

Usage Patterns of a Web-Based Library Catalog

Michael D. Cooper

School of Information Management and Systems, University of California, Berkeley, CA 94720-4600.

E-mail: cooper@socrates.berkeley.edu

This article reports on a model and patterns of use of a library catalog that can be accessed through the Internet. Three categories of users are identified: individuals who perform a search of the catalog, tourists who look only at opening pages of the library catalog's site, and Web spiders that come to the site to obtain pages for indexing the Web. A number of types of use activities are also identified, and can be grouped with the presearch phase (which takes place before any searching begins): the search phase, the display phase (in which users display the results of their search), and phases in which users make errors, ask the system for help or assistance, and take other actions. An empirical investigation of patterns of use of a university Web-based library catalog was conducted for 479 days. During that period, the characteristics of about 2.5 million sessions were recorded and analyzed, and usage trends were identified. Of the total, 62% of the sessions were for users who performed a search, 27% were from spiders, and 11% were for tourists. During the study period, the average search session lasted about 5 minutes when the study began and had increased to about 10 minutes 16 months later. An average search consisted of about 1.5 presearch actions lasting about 25 seconds, about 5.3 display actions, and 2.5 searches per session. The latter two categories are in the range of 35–37 seconds per session each. There were major differences in usage (number of searches, search time, number of display actions, and display time), depending upon the database accessed.

Introduction

The Internet has substantially increased society's access to information resources, including library catalogs. Many libraries offer access to their catalogs through browsers located at libraries or at users' homes or offices. This article develops a methodology for analyzing patterns of usage of a Web-based library catalog and evaluates that methodology with data collected from the transaction logs of the Web interface version of the University of California's (UC)

Melvyl® Systemwide Library Catalog. It is a descriptive statistical analysis with interpretations of the results, not inferential or predictive. The catalog can be accessed at www.melvyl.ucop.edu.

The Context for Transaction Log Analysis

Usage analysis is an important part of monitoring and evaluation studies. It allows collection of information to study user behavior, system design, as well as providing general management information, user account information, and customer support. It also allows analysis of specific research questions (Cooper, 1998).

The study reported here is part of a broad category of research that uses the results of the computer-generated transaction logs to analyze human–computer interaction. The goal of studying human–computer interaction is to model, understand, and improve the way in which a user employs a computer system's hardware, software, and telecommunications facilities.

When transaction logging is employed, each time a user performs an interaction with an information system, such as submitting a query or asking for help, the system writes records to a transaction file describing the nature of the action. These records are date and time stamped, and when reviewed, provide a picture of the user interaction with the system and the system response to the user request.

Transaction log analysis has been used extensively over the last 30 years to evaluate library automation systems. Many studies have employed transaction logs to evaluate user behavior, and many different methodologies have been employed in these studies. Peters (1993) and Borgman, Hirsch, and Hiller (1996) extensively review the past literature from both perspectives. Peters summarizes previous studies according to the system being studied (e.g., Medline, Melvyl, Innopac). He also summarizes the studies by the type of analysis conducted, such as frequency of commands issued, response time, session length, transition analysis between states, error rates, and the like. Borgman et al. summarize studies in many ways including the goal of the study: improve user interface and system design, qualitative

Received August 23, 1999; revised February 2, 2000; accepted May 2, 2000

© 2001 John Wiley & Sons, Inc.

evaluation of the systems, quantitative evaluation of operational systems, and capture task sequences in a session. Kurth (1993) summarizes the problems inherent in transaction logging, such as the complexity of obtaining the data, the large size of the data sets one must work with, the difficulty of replicating studies, the difficulty of adding logging facilities to existing systems, and the ability only to deal with the actions the user takes, not their perceptions, emotions, or background skills.

There are many ways of collecting information about user interactions, not just transaction logging. Alternatives for data collection include transaction logs, questionnaires, interviews, video analysis, and verbal protocol analysis (Henderson, Smith, Podd, & Varela-Alvarez, 1995). Transaction log analysis takes its place as a method for remote analysis without requiring the use of personally administered questionnaires or interviews. Studies of human-computer interaction can be performed in a laboratory or remotely, and data about the interaction can be collected in a variety of ways. A usability laboratory allows controlled experiments and careful monitoring of all user interaction. Remote evaluation can take place at a distance from the user-system interaction (Hartson, Castillo, Kelso, & Neale, 1996; Peters, 1998). Transaction log analysis facilitates remote evaluation, but even so, remote evaluation presents certain problems to the researcher. For example, it is straightforward to capture some events, like the search command used or the records displayed in response to a search. But other events are masked from standard remote logging mechanisms, such as when the user clicks on the *back* or *print* button on the browser software, or cuts or pastes information from one window to another on a client computer. Some software is available for logging these types of events, for example, the *Tracker* research package (Choo, Detlor, & Turnbull, 1998) and the software system from the WinWhatWhere Corporation (www.winwhatwhere.com).

In a previous paper, Cooper (1998) showed that transaction log analysis should use a number of different data sources. One is the traditional log file generated from online library catalogs, which records the user's interaction with the catalog. Another is the log file from the database management system that supports the storage and retrieval of the institutions bibliographic, text, and nontext records. This category of log file records actions performed on the database and the time and resources used to perform them. The third source are log files generated by the operating system that supports the application. These files give information on computer and telecommunication system performance, accounting, and capacity.

Some progress has been made in transaction log analysis since its review in Issue 42, 1993, of the journal *Library Hi Tech* and in Cooper (1998). For example, Yuan and Meadow (1999) codify many of the variables employed in studies of user behavior of information retrieval systems. This type of classification is relevant to transaction log analysis because it provides a uniform foundation for pre-

senting results. Yuan and Meadow's classification system includes six major areas (design of study, participants, learning, searches, variables related to search, and value measures). Within the major category *participants* are social variables (including education level, status, experience, gender, age, and socioeconomic level), and psychological variables (including intelligence, cognitive style, learning style, and symbolic reasoning skill). The major category *searches* includes variables such as query source, query evaluation methodology, searches per question, searches per participant, and information-seeking context. Variables related to search also include process variables (command set, chains of commands, and moves), errors, cycle data, and times. This codification is an important step in evaluating user studies.

Connell (1998) summarizes the literature on another type of taxonomy: user errors. Both user and system errors are encountered in transaction log analysis. Connell suggests that user errors are of two types, a mistake which is "... an error in the intention to act, and a slip, which is an error in carrying out that intention" (Connell, 1998, p. 929). It is important but difficult to differentiate these types of errors in transaction log analysis.

Articles are beginning to appear on the use of digital libraries and the use of the Web as one part of an academic library's research collection. Jones, Cunningham, and McNab (1998) report on usage analysis of a collection of 46,000 full-text computer science technical reports stored in a digital library in New Zealand. They discuss the structure of the transaction log file used to accumulate the data, the number of queries processed, and the frequency of use of Boolean operators in queries. They give information on the number of documents viewed and the ranked position of documents selected for viewing within a retrieved set.

Abramson (1998) employed a commercial Internet monitoring program to collect data on Web use by users at the American University library in Washington, DC. The number of fields of data collected is somewhat limited for each Web use, but the analysis gives data on usage by time of day and by individual URLs, as well as data on quantity of information transferred.

