannot make this assumption. For many people, libraries will continue to provide their first encounter with computers.

Syntax of entering search commands. Online catalogs provide structure in more abstract ways than do card catalogs, and the starting point for browsing must be

stated much more explicitly than pulling out a random drawer. Most online catalogs require the user to enter commands in a precise format and sequence such as the three-parameter structure discussed above to describe the semantic knowledge requirements. This structure is fairly consistent across command-driven online catalogs, although the terminology varies widely among systems (unless the Common Command Language is implemented, but it has not been widely adopted in the U.S.). In these examples we use FIND for action, often abbreviated as FI, and AUTHOR (AU) for author. The label for the author field varies widely and can be found also as AUT, NA (name), or PA (personal author) among others. The format for entering the search terms varies, with some systems being more rigid than others. For example, to search for works by Robert M. Hayes, one needs to know which, if any, of the following search statements are acceptable to the system. Some may match all works in the catalog by this author, some a subset, and others none at all.

FI AU ROBERT M. HAYES
FI AU R M HAYES
FI AU HAYES, ROBERT M
FI AU HAYES, ROBERT M.
FI AU HAYES ROBERT M
FI AU HAYES ROBERT
FI AU HAYES, ROBERT
FI AU HAYES RM
FI AU HAYES R.M.
FI AU HAYES R
FI AU HAYES, R?

The search statements in this example vary by word order, number of words in the search terms, completeness of search term (name or initials), and punctuation; none include explicit Boolean operators. Most newer online catalogs are fairly forgiving on these factors, searching words in author names as Boolean combinations in any order and ignoring punctuation, while older systems (but still in general use) require precise character-by-character matches. The last entry in the above list is an example of truncation, which allows for variant word endings. This feature is similar to flipping through catalog cards before and after the one sought to look for variant endings. The distinction between exact-match and browsing searches in online catalogs can be difficult to explain. Most online catalogs match exactly on the characters entered and do not select variant endings unless explicitly told to do so. The person with minimal computer experience may not realize that an author search for "HAYES R" will *not* match "HAYES ROBERT." To retrieve all variant endings may require a different command that searches the authority files, such as a *browse* search discussed previously.

Commands to manipulate the system. In addition to the commands required to execute a query, online catalogs have commands to control various aspects of the system. We divide these into several groups:

Commands to control the search. Online catalogs include a variety of control commands, such as actions to modify a prior search, start a new search, move forward and backward through lists, and switch between lists of subject headings and lists of bibliographic records.

Commands to control display formats. Card catalogs have only one display format: The card. Online catalogs often have several types of displays under searcher control. Search results may be presented as a summary list of single-line entries, and individual records may be displayed as short or long versions of bibliographic records that vary by which fields are shown (e.g., notes, holdings, circulation status).

Commands to control output of results. One of the weaknesses of early online catalogs was that results still had to be copied by hand onto slips of paper, offering little improvement over card catalogs. Now many online catalogs allow search results to be printed, downloaded onto disks, or even sent by electronic mail. Various commands allow the user to control how this output is formatted and where it is sent.

Making Online Catalogs Easier to Use

Online catalogs should be judged by their success in answering questions rather than by their success in matching queries. In the long term, we need to design systems that are based on behavioral models of how people ask questions. Such a design model could assist in the question-negotiating process, allowing the searcher to pursue multiple avenues of inquiry by entering fragments of the question, exploring vocabulary structures, capturing partial results, reformulating the search with the assistance of various specialized intelligent agents, retaining elements of a search for future sessions, and even transferring elements to other systems. Many researchers are proposing new models for the design of online catalogs, such as Bates' (1989) "berry-picking" and (Bates, 1986) "subject access design model"; Hildreth's (1993) model of the "E^{3}OPAC": Enhanced, expanded, and extended online public access catalog; and the work of Robertson, Hancock-Beaulieu, and Walker (Robertson & Hancock-Beaulieu, 1992; Hancock-Beaulieu, 1989; 1991; Walker & Hancock-Beaulieu, 1991) and Efthimiadis (1992, 1993, 1995) on query reformulation and relevance ranking. We have devoted more than 5 years to the iterative design of an exploratory online catalog based on children's development levels, knowledge, and interests (Borgman et al., 1995). Other researchers are studying searching behavior on extant systems that provide a basis for design models, such as the work of Bates (1990); Belkin, Oddy, and Brooks (1982); Ellis (1989); Ingwersen (1984); Kuhlthau (1988, 1991); Lynch

(1992); Marchionini (1992, 1995); Markey (1980, 1986, 1989); and Shute and Smith (1993). Very little of this body of research has informed the design models of the commercial online catalogs that are in general use around the world.

