

Subject Searching in Online Catalogs: Metaknowledge Used by Experienced Searchers

Tschera Harkness Connell

Kent State University, School of Library and Information Science, 124 Mount Hall, 1050 Carmack Road, Columbus, OH 43210

ABSTRACT: This paper begins to identify and characterize the knowledge used by experienced librarians while searching for subject information in online catalogs. Ten experienced librarians performed the same set of six subject searches in an online catalog. Investigated was the knowledge used to solve retrieval problems. This knowledge represents expertise in the use of the catalog. Data were collected through the use of think-aloud protocols, transaction logs, and structured interviews. Knowledge was defined as knowledge of objects (factual knowledge), knowledge of events (experiential knowledge), knowledge of performance (process knowledge), and metaknowledge. Metaknowledge is the sense of whole derived from the integration of factual, process, and experiential knowledge about the search and the conditions under which it is performed. The focus of this paper is on metaknowledge. For evidence of metaknowledge the data were examined for explanations that participants gave for their actions and observations, and for ways that participants evaluated their own progress during the process of searching. Reasons and explanations given by searchers were related to all phases of the library information retrieval process from the user's receipt of material to policies for collection development, and not just events directly related to the performance of a particular search task.

Introduction

While prior studies of online catalogs have investigated searching behavior of library users (Borgman, 1986; Hancock, 1987; Markey, 1983; Matthews, Lawrence, & Ferguson 1983), this article begins to identify and characterize the knowledge used by experienced librarians while subject searching in online catalogs. This was considered an important topic for investigation because online catalogs are difficult to use, and an understanding of the specialized knowledge used by librarians may facilitate the design of more usable systems.

Catalogs are Difficult to Use

There are a number of inherent characteristics of online catalogs that make them difficult tools to use, especially when someone is seeking subject information (Bates, 1972; Borgman, 1986; Connell, 1991; Matthews et al., 1983). In covering a library's collection, an online catalog includes information on a wide variety of disciplines and subject areas. This variety results in a vocabulary that is diffuse. The long history of library catalogs also affects vocabulary in that the records in online catalogs reflect not only current input standards, i.e., descriptive cataloging codes and subject cataloging traditions, but also standards devised for prior technologies and earlier times.

Indirectly, catalog history affects the catalog because library collections (and therefore library catalogs) represent the output of recorded knowledge throughout the ages. Library collections reflect evolving human views of knowledge—the changing paradigms (in the sense of Kuhn, 1970) of the subject disciplines themselves. This means that even if there were no inaccuracies and inconsistencies in indexing over the years, paradigm shifts in subject disciplines would create variations in the vocabulary (controlled and natural language) used. All these factors make the catalog a complex tool for information access. Yet, despite these complexities we expect inexperienced users to search them independently.

Many authors point out that serious problems can arise when a user does not have an accurate

understanding ("mental model") of the limits and capabilities of a system (Borgman, 1983; Daniels, 1986; Dayton, Lundeen, & Pollock, 1980; Doszkocs, 1986; Hildreth, 1989). The importance of the user's model is obvious when considering the scope of the database. It does not make sense for a user to seek information about dinosaurs in a database devoted to cancer treatment; or, to look for journal articles in most contemporary library catalogs. Less obvious but equally important is the user's model of how the information in the database is indexed. One indexing concept used in catalogs is "specific entry." Specific entry is the subject cataloging guideline to enter a work under the subject heading that is coextensive with its subject coverage as distinguished from the broader, or more general heading that includes the subject of the book. Therefore, a book about jazz will be assigned the subject heading "Jazz", not "Music." Bates (1977) has shown that the awareness of the principle of specific entry influences significantly the success of a search in an online catalog.

Specialist Knowledge

The more one knows about a system the more effective he or she will be in using the system. Does the specialist knowledge that librarians have about library catalogs guide the ways they search those catalogs? Librarians build and provide the structure for the records in library catalogs and yet little research has been done about how librarians use these catalogs. Use studies involving library catalogs have concentrated on the end user. In contrast, it is the skilled user, the intermediary, that has been the focus of studies of bibliographic retrieval systems other than library catalogs. Bates' list of search tactics (Bates, 1979, p. 207), Fidel's "Selection Routine," (Fidel, 1988, p. 26) and Harter and Peters' list of search heuristics (Harter & Peters, 1985, p. 407) are descriptions of procedures for searching information retrieval systems that have resulted from such studies. The research presented here begins to identify and characterize the knowledge used by experienced librarians while searching for subject information in online catalogs. In order to give users more assistance in the use of our catalogs and guidance in choosing alternative approaches to information, it is first necessary to understand which approaches are successful and in what circumstances. If we begin to understand the knowledge that is used by skilled searchers of online catalogs then perhaps we can build that knowledge into our systems so that the infrequent user can benefit from specialist knowledge without having to develop the expertise explicitly.

Method

Ten experienced academic librarians performed the same set of six subject searches in the Illinet Online catalog used by libraries throughout the state of Illinois. Because of their experience, these librarians were considered to be experts in the use of the library catalog. Participants were chosen from two academic institutions in the state. Both catalogers and reference librarians were chosen as participants because each perform different phases of the information retrieval process.

An effort was made to obtain participants who represented a range of years of experience. The years of experience for participants ranged from 12 to 33 years. Years of experience at their current institution ranged from 2 to 25 years. All catalogers had been employed full time as reference librarians for at least 2 years sometime during their career. One reference librarian had worked as a cataloger. The fact that some of the librarians had worked both as catalogers and as reference librarians was not seen as a disadvantage. The purpose of having both catalogers and reference librarians participate in the study was to provide the potential for variety in the knowledge used by the librarians in the study. Librarians who have worked both as catalogers and reference librarians may provide a different perspective than librarians who have had only one kind of experience.

All the participants were experienced with computers. Most had used the online catalog since it first became available at their institution. The shortest period of time listed by any participant for

using the system was 2 years. All participants had experience with other databases. Reference librarians used CD-ROMs for answering patron queries. Catalogers used the Online Computer Library Center (OCLC) for verifying and creating bibliographic records and in-house systems for administrative purposes.

The knowledge investigated was the knowledge used to solve retrieval problems, independent of the particular subject area of the inquiry. This knowledge represents expertise in the use of the catalog. Therefore, techniques of knowledge acquisition, used by knowledge engineers in building expert systems, were the techniques used for knowledge elicitation.

In an article on the quality of expertise, LaFrance (1989) observed that the major difference between expert knowledge and novice knowledge is not in quantity but in the structure and organization of the knowledge. Knowledge acquisition efforts must take into account not just what experts know, but how they organize what they know. Since any one knowledge elicitation methodology is not universally effective for all kinds of knowledge, using multiple methods to elicit knowledge from experts is more likely to result in an accurate and more complete representation of the expertise. Three methods were used to elicit knowledge from the librarians in the study; data were collected through the use of think-aloud protocols, transaction logs, and structured interviews.