In 1986, Sewell and Teitelbaum (1986) analyzed more than 8,000 searches of the Medline system, and showed how patterns of use changed over an 11-year period. In 1991, Cooper (Cooper, 1991a, 1991b) modeled and analyzed changes in usage patterns with an office automation system over a 2-year period. He found that most users employ a small number of system features and functions and do not readily explore new components. He discovered differences in user behavior when data is analyzed by demographics and job functions. In 1997, Yuan (1997) conducted a controlled experiment with 43 subjects, monitoring their searching behavior with a legal information system over 12 months. He found that increased user experience broadened the use of system functions and features and increased user efficiency with the system.

This article continues research into how patterns of use of on-line library catalogs and information systems change over time. It uses transaction log analysis to focus not on individual users but on how overall patterns of use change. Throughout the article references are made to related studies of a command-line version of the Melvyl system by Larson (Larson, 1986, 1991), and Berger (1994) along with comparisons to the Medline system by Cooper (1983) to elucidate how usage has changed over time and with a different interface to the catalog.

A Model of a Session

When a user makes a connection to a library catalog through his or her browser, a *session* is initiated. The session lasts until the user overtly disconnects from the library catalog or until the library system's computer detects inactivity for a predetermined *timeout* period. There are three distinct types of sessions in this model. In the first, called a *real session*, a user performs some work, that is, conducts a search. In the second, the user connects to the system for a minimum specified length of time, performs some set of activities, but does not search any database. These users are called *tourists*, a concept first introduced in Cooper's analysis of user behavior of an office automation system (Cooper, 1991b). Tourists may perform useful functions, such as inspecting the site, deciding whether the interface is comprehensible, bookmarking the site, and evaluating usefulness. But after a brief period, perhaps because of a call to perform other activities, these users leave. In this article, tourists are defined as spending more than 10 seconds connected to the system. In the third type of session, users are believed to be artificial or at least extremely rapid thinkers and typers. These users are called *spiders*, and in this model are defined as spending less than 10 seconds connected to the site, without performing any database searches. These users are generally search engines, such as *excite*, *Yahoo!*, *AltaVista*, *Lycos*, *HotBot*, or the like, probing the site and downloading pages from the site to their own computer system for use in building indexes of Internet resources.

Parts of a Session

Once a session is initiated, the user performs a number of tasks. They include presearch, search, display, other, help and error actions. In this model, all the activities that take place before the initiation of the first search of a database are termed *presearch activities*. Examples include reading about the system and how it works, determining which databases are available for searching, selecting a database for searching, establishing a user profile, and registering. (The latter two activities are not required in the system being studied, but can be initiated if the user desires.) Because many users (spiders and tourists, in particular) will not continue with a session, the most frequently occurring

activity in the presearch phase, is to terminate a session and disconnect from the Web site.

The second category of activity is the *search*. A session can consist of multiple searches, each potentially performed on a different database. In the search process, the user enters a search statement through the Web interface to the catalog. The statement consists of selecting the database to be used, specifying (directly or indirectly) the indexes to be searched, the terms to be used, and any delimiters for the search, such as specific language requirements, data ranges, source institutions, and the like. The search process is iterative—the user starts a search, refines it, and repeats the process until satisfied. To refine a search, the user may take advantage of Web catalog features such as finding more records like the one just retrieved or limiting the number of records.

During the search process, the user often breaks out of the search activity to perform the third category of activity, *display* actions, which allow the user to view the results of a search. The display activity includes viewing the retrieved set in a default display mode, displaying a browse screen full of citations, showing only one citation from the browse screen, or changing the form of the display (from default, say, to showing the full citation plus abstract).

In a fully functional online catalog there are many other actions that can be performed and they are classified as *other* actions. The user can log off, download citations to his or her own computer, or have the results of a search mailed to a specified party. In addition, there are some activities that might normally be presearch activities, that become *other* actions once a search is initiated. They include switching to a different database, establishing a profile, or reviewing the contents of saved sets of records.

The user can request *help* on a variety of topics, and this is tracked as a separate category of activity irrespective of the context in which it occurs (presearch, search, display, other). And finally, users and the system do make errors, and these can be recorded as well.

The Transaction Log

A user connects to a Web-based library catalog with a browser by entering the URL of the catalog. The server computer's program (the *http daemon*) that processes the browser request acts as an intermediary by sending and receiving appropriate responses to and from client and server. As the browser works, it maintains a minimal log of the activities the user performs. Analysis of these log files is a popular pastime for organizations that make money by selling advertising space or that use volume of traffic to justify their presence on the Web.

For purposes of any in-depth analysis of user behavior, these http daemon logs are inadequate. To provide the level of detail necessary to support a comprehensive analysis, a transaction logging system was designed and implemented by the Melvyl Web Interface software development team using guidelines suggested in Cooper (1998). When a ses-

sion begins, a unique identification number is created. Each screen displayed to the user and each user response is recorded in the log file with that unique identification number and a date-and-time stamp. One log file is created for an entire day's worth of transactions. Log records are written in a format that maximizes writing efficiency but make statistical analysis difficult. For this analysis, I used Perl programs to parse, edit, and verify the log files, and a statistical analysis program (SAS) to perform tabulations and synthesis. I used a PC-based office suite to further synthesize the data. The log files serve as the basis for the analysis that follows.

The Database of User Transactions

The University of California introduced the first version of its Melvyl online library catalog in 1981. The complexity and usage of the catalog, which contains holdings for all nine university campuses and many other institutions, has grown rapidly over time. In the mid-1990s the University of California's Division of Library Automation (the unit then responsible for the Melvyl system) began working on a Web-based interface to the catalog to supplement and/or replace the existing command-line interface. This system became operational in early 1997. The analysis of user behavior spanned the period from June 11, 1997 to October 1998, beginning soon after the system was introduced and lasting 16 months. This was a period in which early adopters of innovation could have tried the system, but also one in which a certain amount of stability of searching patterns could have been achieved.

During the study period of 479 days, about 2.5 million sessions were recorded. These consisted of approximately 3.6 million presearch activities, 7.4 million searches, 13 million display requests, 11 million *other* requests and only about 60,000 help requests. The data processed for this article is not a sample; it includes all sessions that took place. The raw log files for the 479 days comprise about 18.2 gigabytes of data, and the data files, after parsing, consist of about 3.5 million characters for the help data, 753 million characters for the display data, and 868 million characters for the search data. This database serves as the basis for this analysis. The time taken to write the logging, parsing, and editing programs was significant, as was that used to optimize instructions to the statistical analysis program to process the parsed data. It was not uncommon for some statistical analysis runs to take more than 9 hours on very powerful Sun computer systems.

