From Access Points to Materials: A Transaction Log Analysis of Access Point Value for Online Catalog Users

Brendan J. Wyly

It is possible to obtain some insights into searchers' judgments by using transaction log analysis to associate online catalog search methods with decisions to retrieve location information. The Mainframe Interface to Libraries Online (MILO) is an interface to the Illinois Library Computer Systems Office online union catalog for 45 academic libraries. MILO provides access to a bibliographic database and directly links to another database with circulation and location records. Because the latter database only provides location and circulation status, searchers' decisions to make links to such data are seen as an indication that the records being linked represent potentially useful material. Via a transaction log analysis, the linked location records were associated with the access points used to retrieve them in order to analyze the value and problems of searchers' uses of specific access points. Transaction logs were analyzed for a 38-day sample of the 1994 logs. Counting records retrieved through the use of multiple access points (making the total greater than 100%), subject fields were used to access over 30%, author fields to access over 19%, and title fields to access over 51% of all records linked to location information. Other fields were used to retrieve very small percentages of linked records.

In an online catalog that requires an additional searcher action to obtain location and circulation information following a search, it is possible to obtain some insights into the searcher's judgments about the results. An analysis of such a system raises two interesting questions: (1) When specific access points are used to find online catalog records, how often do searchers find the results sufficiently valuable that they decide to seek location and circulation information? and (2) In what relative proportion are specific access points in an online catalog used to obtain

records for which location information is subsequently viewed? The answer to the first question reveals one kind of success rate for the use of an access point, regardless of how rare or common that use is. The answer to the second question offers a purely quantitative measurement of the value of each access point compared to all other access points.

Because many academic libraries have been caught between budgetary constraints and a commitment to thorough cataloging (in essence, a commitment to intellectual access), some library leaders have asked: Can we reduce or eliminate subject cataloging or complex name authority control? Large backlogs of uncataloged material have developed. The choice seems to be between providing some lower level of cataloging or allowing the backlog to remain uncataloged, and perhaps even expand, while seeking resources for full cataloging. In either case, some unknown but possibly substantial amount of intellectual access is being lost. Ascertaining the significance of that loss is helpful in guiding the treatment of backlogs and in guiding cataloging priorities. Even in less dire circumstances, establishing the value of the elements of cataloging and indexing is important. This study is an attempt to discover the value of various access points in catalog records by determining their usefulness to searchers in locating items in a collection.

An automated analysis of nearly 800,000 online catalog searches revealed how often the use of various access points resulted in retrieved records for which the searcher went on to find location and call number information. With some qualifications, which are explained below, logged searcher actions of a particular type might provide a strong indication of the usefulness of such access points in searches. For this study, the raw numerical analysis was accomplished by a parsing and tallying program written in FORTRAN. Descriptive statistics are reported in order to cross-tabulate search type, number of records retrieved, bibliographic record display format choices, and decisions to link to location information.

LITERATURE REVIEW

Numerous researchers have provided introductions to the advantages and limitations of transaction log analysis (TLA). Tolle (1984) and Peters (1993) provide a general introduction to TLA research. Peters, Kaske, and Kurth (1993) also have provided a more extensive bibliography and research review. Borgman (1986) provides a valuable review of older research. Both Tolle (1984) and Peters (1993) also provide some cautionary notes about the limits of TLA for extracting information about searchers' intentions. Further discussions of the limits and possibilities of the method are provided by Kaske (1993) and Kurth (1993), who raise helpful design considerations for any TLA.

Studies of searchers' online catalog search strategies are a very common, almost universal, component of online catalog TLAs. Kalin (1991) compared the searching techniques and success rates of remote and in-house searchers through a TLA and duplication of 1,000 searches by each group. The fact that searchers had to take a specific action to pull up a full bibliographic record was used to try to determine searchers' judgments about the relevance of the retrieved records. The present study provides firmer ground for the detection of searchers' decisions about the usefulness of retrieved records. Hancock-Beaulieu, McKenzie, and Irving (1991) conducted a TLA collaboratively at a number of academic and large public libraries in the United Kingdom and the United States. They administered interactive pre- and post-search questionnaires via online catalog terminals to ascertain searchers' intentions. They then compared these intentions with the actual search strategies revealed by the transaction logs. The general conclusion was that subject searching, whether direct or indirect, remains the prime means for use of the online catalog even though searchers face great difficulties integrating controlled vocabulary subject headings with keyword searching possibilities. Their study highlights the benefits of enhancing TLA with other analytical methods; questionnaires allowed measurement of searchers' intentions.

Reading intentions from the transaction logs must be undertaken with extreme caution. Nielsen (1986) challenges the benefits of enhanced TLA, observing that there are differences between what searchers say they do and what they actually do. Nielsen's results were sufficiently conclusive that he strongly warned against relying upon any self-reporting of search strategies. A study of search persistence by Wiberley, Daugherty, and Danowski (1989) also showed self-reported search behavior to be inaccurate. Because the present research focuses on questions relating to searchers' successful use of access points, rather than searchers' intentions, this study avoids the methodological pitfalls identified by Nielsen and others.

Ballard (1994), Barrett and Maticka (1989), Frost (1989), Hunter (1991), Marner (1993), and Peters and Kurth (1991) are just a few who have explored the use of search strategies in online catalogs. One of the most significant treatments of subject access in recent years was reported by Larson (1991), who was also a participant in the Council on Library Resources project on online catalogs. (Important earlier work in the field includes the BOOKS Subject Access Project by Atherton (1978), the final reports of the Council on Library Resources project (Kaske and Sanders 1983; Markey 1983; and Tolle 1983), and Markey's (1984) OCLC Online Computer Library Center, Inc. work.) Wiberley, Daugherty, and Danowski's work (1989, 1995) is also of interest, since they discuss the information overload that Larson perceives to be a particular problem. Drabenstott and Vizine-Goetz (1994) also offer an excellent overview of subject searching. They include an interesting TLA on subject searches that corresponds with many of the broad findings of this study. Drabenstott and Vizine-Goetz explore subject search failure and success in much more detail than this study, but in the present article I aim to explore the searchers' perceptions of success in their use of access points rather than the details of the failed and successful search entries.

Prabha (1990), Hickey and Prabha (1990), and Wiberley, Daugherty, and Danowski (1989, 1995) believe that the threat of information overload should be a central concern for information system designers. Indeed, Larson's research suggests that subject searching has become so difficult for online catalog searchers due to zero-hit searches and information overload that searchers are slowly being driven away from controlled vocabulary subject searching and may move towards even very limited title keyword access as a substitute form of topical access.

While Larson's conclusions about the slow decline of subject searching in the MELVYL system of the University of California appear to be sound, his analysis of causes is based on the rather narrow definitions of subject searching success and failure that he draws from the transaction logs. The difficulties he cites are of two types, the zero-hit subject search and the excessive-hit or information overload subject search. These difficulties afflict any type of online bibliographic searching and, perhaps, are particularly troublesome with subject searching. However, classing all zero-hit and large-hit controlled vocabulary subject searches as failures—failures that have an iterative effect of driving searchers towards alternative topical access methods—is too broad a generalization. An easily conceived example illustrates this point.

A search in a large academic catalog using the subject "Dogs" will likely produce over 100 titles and thus, according to Larson's analysis, our example appears to be a case of failed subject searching due to information overload; it seems to be a case of a searcher not understanding the excessive generality of the subject search executed. However, in many catalogs, one of the first books in a date-sorted list is the latest edition of *The Complete Dog Book: The Photograph, History, and Official Standard of Every Breed Admitted to AKC Registration, and the Selection, Training, Breeding, Care, and Feeding of Pure Bred Dogs*, a volume that is likely to meet the needs of many subject searchers. Similarly, were such a searcher to scan only the first of several hundred screens, find this book, and then leave the online catalog with a promising reference title

and call number, Wiberley, Daugherty, and Danowski (1989, 1995) would classify the search process as a case of information overload because of the nonpersistence in scanning records. However, such a hypothetical searcher would likely consider the search process a success. In other words, ascertaining information overload is a very subjective matter; we need a way of letting the searcher inform us when information overload is a problem and when it is not. The number of hits alone cannot reveal the problem. Though it is perhaps somewhat less obvious, neither can zero-hit searches be clearly judged as failures if the whole search process—instead of the individual search transaction—is considered. For example, a mistyped and failed subject search for "Sgnal processing—Digital techniques" followed immediately by a search for "Signal processing—Digital techniques" appears, based solely on the transaction logs, to indicate a 50% failure rate for this beleaguered subject searcher. However, if asked whether subject searching worked, the searcher might reasonably respond that it did. Another easily overlooked success is the searcher who initially fails with the subject search "Canoeing" and converts it to a title keyword search and discovers the subject "Canoes and canoeing" and then returns to the controlled vocabulary subject approach. This searcher has undergone a perfectly normal process of communication with the bibliographic information system, but according to Larson's analysis of the transaction logs, the searcher appears to have succeeded with the title keyword search 100% of the time and with the subject search only 50% of the time, insofar as zero-hit searches are considered "failures."

The problem is that "failure" is a natural part of success in extracting information from any bibliographic information system. Communication involves "failure" because it necessarily involves feedback and learning. Online catalogs are communication devices that allow searchers to interact with a database. We must analyze them as communications facilitators.

My intent in this study is not to dismiss Larson's findings regarding the decline of subject searching, nor to dismiss his insights on the difficulties with subject searching. Instead, it is simply to note that Larson's study does not establish a firm link between the difficulties of subject searching and the decline in subject searching. Many searchers choosing to use title keyword searching for topical access in MELVYL might only be becoming more intelligent users of all the options of online catalogs. We need more information about searchers' perceptions of the value of all access points and stronger evidence of searchers' problems in using all access points. The present research provides some of this information by inquiring into the searchers' perceptions of merit in the records they retrieve via the use of subject and other access points.