Think-Aloud Protocols

The participants were instructed verbally and in writing to think aloud—to tell everything that passed through their minds as they thought about and performed each search. They were informed that their attempts and discarded ideas were as important as their final approach. These think-aloud protocols were audiotaped using a wireless lapel microphone clipped to the participant's clothing. This means of audiotaping enabled the participants to move around during the searching process and to use any auxiliary tools that they might normally use in conjunction with the online catalog. Typed, verbatim transcripts of the tapes were prepared.

Think-aloud protocols do not provide a complete record of the searcher's thoughts because not everything is verbalized. In fact, there is evidence that the greater the experience with a task, the more automatic the process of performing the task. This means that intermediate steps in the process are carried out without conscious thought and, therefore, without verbalization (Ericsson & Simon, 1984). This is a disadvantage for knowledge elicitation because much of the structure of the expert's knowledge is not verbalized. However, what is verbalized represents the participant's thinking at the time. Think-aloud protocols are effective for revealing the development and progression of the expert's solution to the task at hand (process knowledge).

Transaction Logs

Computer transaction logs recorded the participants' interaction with the online catalog. Illinet Online appears as a single system for the user but it is actually two separate systems: the Library Computer System (LCS) and the Full Bibliographic Record (FBR). LCS is a short record circulation system. FBR contains two files, the Bibliographic File (with complete cataloging records) and the Authority File. The commands entered by participants into both systems were gathered and merged in the proper sequence to create a single file. Activating this file from a personal computer recreated the entire search session, commands, and system responses (what the participant saw on the screen), into a single, complete, transaction log. Once the complete transaction log had been created and printed, the commands were numbered consecutively. The commands were then matched against the audiotape of the corresponding think-aloud protocol to determine when each command occurred. The occurrence of each command was noted on the typed transcript of the verbal protocol.

Transaction logs provide a bridge between a participant's think-aloud protocols and

accompanying actions. They add the potential for increased understanding of the think-aloud protocols because the researcher has access to both searcher keystrokes and system responses.

Structured Interviews

Comparative studies of knowledge elicitation techniques show that the structured interview performs well for eliciting structure and process (Breuker & Wielinga, 1987). The sources of questions for the structured interview were the verbal protocol transcript and the complete transaction log. Because of the time required to obtain the protocol transcript and to create a complete transaction log, the interviews took place 1 week after the initial search session.

The time lapse between the search session and the follow-up structured interview influenced the kinds of questions that could be asked. The time lapse was too great to obtain reliable data on the details of the participant's thinking behind the processes used, but clarification of broad process issues was possible. The structured interview concentrated on clarifying statements and actions made by participants that were not understood by the researcher, on establishing relationships among terms and processes used by the participant, and on the elaboration of allusions to previous search events. The structured interviews were audiotaped and transcribed for analysis.

Document sources were used for explanation and/or clarification of data gathered by think-aloud protocols, transaction logs, and structured interviews. Document sources included the *Library of Congress Subject Headings* (LCSH; 1989) the *LC Subject Cataloging Manual: Subject Headings* (SCM:SH; 1990), and descriptions of and instructions for using Illinet Online.

Search Tasks

In two searches the participants were asked to find material with similar subject coverage as the item described in an abstract. This kind of search task imitates a student's request for a book "like" a faculty-recommended book that is unavailable. The remaining four search tasks evolved from patron reference queries. The queries came from patrons' questions in two departmental libraries at the University of Illinois at Urbana-Champaign. The queries were in final negotiated form; i.e., the queries were in the form that the reference librarian searched after interviewing the patron. Five of the search tasks were selective: Participants were asked to find one or two good items that could help the patron answer the question asked. One was comprehensive: Participants were requested to find everything published during the last 5 years that was related to the topic.

Questions for searching were chosen from the disciplines of economics and psychology. Bates (1972), in her dissertation, observed that because of its geographical emphasis, economics was a more difficult field for students to match subject headings than was the field of psychology. Difficulties were particularly apparent in the application of the principle of specific entry for subject assignment. By comparison, psychology headings were straightforward. The two abstracts used in this study were taken from Bates' dissertation. Using questions that posed difficulties for students could reveal how experienced librarians use their expertise to solve problems. As well, it was assumed that economics and psychology, because of their relationships to many other disciplines, are two fields in which most librarians have had some experience searching.

Data gathering took place in the participants' own libraries. It was hoped that the research would take place in their own work area because then participants would have available any auxiliary tools that they might normally use in conjunction with the online catalog. Unfortunately, for a variety of reasons, catalogers were unable to perform the searches in their normal work areas.

Data Analysis

One useful way to approach the analysis of the data is to distinguish among kinds of knowledge to be elicited. Several authors, including Belkin and Vickery (1985), Fenley (1988), and Hayes-Roth, Waterman, and Lenat (1983), have provided descriptions of the knowledge needed for an expert system. Belkin and Vickery's typology, because of its explanatory detail, is the basis of the operational definition of knowledge used in this study. Knowledge was defined as knowledge of objects (facts), events (experience), performance (process knowledge), and metaknowledge.

The distinctions are not clear cut, as the various kinds of knowledge are interrelated. The purpose of distinguishing kinds of knowledge is to assist in the process of identifying different kinds of evidence of knowledge when examining verbal data.

The analytic process occurred in two main stages. The first stage of analysis involved extracting from the think-aloud protocols evidence of factual and experiential knowledge. Evidence for factual knowledge was primarily the presence of nouns and adjectives used by librarians during the process of searching. Evidence for experiential knowledge was explicit statements indicating that the participant was relating the present event to previous search experience.

The second stage of analysis involved more in-depth analysis of the transaction logs and think-aloud protocols for evidence of process and metaknowledge. Evidence for process knowledge came primarily from the transaction logs, although both the think-aloud protocols and structured interviews served as aids in interpretation. Analysis of the data for process knowledge focused on participant output in terms of a comparison between the comprehensive search task and the selective tasks and on participant search patterns. For metaknowledge, all data were examined for the reasons that participants gave for actions taken and observations made, and for evidence that participants evaluated their own performance during the process of searching. Factual knowledge, experiential knowledge, and process knowledge can be considered building blocks for metaknowledge. The remainder of this paper will focus on a discussion of evidence of metaknowledge.

Metaknowledge

Metaknowledge in a search task context is the sense of whole derived from the integration of factual, process, and experiential knowledge about the search and the conditions under which it is performed. Metaknowledge is fluid. It changes as the searcher considers more facts, analyzes more procedures, and makes more connections with past experience. Metaknowledge is revealed in the searcher's (1) understanding of relationships among terms, tools, and search tactics, (2) evaluation of progress, and (3) selection or rejection of records retrieved.

