CHARLES R. HILDRETH
Chief Consultant
READ Ltd.
Seattle, Washington

Extending the Online Catalog:
The Point of Diminishing Returns

INTRODUCTION

This paper discusses online public access catalog (OPAC) models and milestones. The journey of the networked online catalog has clearly departed from its early, first-generation manifestation as a stand-alone, local "card catalog online" and continues at an alarming pace in its evolution to some fuzzily conceived expanded, transformed entity we may humbly refer to for now as the library catalog information system. What this future information management and retrieval system will be like is not entirely clear, which makes it the topic of much speculation and wishful thinking. Its parameters and features may be decided by, nay, even driven by, new, emergent technologies, or, hopefully, its form and function will be guided by the knowledge attained from a vast body of research and experience with online catalogs and their users.

Technology defines the possible—what can be done with available resources. Technology sets no intrinsic limits on itself and has no magical power to articulate its ideal future. On the other hand, advances in technology may add the spark of innovation to the application of acquired knowledge in the design of new and improved information systems. As we stand on the threshold of an exciting era of expanded information access and delivery, there is another danger equal, perhaps, to the uncritical adoption of new technologies. The danger I speak of is the danger of uncritical devotion to outmoded models (forms, if you like) and long-standing ways of doing things. (I am tempted to coin the phrase, "bibliographic nostalgia," but I shall restrain myself.) Furthermore, the gloss of new technologies may disguise the fact that underneath things remain pretty much the same as they always were.

Let me bring to mind imagery from a not-so-long-ago era in library history. It is a personal recollection of a "hard" reality, far distant, it would seem, from those virtual realms we hear so much about these days. Thirty years ago, while an undergraduate student at Ohio State University, when I went to the main library on campus to look up materials, I accessed the massive public card catalog to locate needed books or bound periodicals. If successful at the card catalog, I then went to a designated "workstation," which supplied small pencils and slips of paper. For each book I desired, I wrote down on one of these slips its author, title, and call number. I then took these slips to the
circulation counter where a clerk confirmed that they contained sufficient data and sent them via pneumatic tubes to "stack runners." You see, the stacks with their organized shelves of books were closed to ordinary folk like undergraduate students (unless one was employed by the library as a stack runner). Bookshelf browsing, as a means of discovering works of interest, was not permitted. In this closed-stack retrieval system, you had to know precisely what you wanted and identify it in the card catalog before it could be retrieved by a stack runner from the bowels of the library warehouse.

The central question I want to pose today is this: By providing network access to our conventional online catalogs using advanced access tools, such as Telnet, Gopher, Z39.50 protocols, Wide Area Information Servers (WAIS), and World Wide Web (WWW)—that is, opening the doors to a vast number of online "libraries without walls"—are we not at the same time "closing the stacks" and eliminating the opportunity for a variety of kinds of meaningful browsing—browsing that often leads to the discovery of previously unknown items of interest? I consider personal interaction with an organized collection of published materials to be a rewarding activity, and I think providing this opportunity should be a service priority of most libraries. As we shift our priorities from building and maintaining physical collections to the provision of electronic access to document collections, are we not in danger of eliminating a qualitative experience and closing the stacks once again? Perhaps today's network access technologies like WAIS and Z39.50 are the digital versions of human stack runners and book retrievers.

THE DEVELOPMENT OF OPAC MODELS AND APPROACHES

OPAC Access Models

Figure 1 presents a framework for understanding 15 years of OPAC history and the significance of recent developments. Two access models informed and motivated the design of the first OPACs. One track of development reflected attempts to emulate the familiar card catalog. Another track adopted the model familiar to online database searchers of commercial search services like DIALOG and BRS. Second-generation OPACs, with their multiple access points, search approaches, and user-friendly display formats, represent the marriage of these two access models.

From the earliest days of second-generation OPACs, the challenge that confronted system developers was to combine the ease of use (or at least, familiarity) of the card catalog with the powerful search capabilities available to trained online database searchers. This task has not been easy, but significant progress has been made.