Sessions

Of the roughly 2.5 million sessions analyzed in this project, about 27% are from spiders, 11% from tourists, and 62% (1,527,971) are *real* sessions. Figure 1 shows the

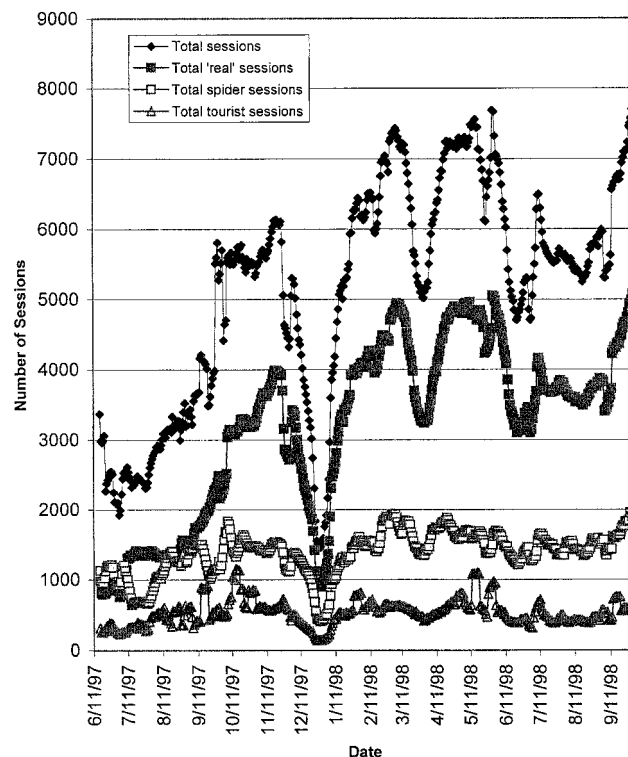


FIG. 1. Number of Melvyl Web sessions.

distribution of the sessions over the 479 day period.¹ In general, the number of spider and tourist sessions holds relatively steady in the range of 1,500 per day for spiders and 500 per day for tourists. It is quite interesting to find so many extremely short connections to the system that seem to be made for the sole purpose of indexing the Web.

There are major changes in the number of *real* searches over time: a precipitous decline over the Christmas and New Year holidays and variations due to start and end dates for semesters and quarters. Overall, the pattern of system use is on an upward trend reflecting the adoption of Web-based searching as an alternative to command-line searching of the same databases. There are major variations in session volume by both day of week and time of day. Overall, the number of sessions on Tuesday and Wednesday of an average week are about 18% each of the total. Larson (1986) found Tuesday's sessions to be about 21% of the total and Wednesday's, 19%, very similar to the results in the current study.

Figure 2 shows that the maximum number of *real* sessions occurs on Tuesday of an average week with about 280,000 sessions and drops off to about 100,000 sessions on Saturday. Evidently, searching gets done midweek.

Even tourist and spider sessions have some weekday variation, but it is relatively small. Hourly variations are

¹ In this and subsequent time-series graphs there is considerable day-to-day variation. The graphs have been smoothed using moving average techniques to clarify the trends.

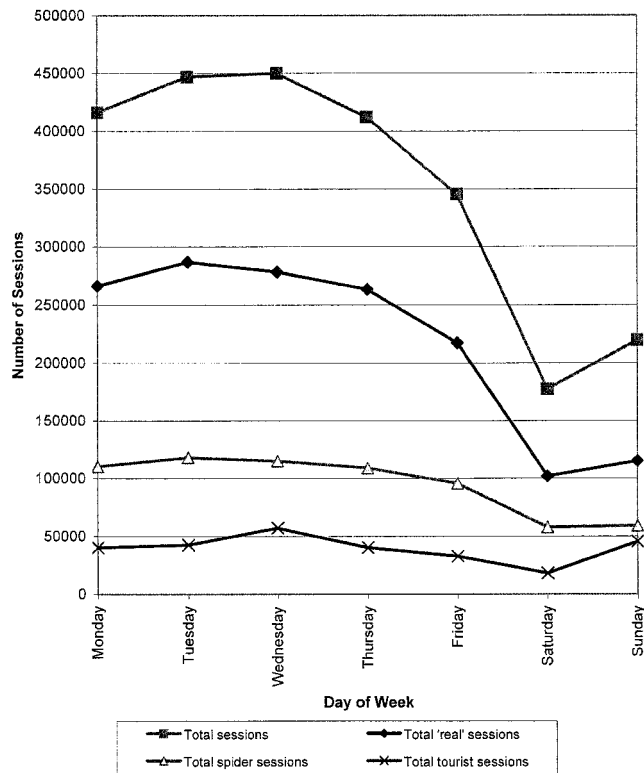


FIG. 2. Number of sessions by day of week.

also quite strong. The computer system that runs the Melvyl Web System is located in Oakland, CA, in the Pacific time zone (GMT-8 hours). Peak traffic occurs between 2 and 3 p.m., and it builds steeply to that peak from 8 a.m. onward. The early rise in traffic could be due to usage from sites east of California.² Relatively large traffic in off hours is a reflection of the popularity of the system throughout the world—Melvyl is recognized as an international resource with its very large database.

The average session length for all sessions is 617 seconds—10.3 minutes, with standard deviation of 1,070 seconds but varies substantially by category of session. In a study of 144 sessions, Berger (1994) found session time to be 8.1 minutes for the command-line version of the Melvyl system, and Cooper (1983) found a session time of 8 minutes for 60,000 sessions of the Medline system.

Session length is relatively constant for tourists, hovering around a mean time of 200 seconds (3.3 minutes), as shown in Figure 3. This is probably just enough time to find out what the site has to offer, and to decide whether to bookmark it.

The length of a *real* session gradually increases over time, starting at approximately 300 seconds (5 minutes) in June 1997, and increasing to about 600 seconds (10 min-

² Time-of-day and day-of-week variations occur throughout all activities discussed here, but are not reported further. The pattern in Figure 2 of low values on weekends, rising to a peak midweek, as well as hourly afternoon peaks are omnipresent.

utes) by October, 1998, a reflection of more sophisticated searching behavior, such as gaining experience with the system, exploring more system features, and using more databases, as we will see later.

There are few surprises in analyzing the session length by time of day and by hour of day. The length of a session holds relatively constant for *real* sessions from 9 a.m. to 9 p.m. and fluctuates considerably at other times. This reflects differences in user characteristics during off-peak periods, such as the possibilities that individuals from countries outside the United States may have more difficulty using the system.

The model defines six categories of actions that take place during a session: presearch, search, display, other, help, and error. Figure 4 summarizes the mean number of actions per session for each category (except help) over the 479-day study period. Display actions occur most frequently (about 5.3 times per session). This is quite a reasonable result, because a user will probably do one search and perform a number of display actions for the same search.

Overall, 10% of all actions took place in the presearch phase, 21% were for searching, 37% for display, and 31% for other actions. The small remainder of 1% was for error and help actions. In Larson's study (1985) of 169,000 sessions, 37% of the actions were search and 19% display. The highly interactive nature of a Web interface, and the increased database size probably explains the increased growth of display actions in the 15 years since the command-line interface was examined.