TABLE 1. Summary of the Illinet Online search commands found in the examples.

| FBR commands | Qualifier |
|---|--|
| <i>F</i> = FIND [in Bibliographic File] | <i>\$all</i> = all libraries on the system. |
| <i>S</i> = SUBJECT [followed by a subject heading] | [Without a qualifier the default is determined by the library location of the terminal used] |
| <i>T</i> = TITLE [followed by one or more title keywords] | |
| Boolean operators | |
| <i>.a.</i> = and | |
| <i>.o.</i> = or | |

There may be more than one solution to a search task problem. The solution chosen for the search task as a whole, or for a mid-search problem, will vary according to the particular combination of factual, process, and experiential knowledge of the searcher. For example, when no items are retrieved from the catalog, a searcher may search more broadly. If one searches more broadly or how one searches more broadly will depend upon the searcher's metaknowledge of the situation. If the patron is a freshman working on a required English essay, the searcher may conceptually broaden the topic by choosing a broader controlled vocabulary term in the local catalog. However, if the patron is a graduate student, the searcher may first try another member library in the system that the searcher knows from experience has a stronger collection on the topic requested. Once the decision for broadening the search has been made, the searcher decides which search tactic should be used. The decision will be made from a combination of factual, process, and experiential knowledge about conceptual options, strengths, and weaknesses of a controlled vocabulary, system capabilities, searcher experience with those capabilities, etc.

Excerpts of two search tasks (one patron query and one abstract) will be used to illustrate participants' use of metaknowledge. These two tasks were chosen for illustration because they produced widely divergent data. Together they illustrate the richness and variety of metaknowledge used by the participants. Both of these search tasks were selective: Participants were asked to identify one or two good items in the online catalog that could help the patron answer the questions asked. After a discussion of the two tasks, the evidence will be presented in the form of metaknowledge "rules" relating to the nature of the search task and to the progress of the search. In the results section, metaknowledge is further summarized in terms of the interrelatedness of the search task to the context in which it is performed. The command syntax used in the examples is the syntax used by the participants and is presented in italics in Table 1.

Search Task: Dreams

How Do Dreams and Daydreams Affect Creativity? This task provides an opportunity for searchers to use a variety of means for handling vocabulary issues. The task was included in the study because the relationship of dreams to creativity and daydreams to creativity seemed to suggest the use of the Boolean operator *and* to combine the two pairs of concepts. All three participants in a pilot study performed explicit Boolean searches for this task; i.e., they used Boolean operators *and* and *or* in their search argument. However, only 4 of the 10 participants in the final study performed explicit Boolean searches. Three other participants used implicit Boolean searches by using multiple terms in a title search.¹

Explicit Boolean Searches. The explicit Boolean searches revealed in the think-aloud protocols of two participants are good illustrations of participants evaluating the search process and progress. One example also illustrates the inefficiencies that can occur when a searcher has

¹ Title searches on Illinet Online use an implicit Boolean *and* between terms.

an incorrect mental model of how the system handles the Boolean search.

Participant F began this search with a *F T dreams .o. T daydreams .a. T creativity*, but the search retrieved more records than expected (132) because the participant failed to recognize that the system processes the Boolean operator *and* before the operator *or* which makes this, in essence, a search for the title word "dreams." The participant tried unsuccessfully to refine the search and was still unable to understand the problem encountered. Participant F was incredulous. During the interview she elaborated.

Dreams or daydreams and creativity. I got 132 items. I did find title dreams and title creativity and I got zero. I did daydreams and creativity and got zero. I don't see how I did that. . . you can see how I kept doing it and I don't see how I got it. . . . On the face of it, it looks impossible.

Later, Participant F tried the same search with alternative vocabulary, substituting the subject heading Creative ability for the keyword term creativity (*F T dreams .o. T daydreams .a. S creative ability \$ail*). The system responded that the search is "too general" (more items retrieved than the system will display). Again, Participant F misunderstood the search algorithm. She assumed the problem was the subject Creative ability. But in order to test that assumption (evaluate the process) she entered the search *F S creative ability \$ail* and concluded, "No, there's 664 of those" (Table 2).

The misinterpretation of the machine process continued to plague Participant F during this search task. She finally performed an implicit Boolean subject search in all libraries in the system for the subjects "Dreams" and "Creative ability" (*F S dreams—creative ability \$ail*).² The participant was satisfied with the two items retrieved.

TABLE 2. Summary of Participant F's Boolean searches used for dreams.

| Search argument | System response |
|--|-----------------|
| <i>F T dreams .o. T daydreams .a. T creativity</i> | Yielded 132 |
| <i>F T dreams .a. T creativity</i> | Yielded 0 |
| <i>F T daydreams .a. T creativity</i> | Yielded 0 |
| <i>F T dreams .o. T daydreams .a. S creative ability \$ail</i> | |
| <i>F S creative ability \$ail</i> | Too general |
| <i>F S dreams—creative ability \$ail</i> | Yielded 664 |
| | Yielded 2 |

Participant F's use of the implicit Boolean subject search is evidence that she understood the indexing algorithm used by Illinet Online for subject headings, and could therefore use it to advantage with the Library of Congress subject heading system. Participant F was the only participant who made deliberate use of this system feature in searching. Others reacted to it when they remarked that they did not like records retrieved from mixing and matching main headings and subdivisions, but Participant F was the only participant consciously to use the feature to

² In Illinet Online each subfield of the MARC subject fields is indexed separately. If the subject indexing policy is known, it is faster to enter *F S dreams—creative ability* (an implicit Boolean formulation in which each component is searched as a subject heading) than it is to enter the Boolean formulation *F S dreams .a. S creative ability*. This approach is contrary to conventional thinking because "Dreams—Creative ability" is nota valid LCSH.

advantage.

In the second example of explicit Boolean searchers, Participant H began the search on the subject "Dreams" in University Y's database and retrieved 266 items. Participant H then added the term "Creativity" using an explicit *and*, and retrieved "zero bibliographic items." In evaluating the system response, she reasoned:

Since "Dreams" and the subject heading "Creativity" didn't come up with anything . . . I'm going to look up [in LCSH] the subject heading for "Creativity", which obviously isn't right, and do another *and* search. [Reading from LCSH:] Creation, creative ability. . . . Creative ability, ok. Find subject "Creative ability" plus "Dreams."

The revised local search retrieved the same two items that Participant F retrieved in all libraries.

How did the Participant H know that Creativity was not "right"? Her assessment may have been an educated guess. She knew that dreams was an authorized heading; she had just retrieved 266 items using "Dreams." However, because the system only reports the hits for a complete search statement and not for each component, "zero bibliographic items" could mean that University Library Y has no items with both the heading "Dreams" and the heading "Creativity" assigned. In fact, this revised search argument (using "Creative ability") in University X's database retrieved no items. However, as her think-aloud protocol illustrates, her evaluation of the process led her to conclude that her vocabulary was incorrect. By consulting LCSH, Participant H was able to check her initial assessment. LCSH gives a cross reference from "Creativity" to "Creative ability" which led Participant H to revise the search argument. If, at this point, there had still been no items retrieved, then Participant H would have known that University Y had no materials with those two headings assigned.