Recent developments with OPACs, including the advent of the networked OPAC, must be kept in proper perspective. Progress has been made along some, but not all, dimensions. Improvements to the user interface have made second-generation OPACs more usable for the untrained user, and new graphical user interface (GUI) techniques like windows and point-and-click buttons hold the promise of rendering the search process both more intuitive and more direct.
In less than five years, escalating demand and professional ingenuity have resulted in the expanded, extended OPAC. Second-generation OPACs are being expanded in information content and coverage almost daily, and access to and from these OPACs has been extended to vast numbers of remote users through network technologies. Why, you may ask, is not this expanded, extended OPAC advanced enough to qualify for the distinction of a third-generation OPAC? I shall address this question in a moment.

**Online Catalog Perspectives**

At about the time second-generation functionality was becoming the operational standard for installed online catalogs in the mid-1980s, librarians and system planners had already begun to expand and extend the traditional access boundaries of these library catalogs. These developments have enriched and enlivened the debate about the proper role, content, and functions of the online catalog. In this continuing debate, several perspectives on this evolving library access system can be identified. The online catalog may be viewed as

- the expanded bibliographic database,
- a sophisticated computerized search and retrieval system,
- a "gateway" to other online catalogs and electronic resources,
- an integrated component of a multifunction "scholar's workstation."
These new perspectives on the library catalog represent the nature and extent of the rising expectations for expanded online catalog data content, access, and functionality. Furthermore, these perspectives are being integrated in our emerging vision of future library information systems, a vision that hopefully will guide the design of these systems.

The $E^3$OPAC

There has been some confusion in the literature of late as to whether the emerging online catalog is the "expanded" or the "extended" online catalog. Some writers view these terms as synonymous and use them interchangeably, while others use them to explain related but different developments. For example, Potter (1989) identifies three complementary "expansion" paths along which online library systems will proceed: (a) more indexes to more sets of collections and more online reference databases; (b) the gradual inclusion of more full text of journal articles and, possibly, books; and (c) "greater connectivity from online library systems to other systems, including other library systems, commercial services, bibliographic utilities, local networks, CD-ROM servers, and other information providers in the community" (p. 104). Mischo and Cole (1992) put it this way: "Recently, the idea of the 'extended' OPAC has been introduced to describe online catalogs containing specific functional or data extensions. Extended third generation catalogs typically provide value-added access beyond the conventional OPAC by providing expanded entry points, augmented information resources, access to locally mounted and/or remote periodical index databases, and gateway functions to local, regional, and national telecommunication networks" (p. 38).

I have introduced the notion of the $E^3$OPAC as a vehicle for elucidating these concepts: the $E^3$OPAC would have enhanced functionality and usability; its indexing, record data content, and collection coverage would be expanded to make it a "full-collection" access tool; and its access would be extended (through linkages, networks, and gateways) to include the collections and resources of other libraries and information centers.

Progress along the road to the $E^3$OPAC may be tracked on a three-dimensional scale, as illustrated in Figure 2. One can redefine or reprioritize the axis values, but such a scale is useful for identifying where progress may be lagging and also for comparisons of different systems. Third-generation systems are defined primarily by their advanced interface search functionality. It is along this dimension that OPAC progress seems to be stalled. It is important to understand why this is the case.

Third-Generation Online Catalogs

I have outlined a general framework for classifying online catalogs into first-, second-, or third-generation systems, each generation having its distinguishing features and functions (Hildreth, 1989). Innovative design work on the user-system interface, including GUIs, has made many of the second-generation online catalogs far easier to use than the conventional, dial-up commercial database search systems after which they were modeled. Third-generation online catalogs are not yet generally available in the mainstream
library system marketplace. Only a few of these more advanced catalogs have been developed, primarily as prototype or demonstration systems. These systems incorporate many of the listed search, matching, and interactive display techniques developed and tested over the past 25 years by information retrieval system researchers. The major functional improvements that will define the next generation of online catalogs are listed below. Third-generation OPACs will include second-generation functionality plus the following features:

- Natural Language Query Expressions (In your own language, what it is you are looking for)
- Automatic Term Conversion/Matching Aids (Spelling correction, Soundex, intelligent stemming, synonym tables, etc.)
• Closest, Best-Match Retrieval (Unlike Boolean queries, doesn’t require exact match to be retrieved as possibly relevant; matching documents are weighted for ranking)
• Ranked Retrieval Output (Many ranking criteria: most likely to be relevant first, most recent, most cited, most circulated, etc.)
• Relevance Feedback Methods (‘‘Give me more like this one,’’ ‘‘What else do you have on this topic? ’’ ‘‘This book is not at all what I want!’’)
• Hypertext, Related-Record Searching and Browsing
• Integration of Keyword, Controlled Vocabulary, and Classification-Based Search Approaches
• Expanded Coverage and Scope (The ‘‘full-collection access tool’’)

There are many ways of describing and classifying these features, and progress will almost certainly occur in incremental steps, but the third-generation online catalog will be a wholly new kind of retrieval system because it will be based on much more representative models of actual user information-seeking behaviors.

The Scholar’s Workstation

The model of the E³OPAC is being extended further—some might say, replaced—by the recent development of microcomputer-based personal or ‘‘scholar’s’’ information workstations. The emergence of several important information technologies has provided researchers and system developers the tools needed to support not only the expanded, extended online catalog, but also this further development of powerful, multipurpose, information workstations. These technologies include powerful microcomputer workstations, optical and advanced magnetic disk storage media, computer graphics and imaging technologies, sophisticated document retrieval and management software, and widespread national and institutional high-speed computer communication networks, gateway, and linking facilities.

The early applications of these technologies center around the development of personal information systems or ‘‘scholar’s workstations’ linked via local networks to both nearby and remote computing and information resources. The workstation is typically implemented on a microcomputer platform that employs a variety of special-purpose software modules to enhance user access to the online catalog and other local and remote information resources. The workstation may be viewed as the center of a client-server access system model that includes a distributed retrieval network of databases on local and remote file servers, with the user interface, gateway, and other ‘‘client’’ software residing on the microcomputer workstation. The University of Illinois’s implementation of this model is illustrated in Figure 3 (adapted from Mischo & Cole, 1992). In this architecture, the information databases may be contained in a variety of storage media and may reside at various locations. Search interface software is used to provide a unified access environment for the end-user.

Some have described this scenario as the ‘‘one-stop, self-service information station.’’ Others call it the ‘‘electronic library without walls.’’ Mischo and Cole (1992) point out that:

From a single workstation, a user will be able to: 1. perform a literature search using the major periodical index databases; 2. identify, retrieve, and read the full text of journal articles, book chapters, etc.; 3. send results to
electronic mailboxes and personal databases as desired; 4. use scholarly software residing on the workstation or provide a gateway to a remote computing facility (such as a supercomputer) for data analysis or preparation; and 5. capture and display the results of the work using the multimedia capabilities of the workstation to prepare presentation materials for the classroom or publication. (p. 43)

With the proper communications software installed, these workstations can provide access to the expanded, extended catalog from within libraries, as well as other locations such as work offices and homes. Some predict that most online catalog searches in the future will be performed on nonlibrary workstations.

**SUBJECT SEARCHING IN OPACs**

In the use of conventional information retrieval systems and second-generation OPACs, we are faced with the paradox of information retrieval: the need to describe that which you do not know in order to find it. You see, this is because these systems are *query-oriented* retrieval systems. Query-oriented retrieval systems require the matching of queries and representations of documents or text. The specifications provided in the query must be satisfied to some extent by any document representations that would make up the retrieval or "results" set. Thus, in a query-oriented retrieval system, one must have a pretty good idea of what one is looking for; what one needs to satisfy the
information need, the "object of one's interest"; and one must be able to describe that object linguistically, at least partially, in a way that can be "understood" by the system. There is good reason to question whether or not this search and retrieval paradigm reflects the way most information seeking is actually carried out by individuals.

For purposes of analysis, I have classified OPAC subject searching approaches as either "querying" or "browsing," which are discussed below.

