

MAGAZINES ONLINE: USERS AND USES OF FULL TEXT

Carol Tenopir, Diane Nahl-Jakobovits, and Dara Lee Howard
School of Library and Information Studies, University of Hawaii at Manoa
Honolulu, HI

ABSTRACT

A user study of online magazine full text databases was conducted to examine: the search purposes of end users; search strategies tried by end users; the relationship between strategy and purpose; and relevance judgments of end users. Subjects were free to pursue their own goals with a monitor present to assist them when necessary. Information was gathered through interviews, transaction logs and tape recordings of end user-monitor verbal interactions. Preliminary results from 11 subjects show that the most common use was to locate articles about a topic. Other uses include browsing, partial document retrieval, and fact retrieval. Search strategies were initially linking, followed by narrowing strategies, including selecting value-added fields. Protocol analysis reveals that online searching is an emotionally charged activity involving affective, cognitive and sensorimotor acts.

INTRODUCTION

Research on user willingness to search full text databases has concentrated on users and databases oriented to the research world [1-6]. Increasingly, interest is being shown in the users and uses of online full text popular periodical databases such as Magazine ASAP and Trade and Industry ASAP, both of which include magazines usually available to and read by the general public [7,8]. Particularly interesting questions about these databases and their users include:

- (1) what types of goals will end users pursue,
- (2) what search strategies will be tried,
- (3) what is the relationship between purposes and strategies in both successful and unsuccessful searches,
- (4) what are appropriate measures of end user searching in these full text databases.

The current research begins to investigate these questions. This report concentrates on (1), (2) and (4).

METHOD

The study was conducted at the University of Hawaii under a grant from the Council on Library Resources. It examined the use of a general purpose magazine full text database by end users in an academic environment. A two hour training session covered a selected set of basic search commands, and an introduction to the technique of "thinking aloud." This involves verbalizing their on going focus, reactions and decisions during the search. Each end user searcher did up to five hours of online searching free of charge. During each search session, a trained monitor assisted with the syntax of the search language, encouraged the searchers in verbalizing their thought processes, and set up the equipment. Searchers chose their own topics. All sessions were both downloaded and tape-recorded. Copies of the downloaded sessions were provided to searchers after each session.

Database and Search System

The searching was done in the Magazine ASAP [TM] database from Information Access Company, using the DIALOG search system accessed via DIALOGLINK telecommunications software. Magazine ASAP contains the full text of over 100 general interest magazines from 1983 to the present. Categories (and examples) of magazines indexed include news (Time and Newsweek), business (Forbes and Money), hobby (Popular Photography and Popular Mechanics), political/commentary (New Republic and Nation), women's (Ladies Home Journal and Redbook), entertainment (Sports Illustrated, Teen and Rolling Stone), and science (Scientific American and Psychology Today). Aimed at the non-specialist reader, the printed versions of these magazines use many visual images, such as photographs, tables, charts, and diagrams. The electronic versions normally include only the text portion of articles and picture caption headings.

The DIALOG search system is representative of the commercially available online access and search systems. For full text searching, it allows access to all words in the text, except stop words, and uses boolean operators, proximity operators, and truncation. For this investigation, selected DIALOG commands were introduced to searchers in a two-hour, hands-on training session. They were taught enough basic DIALOG commands to search, retrieve and view articles and portions of articles. The commands were described to, demonstrated to and practiced by the searchers.

The Searchers

There were 11 volunteer end user searchers, including undergraduates, graduate students, and faculty, mostly from the social sciences and humanities fields. All of the searchers had some experience with microcomputers and all but one had used the University of Hawaii online public access catalog. Some had used a CD-ROM database, or an online literature system once (BRS After Dark, Knowledge Index, DIALOG, and LEXIS). The searchers were familiar with computer based systems but had almost no experience with literature searching.

We assessed the level of familiarity with the periodicals in the database by asking searchers to indicate which magazines in Magazine ASAP they subscribed to or had read at least once. Surprisingly, no one magazine was familiar to all the searchers. Fortune and Sports Illustrated were read by 7, followed by PC Magazine, Time, US News and World Report, and Forbes with 5 readers each. Of the total 106 periodicals in the database, 57 were known to at least one searcher. The largest number of periodicals read by a single searcher was 23, the least, 3.

Data Collection Techniques

Three techniques were used to collect information before, during, and after the search sessions: (1) searchers were

interviewed at the beginning and the end of the investigation to collect their opinions of and reactions to the experience; (2) search sessions were downloaded via DIALOGLINK software to provide a transaction log for later analysis; and, (3) search sessions were tape recorded to capture the searchers' verbalizations.