Early in the study period, the number of display actions is relatively low, perhaps reflecting user inexperience with

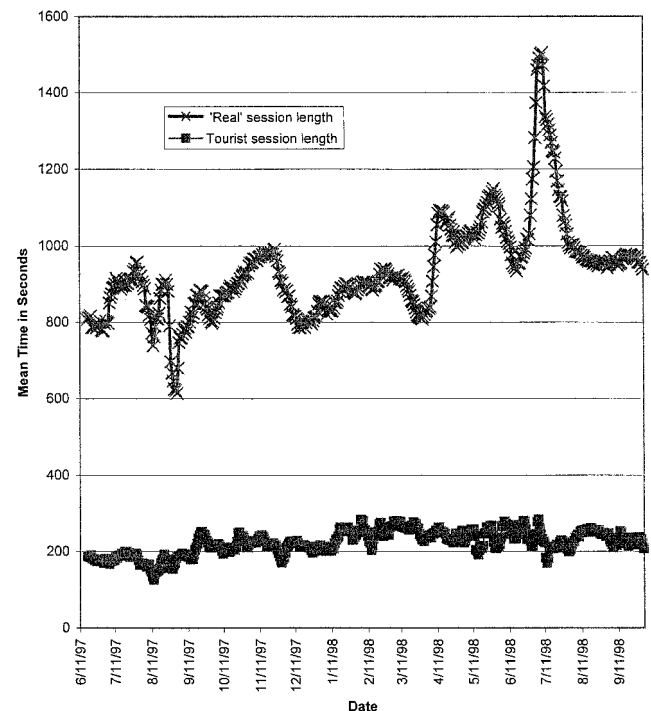


FIG. 3. Session length.

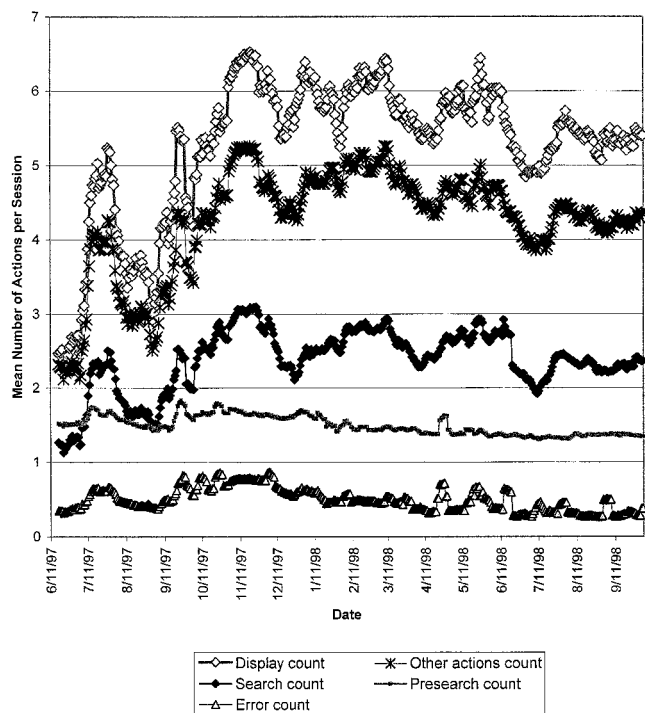


FIG. 4. Session characteristics.

the system, but after about 5 months it stabilizes at five to six actions per session, reflecting increasing user knowledge. Interestingly, the number of search actions is also lower in the first 5 months, but stabilizes, although its variance is generally lower over time. Presearch actions are almost constant, averaging 1.5 per session; error actions fluctuate within a narrow band of about one error every two sessions; and help actions are not shown on the graph because they are so small (mean of 0.03 help requests per session with standard deviation of 0.29). The most interesting part of the analysis of user session behavior comes in the examination of time spent in each of the activities. Users spend about 25 seconds per session performing presearch activities (very stable over the study period) and about 35–37 seconds per session performing each of the display, search, help, and other activities. Although on average, each takes about the same amount of time, each is performed with different frequency within a session.

Presearch Activities

There are about 938,000 presearch actions in the transaction logs for the sample period. As conjectured earlier in the paper, the dominant activity performed in the presearch phase is termination of the session. A full 26% of all presearch actions are just this. Every other action taken comprises less than 7% of the total, a very flat distribution of tasks performed, and will be discussed in subsequent paragraphs. The largest number of terminations come from users who examine the home page of the site and simply exit at that point (such as spiders).

The Melvyl Web home page gives users two categories of options: (1) to link to general information about the site (such as an introduction to the site, instructions for setting a profile, and available resources); and (2) to choose how the first search is to be conducted (namely which database to use and whether the first search is by author, title, subject, or some combination of the three [termed *power search*]). By a small margin, users preferred title over subject searches (6.3 vs. 6.1%), with author searches occurring 4.3% of the time. However, these statistics are for the choices the user makes during the presearch phase and differ from the choices made in the search phase (discussed later).

Few requests for general information are made. Users link to the area where they can establish a profile in 1.6% of the cases but only set up a profile 0.5% of the time. They ask for database information 1.2% of the time and general information 0.5% of the time.

Database Activity

The Melvyl Web Interface provides access to a large number of databases for searching. The *Catalog* database (referred to in this article as *Cat*) is the foundation of the Melvyl system. It contains more than 10 million bibliographic records, representing the holdings of all nine campuses of the University of California system and other institutions as well. The Melvyl Web Interface also makes many other databases available for searching including the *Medlars* medical literature database [produced by the U.S. National Library of Medicine (*Med*)], the *Magazine* database [from the Information Access Company (*Mags*)], *California Periodicals Database* [a database of periodical holdings of the nine U.C. campuses and other California libraries (*PE*)], the *Inspec* database [from the Institute of Electrical and Electronics Engineers (*Inspec*)], and *Current Contents* [from the Institute for Scientific Information (*CC*)].

In the 479-day period, about 7.4 million individual database selections were made. *Cat* accounts for 32% of all searches (about 2.4 million uses), *Med* for 22%, *Mags* for 10%, *PE* for 8%, *Inspec* for 4%, and *CC* for 4%. These six databases account for 80% of all selections (about 5.9 million uses). *Cat* database use is highly variable over time, ranging from a low of about 1,500 uses per day in mid June 1997, to about 9,000 in early March 1998, and ending at about 7,500 in early October 1998. The low initial numbers probably reflect a trend, mentioned earlier, toward adoption of the new search system, and the variations reflect semester and quarter research deadlines.

The University has nine campuses, some with specialized medical, agricultural, and other programs. Obviously, there are major differences in, say, the use of the Medline database across campuses, but these are not recorded in the tabulated data.

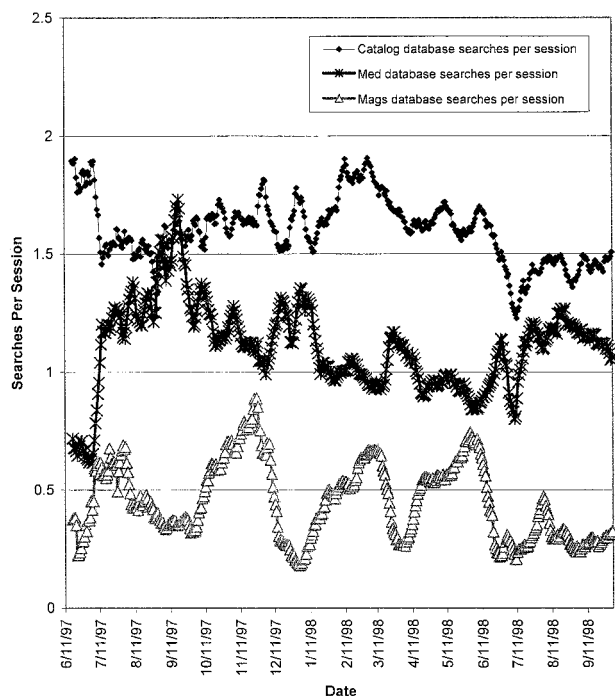


FIG. 5. Searches per session for selected databases.