In the short term, we can help make online catalogs easier to use through improved training and documentation that is based on information-seeking behavior. Current instructional approaches tend to focus on the procedures for query formulation, rather than on the question being asked. Instruction is short and precise, with examples of how to enter an author, title, or subject query, and examples of Boolean operators. The formats usually consist of reference cards, online help, and short training sessions for the few willing to attend. Full instructional manuals aimed at the end user are rare, thus documentation on important features such as implicit Boolean and execution order is hard to locate. Even harder to find is a conceptual framework for the how, when, and why of searching.

Some people need and want more instruction than do others, and no single form of instruction will be adequate or appropriate for all online catalog users. Online catalogs will be easier for most people to use if they can find a level and type of instruction or documentation that suits their needs. Some will prefer sets of documentation similar to that provided for other complex software such as word processing or spreadsheet applications, including instructional manuals, reference manuals, and summary reference cards. Others will prefer short summaries that provide an overview and a set of examples. Still others, including skilled searchers familiar with many retrieval systems, will want detailed documentation on how the system handles each feature and function. Given the increasing portion of users who access online catalogs remotely rather than from the library building, much of the instructional material needs to be online as part of the system. Those who use terminals in the building may be served best by aides who can provide individual assistance and who, in turn, can troubleshoot system problems. Searching assistance can be provided to remote users by an online helper who can respond to questions in real time.

High-quality instruction can overcome some problems in system design, but should not be a substitute for design improvements. Time invested in elaborate help systems often is better spent in redesigning the user interface so that the help is no longer needed. We learned early in our studies of online catalogs that people were not willing to devote much time to learning to use these systems (Matthews et al., 1983). Searchers expect systems to be easy to use with relatively little time investment, which has considerable implications for library services. As O'Brien (1994) notes, we assumed that serious users of card catalogs would invest considerable effort in learning to use them well; we have fewer expectations of online catalog users. Thus, we are not claiming that card catalogs were easier to use or required less instruction. Rather, we have higher expectations of online catalogs' capability to support the information-seeking process than we did of card catalogs. We also have devoted more study to understanding the online catalog searching process, so we have better data on the problems. Our long-term goal should be to design intuitive systems that require a minimum of instruction. Given the complexity of the information retrieval problem and the limited capabilities of today's systems, we are far from achieving that goal.

Summary and Conclusions

Despite numerous improvements to the user interface of online catalogs in recent years, searchers still find them hard to use. Most of the improvements are in surface features rather than in the core functionality. We see little evidence that our research on searching behavior studies has influenced online catalog design. While queries now may be simpler to input, searchers still bear the burden of translating their question into a precise structure that the system can interpret. Online catalogs continue to require searchers to specify a query in terms of actions, access points, search terms, and Boolean operators to begin a search. Query-based systems were designed for expert librarian searchers who have a rich conceptual framework for information retrieval; their expertise lies in translating questions into queries on behalf of end users. Most end users of online catalogs are perpetual novices who lack the requisite conceptual knowledge for searching. They need assistance in the translation process, whether provided by the system itself, by instruction in using the system, or by a search intermediary.

Further improvements in the user interfaces to query-based retrieval systems are likely to have minimal effect on searchers' abilities to answer questions. Instead, we need to incorporate more knowledge of searching behavior into the design of these systems. Online catalogs should assist searchers in employing whatever information they have in hand to obtain an answer to their question. System designers must recognize that the goal of a search may be merely to refine the question, not to obtain an answer, and that a "search" may be conducted over multiple sessions and multiple systems. Assistance and explanation should be offered accordingly. Current systems can be made more effective through training that provides a conceptual framework for searching rather than simply a set of procedures for stating queries, with the caveat that good training is not a substitute for good system design.