The research already conducted in this area has not been large in scope nor has it provided an idea of how searchers themselves perceive the relative value of controlled vocabulary subject searching. Drabenstott and Vizine-Goetz (1994) do provide dramatic insights into subject searching and online catalog design for subject searching. However, they offer little insight into the relative value of all the access points as perceived by users. In order to investigate the reliance of patrons on subject headings for finding material they wanted to check out, Knutson (1986) correlated the number of subject headings for 1,105 records in three Library of Congress classifications and one subclassification with circulation levels by patron type in an academic library. On the whole, Knutson found little positive correlation and some negative correlation between the number of subject headings and circulation levels. However, the scope of the study was so limited that conclusions were very difficult to draw.

Later, Knutson (1991) tracked one year of circulation for three groups of social science essays with three different levels of catalog records, ranging from ordinary treatment to enhanced subject headings and contents notes. Again, the scope of the study was too limited for the results to be generalized, and the results were inconclusive in any case. Charbonneau (1986) also found no significant dif-

ference in levels of circulation when he compared the circulation of materials receiving original cataloging to that of materials receiving copy cataloging at the University of Pittsburgh. He did not focus on the two types of bibliographic records, however, but rather on the differences in the types of materials that received original cataloging as compared to those for which copy was available.

Knutson (1986) recognized the problems with making cataloging decisions based on his results and called for studies with similar intent to be done using protocol analysis or TLA. The current study responds to that call. Much of the uniqueness of the TLA reported here hinges on the claim that when searchers linked to location information for selected records they had made a decision about the usefulness of these records. No attempt is made in this study to discern why searchers judged records to be useful. Rather, the mere fact that they did make a judgment at a particular point in time is acknowledged.

INSTITUTIONAL AND TECHNICAL ENVIRONMENT

The transaction logs, as shown in the ex-

ample in figure 1, provide information on date, time, terminal, response time, number of Boolean operators, number of potential index entries in a search (e.g., "engineers smalltalk" as a subject heading is one index entry, but as a title keyword search two index entries are searched), number of truncations, hits, error codes, and the text of the input for the transaction. Because search, display, and location linking commands all result in logged transactions, the total of nearly 2.7 million analyzed transactions reflected 795,810 searches and their associated linking and display transactions.

The raw logs had records from all of the system's terminals mixed according to the time stamp on the transaction. As shown in figure 1, these raw data were sorted according to terminal codes and time so that specific search sequences could be analyzed. For purposes of this analysis, it was not necessary to ascertain when one searcher left a terminal and another began using it, but time stamp analysis might allow interested researchers to obtain some partially revealing information in this regard. A FORTRAN program was written to detect and tally all the search sequences and attempts to view location information. Thirteen dif-

```
a       b         c    d  e f g  h     i          j
1,1 94/12/31 12.33.34.90 AAI7 0.10 1 1 0  0     0  F  S ENGINEERS SMALLTALK-SC,1
1,2 94/12/31 12.34.11.10 AAI7 0.10 1 1 0  0     0  T  SA ENGINEERS SMALLTALK-NOSCOPE,3
1,3 94/12/31 12.34.11.70 AAI7 0.60 1 2 0  0     0  F  T ENGINEERS SMALLTALK -SC,1S

2  94/12/31 12.44.55.90 AAI7 0.10 1 1 0  0     0  F  S SMALLTALK-SC,1
   94/12/31 12.45.33.20 AAI7 0.00 1 1 0  0     0  T  SA SMALLTALK-NOSCOPE,3
   94/12/31 12.45.33.30 AAI7 0.80 1 1 0  12    0  F  T SMALLTALK -SC,1S

3  94/12/31 12.47.31.20 AAI7 0.10 1 1 0  0     0  F  S SMALLTALK-ALL,1
   94/12/31 12.48.00.30 AAI7 0.10 1 1 0  0     0  T  SA SMALLTALK-NOSCOPE,3
   94/12/31 12.48.00.90 AAI7 0.80 1 1 0  42    0  F  T SMALLTALK -ALL,1S
   94/12/31 12.48.02.20 AAI7 0.80 0 0 0  20    0  S  3-22 -,1
   94/12/31 12.48.03.50 AAI7 0.10 0 0 0  2     0  S  1-2 -,1
   94/12/31 12.50.21.30 AAI7 0.20 1 1 0  1     0  H  2-ALL,5

4  94/12/31 15.57.13.90 AAI7 0.60 1 1 0  26    0  F  T AUTOCAD -SC,1S
   94/12/31 15.57.15.10 AAI7 0.10 0 0 0  6     0  S  1-6 -,1
   94/12/31 15.57.15.10 AAI7 0.10 0 0 0  6     0  S  19 -,1
   94/12/31 15.58.00.10 AAI7 0.10 0 0 0  0     0  DISP 19 -SC
```

1. (1,1) Zero-hit subject search scoped to search one instituteion (SC=SIU, Carbondale), (1,2) converted to zero-hit "see" reference search (all authority file searches are non-institution specific) (1,3) converted to zero-hit title keyword search.

2. Zero-hit subject search converted to 12-hit title keyword search scoped for searching one institution (SC).

3. Same search as (2) scoped to search all institutions. The H transaction links to location information for multi-institution search results. S transactions with multiple hits result in short displays, and those with single hits result in full displays.

4. A single-institution (SC) title keyword search for which one of 26 records was fully displayed (S 19) and linked (DISP 19).

F=(find/initiate search); T=(initiate authority file search); S=(show bibliographic record(s)); DISP=(link to location information, DISP does not result in a hit count so 0 on DISP transactions is normal); H=(link to location information for searches scoped to multiple institutions)

a) date, b) time; c) terminal id; d) response time; e) # of Boolean operators; f) # of index entries searched; g) # of truncations in search; h) # of hits; i) error indicator; j) transaction text

Figure 1. Time and Terminal Sorted Sample of Transaction Logs.

TABLE 1

SEARCH TYPES AND ACCESS POINTS

Search Type	MARC Record Access Points Searched
1. authority-based corporate author	028, 110, 111, 240, 410, 411, 610, 611, 693, 694, 710, 711, 715, 730, 810, 811
2. corporate author keyword	028, 110, 111, 240, 410, 411, 610, 611, 693, 694, 710, 711, 715, 810, 811
3. authority-based author	028, 100, 110, 111, 130, 240, 700, 705, 710, 711, 715, 730, 240, 400, 410, 411, 800, 810, 811
4. author (automatic truncation)	028, 100, 110, 111, 130, 240, 700, 705, 710, 711, 715, 730, 240, 400, 410, 411, 800, 810, 811
5. author+title keyword	028, 100, 110, 111, 130, 240, 700, 705, 710, 711, 715, 730, 240, 400, 410, 411, 800, 810, 811, +245, 246, 247, 653
6. series title (automatic truncation)	440, 490 (for first subfield = 0), 830, 840
7. exact phrase title	245, 246, 247, 653
8. title keyword	245, 246, 247, 653
9. title keyword automatically substituted for failed subject	245, 246, 247, 653
10. subject headings and/or subheadings	052, 600, 610, 611, 630, 650, 651, 690, 691, 692, 693, 694, 695
11. authority-based subject	052, 600, 610, 611, 630, 650, 651, 690, 691, 692, 693, 694, 695
12. miscellaneous (e.g., ISSN, ISBN, accession number)	020, 022
13. undetermined authority-based derived from retrieved record	could be any of above authority-based options

ferent search types were detected, as shown in table 1. The raw and sorted data were compressed and archived for additional analysis and for sharing with other researchers.

SAMPLING

Kaske (1988, 1991) has reported on the variability of search types in online catalogs across time periods in academic institutions and has encouraged analysis of entire transaction populations. However, due to problems with very large data sets, sampling was used for this study. Kaske's concerns were taken into account by paying attention to use cycles in the sampling procedure. The large Mainframe Interface to Libraries Online (MILO) installation typically produces over 140,000 transactions on a high-use day during the

semester. This is a peak rate that is roughly maintained by system limitations, and the peak is fairly consistently reached on semester weekdays.

Because the availability of MILO has been greatest most recently, the sample was limited to 1994. The data could only be sampled based on time. However, the interest was not in sampling based on time so much as sampling based on the lowest unit of analysis, the transaction. The sampling technique needed to allow every transaction an equal chance of being chosen for analysis. Randomly sampling across the 365 days of the year meant that any transaction would have an equal n:365 chance of being chosen. However, the distribution of transactions per day across the days of the year was extremely nonnormal. Thus, a purely random sample based on time might not have a representative

number of transactions from low-use, high-use, semester, or non-semester days. According to Kaske (1988), the failure of a purely random sample to be truly representative might be significant, because search methods vary with levels of use. The sampling method had to assure an overall representative sample in this potentially diverse population of days.

Five subpopulations were chosen from 1994: high transaction-count semester days, low semester, high non-semester, low non-semester, and exceptionally low transaction-count days having below 10,000 transactions. The latter were mostly holidays in Illinois' institutions. The division between semester and non-semester periods was based on the data rather than on an academic calendar because a number of academic calendars are at work in the ILCSO system. The divisions were made on days when transaction counts fell to non-semester norms and did not rise again until the next semester, or vice versa. Within the semester and non-semester subpopulations, further divisions were made between high-count days and low-count days. For the semester periods, the division between high and low days was made at 120,000 transactions and seemed rather clear-cut. For the non-semester periods, the division between high and low was less obvious but placed at 40,000 transactions. Given these divisions, 111 days were high semester days. One-hundred-thirteen days were low semester days. Ninety-one days were high non-semester days. Forty-two days were low non-semester days. And 8 days were

extreme lows. Altogether these latter 8 days only totaled 40,714 transactions and were all included in the sample rather than have a very weak intra-type sample for what might be an interesting subpopulation. In the other categories, random sampling in each category selected 9 high semester transaction-count days, 9 low semester, 8 high non-semester, and 4 low non-semester. This total of 30 days was roughly in proportion to the 111:113:91:42 split of the total population of day types.