I. QUERYING
   A. Phrase Matching
      (Text strings or controlled vocabulary)
   B. Keyword Matching
      (Discrete words, with Boolean or proximity formulations)

**Query search requirements: Search aim/criteria known and can be expressed with relative precision and completeness

II. BROWSING
   A. Pre-sequenced, linear, inflexible
      (Typically, lists of index terms, headings, descriptors, or brief titles)
   B. Nonlinear, multidirectional, flexible
      ("navigation," "chain," "bridge," "relational," "hypertext," "serendipitous" browsing)

**Browse search requirements: Search aim/criteria not specific, not known, and/or cannot be expressed in appropriate query/indexing language

Query Searching

There are two kinds of query searching: phrase matching and keyword matching. A query consists of a term or terms (e.g., a character, number, word or words, or a phrase) and the specification, sometimes called the query "formulation," which defines how the component term(s) of the query are to be interpreted or related for matching purposes (e.g., word truncation, Boolean combinations, word adjacency). The matching function of an online catalog is the mechanism through which the retrieval software makes a comparison between index terms that represent documents and query terms to effect retrieval. The matching criteria are specified through the query by the user or applied automatically by the system. Query searching of either kind (often called just "searching," to distinguish it from an online catalog's browse mode) utilizes an exact matching function on the part of the system, regardless of the manner in which the matching criteria are specified.

In this all-or-nothing approach, documents (bibliographic records in online catalogs) will be retrieved in response to a search only if an exact match of the query is found. The query may consist of a pre-coordinated phrase (with or without truncation) or a post-coordinated Boolean expression of keywords. In either case, the query search matching requirements are precise and rigid. The process is purely mechanistic. The burden is on the searcher to enter terms that will match the entry (index) terms in the database and to specify appropriate proximity or term relationship logic. Bates (1986) criticizes this predominant
approach to subject searching for requiring a "perfect 'pinpoint' match" on the one best term (p. 373). No match means no retrieval, as viewers of empty online catalog screens witness too often. The search may fail (i.e., not identify relevant documents that are in the collection) unless the searcher knows or guesses the exact way the term (word or phrase) appears in the subject index.

In keyword, Boolean queries, the system's matching mechanism makes a binary (yes/no) split of the database into bibliographic records that conform exactly to the requirements of the query, and all the rest. Only the former are retrieved as "hits." Partial or "closest" matching operations are generally not supported in second-generation online catalogs and conventional retrieval systems.

Query searching is an appropriate, useful search option when the aim of the search is specific, when the searcher knows precisely what he or she wants, and when this request can be expressed in the language of the database. Even in subject searching for books or articles on a topic, the searcher may know his or her topic exactly and may be able to express it in the language of the system (e.g., the assigned subject headings or descriptors).

**Browse Searching**

Browsing in online catalogs can take many forms. Typically, the system displays ordered lists of terms, descriptors, or brief bibliographic records for scanning by the searcher. Lists of index terms are usually presented in alphabetical order. The arrangement of brief citation records may be according to date, and some systems support short record browsing in shelf-list order. Usually the only "navigation" option for browsers is to go backward or forward through the list in a constrained, linear manner. Cross-references, if included, represent a way of jumping out of the sequence and over to related areas of the database. Hypertext operations, which permit navigation throughout the database's network of related terms and records, and the dynamic definition of "related areas and interests," have not been implemented in second-generation online catalogs. Conventional browsing assumes a vocabulary aim on the part of the searcher. It assists in identifying the correct form of a term and any related terms. Other forms of browsing, rare in today's online catalogs, support related record or document discovery through nonlinear explorations of the database.

Browse searching is the most useful and preferred approach when the search aim is not specific (regarding, for example, discipline or topic, type of publication, level of treatment, perspective, etc.), the desired results are not precisely known in advance, or the correct terms for representing the user's query (which may be vague) are not known at the outset. One or more of these circumstances may be present in most subject searching activities.

**Current Status of Subject Searching**

Most of today's operational information retrieval systems and second-generation online catalogs use exact match retrieval techniques, featuring keyword, Boolean, proximity, and string searching. Search field specification, truncation, and/or wild card searching is usually supported as well. These
exact match techniques require that the specifications of the query (e.g., the search terms and their specified logical or textual relationships) be satisfied precisely by any and all document representations that would make up the retrieval set.