Online catalogs are the most widely-available automated retrieval systems and the first that many people encounter. They have become the core of public access systems that provide a range of databases and other information services. After nearly 30 years of experience

with commercial retrieval systems and 20 years with online catalogs, we would expect them to employ the latest in interface design, yet we continue to find them difficult to use. As software for other educational, business, and entertainment applications becomes simpler and more powerful, expectations of the online-catalog-user community will continue to increase, placing even greater demands on the library community to provide systems that are at least as refined as today's desktop workstations. If libraries are to provide primary information services for the networked world, they need to rise to this challenge, putting research results on the information-seeking process into practice in designing the next generation of online public access information retrieval systems.

Acknowledgments

Portions of this article are based on a talk presented in the Wayne State University Libraries Distinguished Lecture Series in April 1994. The author is most grateful for extensive commentary on earlier drafts of the paper by Donald Case, Davina Klein, and Katalin Radics of UCLA, and Charles Hildreth of the University of Oklahoma. Nadia Caidi of UCLA provided expert bibliographic assistance. All errors, omissions, and interpretations remain the responsibility of the author, of course.

References

American Library Association. (1942). *A.L.A. rules for filing catalog cards.* Prepared by a special committee; Sophie K. Hiss, chairman. Chicago: American Library Association.

American Library Association. Filing Committee. (1980). *ALA filing rules.* Chicago: American Library Association.

American Library Association. Subcommittee on the ALA Rules for Filing Catalog Cards. (1968a). *ALA rules for filing catalog cards.* Prepared by the ALA Editorial Committee's Subcommittee on the ALA Rules for Filing Catalog Cards. Pauline A. Seely, chairman and editor (2nd ed.). Chicago: American Library Association.

American Library Association. Subcommittee on the ALA Rules for Filing Catalog Cards. (1968b). *ALA rules for filing catalog cards.* Prepared by the ALA Editorial Committee's Subcommittee on the ALA Rules for Filing Catalog Cards. Pauline A. Seely, chairman and editor (2nd ed., abridged). Chicago: American Library Association.

Baker, N. (1994, April 4). Discards. *The New Yorker,* 64–86 passim.

Barlow, J. P. (1994). The new economy of ideas. *Wired, 2,* 84.

Bates, M. J. (1986). Subject access in online catalogs: A design model. *Journal of the American Society for Information Science, 37,* 357–376.

Bates, M. J. (1989). The design of browsing and berry-picking techniques for the online search interface. *Online Review, 13*(5), 407–424.

Bates, M. J. (1990). Where should the person stop and the information search interface start? *Information Processing & Management, 26*(5), 575–591.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation, 38*(2), 61–71.

Borgman, C. L. (1978). *The role of technology for the Dallas Public Library in long range planning.* Educational Resources Information Center, September, 1978. (ERIC Document Reproduction Service No. ED 153 698)

Borgman, C. L. (1986a). Why are online catalogs hard to use? Lessons learned from information retrieval studies. *Journal of the American Society for Information Science, 37*(6), 387–400.

Borgman, C. L. (1986b). The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *International Journal of Man-Machine Studies, 24*(1), 47–64.

Borgman, C. L. (1995). Information retrieval or information morass? Implications of library automation and computing networks in central and eastern Europe for the creation of a global information infrastructure. In T. Kinney (Ed.), *Proceedings the Annual Meeting of the American Society for Information Science, 32,* October 9–12, 1995, Chicago (pp. 27–34). Medford, NJ: Information Today.

Borgman, C. L., Case, D. O., & Meadow, C. T. (1989). The design and evaluation of a front end user interface for energy researchers. *Journal of the American Society for Information Science, 40,* 86–98.

Borgman, C. L., Gallagher, A. L., Hirsh, A. G., & Walter, V. A. (1995). Children's searching behavior on browsing and keyword online catalogs: The Science Library Catalog Project. *Journal of the American Society for Information Science, 46,* 663–684.