The sample contained a total of 2,689,396 transactions out of 33,386,286 transactions in 1994, or 8.06% of all transactions. The sample contained transactions from the different semester/non-semester and high/low transaction count days in the proportions shown in table 2. Except for skewing caused by the inclusion of all of the extreme low days, this random sampling within subpopulations resulted in a fairly representative overall sample. The extreme low transaction count days were over-represented but were such a small portion of the total sample that they could not significantly alter the overall results. Including the low count days at this level was worthwhile simply to allow comparisons of interesting day-type distinctions that might emerge.

THE MILO SEARCH ENVIRONMENT

The author of this study examined transaction logs that recorded the use of MILO by searchers of the catalogs of the 45 uni-

TABLE 2
TRANSACTIONS IN STRATIFIED SAMPLE FROM 1994

	Days $N=38$ $(N=365)$	Sample % of Transactions $N=2,689,396$	Pop. % of Transactions $N=33,386,286$
High-count semester days	9 (111)	46.59	46.89
Low-count semester days	9 (113)	27.83	29.68
High-count non-semester days	8 (91)	21.30	20.44
Low-count non-semester days	4 (42)	2.77	2.87
Extreme low-count days	8 (8)	1.51	0.12

versities, colleges, and community colleges which belong to ILCSO. MILO provides an online catalog interface to the Full Bibliographic Record (FBR) database, which is a catalog of the material acquired by hundreds of libraries in Illinois over a broad period of time. The ILCSO libraries' holdings are included in the database with complete title and author access and fuller bibliographic access for more recently acquired materials. Though precise dates vary by institution, most materials within ILCSO have been fully cataloged since the 1970s. This fact introduced an important limitation in the study: MILO provides access to a wider range of materials through author and title fields than through other fields. Searchers using the former access points retrieved more records and had greater chances of successfully fulfilling their needs insofar as those needs might be met by older material. Thus, the apparent success of searching based on author and title access points was inflated in this TLA. This problem does not affect recently acquired material and does not affect all ILCSO institutions.

Through the use of the Library Computer System (LCS) database, MILO also provides an interface to some ILCSO volume holdings and all copy holdings records and the circulation components of the catalog. MILO allows linking from a record in the FBR database to the holdings and circulation information in LCS. These links must be made if a searcher is to obtain the location and circulation status of an item found by searching FBR. All linking between an FBR record and the associated LCS record is recorded as a transaction in the separately maintained transaction logs for transactions occurring through the MILO interface. From these logged linking actions, the TLA software written for this research traced back to discover the types of searches that retrieved the original records.

Many non-ILCSO libraries also have records in FBR; a total of over 800 libraries throughout the state of Illinois are represented at varying levels in FBR. One can use MILO to search for non-ILCSO materials contained in FBR. Upon finding a record from such a non-ILCSO institution, a searcher who attempts to obtain location information can determine only which libraries hold the item but neither the call number nor the circulation information at non-ILCSO libraries. The retrieval of any location information by ILCSO searchers of non-ILCSO holdings was detected as a linking effort in this TLA.

There is a command-mode interface to the FBR database that provides more flexibility than the MILO interface—in particular advanced Boolean searching—but all of the more commonly used capabilities present in FBR command mode are also available via MILO. This command-mode interface is available as a bypass option under the MILO menu system. Searches entered in this way are logged with the MILO searches and were thus analyzed for this study. All MILO-based searches are converted into a set of command-mode FBR searches. By analyzing transaction sequences it was possible to detect MILO-automated search types such as automatic uses of the authority files and automatic conversions of failed subject searches into title keyword searches. The thirteen search types analyzed are listed in table 1. Call number searching takes place directly in LCS and is therefore not in the analyzed FBR logs for MILO.

In the MILO interface, subject searches that retrieve no records are converted first to searches for a subject see reference in the authority file and, if that fails, to title keyword searches. These conversions are done without notifying the searcher as to what specifically is taking place. The searcher is only queried as to whether a multi-step subject search is desired. This typically occurs when a searcher enters a non-Library of Congress (LC) subject heading, but it can happen if there are no records in the bibliographic database for a valid LC heading. The former is a very common occurrence. This type of conversion of attempted subject searches can be detected in the logs by detecting specific sequences of transactions, as shown in figure 1. In this study, I treated such converted searches as a spe-

cial subject-converted-to-title-keyword search type.

I also tracked separate categories of searches that were initiated based on authority file information. Authority-based author, subject, and corporate author searches were separately tracked. A miscellaneous category dealt with quasi-authority-based searches that occurred due to the system's capability to initiate a search based on an access point of a retrieved record. This capability exists at any time in command mode bypass from MILO, but is only available in the MILO interface when a subject search has been converted into a title search. This miscellaneous category was also used to keep track of other, very rare types of authority-based searching such as authority-based series title searches and authority-based uniform title searches. The authority-based search categories were used only when the results of a search in the authority file were directly used to drive a search in the bibliographic file. Searches that were actually keyed in based on the information read from the authority file were not detectable as being authority-based searches.

Searchers must conduct their search via the subject authority records if they wish to execute exact subject heading and subheading combination searches. Even a precisely entered heading and subheading combination retrieves records that contain the elements in any combination. For example, an entered search of "Nutrition—Economic aspects" also results in records containing a combination of subjects such as "Agriculture—Economic aspects—Mexico" and "Working class—Mexico—Nutrition." A search executed in the authority file to find "Nutrition—Economic aspects" could be used to drive a search in the bibliographic file for records containing the exact combination. However, the only access to the authority file in MILO occurs when a searcher enters a heading with many subheading combinations—such as "Nutrition"—or when a searcher retrieves records through a title keyword search as a result of a failed subject search. The default of searching for headings and subheadings in any combi-

nation is sometimes helpful to searchers, but it turns subject searching into phrase searching in the subject headings and subheadings. Finally, it should be noted that MILO does not allow keyword searching of the subject headings.

The most fundamental assumption in this study is the assumption that, generally, searchers intentionally decided to link from the display of bibliographic records to the location information in the associated LCS records. Some unknown number of links were no doubt errors or confusions in the use of the interface rather than genuine expressions of a desire for location information. Through a TLA, there was no way to determine how often these unintentional links occurred. Obtaining such information about searchers' intentions would be extremely difficult to obtain by any means. However, the linking command keystroke on the menu is clearly labeled as "Location." It might be reasonable to assume that such unintended actions afflict all types of searches in roughly equal proportions, so that at least the proportional effects on the results have been minimized.

Transaction logs were available from October 1991, when MILO began operation as an FBR/LCS interface. Though MILO was not widely used at all ILCSO member libraries immediately after its implementation, the system is now very widespread, with only the University of Illinois at Urbana-Champaign (UIUC) still running a significant number of public terminals with older interfaces to FBR with separate transaction logs. Now even most of UIUC's public terminals and all its remote logins run MILO. Note that the introduction date of MILO does not in any way correspond either with the varying dates on which ILCSO members began creating full catalog records in the FBR database in the 1970s or with dates to which they have conducted retrospective conversions. Also note that only those searches conducted through MILO and through the MILO bypass to command-mode were analyzed in this study. Other interfaces to FBR store their transaction records in separate transaction logs.

AUTOMATED ANALYSIS

A FORTRAN log analysis program recorded an entry in a data table for every search sequence analyzed. A search sequence consists of:

- A search command that drives a search of the bibliographic file;
- Any automatic conversions of a failed subject search to a see-reference search and a title keyword search; and,
- All the transactions to display the search results and link them to location information.

A search sequence is ended by the initiation of another search from the same terminal. Thirteen search types were analyzed as shown in table 1. Some search types make different uses of the same access points (e.g., title keyword searches converted from failed subject searches as compared to searcher-initiated title keyword searches). In addition to the search type, each element in the data table recorded three additional characteristics of the searches in any given search-type category:

- The number of retrieved records (hits) between 0 and 1,999, listed in intervals of 1, and a single category for greater than 1,999 hits;
- The number of links to location records for a given retrieval, between 0 and 19, listed in intervals of 1, and a single category for greater than 19 links;
- A display code to reflect the bibliographic display modes. The display code category was broken down into: (a) zero-hit searches, i.e., no display; (b) searches with or without linking in which only the short display was used; (c) single-hit searches with or without linking that resulted in an automatic full bibliographic display for the single hit; (d) searches in which the searcher chose to force full displays of some records and in which any linking occurred from those full displays; and (e) cases in which the searcher linked some records only displayed in short form but also chose to force full displays of some records whether linked or unlinked.

It would have been desirable to track whether a given linked record was linked from a full or short bibliographic display, but the unit of analysis was the entire search-display-link sequence rather than individual hits. Therefore, some compromises had to be made in recording the display type used when a searcher linked from mixed display types. A future analysis might treat each hit as the unit of analysis. The short display includes author, title, and publication date. The full display includes all the elements of the MARC record typically displayed in on-line catalog full displays (i.e., title, author, notes, physical description, publication information, subject headings, standard numbers, continuation notes, etc.).

Recorded transactions in the logs allowed all these characteristics of a given search sequence to be determined. Because it is possible for a searcher in command mode to restore the immediately preceding search, this study treated the restoration of a previous search result as a new search identical to the old search rather than as a continuation of the previous displaying and linking sequence.

Having assembled a 13 x 2,001 x 21 x 5 (search-type x hits x links x display-code) table that contained the tallies for any given run of transaction logs, the table was analyzed by the same program that was used to extract the tallies from the transaction logs.

The program was run to analyze all 38 sample days individually and then to analyze collectively the high-count and low-count sample days for semester and non-semester periods as well as the extreme low-count days. Finally, the program was run against all 38 days as a whole to assemble the overall results that are the main focus of this report on the research.