Implicit Boolean Searches. Three participants who approached the search task by title (implicit Boolean searching) performed very similar searches. Their initial attempts using keyword title searches resulted in "zero bibliographic items" in University X's database. The participants then tried a title search on just the single word "dreams." None were eager to browse through the 132 items retrieved. One participant consulted LCSH. Another tried the term "daydreams" (retrieved one item) and the term "creativity" (retrieved 133 items), before going back to "dreams" and browsing. The third participant noticed that several records used the subject heading "dreams." Searching on the subject heading "dreams" produced "66 items which is half of the title listings."

For this search task in University X's database, searching on a single heading or a single keyword term was the only approach that had the potential of working. None of the combination of terms discussed above (subject headings or title keywords) retrieved any items in the University X database.

This search example also provides an illustration of different decisions made by two participants in almost identical circumstances. For these two participants the difference lies in the searchers' overall experience and their familiarity with LCSH. Both are reference librarians; however, one had also served as a professional cataloger for many years. At University X, LCSH was not located at the reference desk. One of the two participants never used it any of the six searches performed. The participant with cataloging experience would "start with a keyword search as a shortcut," then switch to LCSH if the shortcut did not work. The inconvenient location of LCSH, on the one hand, and one participant's extensive familiarity with the tool, on the other, contributed to different metaknowledge for the two participants.

Participant J's use of an implicit Boolean search is uncharacteristic of his normal search patterns; yet, for the "Dreams" search task he ended up using a title keyword search. Participant J began this search by consulting LCSH and discovered that the subject heading for creativity is "Creative ability." He searched the subject "Creative ability—Psychological aspects" and retrieved seven items. He remarked that the items might be useful but that they did not say anything about dreams.

Ok. One shortcut. Instead of doing a subject search, I'm going to try a title/keyword search using creativity and dreams. See if there are any books with both those words in the title. ... Ah hah. So there is a book. We got a book using a title/keyword search on this one. ... Ok. Let me try Creativity and daydreams. Ok, got a big zero on that one. Let me try it with fantasy. Ah hah. There is a book. So using title/keyword I got two books. ... [That's all I need.] Title/keyword worked a lot better on that one.

It did work for this participant, but he was working in University Y's database. As was noted above, these same search arguments retrieved no items in University X's database.

The title keyword searches just described are the only three title searches performed by Participant J in all six search tasks. When asked during the interview to extrapolate a general rule of when to use a "find title" for keywords, and when to use a subject search, he replied:

For one thing it wasn't a comprehensive search, and there was ... no single subject term or subject string that I could create using those two words. I figured using a keyword search I could surely come up with two or three books that had "creativity" and "dreams" in the titles and there would be a real good chance that these books would be about creativity and dreams.

On a comprehensive search I'd probably do [a subject search] on "Dreams" figuring that there might be some books that don't have "creativity" and "dreams" in the title, even though the book itself might include creativity and dreams—because the title does not always describe the whole subject. ... Sometimes the title is very misleading.

Participant J's knowledge of LCSH and of keyword searching placed in context of a selective search resulted in metaknowledge that caused him to search differently than he had for other tasks.

Output. The participants' use of metaknowledge is also evident in the explanations given for choosing or rejecting records found. For the Dreams search task, explanations could be identified for two cases when a record was chosen by one participant yet rejected by another. Examination of these cases gives some insight into the differences in metaknowledge of the participants. In the first case both participants searched *FS dreams* and initially chose Cartwright, *A Primer on Sleep and Dreaming* (1978) from the abbreviated display. Participant A considered the patron to be an undergraduate working on a class assignment and liked this item because it would offer an introduction to the topic. Participant C examined the full bibliographic record then rejected the item because the subdivision used with "Dreams" was "Physiological aspects." She reasoned that the patron was more interested in the psychological aspects than the physiological aspects of the topic. Subject headings (i.e., factual knowledge) identified through the use of an additional search command (i.e., process knowledge) placed in the context of participant C's understanding of the patron and the topic resulted in a change of metaknowledge for C and, in this case, she

changed her mind about the appropriateness of the item. The other item both selected and rejected was Moss *Dreams, Images and Fantasy: A Semantic Differential Casebook* (1970). One participant rejected it outright saying that it had nothing to do with creativity. The other was not so sure. This participant examined the full record, discovered that in addition to "Dreams," "Hypnotism," and "Psychotherapy" were subject headings assigned. The participant suggested the patron examine the index of the book to see if creativity is covered. The recommendation for this item was weak, but the participant was not pleased with how the search was progressing, and mentioned that she felt she had spent too much time on the search: "I feel as if other patrons are waiting."

Search Task: Southwest

For the following search task the searcher was instructed to find another book or books on the same subjects) of the abstract.

California and the Southwest. Thirty-two sociologists, each writing separate chapters, have produced this economic and cultural geography of California, Utah, Arizona, and Nevada treated as a region. The book is profusely illustrated and includes discussion of the physical characteristics of the region, patterns of settlement, productive activities, transportation and trade, population, and certain aspects of its political geography.

One difficulty with this search task is the problem of matching the scope of the abstract to the scope of various LCSH headings that can be used for this geographical area. LCSH offers at least two choices for the geographical area described in the abstract. One participant performed the search task by using each topical subdivision he had chosen, with both headings: "Southwest, New" and "Southwestern states." In the interview, he explained that:

You have to use both because some cataloger might have put either one down. It seems that both of them are just about right. I got up to look at the list to see exactly which [states] each one covered. But they seem pretty much synonymous.

"Southwest, New" is the more specific heading and probably the more appropriate for the region described in the abstract. A listing of states in a scope note for this heading includes California; however, many people do not visualize California as part of the Southwest. Some participants searched for the very broad heading "The West (U.S.)." Some participants also searched for "California" and "Southwest, New" separately. One participant even combined the terms "California" and "Southwest, New" in a Boolean *or* search. Those participants who used the bibliographic record for the title provided knew that the original book was assigned the geographic area "Southwest, New."

Using the bibliographic record of the original book is an efficient way of identifying an applicable subject heading. Six of the 10 participants tried this approach near the beginning of their search, yet initially, only 2 found a record for the item. Three of the six searched outside their own institutions assuming a larger pool of records would increase the chances of finding the item. Searching all institutions on the system retrieved 55 records. Those that missed the original book handled the list of 55 in different ways. One participant selected several items and stopped before coming to the original title. Another examined 10 records but did not want to take the time to look at 45 more. Still another stopped because the items examined did not seem to have a sociological approach.