Borgman, C. L., & Kaske, N. K. (1980). Online catalogs in the public library: A study to determine the number of terminals required for public access. *Proceedings of the 43rd American Society for Information Science Annual Meeting,* October 4–9, 1980, Anaheim, CA (pp. 273–275). White Plains, NY: Knowledge Industry Publications.

Borgman, C. L., & Siegfried, S. L. (1992). Getty's Synoname™ and its cousins: A survey of applications of personal name matching algorithms. *Journal of the American Society for Information Science, 43,* 459–476.

Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science, 42,* 351–360.

Buckland, M. (1992). *Redesigning library services: A manifesto.* Chicago: American Library Association.

Carlton, J. (1994, March 1). Befuddled PC users flood help lines, and no question seems to be too basic. *Wall Street Journal.*

Carothers, D. F. (1981). *Self-instruction manual for the filing of catalog cards.* Chicago: American Library Association.

Chen, H., & Dhar, V. (1990). User misconceptions of information retrieval systems. *International Journal of Man-Machine Studies, 32,* 673–692.

Crawford, J. C., Thom, L. C., & Powles, J. A. (1993). A survey of subject access to academic library catalogues in Great Britain. *Journal of Librarianship and Information Science, 25*(2), 85–93.

Cutter, C. A. (1904). *Rules for a dictionary catalog* (4th ed). Washington, DC: Government Printing Office.

Efthimiadis, E. N. (1990). Online public access catalogs: Characteristics of the literature. *Journal of Information Science, Principles, and Practice, 16*(2), 107–112.

Efthimiadis, E. N. (1992). *Interactive query expansion and relevance feedback for document retrieval systems.* Unpublished doctoral dissertation, City University, London, U.K.

Efthimiadis, E. N. (1993). A user-centered evaluation of ranking algorithms for interactive query expansion. In R. Korfhage, E. Rasmussen, & P. Willett (Eds.), *Proceedings of the 16th International Conference of the Association of Computing Machinery, Special Interest Group on Information Retrieval,* June 1993, Pittsburgh, PA, (pp. 146–159). New York: ACM Press.

Efthimiadis, E. N. (1995). User choices: A new yardstick for the evaluation of ranking algorithms for interactive query expansion. *Information Processing and Management, 31*(4), 605–620.

Ellis, D. (1989). A behavioural model for information retrieval system design. *Journal of Information Science, 15,* 237–247.

Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models.* Hillsdale, NJ: Erlbaum.

Hancock-Beaulieu, M. M. (1989). Online catalogues: A case for the user. In *The online catalogue: Developments and directions* (pp. 25–46). London: Library Association.

Hancock-Beaulieu, M. M. (1991). User friendliness and human-computer interaction in online library catalogs. *Program, 26*(1), 29–37.

Hildreth, C. R. (1987). Beyond Boolean: Designing the next generation of online catalogs. *Library Trends, 35*(4), 647–667.

Hildreth, C. R. (1993). *An evaluation of structured navigation for subject searching in online catalogues.* Unpublished doctoral dissertation, Department of Information Science, City University, London, UK.

Hoffman, H. H. (1976). *What happens in library filing?* Hamden, CN: Linnet Books.

Ingwersen, P. (1984). Psychological aspects of information retrieval. *Social Science Information Studies, 4,* 83–95.

Kuhlthau, C. C. (1988). Developing a model of the library search process: Cognitive and affective aspects. *RQ, 28*(2), 232–242.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*(5), 361–371.

Lancaster, F. W., Connell, T. H., Bishop, N., & McCowan, S. (1991). Identifying barriers to effective subject access in library catalogs. *Library Resources & Technical Services, 35*(2), 377–391.

Larson, R. R. (1991a). Between Scylla and Charybdis: Subject searching in the online catalog. In *Advances in Librarianship, 15* (pp. 175–236). San Diego: Academic Press.

Larson, R. R. (1991b). Classification clustering, probabilistic information retrieval and the online catalog. *Library Quarterly, 61*(2), 133–173.

Larson, R. R. (1991c). The decline of subject searching: Long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science, 42*(3), 197–215.

Leazer, G. H. (1994). A conceptual schema for the control of bibliographic works. In D. L. Anderson, T. J. Galvin, M. D. Giguere (Eds.), *Navigating the networks: Proceedings of the American Society for Information Science Mid-Year Meeting* (pp. 115–135). Medford, NJ: Learned Information.