FINDINGS

CLASSICAL TLA RESULTS

The analysis extracted typical information about searchers' uses of various search types and access points. Although the focus of the present study was on the logged decisions by searchers to link to location information, the general results are offered here for comparative purposes.
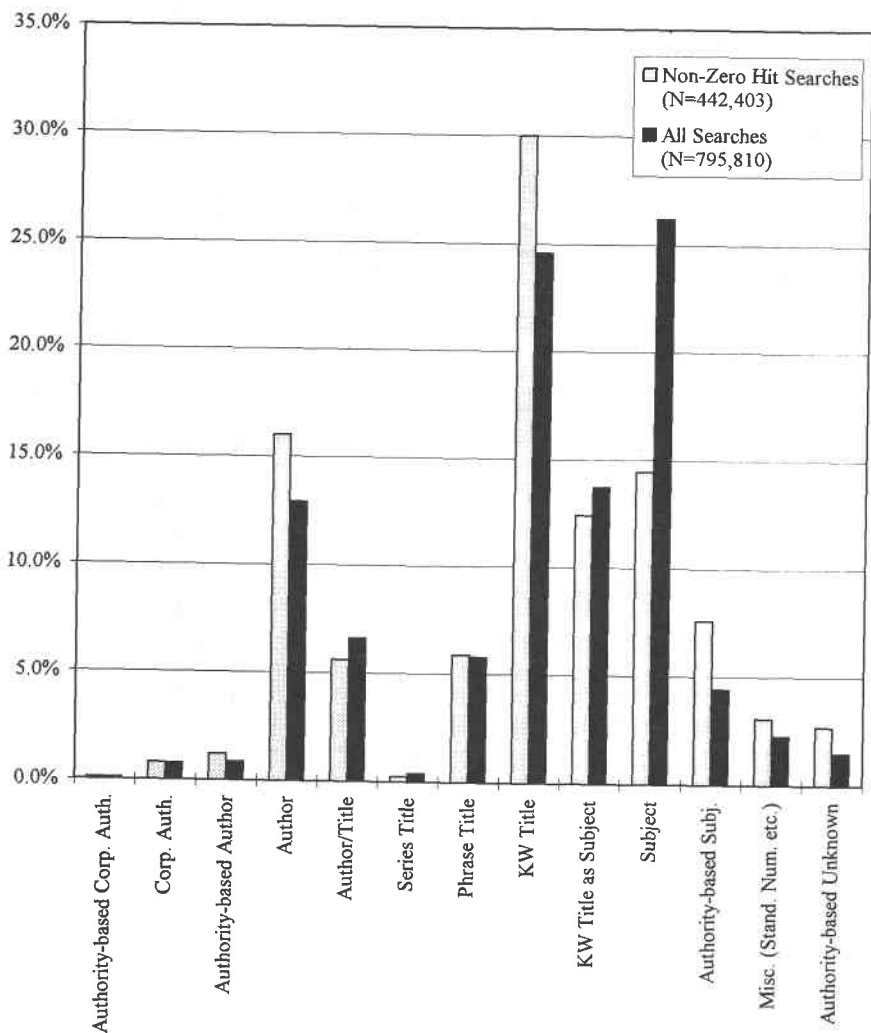
Figure 2. (A) Non-Zero Hit Searches by Type as a Percentage of All Non-Zero Hit Searches; (B) Searches by Type as a Percentage of All Searches.

Some of the same patterns reported by Larson (1991) were found. All types of explicitly topical searching (i.e., subject, converted subject to title keyword, and authority-based subject) comprised about 45% of the searches. This is roughly in accord with Larson. The breakdown by search type for all searches and for searches with non-zero hits is given in figure 2. By comparing percentages that are not equal, we can see which search types had the most or least zero hits. For example, zero-hit searches were very common for subject searchers using controlled vocabulary. Subject searches dropped from over 26.2% of all searches to under 14.5% of all searches that retrieved one or more items, i.e. non-zero hit searches. This drop results from the nearly 70% of such searches that retrieve no records as shown in figure 3. However, even in MILO, which automatically
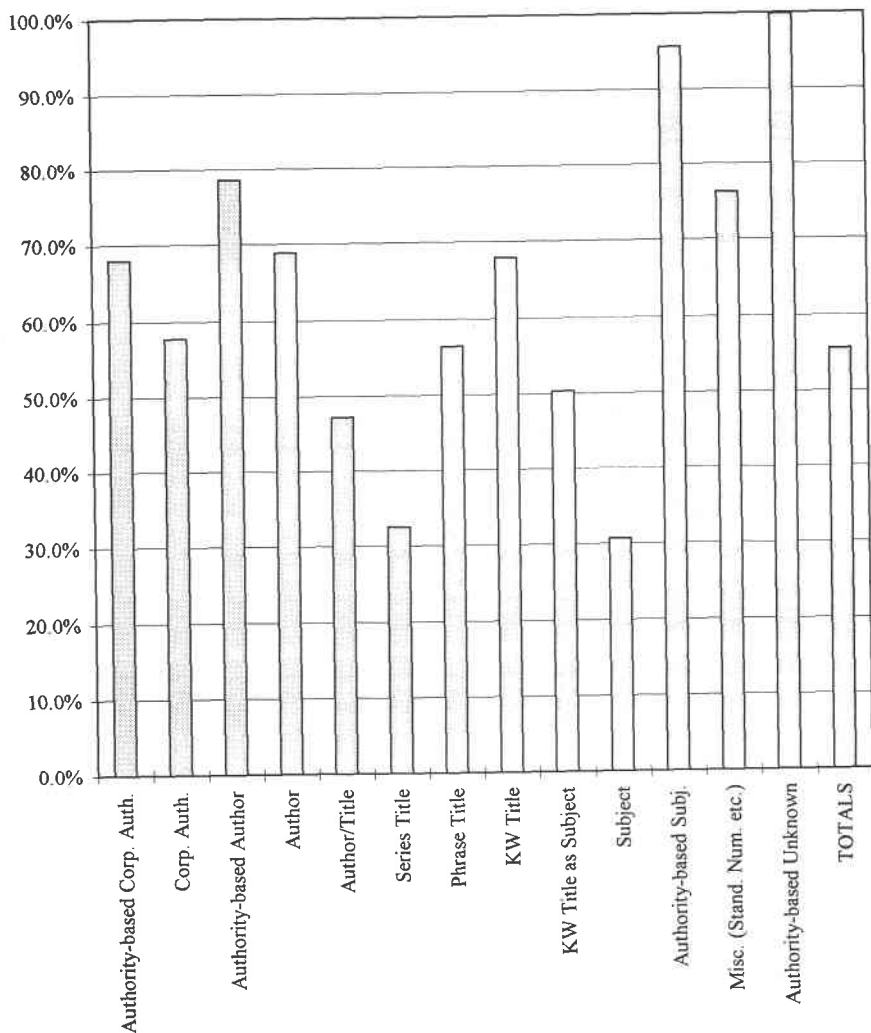
Figure 3. Percentage of Each Search Type Resulting in Non-Zero Hits.

searches for see references in the author-ity file and—failing to find any—then con-verts failed subject searches into title key-word searches, these converted topical title keyword searches also dropped from 13.7% of all searches to 12.4% of non-zero hit searches. The shifts in proportion when comparing all searches to non-zero hit searches were towards title keyword, author, and authority-based subject searching rather than towards failed sub-ject searches automatically converted to title keyword searches. These shifts in proportion result from variations in the proportions of zero-hit searches that are a part of all search strategies to the varying degrees shown in figure 3. However, fig-ure 3 does not contain information about how common a search type was—only how likely it was to retrieve some results when used. Figure 2 includes information both about how common a search type

was and how often it retrieved one or more records.

FUNDAMENTAL NEW TLA RESULTS

The most fundamental measure offered by this study is presented in figures 4 and 5. Both figures contain plots of the various subpopulations as well as the results for the overall sample. The overall sample is plotted as a bar and the subpopulations as scatter graphs. One measure of the value of various access points—which considers the average number of links to location information that were made for each non-zero search—is plotted in figure 4. This addresses the first research question about the value of records in a retrieval set. We see that subject searches resulted in more linked records per non-zero hit search than any other search type, including "converted" subject searches, authority-based subject searching that resulted from "see" references, and searches on large headings that resulted in a display of subheading combinations by MILO. When searchers and catalogers were speaking the same subject language from the outset, the results were best, as measured by quantities of linking decisions. When searchers relied on writers to reveal topics in their titles, the number of linked records for the retrieved sets dropped. When authority-based subject cross referencing was used to try to redirect searchers, they linked fewer of the resulting records as they were forced away from the original articulation of their query.

When examining the scatter of the different subpopulations in figure 4, two noticeable characteristics emerge. First, series title searching was very helpful for non-semester searchers in terms of their link rate, and in contrast, this type of searching had a very low link rate for semester searchers and searchers on extreme low-count days. The problems with series title searches are not surprising, but the extreme differentiation in link rates between high-use and low-use periods suggests that searchers who dominate during high-use academic cycles need help in knowing when and how to use this special type of title searching. Note however that only about 0.4% of all searches were series title searches.

Subject searching is dealt with in the second and more important scatter phenomenon among the sample subpopulations. In figure 4, one of the search types with significant scatter is the undifferentiated authority-based search type that results from searchers' ability to use the fields of one record to initiate a new search for other records containing identical access points. This type of searching is probably dominated by subject searching because the option is only presented to users who have had a subject search automatically converted to a title keyword search. However, this assumption cannot be definitively tested. Nevertheless, the authority-based subject searching, the ordinary subject searching, and even the subject searches converted to title keyword searches all display similar, albeit less dramatic, scatter patterns among the different subpopulations. In all of these cases, the highest link rates occurred during semester periods. At first this might not seem particularly interesting given that we might expect subject searching to be relatively more prevalent during semesters. However, this is a graph not of the volume of linking for a particular type of search compared to all linking; rather, it is only the number of links per non-zero search of the specific type. Thus, for the subject searching that was done during the non-semester periods, the frequency with which the search results were linked to location information was lower than during the semester periods. In other words, subject searches were not only less frequently used by the types of searchers who utilized the catalog during non-semester periods, subject searches were also less successful when they were used, insofar as success is measured by linking decisions.