PRELIMINARY FINDINGS

Uses of the Database

Table 1 lists some of the searchers' actual uses. These findings show that the major use was to search for full documents on a specific subject. Upon determining the potential relevance of a cited article by using KWIC or viewing the article online, most people chose to download the article for later examination rather than read it completely online. Since searchers were not charged for the sessions, this may have influenced their decision to download the full text. Responses to the post-search interview question concerning the use made of the printouts of the articles or lists of citations show that 2 searchers had not yet looked at the printout; 1 person used the downloaded disk on his wordprocessor; and 4 people located the magazines in the library and obtained photocopies of articles.

Several uses of parts of the documents were found. One was to find initial or background information by browsing through the full text of the article. Less commonly, searchers retrieved articles so that the information could be used to answer a different question from that posed in their query. In such cases searchers conceptualized a general strategy to accomplish a specific goal. For

example, one searched for research in her area being done at university medical centers. Her real interest was in finding which medical centers were doing research in her area. A third partial document use was for fact retrieval, though this was not tried very frequently. The tape recordings reveal that the retrieval of documents judged to be partially relevant was sometimes a hoped for result, and some times a pleasant surprise.

Search Strategies

The most common search strategy used was linking of concepts by boolean operators. To broaden a search, truncation was liberally used, but rarely were synonyms added, or the initial search words changed. Searchers repeatedly used the same words for the concepts they were searching. To narrow a search, the most frequently used technique was to successively restrict the scope of the search with proximity operators. Initially, (S) was used, followed by (10N) and (5N), as had been taught in the training session. Following this, several searchers employed value-added fields such as descriptors or SIC codes, usually as an aid to increase precision. Several people began new searches after discovering this tactic, using the value-added field to restrict their broader concepts (e.g. dance) rather than their narrower ones (e. g. China).

Protocol Analysis of Tape Recordings and Transaction Logs

A behavioral approach to the study of end user searching views using or searching as a stream of actions or decisions guided by a goal. This is analogous to a computer program; the behavior is analogous to the program, and the program steps are like the end users moment to moment

TABLE 1. TYPES OF USES OF MAGAZINES IN FULL TEXT

USE	EXAMPLE
document retrieval (online)	find and download articles about Chinese dance
document location (print)	find citations for relevant articles that our library has
browsing for background information	read or scan articles or portions of articles about telecommunications
browsing a magazine's issues	scan articles from Science issues from last year
fact retrieval	what was the date and time of the Challenge explosion?
find peripheral mention of something or someone	find all mentions of Ellison Onizuka or a certain court case
word in context retrieval	locate articles that use a set of vocabulary words for students to read in an English as a Second Language class
special features	find articles about "computer use with physical therapy" that have photographs or pictures in the print version so I can show them in a class report
specific items	find recipes with certain ingredients or for quiche

decisions. In addition, it views decisions as reactions to an immediate problem that has arisen, and for which a solution must be found in order to reach one's goal, the last step in a program sequence. Finally, the behavioral view provides a three-way classification for all human behavior (feelings, thoughts and actions) [9].

An example from Table 2 in lines 5 and 6 illustrates how the behavioral approach analyzes a single end user act within the searching stream. The monitor provides the **stimulus** which interrupts the stream of search behavior. In this case, the monitor says, "K W I C space." To the end user, this stimulus is a problem to be solved by **reacting** appropriately. This reaction occurs in three functionally corresponding domains. Affectively, the

searcher wants to understand the problem. This produces homeostatic stress which subsides only when a solution is found. Cognitively, the person's plan is to request confirmation of her comprehension of the monitor's instruction. Finally, in the sensorimotor domain, the person repeats aloud "K? W?" and keys in the characters. All search behavior can be examined through this behavioral approach.

The behavioral approach has evolved the think aloud technique to obtain information on search acts. Think aloud protocol data have been used to study online catalog searching [10], reference interview negotiation [11], and online periodical databases [12]. The think aloud approach yields information about searching that is different from

TABLE 2. PROTOCOL ANALYSIS OF A SEGMENT OF A SEARCH SESSION

M= Monitor S= Searcher		
[A] = Affective Behavior [C] = Cognitive Behavior [S] = Sensorimotor Behavior		

VERBALIZATIONS	DOMAIN	BEHAVIORAL ACT
1] M: Try "set"	[C]	suggests command
2] M: Try type in the word "set space"	[S]	states keying instructions
3] S: S-E-T space?	[C]	requests confirmation
	[S]	performs keying
4] M: Mm hmm.	[A]	gives confirmation
5] M: K-W-I-C space	[C]	further states command
6] S: K? W?	[S]	repeats command part, performs keying
	[C]	requests confirmation
	[A]	vocal stress
7] M: I-C	[S]	restates command part
8] S: K? W? I? SPACE	[S]	repeats command part incorrectly
	[C]	requests confirmation
	[S]	performs keying
	[A]	vocal stress
9] M: "C" after the "I"	[S]	re-phrases command part
10] S: Hmm?	[A]	vocal stress
	[C]	confusion
11] M: Can you, can you add the C after I?	[C]	re-phrases instruction
12] M: Try "C space"	[S]	restates command part
13] S: (Heavy sigh)	[A]	vocal stress
14] M: And add C after I, see it's KWIC	[C]	re-phrases instruction, gives context
15] S: (Giddy voice) after I, let's try again	[A]	flooding out--emotionality
16] M: Let me do it. I'm not supposed to be doing it anyway, but I'm lazy	[A]	flooding out--emotionality
	[S]	takes over the keyboard
17] S: (Laughing) Oh "C! C!" A little longer "el"	[A]	relief
	[C]	insight
18] M: I set KWIC, ...so it should be typed with a C	[C]	states rationale