Search Activity

Users of the Melvyl Web Searching System average about three searches per session. That same average search employs 1.7 Boolean *and* operators, 1.3 *and not* operators, and 1.5 *or* operators. User behavior varies significantly depending upon the database being searched. Figure 5 shows details for three selected databases. Over time, users average more than 1.5 searches per session with the *Cat* database, and more than one search per session for the *Med* database. In Larson's study of 5.3 million sessions (1991), he found 2.88 searches per session of the *Cat* database using the command line interface. Berger's study (1994) of 144 sessions of *Cat* found 4.9 per session. It appears that the Web searcher is much more purposeful, perhaps because it is much easier to log onto the system from home or work than it was in 1991 or 1994 when the user usually had to go to the library to conduct a computer search.

There is considerable variation in the number of searches per session ratio for the *Mags* database over the 479-day period. The high number for the *Cat* database can be explained, perhaps, by familiarity. Users have had access to the *Cat* database via other interfaces for a considerable time, and it is not new to them. The single *Med* search per session is not surprising, because medical searchers are likely to be searching for a specific answer than to be browsing.

The length of time a user spends performing search activities also varies considerably by database. Figure 6 shows search time for the *Cat* and *Med* databases over the study period. Although there are some variations, search

time for the *Cat* database reaches a relatively steady state of about 110 seconds (1.8 minutes) by the end of the period. Likewise, search time for the *Med* database stabilizes at about 175 seconds (2.9 minutes). Both graphs show search time diminishing over the study period (by about 40 and 25 seconds, respectively, for the *Cat* and *Med* databases). The most likely explanation is user experience with the interface: as time goes on, users gain more familiarity with the system becoming more efficient.

Subjectively, the *Med* database is more complex to search because of its use of a sophisticated thesaurus and the thesaurus' associated *explode* and *tree* commands to view and explore the thesaurus. This probably explains the longer search time for the *Med* database along with potentially more complex searches.

Many options are available to the user in formulating a search, but for the most part, users accept the default. The system allows the user to look for newly added material, but only 0.3% of all searches incorporate this feature. Users can select the form of material (e.g., books, video, dissertations) for their search, but again only 5.9% of the searches use this option. Users can select date ranges for materials to retrieve, but only 1.3% of searches use this capability. About 15% of all users specify the language of the material to be retrieved. Of those 15%, 98% say they want English materials. Most searches (40%) are power searches (ones that allow a combination of author, title, and subject to be specified along with Boolean operators and qualifiers). That is followed by subject (27%), title (18%), and name (15%) searches. Although it is difficult to compare because of category problems, Larson's (1991) results indicate 25% of the searches were subject searches, 27% title, and 16% name. Larson argued that subject searching was declining over time in his sample. The current results indicate a possible

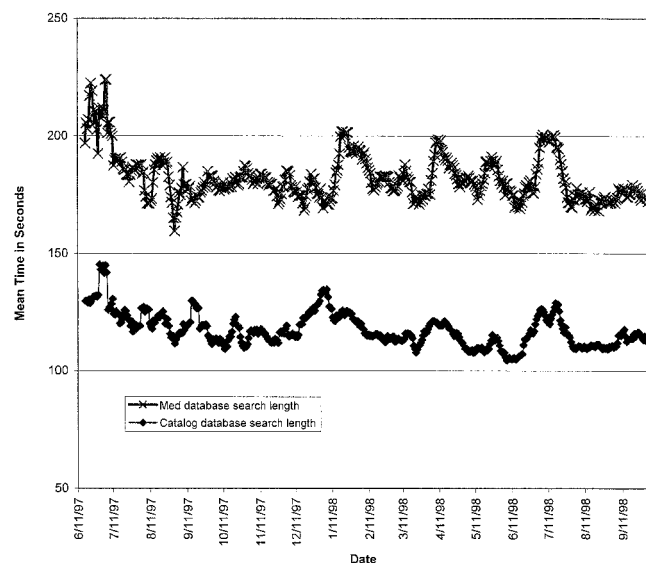


FIG. 6. Search length by database.

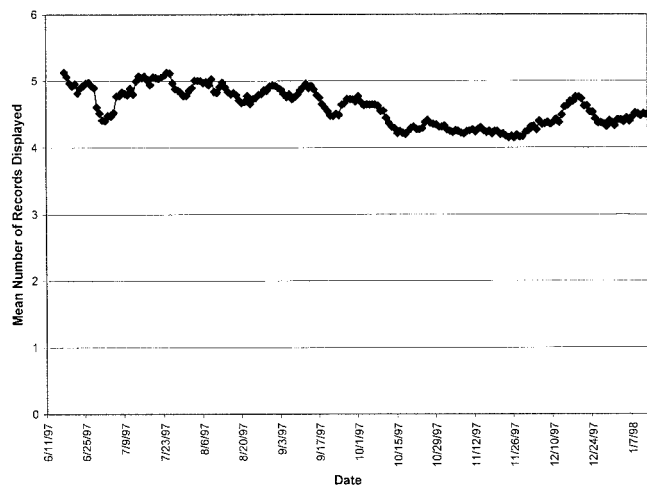


FIG. 7. Number of records displayed.

reversal, but no trend analysis was conducted to see the effect over time.³

From a systems design standpoint, the numbers reported above confirm a well-known phenomenon: most users are satisfied with the most standard features of a system, and a very small number take advantage of (what they consider) advanced features. One only has to view the bar menus of current word processing/spreadsheet/database software to realize how many more features of the software there are that one does not use frequently, if at all.

Display Actions

The search process is iterative. Users locate materials and display the results of their search. In Figure 4, presented earlier, it was shown that display actions per session were consistently the most frequently performed actions in the model. That is reasonable, because a user is likely to perform one search and then issue multiple requests to view the results.

Once a search is completed, the system presents the user with a summary screen of the number of hits obtained for the search, which is termed the *search history*. The user has the option of displaying the citations and/or text, modifying the existing search, or initiating another search. As was the case with the search actions, users have a variety of options available for displaying the results. Of the approximately 12 million display actions, users accept the defaults of author, title, publisher, date, pagination, and library holding the item 55% of the time. The remainder of the time, users change the display format. Three alternatives dominate their choices: a display including the citation and abstract (33%),

the citation alone (21%), and a complete display of all parts of the citation (19%).

In Larson's study (1986) of 43,893 display actions, 60% were considered *brief*, equivalent to the default (55%) in the current study. This difference does not seem to be significant. The relatively high percentage in both cases simply reflects the user willing to use the quickest and easiest option to display results. In Cooper's study of the Medlars database (1983), 41% were *compressed* or equivalent to the default here.