In figure 5, quantities of linking within a given retrieved set (as in figure 4) are given with information on the frequency of use of various access points. Thus, the graph reflects the percentage of all linked records that resulted from the use of particular access points, directly addressing the first of the two research questions
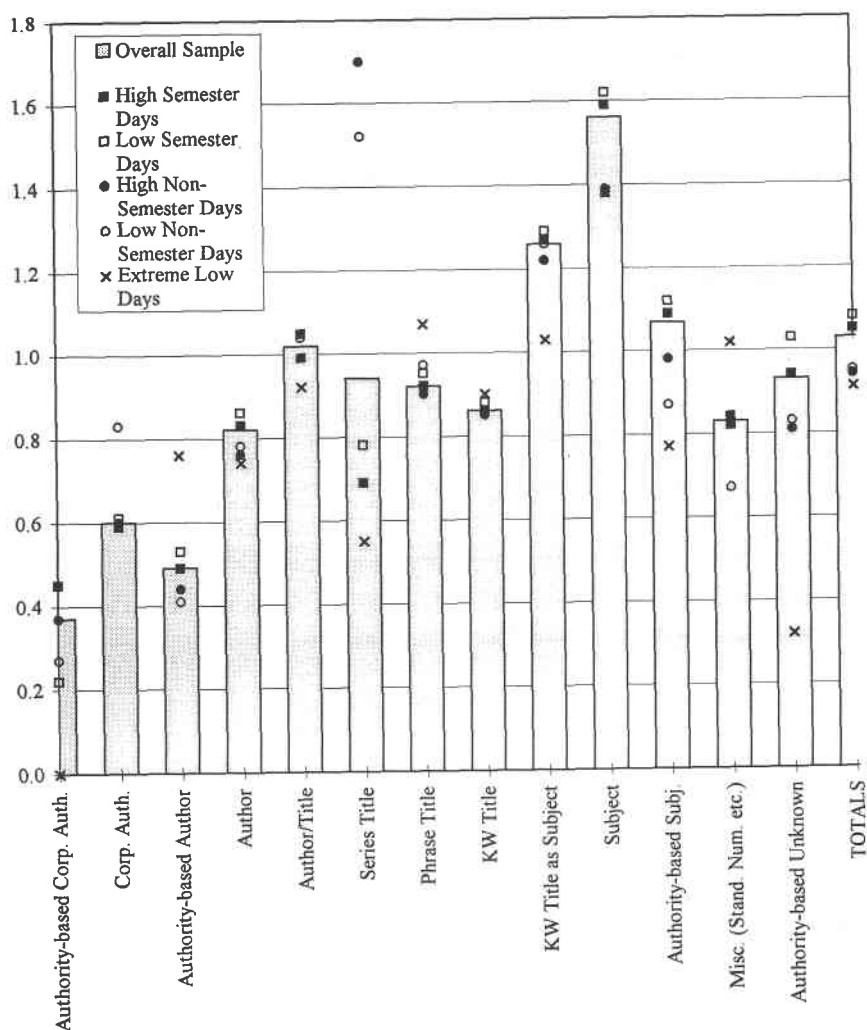
Figure 4. Average Links per Non-Zero Search by Search Type.

posed early in this article: Namely, what were the relative proportions of the specific access points used to obtain these potentially useful records? Even in the MILO environment, which converts zero-hit subject searches to title keyword searches, 29.9% of all the retrieved records for which decisions were made to obtain location information were retrieved by relying upon the subject headings as access points. Subject searches

were used directly to retrieve 22.0% of these linked records, and 7.9% were retrieved by going through the subject authority file. A small proportion, 2.4% of the linked records, represented authority-based searching that could not be specified. These probably were also subject searches; for example, they could be searches based on access points in records received as a result of a subject search that was converted to a title keyword search.
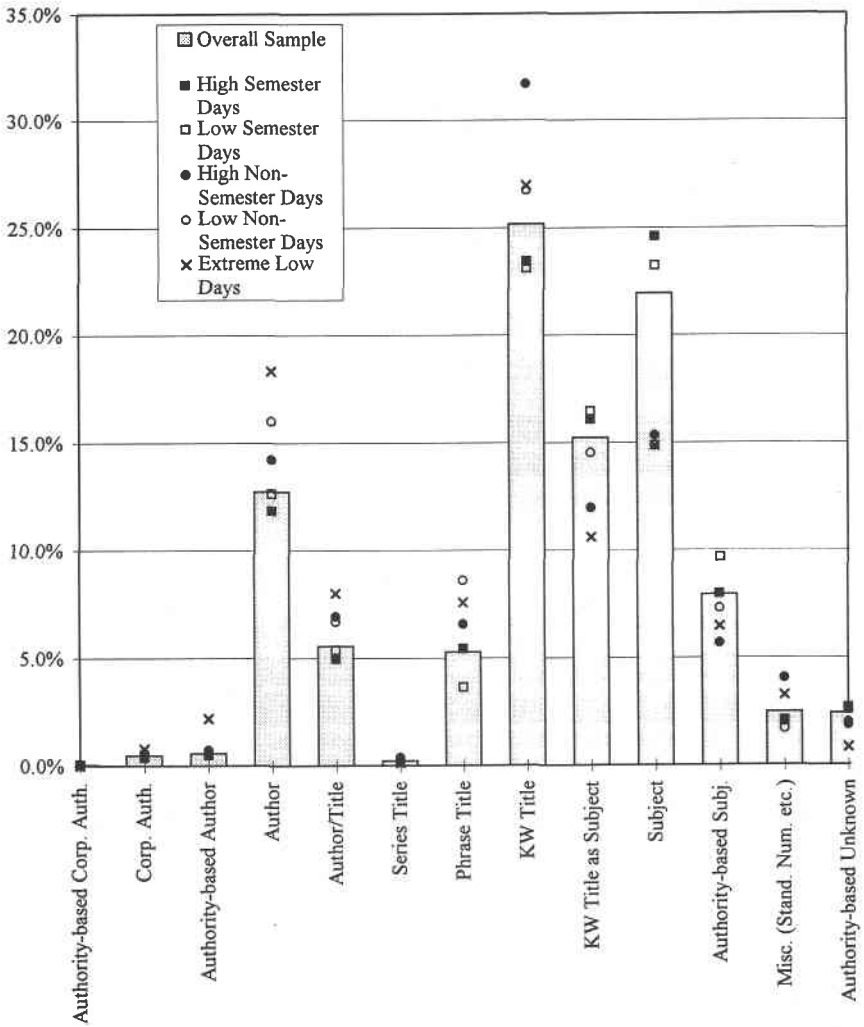
Figure 5. Percentage of All Links Made from Each Search Type.

On the other hand, over 65% of all linked records were retrieved by searchers using the author and title fields. 45.9% were retrieved by using the title or series title access points, 13.8% by using the author or corporate author fields, and about 5.6% by using both the title and an author field. Given that the ability to find material via author and title searching in MILO is enhanced by the greater historical coverage of those indexes, this 65% figure is almost certainly inflated. It would likely be smaller in a homogenous database in which all records had all types of access points. However, even at the somewhat underrepresented figure of 29.9%, the value of the controlled vocabulary subject access points for MILO searchers is very high. Perhaps having this information can be of some help to technical services managers as an indication of how searchers use access points to find mate-

rials that they believe are worth examining.

The scattering of subpopulations in figure 5 is an indication both of the prevalence of each search type and the likelihood within the subpopulation that the results of a search type would be linked. The scatter must be read differently from that in figure 4, because the two factors influencing the subpopulation scatter—prevalence of search type and linking of search type—cannot be separated in this plot. However, even during non-semester periods, over 20% of all linked records were retrieved due to the presence of subject access points.

## ZERO-HIT, INFORMATION OVERLOAD, AND OTHER SEARCHING PROBLEMS

Figure 6 is similar to figure 4, but here data are presented not on the total linking volume but on whether any links were made for a specific search. Some search goals are likely to culminate in only one link decision rather than multiple link decisions, and it is important to present both perspectives on the search types. Over 50% of all non-zero-hit subject searches were followed by some linking decisions. Of course, zero-hit searches cannot be followed by linking, and thus the link rate falls to 15% for all attempted subject searches. Figure 6 is relevant both in identifying the severity of zero-hit search problems and in addressing Larson's information-overload problem. However, several other problems confronting the searcher are revealed as well, such as the communications complexity of subject searching as compared to title searching, and the inaccuracy of title keyword searching as compared to combined author and title searching. In examining figure 6, the interactions of all these problems should be kept in mind.

Having said that, figure 6 provides a new measure of the relative ease or difficulty with which different access points are used when we consider the percentage of non-zero hit searches linked by search type. That new measure is based on the searchers' indications of success reflected in decisions to link to location information. It will come as little surprise to public service librarians that low linking rates indicate the relative difficulty of corporate author and series title searching. Also noteworthy is the relative difficulty of authority-based author and authority-based corporate author searching reflected in low percentages of linked retrieved sets. This alternative measure of relative search difficulty reveals an interesting surprise. Once the zero-hit search problem is overcome, a topical search, whether a subject heading search, an authority-based subject search, or a converted title keyword search, was slightly more likely than an author search to produce results that led to linking with location information. Of course, this is partially explained by noting that author searches as known item searches often fail to retrieve the desired item due to collection limitations. However, it is noteworthy that if we take non-zero results as a first indicator of success, then figure 3 shows author searching to be slightly more "successful" than keyword title searching. But in figure 6, we have an indication that author searching is more prone than topical searching to post-retrieval disappointment in which the searcher finds that the retrieved set does not contain desired records.