Lynch, C. A. (1992). The next generation of public access systems: Lessons from 10 years of the MELVYL system. *Information Technology and Libraries, 11*(4), 405–415.

Marchionini, G. (1992). Interfaces for end-user information seeking. *Journal of the American Society for Information Science, 43,* 156–163.

Marchionini, G. (1995). *Information seeking in electronic environments.* Boston: Cambridge University Press.

Markey, K. (1980). *Analytical review of catalog use studies.* Dublin, OH: Online Computer Library Center. OCLC Research Report No. OCLC/OPR/RR-80/2.

Markey, K. (1986). Users and the online catalog: Subject access problems. In J. R. Matthews (Ed.), *The Impact of Online Catalogs* (pp. 35–69). New York: Neal-Schuman Publishers.

Markey, K. (1989). Alphabetical searching in a online catalog. *The Journal of Academic Librarianship, 14*(6), 353–360.

Markey, K., & Demeyer, A. N. (1986). *Dewey decimal classification online project: Evaluation of a library schedule and index integrated in the subject searching capabilities of an online catalog: Final report to the Council on Library Resources.* Dublin, OH: OCLC.

Matthews, J. R., Lawrence, G. S., & Ferguson, D. K. (1983). *Using online catalogs: A nationwide survey.* New York: Neal-Schuman.

McGarry, D., & Svenonius, E. (1991). More on improved browsable displays for online subject access. *Information Technology and Libraries, 10*(3), 185–191.

Micco, M. (1991). The next generation of online public access catalogs: A new look at subject access using hypermedia. In D. A. Tyckoson (Ed.), *Enhancing access to information: Designing catalogs for the 21st century* (pp. 103–132). New York: Haworth Press.

Miller, S. L. (1979). The evolution of an on-line catalog. In R. D. Stueart & R. D. Johnson (Eds.), *New horizons for academic libraries* (pp. 193–204). New York: K. G. Saur.

O'Brien, A. (1994). Online catalogs: Enhancements and developments. In M. E. Williams (Ed.), *Annual review of information science and technology, 29* (pp. 219–242). Medford, NJ: Learned Information. (Published on behalf of the American Society for Information Science).

Peruginelli, S., Bergamin, G., & Ammendola, P. (1992). Character sets: Towards a standard solution. *Program, 26*(3), 215–223.

Press, L. (1993). Before the Altair: The history of personal computing. *Communications of the Association for Computing Machinery, 36*(9), 27–33.

Robertson, S. E., & Hancock-Beaulieu, M. M. (1992). On the evaluation of IR systems. *Information Processing & Management, 28*(4), 457–466.

Rosenberg, J. B., & Borgman, C. L. (1992). Extending the Dewey decimal classification via keyword clustering: The Science Library Catalog project. *Proceedings of the 54th American Society for Information Science Annual Meeting, 29,* October 26–29, 1992, Pittsburgh (pp. 171–184). Medford, NJ: Learned Information.

Shneiderman, B. (1992). *Designing the user interface: Strategies for effective human-computer interaction* (2nd ed). Reading, MA: Addison-Wesley.

Shute, S. J., & Smith, P. J. (1993). Knowledge-based search tactics. *Information Processing & Management, 29*(1), 29–45.

Tagliacozzo, R., Kochen, M., & Rosenberg, L. (1970). Orthographic error patterns of author names in catalog searches. *Journal of Library Automation, 3,* 93–101.

Taylor, A. G. (1984). Authority files in online catalogs: An investigation of their value. *Cataloging & Classification Quarterly, 4*(3), 1–17.

Tillett, B. B. (1991). A taxonomy of bibliographic relationships. *Library Resources & Technical Services, 35*(2), 150–158.

Tversky, A., & Kahneman, D. (1974). Judgements under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Walker, S. (1988). Improving subject access painlessly: Recent work on the Okapi online catalogue projects. *Program, 22*(1), 21–31.

Walker, S., & Hancock-Beaulieu, M. M. (1991). *OKAPI at City: An evaluation facility for interactive IR* (British Library Research Report No. 6056). London: British Library.