The number of records displayed during each display activity is relatively stable over the study period, with a mean ranging from four to five citations. There is no appreciable increase or decline over the more than 16 months of the study (Fig. 7). There are, however, differences in user behavior between databases. Figure 8 shows the trend for the *Cat*, *Med*, and *Mags* databases. On average, more records are displayed when the user is searching *Cat* than *Med* or *Mags*. It is possible that the reason the figure for the *Med* database is lower is that there is more precision in the indexing of the database and more sophistication in the user population. The low figure for *Mags* could be because the database size is small, and the database itself is new, with potentially unknown coverage, to the user.

Aside from an early dip at the beginning of the study period, the number of records displayed in the *Med* database decline over time, again probably reflecting increased user searching sophistication. The pattern is much more erratic for the *Mags* database, where fluctuations have been high.

One measure of user effectiveness at weeding materials in a search is the ratio of number of items displayed divided

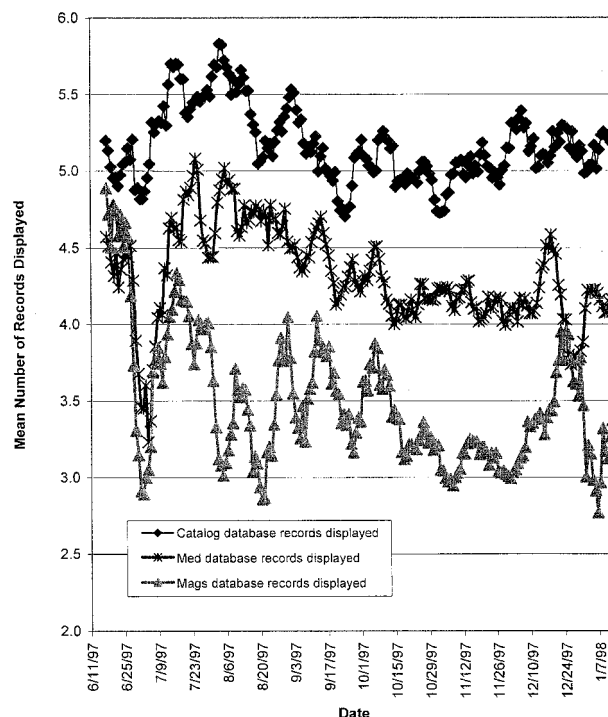


FIG. 8. Records displayed by database.

³ Recall that some of the selection of type of search takes place in the presearch phase in this model. The presearch percentages are within 2% of one another for subject, title, and name, and thus do not compensate for the decline in title searching reported here.

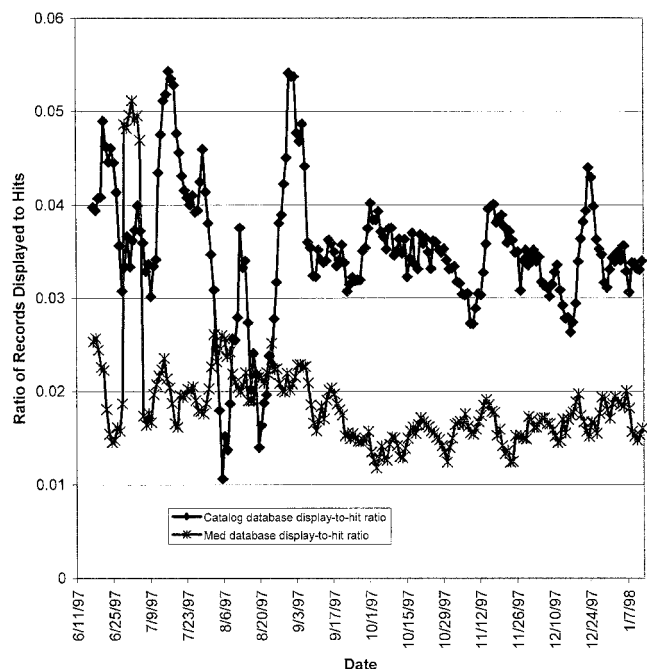


FIG. 9. Display-to-hit ratio for selected databases.

by number of items found in a search (hits). If this ratio declines over time, it tends to indicate that the user is becoming more effective. If the ratio for one database is higher than for another, it indicates more specificity in searching.

Figure 9 shows the pattern for the *Cat* and *Med* databases. Users of the *Med* database have a lower display-to-hit ratio over the entire study period than users of the *Cat* database. In general, *Med* database users display two citations for every 100 they retrieve, while *Cat* users display about 3.5 citations for every 100. These numbers again reflect a more focused user community of medical searchers.

Even though there are differences in the number of records displayed, the number of hits, and the display-to-hit ratio between databases, the time spent displaying records for two of the databases (*Cat* and *Med*) are remarkably similar (see Fig. 10). The two graphs are relatively stable over the study period, showing display time in the 30–40-second range, and they follow a similar pattern over time. Not shown are the patterns for two other databases (*Mags* and *Inspec*), which have more daily variation, but again, follow the same trend as the previous two databases. For all databases searched, the trend is toward a very slight reduction in time spent in this activity (starting at 40 seconds ending at about 38 seconds).

These results are quite interesting: even though users of some databases spend more time searching, retrieve more or less materials, and display more or less materials, the time spent viewing the results is relatively constant over databases. One explanation is that irrespective of the database, there are certain motor limits in place when an individual

scans citations on a screen that keep the time relatively constant.

Other Actions

The category of *other* actions includes functions that are not directly related to the search or display activities. Among these are events like viewing the history of all the searches performed, saving citations for downloading or mailing, formatting citations for printing with the browser's *print* button, and mailing citations to yourself or others. These are not new activities, but the Web browser interface makes the process of performing them much easier.

The number of these actions vary significantly over the study period. In June 1997, the total number per day was around 5,000, and by October 1998, it had climbed to 30,000, with significant swings in the interim. When the data is analyzed in terms of *other* actions per search, the variations are reduced. Figure 11 shows the number of *other* actions per search for the *Cat*, *Inspec*, and *Med* databases as relatively stable across the entire period. Although the total number of actions changes significantly, when normalized in terms of searches, there is little variation. The mean time spent performing this activity is roughly 50 seconds for all databases.

Clearly, there are important activities that the user performs in the search process that do not neatly fit into presearch, search, display, or help requests. These *other* actions are performed with a consistent frequency in all searches, and take a consistent amount of time to perform.

Help Activity

The Melvyl Web Catalog offers help to its users through links to descriptions of system features and functions. Ex-

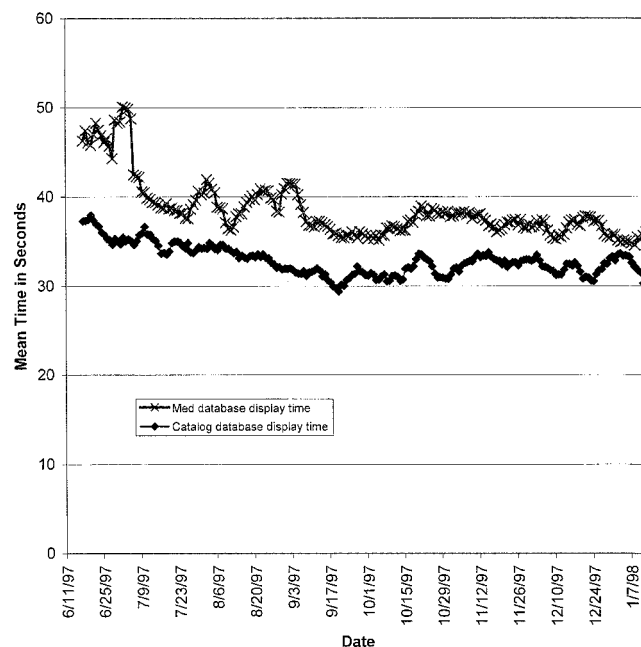


FIG. 10. Display time by database.