The most difficult part about this search task is that the topic is diffuse. Participants either had to search for very general materials about the region or had to make decisions on the relative importance of the many topics mentioned in the abstract and then limit the search to those topics judged most important. Participants approached the topic from cultural, political, and historical perspectives. Participant also emphasized general economic and social conditions, the physical characteristics of the region, and population statistics. Two participants looked for books that were extensively illustrated. One looked for proceedings of regional conferences, arguing that that was one way to find such diversity within the covers of one book.

Formulating headings for some of these topics can be confusing. For example, at the time of this study was performed, the Library of Congress subject heading for cultural geography was "Anthropo-geography." The Library of Congress subject heading for political geography is "Geography, Political." The heading for economic geography is "Geography, Economic," but according to the scope note in LCSH, the heading is used for works on a branch of geography, not for works on the economic geography of particular countries, or states.

For this search task the appropriate subject heading is "Southwest, New—Economic conditions." Without careful consultation of LCSH or an authority file, it is doubtful that a user would formulate headings for these three concepts (i.e., cultural geography, economic geography, and political geography) correctly. None of the headings can be used as a pattern for another.

Four participants did use "Southwest, New—Economic conditions" at some point during the search. Both Participants A and E used it at the end of their searches. Participant A in particular had a lot of output with which to compare by the time he used this heading; he had entered 46 commands in searching this question prior to using this heading. Participant E had used far fewer commands but liked the items retrieved and did not think other approaches warranted the time. Both Participants A and E searched locally stating that the patron would not want to use interlibrary loan. Participant I liked what was found locally but decided to see whether the largest institution in the state could offer more. Participant J wanted to try another aspect (another subdivision) of the topic.

A very efficient way to approach this task is to search by classification number. Participant H searched by inputting search commands using Dewey Decimal classification numbers. One of the pilot study participants also took this approach. Participant H began the search by inputting "the historical number for the Far West, which is 979 period." Later she performed a classification search "by geography just to see, because that would tend to be more illustrated. . . ." She accomplished her search by searching in the 330s (Economics) and adding the historical number for the region (330.979). She selected items retrieved by both the historical and the geographical approaches to the topic.

Output. The data for the Southwest search task reveal six cases where participants evaluated the same record but disagreed on whether to accept or reject it. Reasons given by participants for accepting items mostly had to do with the participants' assessments of the topic. Items were too restricted or too broad, lacked a sociological perspective, had too much of a travel slant, or needed text not just pictures. One item was rejected because it was too old: "Patrons are not interested in looking at material that is over fifty years old." The participant who chose this item acknowledged its age but thought that since it was an early planning document for the region it might be interesting to look at. Reasons related to topic were usually given for accepting records as well. Five participants accepted records because the title looked good and they had the right subject heading (Southwest, New—Economic conditions). Several chose items on the basis of a good geographical match. One item was chosen on the basis that it is "profusely illustrated."

However, while time pressure was never used as a reason to reject an item, in several instances items were chosen to get someone started or because it was the end of a long search and the records chosen were the best seen so far. Participants pointed out that they do not normally have the "luxury of this much time to spend with a patron."

Metaknowledge Rules

As was discussed above, metaknowledge includes an assessment of potential approaches to finding information in light of the overall conditions of the search (contextual factors). For evidence of metaknowledge the data were examined for explanations that participants gave for their own actions and observations and for ways that participants evaluated their own progress during the process of searching. Evidence is presented below in the form of metaknowledge "rules" relating to the nature of the search task and to the progress of the search. The rules have flexibility by providing several approaches to a problem. They illustrate how participants responded to the conditions (contextual factors) of each search task. The metaknowledge rules are further amplified by examples of choices participants made in the Dreams and/or the Southwest search tasks.

Partial List of Metaknowledge Conditions/Rules from the Dreams and Southwest Search Tasks

Nature of the Search Task.

When multiple topics are part of the question, combine two or more topics in a single search.

When multiple topics are part of the question, cover all topics.

When multiple topics are part of the question, search the one considered most important.

Contextual factors: Patron goals and needs; collection strengths and weaknesses; arrangement of the collection; availability of tools; strengths and weaknesses of tools; searcher comfort and familiarity with tools; ease of use of tools; time pressure. Examples: Is the patron an undergraduate student? Is there a facet of the topic, or a perspective that the patron thinks should be emphasized? Is the collection stronger for one perspective than another? Are materials for the topic shelved together? Is Boolean searching available on the system? Is LCSH nearby? Does LCSH have any precoordinated headings that combine the topics? How comfortable is the searcher with Boolean searching? How familiar is the searcher with LC headings and structure? Is searching by classification a viable approach to the topic? How much time is available to perform the search?

For the Dreams search task there were several ways that participants combined two or more topics in a single search. Participant F attempted a complex Boolean search argument using keywords. Later, this same participant used an implicit Boolean search on two Library of Congress subject headings. Several participants used title keyword searching. One participant used precoordinated Library of Congress subject headings (Dreams in art; Dreams in literature).

For the Southwest search task, those participants who chose to look for general materials covering the topic did so because they thought that the original book was an overview and that the student might benefit from an introduction. Participants who took this approach looked for primers, encyclopedias, or general bibliographies that the student could use as a guide to other materials. The participants who searched the part of the topic judged most important often selected different "important" points. The book described in the abstract covered many topics. One participant looked for proceedings of conferences as a way to replicate the diversity.

Another looked for books with a geographical emphasis, believing geographical books were likely to be illustrated. Another who took the geographical approach performed a shelf position search using the Dewey Decimal classification number for the Southwest because of the geographical emphasis in the structure of Dewey.

When looking for only one or two items, design the search without concern for recall.

When looking for only one or two items, search specifically.

Contextual factors: Patron status and goals; time available; availability of tools; strengths and weaknesses of tools. Examples: Is the patron an undergraduate working on a class assignment? Would the patron benefit from an introduction to the topic? How much time is available to work on this question? Does LCSH have a specific heading that is coextensive with the topic? Is Boolean searching supported by the system?

For the Dreams search task, Participant J performed title keyword searches on pairs of terms mentioned in the query. His reason for title searching was that he needed only a couple of good items; it did not matter if he missed some items along the way. Participant E noted that the goal was to get in, find something helpful, and get out. Keyword searching on specific terms or combining subject headings sometimes works for finding something quick. She noted that she often worked with 10 to 15 patrons in a 45-minute time period.

Progress of Search

When large numbers of items are retrieved, examine output.