We can also compare in figure 6 each pair of bars to obtain an idea of the scope of the zero-hit difficulty for searchers using a given search type. In combination, the two elements show that zero-hit subject searches were very common and can be helpfully addressed by converting the searches to title keyword searches automatically as MILO does. Only 15.4% of all subject searches resulted in linking decisions as compared to 25.9% of all failed subject searches converted to title keyword searches. The degree of difference between each bar pair in figure 6 shows the degree to which zero-hit searches had an effect on the desired outcome of finding useful records via various access points. However, we should not be entirely swayed by this illustration of the difficulty of subject searching, because, as we saw in figure 2, a very large proportion of all linked records were drawn from the
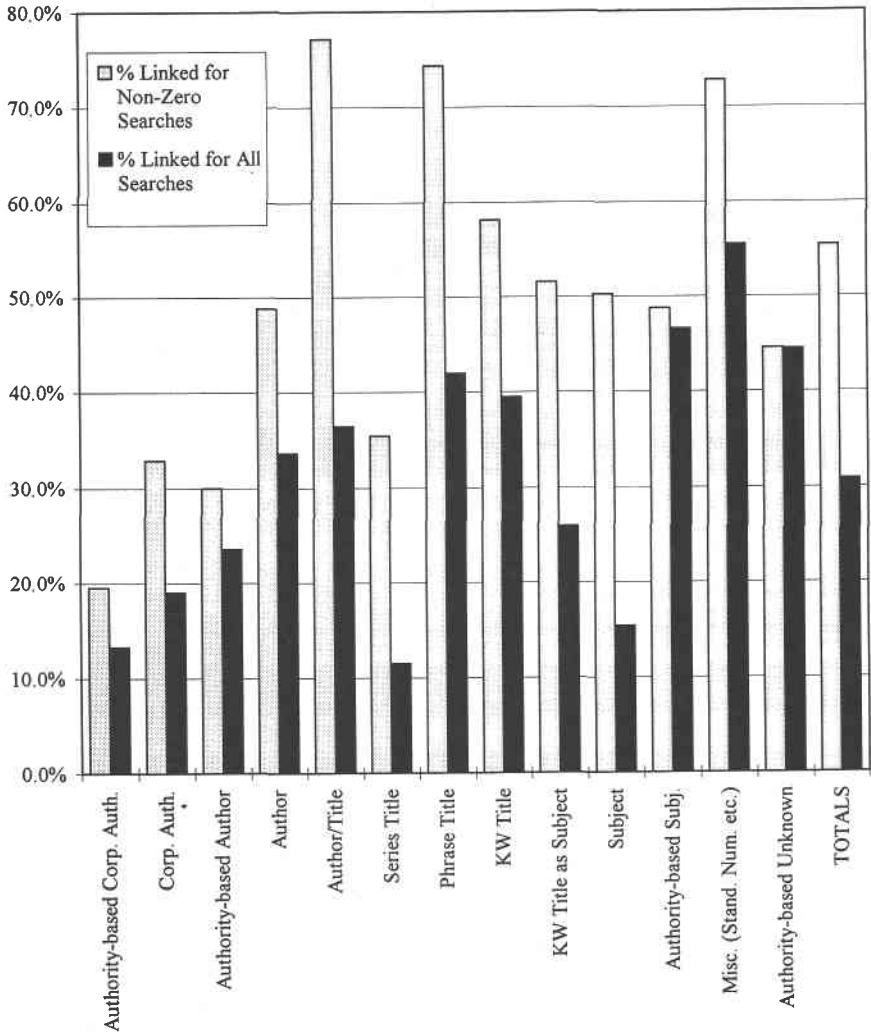
Figure 6. (A) Percentage of Linked for Non-Zero Hit Searches of Each Type; (B) Percentage Linked for All Searches of Each Search Type.

relatively small proportion of subject searches chosen for some linking and illustrated in figure 6. Many subject searches do not retrieve any results and those that do often retrieve results judged not worth linking to location information. However, the data in figure 2 show that subject searches that "succeed" often result in decisions to obtain location information for several items. Zero-hit failure in a single search, which is illustrated

here, is not the equivalent of search-type failure. When considered in combination, figures 2 and 6 illustrate the fact that subject searching is a difficult communications process, but may be worthwhile for searchers.

The other types of searches that figure 6 shows to have been particularly prone to zero-hit search problems are series title, corporate author, combined author and title, and exact phrase title searches.

These types of known-item searching might be failing because of collection limitations, but it might be that many searchers over-specify information with these search types. Keep in mind, however, that figure 6 contains no information about how prevalent a search type was, so the scope of the problem is not as severe as it might appear; figure 2 offers a more general overview of the scope of the problem.

In order to obtain insights into the information overload problem described by Larson (1991) and also investigated by Prabha (1990) and Hickey and Prabha (1990), modes, means, and median hit sizes were calculated for linked and unlinked searches of all types. For the overall sample (excluding zero-hit searches), the hit size mode for all search types—whether linked or unlinked—was one. This uniformity only reveals that for better or worse a single retrieved record was the most common result for all search types and was the most common hit size to be linked or to be left unlinked. The mean number of hits for all non-zero searches and for those non-zero searches with fewer than 2,000 hits is shown in table 3. That table also includes the median number of hits for linked, unlinked, and combined linked-and-unlinked searches. For a given search type, large differences between the mean hit size for all non-zero searches and the mean hit size for non-zero searches with fewer than 2,000 hits indicate that searchers using that search type were more prone to retrieve very large sets. Based on this analysis, rather than subject searching having been particularly prone to the information overload problem as asserted by Prabha (1990), title keyword and corporate author searching (a keyword search on MILO) appear to have been prone to the problem. However, mean hit size forms an extremely uncertain basis for analysis because the standard deviations in retrieved set size were very large in all cases. Therefore, it is helpful to compare median hit sizes for linked and unlinked retrieved sets to see whether sizeable differences exist. If a particular search type exhibited an unusually large unlinked median retrieved set size, that might indicate that

information overload was a problem for the search type because searchers were being discouraged by set size from obtaining location information. Further insight into the information overload problem is possible by taking a higher percentile measurement on hit size rather than the 50th percentile median. The 90th percentile hit sizes for linked, unlinked, and overall searches of all types are also shown in table 3 to allow this comparison. The hit numbers in the 90th percentile columns of table 3 reflect the hit sizes that were greater than or equal to 90% of all search set sizes for a given search type. The aim is to discover the search types in which information overload posed particular problems.

In looking at the median hit size figures in table 3, and noticing which search types have substantially lower medians for linked, as compared to unlinked, hit sizes, it appears that the only search types prone to information overload are the corporate author search—which is executed as a keyword search by MILO—and the title keyword search. None of the median hit sizes was larger than 20 for either linked or unlinked retrieved sets, and most observers would agree that 20 retrieved records does not constitute information overload. Ultimately, however, the information in the median numbers is an indirect rather than a direct indicator of the information overload problem. We can only directly say for sure that half the searches not plagued by the zero-hit problem also escaped information overload problems.

However, by looking at differences between linked and unlinked median hit sizes, indirect information can be gained about the incidence of information overload with a given search type because a large difference indicates that more large retrieved sets were unlinked than were linked. At the median level, only the corporate author keyword and, possibly, the title keyword searches seem to have displayed potential information overload problems. Moreover, in 4 catagories (subject searches converted to title keyword searches; authority-based subject searches; authority-based author searches; and the miscellaneous author-

TABLE 3

SIZE OF NON-ZERO RETRIEVED RECORD SETS BY SEARCH TYPE AND LINKED STATUS

Search Type	Mean Hits (Non-Zero)	Mean Hits (Non-Zero <2000)	Median Hit Size			90th Percentile Hit Size		
			Linked	Unlinked	Total	Linked	Unlinked	Total
Authority/ Corp. Auth.	69.1	51.6	19	19	19	93	118	118
Corp. Auth.	247.4	112.0	12	19	17	151	665	436
Authority/ Author	29.8	25.5	6	4	5	46	48	47
Author	61.8	45.2	8	8	8	53	171	92
Auth./Title	4.9	4.9	1	2	1	6	16	8
Ser. Title	90.7	81.5	19	16	16	125	258	211
Exact Title	5.8	5.8	1	2	1	8	18	10
KW Title	116.8	52.6	3	7	4	51	346	120
KWT as Subject	54.9	40.63	8	5	6	72	120	87
Subject	90.7	64.5	18	18	18	129	180	152
Authority/ Subject	66.7	52.1	5	2	3	118	166	137
Misc.	12.1	5.6	1	1	1	3	8	3
Authority/ Undetermined	52.3	40.3	6	3	4	77	116	96
Totals	74.9	44.6	4	6	5	63	183	101

ity-based searches, mostly derived from retrieved records and probably mostly consisted of subject searches, the median hit size was actually larger for the linked retrieved sets than for the unlinked sets. Based on the median data, searchers appear to have preferred larger retrieved sets for those search types, at least up to a point, and at least half of these searchers seem more likely to have been retrieving too few records rather than too many.

This information is relevant to catalog design, but it may have even more relevance for decisions about collection building. As will be seen during consideration of the 90th percentile hit size figures, there is a point with all search types at which retrieved sets can become too large and the likelihood of linking decreases; but, what is fascinating is that these quantitative data also suggest that there is a point with certain search types at which retrieved sets can become too *small* and,

once again, the likelihood of linking decreases. Note that the median retrieved set size is markedly lower for authority-based subject searches than for general subject searches; this is so partly because authority-based searches in MILO search only for precise subject heading and subheading combinations or subject headings standing alone, whereas general subject searches search for any occurrence of subject headings or subheadings in any combination.

Turning to a consideration of the 90th percentile retrieved set size in table 3, some interesting insights emerge. For all search types, the unlinked 90th percentile hit sizes are larger than the linked; beyond a certain number of hits, there was an obvious preference for smaller retrieved sets when making decisions to locate some items. This is in accord with Larson's analysis. However, there are important subtleties that should be noted in the data.

First, there were some very large retrieved set sizes about which searchers were willing to make some judgments. Because MILO allows results to be sorted by a number of criteria, including publication date, it is possible to make certain types of judgments fairly easily about some very large retrieved set sizes. In particular, note that table 3 indicates that 10% of all linked subject searches retrieved more than 129 records and that 10% of all linked authority-based subject searches retrieved more than 118 records. Ten percent of all subject searches that were converted to title keyword searches and subsequently linked retrieved more than 72 records.