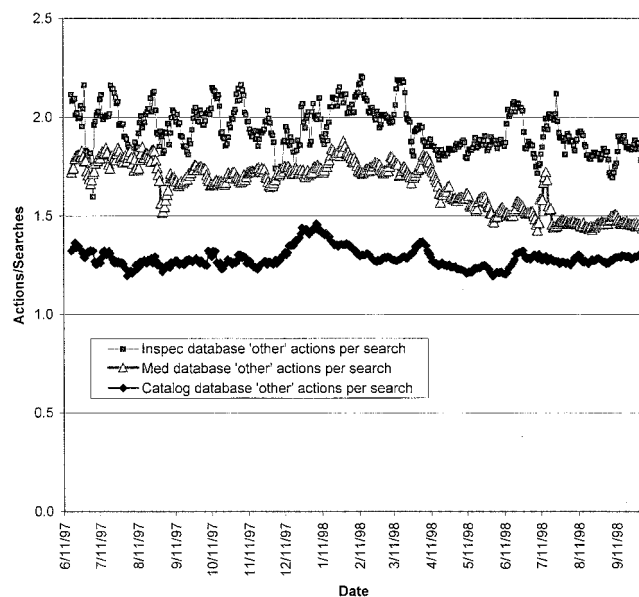


FIG. 11. "Other" actions per search.

amples include providing information on the types of electronic resources available in the system (which databases are accessible and descriptions of each), how to use the system, and how to perform common functions. The most common request is to describe the databases available, which constitutes 42% of the roughly 60,000 help requests.

The number of requests that users make for assistance is small relative to the total number of sessions and searches performed. This study sample consists of roughly 2.4 million sessions; obviously 60,000 is a very small percentage of that total (3%).

What explains the small number of help requests: the system is relatively easy to use, the interface conforms to the normal protocols and offers no surprises, the on-screen explanations are clear, and the user may not be encumbered by previous experience with command-line systems.

At peaks during the study period there were 250 requests for help in 1 day and a minimum of 50 in another. Generally, the number of requests range from 125–175 per day. When users do request help, the time they spend viewing help screens is stable in comparison to other activities, averaging about 37 seconds. The standard deviation of viewing time is high, 84 seconds, reflecting individual variations.

Even though help usage was small, an attempt was made to ascertain if there were any distinguishing characteristics of a search that might lead users to request help. Viewed by the number of hits obtained when a request for help is made, the pattern is not significantly different from that seen before. Likewise, viewed by the number of records displayed, the pattern is also similar. There is no significant difference between the number of help requests made by time of day and day of week and the number of sessions in the same time of day or day of week, thus disputing the idea that there might be more help requested at certain times of the day or week (see Fig. 12).

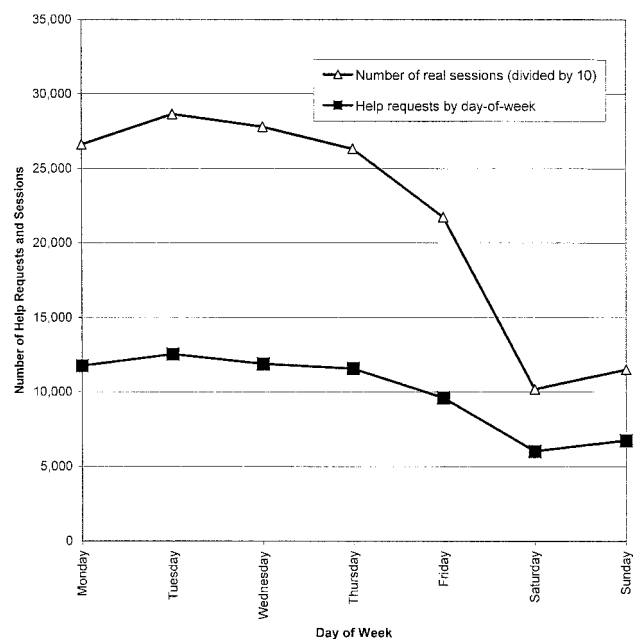


FIG. 12. Number of sessions and help requests by day of week.

When the number of help requests is broken down by database, there are significant differences. Figure 13 plots the number of help requests as a percentage of database use for two databases (*Cat* and *Med*). Although not shown on the graph, the trend for the *Inspec* and *Mags* databases are so close to the *Med* database line that they would overlap if plotted on the same graph. One possible explanation why the number of help requests for *Med* is lower than *Cat* is that users of the former database

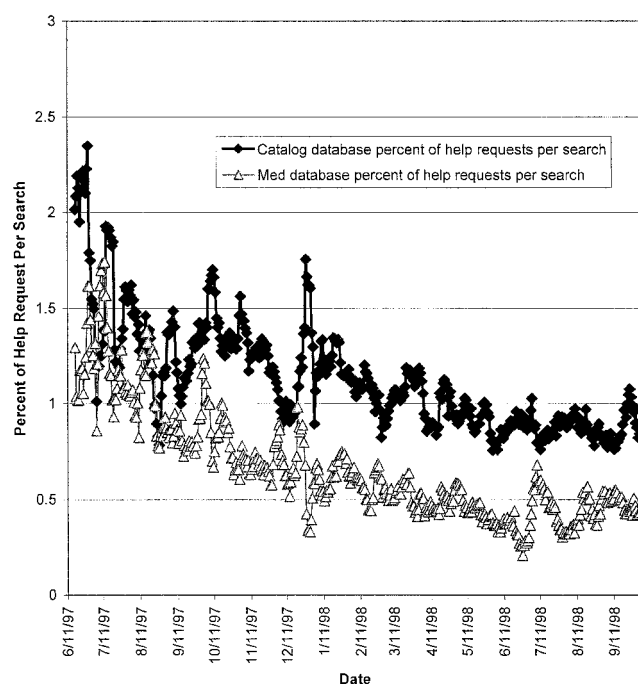


FIG. 13. Help requests as a percentage of database use.

have more experience with its command-line version or more general searching experience. Interestingly, the graph shows a general decline in help requests over time. This conforms to a reasonable assumption about the search process: as users gain more experience with a system, they are less likely to require help, so more requests for help will be made when users are new to the system.

Summary, Conclusion, and Future Research

This article developed a model of user interaction with a Web-based library catalog and evaluated the model with data on catalog use collected over 479 days. It presented a first look at some extremely rich observations of user behavior within an information system.

The model defines types of users, including *real* sessions, *tourists*, and *spiders*. The latter is defined as a program that connects to the system for 10 seconds or less, presumably to collect data for use in indexing the Web. Tourists are defined as individuals who view the site, stay longer than 10 seconds, but do not perform a search.