Although the object of widespread criticism by researchers and many librarians, exact match searching remains the paradigm for operational online information retrieval, CD-ROM, and online catalog systems. There is much discussion and debate in the research literature regarding the reasons for this situation and why it continues. Two explanatory factors should be mentioned in brief: first, some techniques have been employed by system designers that relax the constraints of exact match searching, for example, stemming of query or index terms and the provision of “wild card” searching; and, second, the conventional Boolean, post-coordinate search approach is preferred by professional searchers of online databases because its precision and flexibility permit them to express an information need accurately. In other words, it can be plausibly argued that Boolean propositions provide the flexibility and finesse to represent fine aspects of a user’s information with great precision. Researchers and designers have given database searchers post-coordinate searching tools that are both powerful and flexible for constructing expressions of users’ information needs.

Designers of second-generation online catalogs implemented this model in the 1980s largely because it was the model incorporated by the major commercial online search services and because it was preferred by the librarians who had become the trained, experienced users of those services. Willett (1988) points to the inertia factor: “Boolean systems have been with us for many years now and there is a natural disinclination on the part of both users and system providers to develop new techniques” (p. 11).

After end-users' difficulties with Boolean query systems began to be widely reported and discussed, some online catalog designers implemented various techniques aimed at reducing the difficulties associated with formulating and entering complex queries. For example, menus were provided for command selection, and users of these system interfaces had only to enter search terms and optionally specify a type of search or field to be the target of the search. The online catalog software then “constructed” the query and supplied the Boolean or proximity operator to coordinate the terms entered by the user. The default or “implicit” operator used to specify the relationship between the search terms could, in many cases, be changed by system managers if they felt it was necessary to change the logic of the relationship between search terms. For example, changing the system-supplied implicit operator “between” search terms from adjacency to the Boolean “AND” would likely broaden the search and usually yield a larger results set or reduce the number of no match, “no hit” search failures. This change was found to be necessary when users began to complain of not being able to find titles of books they knew were in the collection, and consultation with transaction logs confirmed the problem. Searchers typically remember and enter two or three significant words in a title, rather than the complete title or precise order of words in the title.

Explicit and Implicit Online Catalogs

With or without these “user-friendly” techniques, most online catalogs in operation are still Boolean query or string-matching, exact match retrieval
systems. One might refer to these two kinds of second-generation online catalogs as "explicit" and "implicit" exact match systems. In implicit online catalogs, the query formulation requirements placed upon the user are greatly reduced or removed altogether. In the former case, the searcher is required to enter a term or terms that represent his information need and, perhaps, specify a type of search or search field by selecting it from a menu or by using a simplified command language (e.g., FIND TITLE medieval art). The system then supplies the combinatorial logic which specifies a relationship between the terms to be assumed and acted upon in the matching operation. Implicit truncation, for example, might also be applied to the terms such that a match could occur on both "medieval" and "medievalist," or "art" and "artists."

Such implicit online catalogs leave the user entirely in the dark about the term combinatorial logic, truncation (if any), and matching functions they automatically employ. As a consequence, most searchers will not have a clue as to why some searches fail to retrieve any documents, or why other searches retrieve large numbers of nonrelevant documents. Thus, they have no information feedback to aid in the modification or reformulation of their search queries for a second or third try. Even if they guess that the online catalog they are using searches on "medieval art" as a unitary string of contiguous characters, these implicit online catalogs generally do not provide the means for a searcher to respecify the request as, for example, "medieval AND art."

Another category of implicit online catalogs includes those that remove the requirement to formulate and enter a query altogether. Using these online catalogs, the searcher may optionally select a type of search from a menu (e.g., author, title, subject, etc.) or proceed directly to a display of index terms or brief document titles usually presented in an alphabetically ordered list. (Some online catalogs display title lists in class number order.) Markey (1989) calls this approach "alphabetical searching." This approach closely mimics the way searchers access and scan document records in the earlier manual card catalogs. Searchers choose a location in the displayed alphabetical list (or drawer of cards) of "headings" terms as an entry point to the database, then scan nearby terms or the bibliographic records filed under them. In the online catalog, a selection of a single term from the list (terms can be keywords extracted from text or pre-coordinated phrases from a controlled subject vocabulary) will typically call up a display of all bibliographic records associated with the selected term. These usually abbreviated document "title" records may, in turn, be scanned for further selection, fuller display, and assessment.