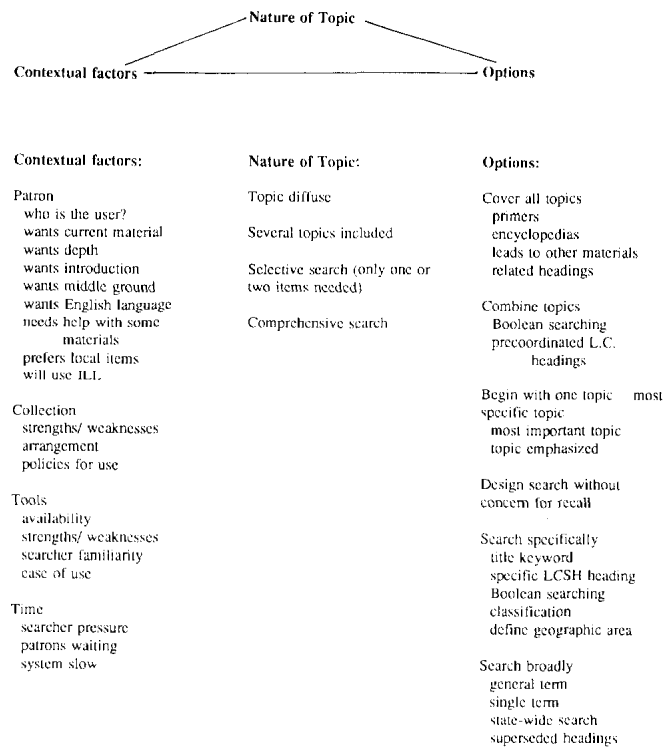
When large numbers of items are retrieved, narrow the search.

When large numbers of items are retrieved, modify the topic.

Contextual factors: Patron goals; familiarity with tools. Examples: Is the patron an undergraduate needing only a couple of items? Would the patron be as happy with a slightly different perspective on the topic? Do large sets indicate that the search vocabulary was too general? Is controlled vocabulary term likely to produce fewer items? Are there applicable narrower terms available?

For the Dreams search task, one participant switched from a title search on "dreams" to a controlled vocabulary search on "Dreams," reducing the number of items retrieved from 132 to 66. Rather than browse through the 132 items retrieved for a title search on "dreams," several participants modified their search topic and looked for material on "daydreams." Others selected from the large set or used Boolean searching to narrow the search.

FIG. 1. Summary of metaknowledge relationships regarding search task topic.



Similar decisions were made when participants searched the Southwest search task. Several participants examined large sets to determine whether the search argument they used produced the results they wanted. Some chose from the records examined. Others narrowed search arguments by adding terms in a keyword search, or by adding a LC subdivision to a subject heading search. Some participants modified the topic by selecting a narrower facet to emphasize.

When no items are retrieved, broaden the search.

Contextual factors: Patron goals and needs; collection strengths and weaknesses; familiarity with tools. Examples: Is the patron a doctoral student working on a dissertation? Can the patron wait for interlibrary loan? Does the local collection have material on the topic? Are other collections more likely to have material on the topic? Are there broader terms that can be used?

Participants tended to view the patron as an undergraduate and the topic of Dreams and creativity as one that someone would likely have written about. One participant commented that she would never send the patron away with nothing. The first inclination was not to rely upon interlibrary loan, but participants from University X, especially, searched outside their local collection when they received no hits for a specifically designed Boolean search. They knew the topic was narrow, and their collection was relatively small. Several participants searched on two keywords in a title search, for example, "dreams" and "creativity." When no items were retrieved, they dropped a term from the search argument.

When search argument produces unexpected results, test the query and/or its component parts.

Contextual factors: Familiarity with collection; familiarity with tools. Examples: Does the collection have materials on the topic? Is the term used likely to produce a match? Is LCSH likely to have a subject heading covering the topic?

Most participants checked spelling when zero items were retrieved. When Participant F received the response of "zero bibliographic items" from a Boolean search, she implemented this rule by searching each term in the Boolean expression separately. Another participant verified a subject heading in LCSH.

When uncertain, seek help.

Contextual factors: Patron needs; availability of tools; familiarity with tools; availability of colleagues. Examples: Is the patron available to give more insight to the topic? If terminology is a problem, is LCSH available? Is there an authority file? Are there some subject specific indexes or bibliographies that can be used to suggest terms? Is there a colleague available to help?

Repeatedly, participants indicated that if the patron were present, they would ask the patron for additional input. With the Southwest search task, terminology was a major problem. Several participants implemented this rule by consulting LCSH to determine the scope of the headings "Southwest, New" and "Southwestern states." The geographical areas covered by these headings are very similar. Some of the participants chose to search on both headings. One participant used the Public Affairs Information Service Bulletin (PAIS) for ideas. Another participant mentioned that he would like to compare notes with a colleague.

FIG. 2. Summary of metaknowledge relationships regarding progress of search task.

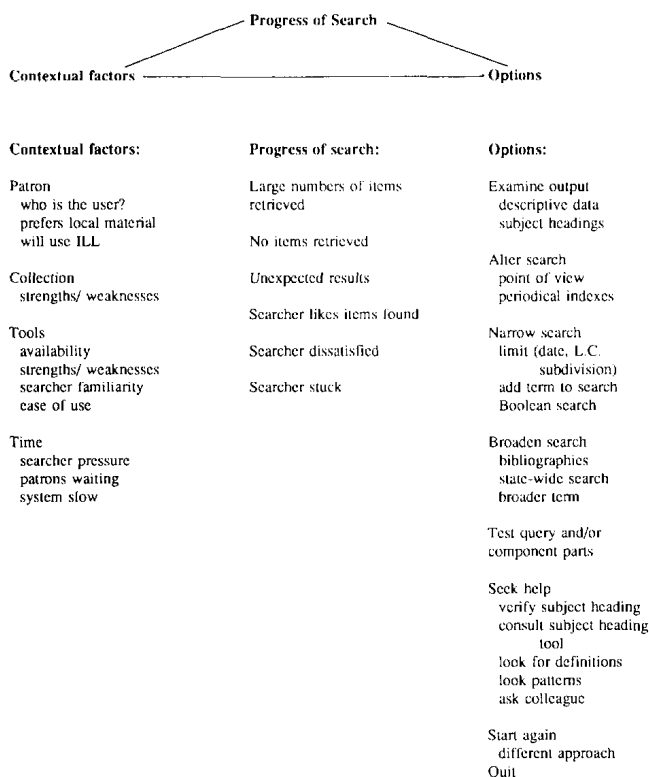
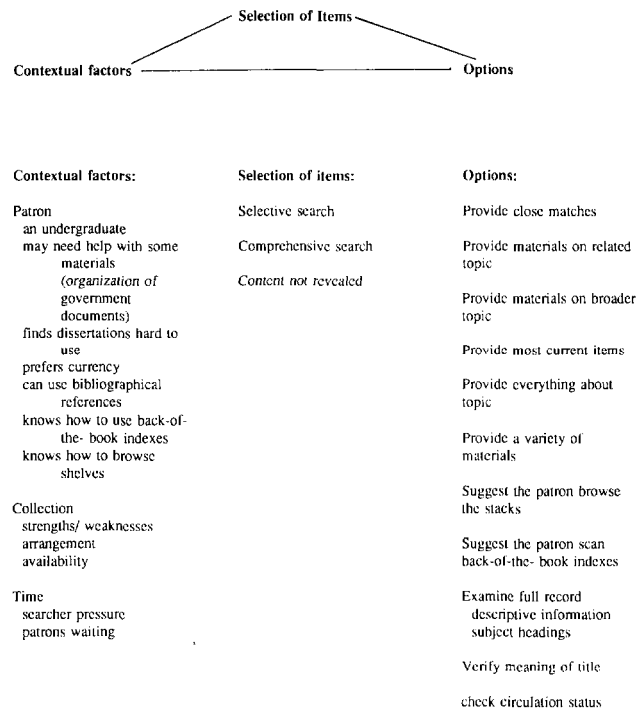


FIG. 3. Summary of metaknowledge relationships regarding selection of items.