Still more informative is the comparison, among all search types, between linked and unlinked records at the 90th percentile of hit sizes. Where the proportional differences are small between linked and unlinked 90th percentiles, information overload appears not to have been a significant problem, except, perhaps, for those retrievals of very large sets of records that constituted less than 10% of retrieval sets. The most noteworthy findings based on this view of the data are that subject searching and authority-based subject searching do not seem to have been as prone to the information overload problem as were title keyword searching, author searching, corporate author searching, and subject searches converted to title keyword searches. Ninety percent of all subject and authority-based subject searches retrieved fewer than 152 and 137 records respectively, and 90% of all linked retrieved sets for these search types retrieved 129 and 118 records respectively. The relative closeness of these two sets of numbers, as compared to the large differences for other search types, indicates that the retrieved sets for these subject search types were, in comparison to other search types, less often so large that they discouraged searchers from making any judgments. Subject searches converted to title keyword searches were somewhat more likely to discourage searchers with large search sets, but the problem was not nearly so severe as it was with general title keyword

searching, author searching, series title searching, and especially corporate author keyword searching.

USE OF BIBLIOGRAPHIC DISPLAY MODES

Findings regarding the relationships among search type, bibliographic record display choices, and subsequent decisions to link to location information are found in figure 7. Note how often the use of a given display mode for the results of a given search type was followed by a decision to link to location information. There is a fourth display mode not shown in figure 7 because it only occurs when some linking occurs from a short display following the linked or unlinked display of a full record. In figure 7, which tabulates linked searches only, this mode would therefore always be 100%. Figure 8 does include this mixed display data. Bear in mind that figure 7 contains no information about the number of linked searches made using various display modes but only presents the likelihood that, when used, a given display mode will be followed by a linking decision. Figure 7 can be taken as an illustration of the effect of obtaining more bibliographic data upon subsequent decisions to obtain location information. For most search types (i.e., keyword title, author, corporate author, subject, subject converted to keyword title, and series title), the searchers' decisions to obtain additional information about items through the full bibliographic display increased the likelihood of a subsequent decision to also obtain location information. For other search types (i.e., exact phrase title, combined author-and-title, and standard number), a decision to display a full record seems to have been an indication that the searcher doubted whether an item matched a known-item need and was looking for hopeful signs but did not find them in the full displays. In those cases, full displays actually resulted in fewer linking decisions. As one might predict, these types of searches, as well as title keyword searches, were most likely to be linked when they resulted in only one hit. Conversely, figure 7 shows that a single-
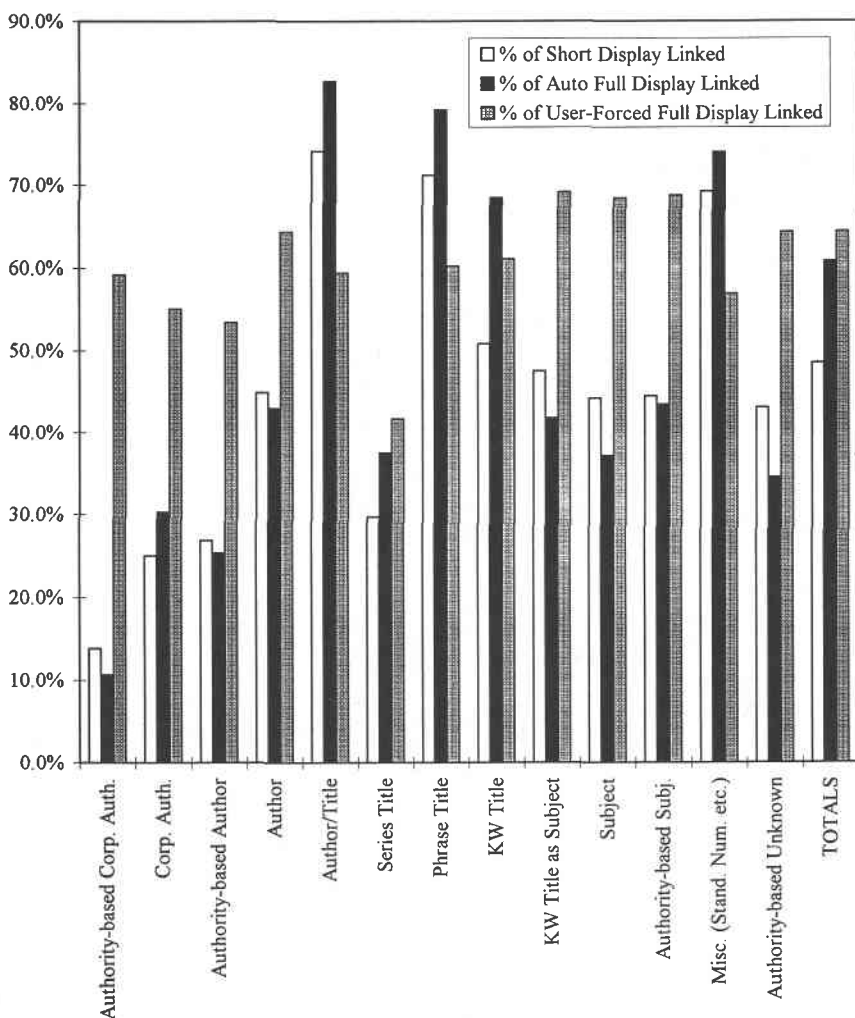
Figure 7. Linking to Location Information vs. Display Mode and Search Type.

hit result for a broader search (e.g., a subject search that results in an automatic full display) was unlikely to be linked to location information because the rare single-hit retrieval in a subject search of a large collection is unlikely to address a need.

See figure 8 for information regarding the actual number of link decisions made from each display mode; information that is excluded from figure 7. The distinction

between the information content of the two figures should be carefully maintained. Figure 7 can be used to illustrate the effects of obtaining additional bibliographic information on linking decisions. Figure 8 cannot be used this way because it illustrates the proportion of all linking done from each display mode for each search type, but contains no information about how often the display mode was used. Figure 8 only serves to illustrate the
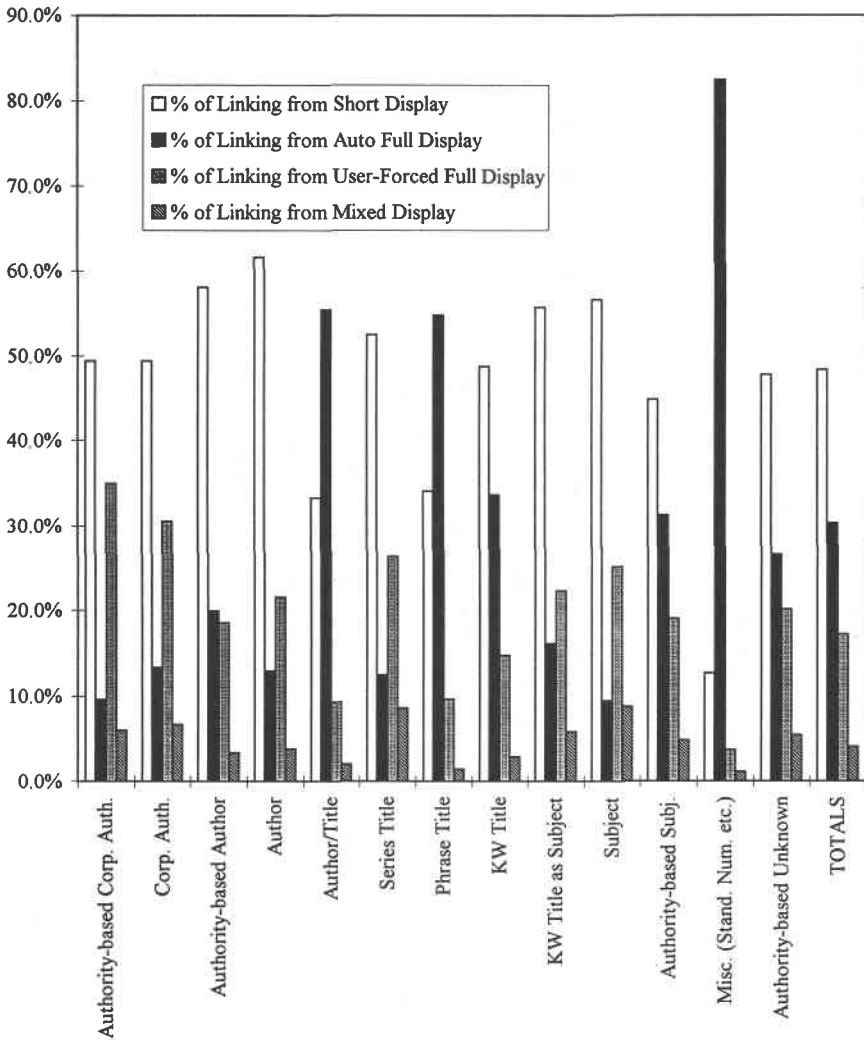
Figure 8. Percentage of All Links for Each Search Type Made from Each Display Mode.

degree to which various bibliographic displays were used when linking was the end result. For example, for all search results chosen for linking, linking to location information was based on short displays 48.4% of the time, on automatic full displays for retrievals of single records 30.3% of the time, on user-forced full displays 17.2% of the time, and on short displays following long displays (from which links might also have been made) 4.1% of the

time. Clearly, searchers relatively seldom chose to obtain additional bibliographic information in order to make a decision to obtain location information.

CONCLUSION

The purpose of this study was to provide a richer base of knowledge from which to work in understanding the process by which searchers communicate with online

catalogs through access point utilization. The primary objective of this research was to provide some quantitative information about the value of various access points not only as they were used for finding material, but as they were *successfully* used. This study measured "success" through searchers' decisions to obtain location information. The findings can be summarized as follows:

- Topical searches, including zero-hit subject searches automatically converted to title keyword searches, constituted 44.4% of searches and nearly 34.5% of non-zero-hit searches.
- Searches using subject access points constituted 30.6% of all searches and 22.1% of all non-zero-hit searches.
- Searchers retrieved records in 55.6% of all searches, ranging from a low of 30.7% for subject searches to a high of 95.6% for authority-based subject searches.
- An average of 1 link to location and circulation information was made for every non-zero hit search, ranging from 0.37 links per authority-based corporate author search to 1.56 links per subject search directly using the subject headings.
- Of all linked records, 29.9% were retrieved by use of the subject headings, 45.9% by using the title or series title access points, 13.8% by using the author or corporate author fields, and 5.6% by using both the title and author fields.