This list scanning and selection approach to searching, found in many online catalogs, is often named the "BROWSE" mode or searching option. The only search approach offered in a few online catalogs, in most second-generation online catalogs this approach is offered as a search option, along with a keyword, Boolean search option (explicit in some, implicit in others). Thus it is that we have identified three types of operational online catalogs: (a) explicit, exact match systems (usually Boolean and string searching systems); (b) implicit Boolean exact match systems (in which the system software defines the term relationships); and (c) "browse" online catalogs that feature alphabetical searching of index terms or citation lists.

The "browse" online catalogs make the least demands on the searcher with regard to the process of query formulation and entry. The searcher merely scans a list, selects a term from the list (rather than entering one of his own), then
sees what document records are retrieved. The searcher may have a term or terms in mind, of course, then consults the system's lists to find it or one like it in some sense and thus suitable for searching. When a term has been selected, the system carries out the "built-in" matching and retrieval operations. Such browse online catalogs may still be classified in the category of exact match systems.

In all three types of operational online catalogs—and most are mixed, hybrid systems—effective subject searching requires the user to express his need for information in a form or terminology acceptable to the system. This means that users must not only specify their need in advance, but think about what sort of documents will satisfy their need, and also translate these concepts into the terms used in the indexing vocabulary of the system. These terms may then be used in a formal query, if the particular system requires one, or sought for in an alphabetical list displayed for this purpose. The system then takes the query or selected term and applies a matching function to determine which records are to be retrieved for display and evaluation by the user.

Larson (1991) explains that the process of query formulation or term selection from lists required in conventional information retrieval systems and online catalogs "involves predicting which terms in the indexing language of the system have been used to index the documents that the user would want to retrieve" (p. 5). He goes on to state that evidence indicates that online catalog users do not conceive of subject searching in this way, and that when required to, they usually do not do a very good job of predicting or guessing the terms used to index the desired or potentially useful documents. Some of the guessing required may be reduced in systems that permit or require searchers to scan lists of index or thesaurus terms to identify search terms. However, in large online databases, the length of these lists, or the complex structure of lists such as thesauri, may place an unreasonable burden on the untrained, infrequent user.

BROWSING AS AN INFORMATION-SEEKING METHOD

Browsing is not one but many kinds of activities, any one of which may be observed in actual searching behavior. All information retrieval systems, including online catalogs, support some form of browsing. In traditional, query-oriented systems, browsing plays a subordinate, supporting role in assisting with the formulation or modification of a query that is to be matched exactly or partially with document representations. This probably explains why some people view browsing as a secondary activity and not as real searching. Some forms of browsing are quite different from this and may serve as the primary information-seeking method used by most people in real-life searching situations. In light of this, some researchers have suggested that a browsing paradigm for searching replace the query-matching paradigm in the design of information retrieval systems. Before commenting further on this point of view, it will be well to examine more closely the concept and types of browsing.

The Concept of Browsing

A browse is an edible in the eyes of a young animal. It may be a tender twig, leaf, or shoot of a plant that is fit and easy to eat. These delicacies must be sought for and are the object of selective review, that is, browsing. Browsing
takes place in a patch of interest and is characterized as tentative nibbling, at least at the start. Human browsing activity has many connotations. In the context of information seeking and library use activities, probably the most visible and commonly understood browsing activity is the behavior of roaming among the shelves of a library or bookstore to scan materials of potential interest or utility. Books and other materials are casually perused in order to decide what we want to buy or borrow, if anything at all. Librarians have long recognized that users who come into the library enjoy browsing among the shelves, and thus they make special efforts to display groups of related books of potential interest in noticeable, easy-to-browse ways. Research studies of library users confirm this experience and show further that many library browsers prefer to browse the organized materials on the shelves than search and browse in the library catalog (Hyman, 1971; Hancock-Beaulieu, 1989).

From our ordinary experiences, we recognize that both the focus of our browsing interests and the strength of our motivation to discover relevant items vary from time to time. When browsing, we may employ a variety of techniques ranging from the casual and undirected to the planned and systematic. As Marchionini (1987) explains, “These techniques are dependent on the object sought, individual searcher characteristics, the purpose of the search, and the setting and context for conducting the search. The objective of browsing may be well-defined (e.g., a particular antique chair to match a desk), or ill-defined (e.g., an interesting wall hanging for a favorite room)” (pp. 69-70). In the latter category, I prefer the example of a tourist on the last day of an island holiday searching about for a souvenir suitable as a memento of the trip.