Results

For evidence of metaknowledge, the data were examined for explanations that participants gave for their own actions and observations, and for ways that participants evaluated their own progress during the process of searching. The evidence has been presented in the form of metaknowledge "rules" relating to the nature of the search task and to the progress of the search. The metaknowledge rules were further amplified by examples, taken from the Dreams and Southwest search tasks, of choices participants made in implementing the rules. Reasons and explanations given by searchers were related to all phases of the library information retrieval process from policies for collection development to the user's receipt of material, and not just events directly related to the performance of a particular search task. Figures 1-3 show the interrelatedness of contextual factors and options as they relate to the nature of the topic, the progress of the search task, and the selection of items. Each figure includes a summary of the relationships mentioned as participants developed their metaknowledge of the situation. The figures include data from all six search tasks used in the study.

Discussion

The description of subject searching derived from observation of participants in this study has a number of implications for online catalog design, and for library and information science education.

Implications for Online Catalog Design

The breadth of knowledge used in subject searching is an important consideration in constructing aids to assist users in seeking information in the online catalog. Searching is

performed in a context. Systems that isolate the search task and concentrate on mechanical aspects of searching (e.g., automatic truncation, spelling checkers), to the exclusion of intellectual aspects (e.g., structural relationships among records, controlled vocabulary maintenance), will mask important connections to the larger picture of information production, storage, retrieval, and dissemination. Mechanical solutions are important because they free the searcher to concentrate on the intellectual aspects of the search. However, it is also important that a relational structure be present in the catalog so that the user can exploit the materials represented in the database. Participants in this study formulated search arguments in a broad context including, for example, what they knew about the kinds and quantities of information likely to be published on the topic or the holdings of the local collection. Systems that ignore the larger contexts are likely to lead the user to simplistic and possibly unsatisfactory solutions.

Because of the complexity of the online catalog, it is difficult even for experts to remember everything that may be useful to a search at any given time. There may be ways to trigger the knowledge of experienced users and offer assistance to inexperienced users. For example, if the system had offered a controlled vocabulary term after participants completed a keyword search for dreams, it might have reminded the librarians to consider controlled vocabulary. Another example would be context sensitive "help" that could provide a conceptual diagram of a user's Boolean search. Had this feature been available to Participant F during the search task "Dreams," she might have recognized the error in the logic of her search argument. Inexperienced users do not often use "help," possibly because they do not understand how "help" can be useful to them. The participants in this study repeatedly expressed a need for more information—most frequently in the areas of controlled vocabulary and system syntax. Most sought help from outside printed sources at some point during the search tasks. They knew what information they wanted and sought to find it. It would be more efficient if they could find it while online.

The online catalog used in this study was not regarded by most participants as a stand alone tool. Eight of the 10 librarians in the study consulted auxiliary tool in order to search the catalog. LCSH was the auxiliary tool most frequently used. However, LCSH was often inadequate for answering participants' questions. Participants needed more definitions, more explanations of LC assignment policy, and clearer delineations of term relationships and hierarchical structures. More definitions and more explanations could be provided if LCSH incorporated some of the information found in the SCM: SH into scope notes. However, clearer delineations of term relationships and hierarchical structures would involve major revisions to LCSH.

A well-designed authority file which included scope notes on usage, as well as relational headings, might replace the need for using a separate tool such as LCSH. However, the participants were unanimous in their dissatisfaction with both the content and design of the Authority File on Illinet Online. One of the most frequent criticisms of the online file was that the only way to maneuver through the file was by browsing in a linear fashion through the alphabet. Even though LCSH is in book form with entries arranged alphabetically, its layout enables the user to glance quickly at two pages and get a sense of relationships among headings in general alphabetical proximity. The use of windows in an online authority file might be a way to enable the user to make spatial, rather than linear and alphabetical, connections among headings.

The suggestions made here build upon the expertise of the knowledgeable user: helping him or her use the knowledge already possessed. The metaknowledge rules derived from experts in this study also have the potential to guide inexperienced users toward solutions of searching problems. While the experienced user may only need a reminder (as in the case of using controlled vocabulary above), the inexperienced user can be guided toward the concept and the use of a controlled term. For example, if the user chooses to use a keyword title search, a

message such as "You have retrieved 132 items with this search of words in titles. 'Dreams' used as a subject heading is a close match to your term. A subject search on 'Dreams' reduces the number of items retrieved to 66. Do you want to try this search?" Although the system could perform the narrower search automatically, there is likely to be some confusion unless an explanation is provided for the user.

Implications for Library and Information Science Education

Participants in this study formulated search arguments in a broad context including, for example, what they knew about the kinds and quantities of information likely to be published on the topic or the holdings of the local library's collection. Recognizing under what conditions such knowledge is effective can influence how we teach both cataloging and reference. The knowledge that proves to be especially useful in solving search problems can be emphasized.

Making the Catalog. Frequently, cataloging is only thought of as providing individual document representations. This study illustrates the importance of emphasizing that cataloging also involves building a structured tool by providing links between documents and showing the relationships among the documents of the collection. To go beyond individual document representation may involve verifying the accuracy and consistency of headings that have been formulated, providing scope notes to define a particular topic, history notes to provide context for a change in terminology, and/or references to link related headings. As frequently as these aids proved useful to the participants in the study, their lack was also noticed. The catalogers' desire to consult the SCM:SH was one way of expressing a need for more information. Several participants, both catalogers and reference librarians, commented that they wished they could find a scope note in LCSH.

Rather than creating the records in-house, most libraries take their cataloging data from bibliographic utilities such as the OCLC. Cooperative cataloging efforts provide basic descriptions of much of the published literature in library collections. However, many of the structural elements (scope notes, history notes, references) of the catalog are not the kinds of data that can be obtained in prepackaged form. Building a structure of a catalog is dependent on the local library collection. For example, providing a cross reference from "daydreams" to "day dreams" makes no sense if there are no materials in the collection to which "day dreams" has been assigned. The same is true for scope notes and history notes. Loading references and lists of headings in catalogs that are not used locally causes confusion for users who do not understand the system. Teaching librarians how to provide the structure among the bibliographic records in the database is a very important part of cataloging instruction.