Clearly, a significant proportion of all relevant record retrievals were based on access points other than the minimal author and title fields. However, it is equally clear that those minimal fields were very useful to a large number of catalog searchers for retrieving a majority of the records judged sufficiently interesting to warrant linking to location information.

There are indications that access points far outweigh description as valuable attributes of the catalog records. In light of actual searcher practices, if cataloging resources are constrained, searchable access points would seem to be more important than rich description for facilitating the combined process of finding and judging materials. Access points are not separable from descriptive cataloging, but this point is worth recalling as the structure and value of catalog records is debated.

There might be an important message in the finding that subject searching appears to have been less successful for non-semester searchers than for semester searchers. Subject access of one kind or another is likely to become more important as information resources expand. In information environments too large to be encompassed by individuals' knowledge, the volume and value of unknown-item topical searches increases relative to the volume and value of known-item searches.

It is interesting that, for the information system analyzed here, searchers during non-semester periods were less successful with subject searching than were searchers during semester periods. Non-semester searchers were probably more likely to be searching for specialized or specific information. Graduate students and faculty constitute a larger percentage of campus populations during non-semester periods and that shift in proportions is almost certainly reflected in online catalog searcher populations. Graduate and faculty searchers are typically assumed to search information systems for more specialized information due to the nature of advancing levels of academic inquiry.

Non-semester searchers' lower success rates might be a message about the utility of online catalogs and other information systems. Catalogs have historically aimed to satisfy only a certain fairly general approach to subject searching. If more specialized subject searching increases in quantity, online catalogs might begin to frustrate an even larger number of subject searchers.

While the data in this study, and particularly in figure 6, offer only one measure of the difficulty and ease of use of access points, the reality is that all types of access come with constraints, including constraints of the collection. We do not know to what extent subject searching is hampered by collection constraints. We

must look at the searchers and their interactions with the records in many different ways to understand the constraints and difficulties, but we should be careful not to assume that difficulties in access point utilization necessarily indicate a lack of usefulness of those access points.

Ease of use is not the only indicator of the utility of a given access point. Another way of stating this is that difficulty in utilization (i.e., failure in a given single query) is not the only indicator of success rates in the total utilization of an access point. Failure of a particular query through zero-hit sets or information overload is a natural part of the communication process in the overall search.

Communication depends on this feedback process. We should be careful not to stifle communication with the catalog by designing in assumptions about what the catalog should say to the searcher. Sometimes it is very important for the catalog to communicate the large number of records related to a given search.

The median and 90th percentile hit size numbers in table 3 raise serious questions about how concerned we should be with the information overload problem in access point utilization. These figures indicate that we should not make sweeping assumptions about what constitutes information overload for searchers. While the data give no indication of how well searchers were ultimately served by retrieval sets of varying sizes, the willingness to obtain location information for varying retrieval set sizes suggests that searchers were not immediately defeated by information overload problems as we might have expected.

A caveat about this analysis: Searchers can use an online catalog to find some types of useful information without needing to link to location information. This is the case, for example, when bibliographic information is needed to clarify a citation. These are common uses of catalogs, and I was not able to detect and report such hits as "successes." It is not safe to assume that this problem afflicts all search strategies equally. In traditional local usage, subject searches seem much less likely to be put to this use. Moreover, it is also becoming

more common for the Internet-accessible catalogs of large academic libraries to be searched by nonborrowers who turn to other borrowing sources and so do not link to location information but only use the catalog as a guide to the extant literature. In this study, I made no attempt to estimate the number of searches that do not necessitate linking to location information as part of the information-seeking process. Because a major aim of the study is the determination of the relative value of various access points, this is an unfortunate, though probably minor, limitation.

Systems designers must not only seek to help the searcher interact with the catalog as Wiberley, Daugherty, and Danowski (1989, 1995), Hickey (1990), and Prabha (1990) suggest. Systems designers must also be careful not to get in the way of a communication process that necessarily involves feedback (even if that feedback looks like failure). Transaction logs must be investigated with a subtle understanding of the communication processes they reflect. Designers need to understand that what appears to be information overload is often a matter of the system informing the searcher about the nature of its content. Some searchers will take the first few postings from the system and declare their search a success. Some searchers will take the feedback and understand that they can potentially enter more specific searches to match their interests more closely.

While a decision to retrieve location information is not an indication of ultimate success in the information retrieval process, providing information that facilitates that decision can improve the process. Therefore, the data in this study can be taken as a useful guide to how access points serve searchers.

WORKS CITED

Atherton, Pauline, and others. 1978. *Books are for use: Final report of the Subject Access Project to the Council on Library Resources*. Syracuse, N.Y.: School of Information Studies, Syracuse University.

Ballard, Terry. 1994. Comparative searching styles of patrons and staff. *Library resources & technical services* 38: 293–305.

Barrett, Beverly, and Margaret Maticka. 1989. An analysis of user failure in subject searching. In *Garbage in—garbage out: The need for quality in the age of automation: Proceedings of the Australian Library and Information Association National Cataloging Conference,* ed. A. Bundy and J. Bundy, 38–49. Adelaide: Auslib Press.

Borgman, Christine 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.

Charbonneau, Gary. 1986. A comparison of rates of patron utilization of library materials receiving original cataloging and materials receiving copy cataloging. *Collection management* 8, no. 1: 25–32.

Drabenstott, Karen Markey, and Diane Vizine-Goetz. 1994. *Using subject headings for online retrieval: Theory, practice, and potential.* New York: Academic.

Frost, Carolyn O. 1989. Title words as entry vocabulary to LCSH: Correlation between assigned LCSH terms and derived terms from titles in bibliographic records with implications for subject access in online catalogs. *Cataloging & classification quarterly* 10, nos. 1/2: 165–79.

Hancock-Beaulieu, Micheline, Lorna McKenzie, and Avril Irving. 1991. *Evaluative protocols for searching behaviour in online library catalogues.* British Library Research and Development report no. 6031. London: British Library Research and Development Department.

Hickey, Thomas B., and Chandra G. Prabha. 1990. Online public catalogs and large retrievals: Methods for organizing, reducing, and displaying. *ASIS '90: Proceedings of the 53rd Annual Meeting of the American Society for Information Science* 27: 110–16.

Hunter, Rhonda N. 1991. Successes and failures of patrons searching the online catalog at a large academic library: A transaction log analysis. *RQ* 30: 395–402.

Kalin, Sally W. 1991. The searching behavior of remote users: A study of one public access catalog. *ASIS '91: Proceedings of the 54th Annual Meeting of the American Society for Information Science* 28: 178–85.

Kaske, Neal K. 1988. The variability and intensity over time of subject searching in an online catalog. *Information technology and libraries* 7: 273–87.

———. 1991. The variability of subject searching in an online catalog over an aca-
demic year. In *Interfaces for information retrieval and online systems: The state of the art,* ed. Martin Dillon, 191–205. Westport, Conn.: Greenwood.

———. 1993. Research methodologies and transaction log analysis: Issues, questions, and a proposed model. *Library hi tech* 11, no. 2: 79–86.

Kaske, Neal K., and Nancy P. Sanders. 1983. *A comprehensive study of online catalogs: An overview and application of findings.* Final report to the Council on Library Resources, vol. 3. Dublin, Ohio: OCLC Online Computer Library Center, Office of Research.

Knutson, Gunnar. 1986. Does the catalog record make a difference? Access points and book use. *College & research libraries* 47: 460–69.

———. 1991. Subject enhancement: Report on an experiment. *College & research libraries* 52: 65–79.

Kurth, Martin. 1993. The limits and limitations of transaction log analysis. *Library hi tech* 11: 98–104.

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

Markey, Karen. 1983. *Online catalog use: Results of surveys and focus group interviews in several libraries.* Final report to the Council on Library Resources, v. 2. Dublin, Ohio: OCLC Online Computer Library Center, Office of Research.

———. 1984. *Subject searching in library catalogs: Before and after the introduction of online catalogs.* OCLC Library, Information, and Computer Science Series, 4. Dublin, Ohio: OCLC Online Computer Library Center.

Marner, Jonathan C. 1993. Measuring the success of keyword search strategy in an online catalog. *Technical services quarterly* 11, no. 2: 1–11.

Nielsen, Brian. 1986. What they say they do and what they do: Assessing online catalog use instruction through transaction monitoring. *Information technology and libraries* 5: 28–34.

Peters, Thomas A. 1993. The history and development of transaction log analysis. *Library hi tech* 11, no. 2: 41–66.

Peters, Thomas A., Neal K. Kaske, and Martin Kurth. 1993. Transaction log analysis. *Library hi tech bibliography* 8: 151–83.

Peters, Thomas A., and Martin Kurth. 1991. Controlled and uncontrolled vocabulary

subject searching in an academic library online catalog. *Information technology and libraries* 10: 201–11.

Prabha, Chandra G. 1990. Improving the quality and usability of large retrievals. *National Online Meeting, 11th*, 339–44.

Tolle, John E. 1983. *Current utilization of online catalogs: Transaction log analysis*. Final report to the Council on Library Resources, vol. 1. Dublin, Ohio: OCLC Online Computer Library Center, Office of Research.

Tolle, John E. 1984. Monitoring and evaluation of information systems via transaction log analysis. In *Research and develop-ment in information retrieval: Proceedings of the Joint BCS and ACM Symposium*, ed. C. J. van Rijsbergen, 247–58. Cambridge: Cambridge Univ. Pr.

Wiberley, Stephen E., Jr., Robert A. Daugherty, and James A. Danowski. 1989. *User persistence in scanning LCS postings: A report to the Council on Library Resources*. ERIC Document Reproduction Service No. ED 309 775. Chicago: University of Illinois-Chicago Library.

———. 1995. User persistence in displaying online catalog postings: LUIS. *Library resources & technical services* 39: 247–84.