Browsing can thus be viewed as a family of information-seeking activities. As Herner (1970) concludes, browsing is not one but many things:

> It is sometimes a purely random, unstructured, and undirected activity. Other times it is closely directed and structured, where, although the final sources or media may not be known, the desired product or goal is clear. Then again, specified media and sources may be browsed or consulted on a regular basis, not necessarily to produce answers to concrete queries, but because it is highly probable they contain items of interest. (p. 414)

**Browsing Aids**

This brief reflection on the varieties of browsing activities is useful because it invites us to expand our traditional understanding of browsing. Browsing may be more or less planned and directed, or it may proceed from an information need or interest that is more or less well defined at the start. In addition, browsing may be carried out in a variety of information media, packages, and bibliographic tools, both manual and online. Many of these media and tools have been systematically designed and structured to facilitate browsing. They employ structural, semantic, and navigational aids for this purpose. The library itself can be such a tool if its collection of materials is stored and maintained in any way other than a random manner. When direct access to the shelves is permitted, the arrangement of books on the shelves according to a subject scheme or some other classification (e.g., author, genre) facilitates browsing by library users.

A book or periodical journal is typically organized and structured to promote browsing. Such devices as the tables of contents, indexes, prefaces or introductions,
and lists of references both encourage and enhance browsing. Whatever the user’s level, specificity, or area of interest, such devices permit the easy and convenient gathering and perusal of information needed to make preliminary decisions about the relevance or potential usefulness of the documents.

Various forms of library catalogs, and indexing and abstracting publications or services, manual or online, incorporate devices and features that permit browsing of one kind or another. These sources utilize structure, recognition, and navigation devices to assist and guide the user looking about for items of interest or pointers to such items. Browsing is essentially visual and depends more on recognition than on recall or a priori formulations of need. A good browsing tool, source, or system exploits the human ability to recognize items of interest, a cognitive ability that is faster and easier than juggling concepts to specify a need and describing relevant items in advance (Card, Moran, & Newell, 1983).

Searching or Browsing?

Marchionini (1987) discusses three primary reasons why people browse:

First, they browse because they cannot or have not defined their search objective; they have what Belkin, Oddy, and Brooks have called anomalous states of knowledge (1982). . . .

Second, people browse because it takes less cognitive load to browse than it does to plan and conduct an analytical, optimized search. . . .

Third, people browse because the information system supports and encourages browsing. . . . Particular information sources like encyclopedias invite browsing by supplying indexes, outlines, section headings, tables and graphs, which help users quickly filter information. (p. 70)

Searchers often have difficulty defining and expressing their information needs. The database structure and vocabulary requirements of the search system may be unknown to the searcher. For such searchers, looking is more inviting than formulating. Browsing is inherently active and engaging, and many users seem to prefer action and encounter to reflection and analysis. It could be said that good browsing systems and sources attract such users, but there are not enough good online browsing systems in operation to justify this claim at this time.

Reflection on the reasons and circumstances in which people browse should yield a new understanding of the importance of this activity. These insights should inform the design of information retrieval systems and lead to improved browsing capabilities in these systems. In the past, browsing has often been viewed as a secondary or supplemental search strategy or technique to primary, query-oriented, directed, structured searching. Bates (1989) suggests that there may still be a “lingering tendency in information science to see browsing in contrast to directed searching, to see it as a casual, don’t-know-what-I-want behavior that one engages in separately from ‘regular’ searching” (p. 414).

Searching by browsing is a natural, preferred searching technique for many people, especially when they are engaged in “general purposive” information seeking. Ellis’s (1989) research on the information-seeking behavior of social scientists shows that various forms of browsing are a standard component of
their research and "keeping aware" activities. He recommends that browsing of a variety of types of information that supplement the standard bibliographic record be provided in online retrieval systems. Liebscher and Marchionini's (1988) research has demonstrated that browsing can be as effective in its results as structured, query-oriented Boolean searching for novice searchers of full-text documents. Marchionini (1987) argues that because of the massive amounts of poorly organized information available in electronic form, browsing is even more important in electronic environments than in traditional environments like those presented by open-access libraries.