Using the Catalog. The difficulties that experienced librarians had with using the catalog in this study suggest that instruction for librarians in the use of the catalog would be helpful. Most library school programs offer opportunities to learn how to search online bibliographic databases available on commercial services; yet, in many ways, the library online catalog is more complex than these tools. In an online library catalog, the problems of vocabulary caused by covering many disciplines and subject areas, and by having records produced by a variety of input standards over many years, magnify what may be only minor problems in newer more discipline-specific databases. The frequency with which some of the participants of this study consulted LCSH to plan their search strategies suggests that knowledge of LCSH would be useful. Students in library schools frequently learn to consult thesauri for searching specialized databases. Students could also learn to consult LCSH and the SCM:SH for searching in the online catalog. For those students who elect not to take more than an introductory course in cataloging, the design of LCSH and the structure of the catalog

may not be familiar. As well, students may not be aware of varieties in system design and how the designs could affect their searches. Students could learn how to take advantage of various system design features in planning a search. The experienced librarians in this study used a variety of techniques and knowledge while subject searching in the online catalog. Once the metaknowledge rules are tested and refined, the relative probability of their success could be taught as a means to approach particular searching problems. There is evidence that even when a system is user-friendly, it is necessary for librarians to assist end-users (King, 1991). Learning to use the library catalog is a very important part of reference instruction.

Conclusions

The breadth of knowledge used by just 10 participants indicates that it is important not to define the task of "subject searching" too narrowly. Much more needs to be known about what knowledge is used and what knowledge is effective when subject searching in online catalogs. So as not to be constrained by existing systems, it would also be useful to know what additional knowledge experts would find helpful while performing a search.

During the planning stages of this study, the institution where the searches were to be performed was not considered a significant factor in the design. All of the institutions considered for the study used the same online catalog. However, the librarians who participated in the study were noticeably affected by their environment. The reference librarians, in particular, were conditioned by a very busy reference environment. Participants from University X were influenced by a relatively small collection. Because of the importance of context to the performance of the search task, data need to be gathered from librarians in a variety of settings.

The metaknowledge rules identified in this study need amplification. Are the rules helpful in thinking through the issues that users encounter when subject searching? Under what conditions are the rules applicable? The situations where the different contextual assessments led to different solutions provide good starting points for comparing and evaluating the different options. What factual and process knowledge is most useful for solving a problem within the contextual factors identified?

The results of the study begin to identify and characterize the knowledge used by experienced librarians when subject searching in online catalogs. A next step is to test that knowledge for its usefulness in finding material in the online catalog. Further research can examine ways to include in our systems the knowledge that proves effective for retrieval so that the infrequent and inexperienced user can benefit from specialist knowledge without having to learn it explicitly.

References

- Bates, M. J. (1972). *Factors affecting subject catalog search success*. Unpublished doctoral dissertation, University of California, Berkeley.
- Bates, M. J. (1977). Factors affecting subject catalog search success. *Journal of the American Society for Information Science*, 28, 161-169.
- Bates, M. J. (1979). Information search tactics. *Journal of the American Society for information Science*, 30, 204-214.
- Belkin, N., & Vickery, A. (1985). *Interaction in information systems: A review of research from document retrieval to knowledge-based systems*. London: British Library.
- Borgman, C. L. (1983). *The user's mental model of an information retrieval system: Effects on performance*. Unpublished doctoral dissertation. Stanford University.
- Borgman, C. L. (1986). Why are online catalogs hard to use? Lessons learned from information retrieval studies. *Journal of the American Society for Information Science*, 37, 387-400.
- Breuker, J., & Wielinga, B. (1987). Use of models in the interpretation of verbal data. In A. L. Kidd (Ed.), *Knowledge acquisition for expert systems: A practical handbook* (pp. 17-44). New York: Plenum Press.
- Connell, T. H. (1991). *Subject searching in online catalogs: An exploratory study of knowledge used*. Unpublished doctoral dissertation, University of Illinois. Urbana-Champaign.

- Daniels, P. J. (1986). Cognitive models in information retrieval—An evaluative review. *Journal of Documentation*, 42, 272-304.
- Dayton, D. L., Lundeen, J. W., & Pollock, J. J. (1980). Automated techniques for online search guidance: A review. In *4th International Online Information Meeting* (pp. 317-333). Oxford: Learned Information.
- Doszkocs, T. E. (1986). Natural language processing in information +retrieval. *Journal of the American Society for Information Science*, 37, 191-196.
- Ericsson, K. A., & Simon, H. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Fenley, C. (1988). *Expert systems: Concepts and applications*. Washington, DC: Cataloging Distribution Service, Library of Congress.
- Fidel, R. (1988). *Extracting knowledge for intermediary expert systems: The selection of search keys* (Report No. IST 85-09719). Washington, DC: National Science Foundation, Division of Information Science and Technology. (ERIC Document Reproduction Service No. ED314059).
- Hancock, M. (1987). Subject searching in an online catalog. *Information Technology and Libraries*, 6, 60-63.
- Harter, S. P., & Rogers Peters, A. (1985). Heuristics for online information retrieval: A typology and preliminary listing. *Online Review*, 9, 407-424.
- Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (Eds.). (1983). *Building expert systems*. Reading, MA: Addison-Wesley.
- Hildreth, C. R. (1989). *Intelligent interfaces and retrieval methods for subject searching in bibliographic retrieval systems*. Washington, DC: Cataloging Distribution Service, Library of Congress.
- King, N. S. (1991). Search characteristics and the effects of experience on end users of PaperChase. *College & Research Libraries*, 52, 360-734.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed., enlarged). Chicago: University of Chicago Press.
- LaFrance, M. (1989, April). The quality of expertise: Implications of expert-novice differences for knowledge acquisition. *SIGART Newsletter*, 108, 6-14.
- Library of Congress. Office of Subject Cataloging Policy. (1990). *Subject cataloging manual: Subject Headings*. Washington, DC: Author.
- Library of Congress. Subject Cataloging Division. (1989). *Library of Congress subject headings* (12th ed). Washington, DC: Author.
- Markey, K. (1983). *The process of subject searching in the library catalog: Final report of the Subject Access Research Project* (OCLC Research Report No. OCLC/OPR/RR-83/1). Dublin, OH: OCLC.
- Matthews, J. R., Lawrence, G. S., & Ferguson, D. K. (Eds.). (1983). *Using online catalogs: A nationwide survey, a report of a study sponsored by The Council on Library Resources*. New York: Neal-Schuman.