The model also defines stages or parts of a search, namely *presearch*, *search*, *display*, *other*, *help*, and *error*. The presearch phase comprises all activities performed before the user begins the first search. The search and display phases can happen iteratively: the user performs one search and then displays the results of that search. Actions not directly related to the search or display process that happen after the first search has begun are classified as *other*, and include selecting citations for saving, mailing results of a search to a user, setting up a profile of search preferences, and the like.

During the study period, users conducted about 2.5 million sessions, during which about 7.4 million search statements were executed. Not all the sessions resulted in the performance of searches. About 27% of the sessions are believed to have come from spiders. Another 11% of the sessions from tourists. Over the study period, the number of spider and tourist sessions remained relatively constant, but the number of *real* sessions continued to grow, albeit erratically, with semester and quarter deadline fluctuations. The length of time *real* searchers spent at a session grew gradually from about 6 to 10 minutes in the more than 16 months of the study period. Users made very few errors and made fewer requests for help over time. Numbers of searches per session and numbers of display actions per session were consistent once the initial learning period with the system was over (about 6 months). There were significant differences in user behavior when searching different databases. Users performed about one search per session of the *Med* database, while they performed about 1.5 on the *Cat* database. The time they spent searching each database also varied, from about 3 minutes per search for the *Med* database to about 2.2 minutes for the *Cat* database.

Users displayed a range of four to five citations for each search they performed, but the range was quite different when broken down by database. The length of time users spent displaying results was not substantially different—generally in the 30–40-second range throughout the study period for the *Cat* and *Med* databases. No substantial differences were found in the number of help requests made by day of week or time of day. Nor was there any difference in the number of help requests made when normalized for the number of database uses.

The model, data, and statistical analysis reported here provide a foundation for further significant research. In this article the general characteristics of user behavior have been outlined, and trends over time discovered. We now know average times for a search, frequencies with which actions occur over a large number of observations, the likelihood of the user asking for help or making an error, the average complexity of a search statement, and the like. These numbers provide a benchmark for fitting equations and calculating deviations and residuals.

There are a number of future steps in this research. They begin with examining individual user behavior. The ultimate goal is to catalog session patterns so as to be able to explain, and then predict, which patterns lead to productive searches. This is not an unreasonable goal even without collecting demographic data about the user. There are variables inherent in the Web-browsing process that show considerable promise in explaining and predicting user search success.

The pattern-detection process can begin with cluster analysis of individual sessions, forming the sessions into groups based on behavior observed in these data files. Subsequent analysis will use techniques such as directed graphs, Markov models, and Bayesian models to build patterns. As searchers, we intuitively know that there are patterns to our work, and these models will detect them.

Additional analysis of the same data will examine how a user changes his/her pattern of use of the system over time. This analysis will include measuring the similarity of the terms the user employs to search the system. The hope in this phase is to be able to detect stepwise refinement of the search, vocabulary problems, and differentiate when there are multiple purposes of a session. This later problem is a significant impediment in understanding user behavior because the session may be composed of searches that are totally different in structure and defy normal analysis. Once term similarity measurement has been accomplished, it will be used to analyze overall patterns of user session behavior over time.

Acknowledgments

I am grateful for assistance throughout this project from Lynne Grigsby-Standfill, Michael Berger, and Bob Brandriff from the University of California's Division of Library Automation in Oakland, CA. They answered innumerable

questions and clarified many issues. At Berkeley, I have benefitted from large quantities of sage advice from Hui-Min Chen, Aylin Küntay, and Walter Wei-Ti Huang. Also, my thanks to Elisabeth Beller for her suggestions to improve the article itself.

References

- Abramson, A. (1998). Monitoring and evaluating use of the World Wide Web in an academic library: An exploratory study. *ASIS '98. Proceedings of the 61st ASIS Annual Meeting*, Vol. 35, (pp. 315–326). Medford, NJ: Information Today, Inc.
- Berger, M.G. (1994). Information-seeking in the online bibliographic system: An exploratory study. Ph.D. Dissertation, School of Library and Information Studies, University of California, Berkeley.
- Borgman, C.L., Hirsh, S.G., & Hiller, J. (1996). Rethinking online monitoring methods for information retrieval systems: From search products to search process. *Journal of the American Society for Information Science*, 47, 568–583.
- Choo, C., Detlor, B., & Turnbull, D. (1998). A behavioral model of information seeking on the Web: Preliminary results of a study of how managers and IT Specialists use the Web. *ASIS '98. Proceedings of the 61st ASIS Annual Meeting*, Vol. 35, (pp. 290–302). Medford, NJ: Information Today, Inc.
- Connell, I. (1998). Error analysis of ticket vending machines: Comparing analytic and empirical data. *Ergonomics*, 41, 927–961.
- Cooper, M.D. (1998). Design considerations in instrumenting and monitoring Web-based information retrieval systems. *Journal of the American Society for Information Science*, 49, 903–919.
- Cooper, M.D. (1991a). Failure time analysis of office system use. *Journal of the American Society for Information Science*, 42, 644–656.
- Cooper, M.D. (1983). Usage patterns of an online search system. *Journal of the American Society for Information Science*, 34, 343–349.
- Cooper, M.D. (1991b). User skill acquisition in office information systems. *Journal of the American Society for Information Science*, 42, 735–746.
- Hartson, H.R., Castillo, J.C., Kelso, J., & Neale, W.C. (1996). Remote evaluation: The network as an extension of the usability laboratory. In *proceedings of the CHI '96, 1996, Vancouver, BC, Canada, April 13–18, 1996* (pp. 228–235). New York: ACM.
- Henderson, R., Smith, M., Podd, J., & Varela-Alvarez, H. (1995). A comparison of the four prominent user-based methods for evaluating the usability of computer software. *Ergonomics* 39, 2030–2044.
- Jones, S., Cunningham, S.J., & McNab, R. (1998). Usage analysis of a digital library. In: *Digital Libraries 98, Third ACM conference on digital libraries*. Pittsburgh, PA, June 23–26, 1998 (pp. 293–294).
- Kurth, M. (1993). The limits and limitations of transaction log analysis. *Library Hi Tech*, 42, 98–104.
- Larson, R.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.
- Larson, R.R. (1985). Workload characteristics and computer system utilization in online library catalogs. Ph.D. Dissertation, School of Library and Information Studies, University of California, Berkeley.
- Peters, T.A. (1993). The history and development of transaction log analysis. *Library Hi Tech*, 42, 41–66.
- Peters, T.A. (1998). Remotely familiar: Using computerized monitoring to study remote use. *Library Trends*, 47, 7–20.
- Sewell, W., & Teitelbaum, S. (1986). Observations of end-user online searching behavior over eleven years. *Journal of the American Society for Information Science*, 37, 234–245.
- Yuan, W. (1997). End-user searching behavior in information retrieval: A longitudinal study. *Journal of the American Society for Information Science*, 48, 218–234.
- Yuan, W., & Meadow, C. (1999). A study of the use of variables in information retrieval user studies. *Journal of the American Society for Information Science*, 50, 140–150.