There are a variety of information-seeking needs, aims, and strategies that would seem to require searching by semidirected exploration, recognition, and discovery, in a word—browsing—rather than searching by explicit query formulation-matching operations, whether aided or not by relevance feedback, query expansion techniques. Thus, it seems self-evident that users would greatly benefit from the development of computer-based information systems that support and encourage searching and exploration of electronic information resources via browsing or "berrypicking" (Bates, 1989).

Designers of information retrieval systems and online catalogs must expand their knowledge of the browsing requirements of searchers and provide capabilities and search options in their systems that will support these requirements. Most information retrieval systems support some aspects of browsing, but most still implement the paradigm of direct, query-matching retrieval. Browsing also provides a suitable paradigm for information system design and, perhaps, an even more representative one, given the many varieties of information needs and searching behavior.

CONCLUSION

Hundreds of OPACs can now be accessed with relative ease via the Internet, either by telnetting to an OPAC's network address or by selecting one from a Gopher directory menu. Almost all of these OPACs can be characterized as second-generation, query-oriented online catalog retrieval systems. They have very limited relevance feedback, query refinement, and browsing capabilities. As such, they place intrinsic limits on the potential of new network-based access and search technologies, like the Z39.50 protocol, to provide more effective and more appropriate search environments for many kinds of search needs and behavior.

The Z39.50 NISO standard search and retrieval protocol, and its International Standards Organization counterpart standards, ISO 10162 and ISO 10163, conform to the query-oriented paradigm embedded in so many second-generation online catalogs. These specifications have recently achieved draft international standard (DIS) status. The search and retrieval protocol is designed to function as an application-level (layer 7) protocol within the Open Systems Interconnection (OSI) protocol suite for the connection and interfunctioning of different computer systems. Among other things, the ISO search and retrieval protocol specifies a canonical search format through which searches can be transmitted from one computer (the "client") to another (the "server"). This
format consists of a series of predicates linked by Boolean operators such as OR and AND; the predicates are composed of field names, relational attributes, and values (for example, AUTHOR—lastname value; or TITLE—keyword-of value). Both the field names and relational attributes are selected from a predefined and registered (i.e., officially sanctioned) attribute set that forms part of the context of a connection between a "client" computer and a "server" computer. The current "working set" of attributes represent fields in MARC or MARC-like bibliographic records.

Although some uniformity will be introduced in the searching of networked online catalogs when Z39.50 and the ISO search and retrieval protocols are available for use (conforming software must be developed and installed on hosts and/or network servers), limited search and browse functionality will be supported by the standard approach, and no assistance to the user having search problems during a session will be provided by the new protocol-based search interaction. For example, this approach will not inform the user why a search resulted in no matches. The search can be repeated with new attributes and/or values, but it will be transmitted and processed in the same predefined and rigid manner. The assistance of a friendly local user interface will be excluded from this process.

WAIS, a distributed database retrieval system also based on the client-server model, provides a rudimentary kind of weighted-term/document retrieval from indexed databases on the Internet. WAIS displays retrieved documents in ranked order, with those documents most likely to be relevant to the query listed first, and WAIS offers some opportunity for relevance feedback on retrieved documents so that the search can be refined and extended. WAIS is an application of the probabilistic theory of information retrieval promoted for many years by researchers as a better alternative to conventional Boolean retrieval systems. However, WAIS is still a query-oriented approach that provides only limited, rather linear browsing capabilities.

A fundamental shortcoming of the client-server approach is that any client, no matter how usable, is limited by what the server search and retrieval "engine" can do.

The first network access technology with promise for inveterate browsers is the World Wide Web (WWW or W3). WWW supports hypertext retrieval and browsing among selected, specially organized databases on the network. Through the multiple linking of related data entities, textual units, or entire documents, the hypertext approach offers the user a network of alternate paths for self-directed, nonlinear browsing and exploration of bibliographic and other document spaces. At this time, there are few if any hypertext OPACs accessible via the Internet.

To end on an optimistic note, there is good reason to expect that these new network access and retrieval technologies will have an impact on OPAC vendors and developers, motivating at least some of them to venture beyond second-generation functionality. The Internet is marvelously hospitable to innovators, and it is a wonderfully public and influential medium.

REFERENCES