transaction logs and post-session interviews. Concurrent verbalizations are self-witnessing accounts. They are objective descriptions of what searchers perceive themselves to be doing while engaged in a problem solving activity. The accuracy of this kind of description improves with training, until it yields protocol sequences that enable researchers to create artificial intelligence programs that simulate the steps of human problem solving [13].

The audio tapes were played back while reading the transaction logs. The attempt was to identify categories of search behavior. To do this, the three broad domains of human behavior were applied to the actions of the subjects, namely, affective, cognitive, and sensorimotor. This psychological classification provides a revealing analysis of how online searching involves three domains of behavior [14]. Affective online search behavior refers to emotionality, stress, frustration, excitement, surprise, or disappointment. Cognitive online search behavior refers to strategy decisions and command choices. Sensorimotor online search behavior refers to keyboarding, command formatting, and locating information on screens.

Table 2 is a portion of the transcript of the first search session by a graduate student. This example shows the interaction between the searcher and the monitor, as well as between the searcher and the database and the online environment. The behavior of the monitors tends to vary in terms of how much they include the end user in their decisions. When explaining or providing confirmation, they either tell the subject directly what to do (exclusion) or they try to talk it through to make it a joint decision, or to provide the reason for the instruction (inclusion).

The degree of inclusion/exclusion is an index of rapport, or affective communication. Such communication is a significant part of the search. For example, in Table 2, lines 1 through 18, the searcher repeatedly requests confirmation when the monitor gives instructions for a command format. The searcher's confusion (cognitive error) is evident in the tape recording by the questioning and rising tone of voice. In the end the situation is resolved

when the monitor takes over the keyboard, giving up the attempt to communicate the instruction verbally. Lines 14 and 18 are the only instances where the monitor attempts to include the searcher in the meaning of the instructions. Inclusion may be an important strategy to facilitate end user searching. Affective reactions by both searchers and monitors influence the progress of the search, as well as searchers' satisfaction with results. We are currently performing more behavioral analyses on the tapes and transaction logs. They reveal that online searching is not simply a systematic cognitive process, but that affective search skills are essential to successful online searching.

Users of Full Text Databases

The preliminary impressions reported by users are based on the responses to the pre- and post-search interview questions. In both interviews, one set of questions queried the circumstances under which searchers would spend (1) \$5-20 and (2) more than \$20 of their own money on an online search. The responses to both pre and post-search questions indicated that searchers stated they would pay up to \$50 for a work related search, if it would meet an urgent, important, and unique need. For example, several searchers indicated they would pay to find information for a valuable need which could not be found by other means.

Responses identified six general motivations for using an online database: (1) personal use (2) work-related, (3) cost, (4) database contributions, (5) value, i.e., importance to user, and (6) time. A sampling of the remarks is shown in Table 3. The majority of the searchers were strongly influenced by the relationship of the information need to their work. Only one searcher said she would use online searching for personal use. Cost was seen by some as a limiting factor. Several searchers expressed an upper limit on the amount they would pay for a search. However, since their searches were without charge, it is unlikely that the searchers had an accurate understanding of the real costs of an online search

Remarks on the anticipated relevancy of the citations and

TABLE 3. SELECTED RESPONSES TO QUESTIONS ON PAYING FOR ONLINE SEARCHES.

F = Faculty G = Graduate Student U = Undergraduate Student

<u>USER STATUS</u>	<u>COMMENTS ON PAYING FOR A SEARCH</u>
(U)	To have access to information for research or personal use (work-related, personal use)
(G)	Research project--master's thesis or long paper. If more than 5 hours to do, then would do online. (work-related)
(F)	If there is a financial benefit, would spend but not just for personal interest (value, work-related)
(F)	Up to \$50; to know precise title and price of a software package I needed in next week; temporal, specific need, couldn't get answer elsewhere and don't know who else to ask (cost, time, DB contribution)
(G)	It would depend on the time it would save me (time)
(F)	If I were confident the database was relevant and contained important information (DB contributions)
(U)	I won't go over \$20 (cost)
(G)	No other source for information (DB contribution)

the uniqueness of the database as a source were mentioned as influences, particularly in the post-search survey. A common position expressed was that the online database would be used if no other source was available. An interesting follow-up probe would be to examine what this really means, since most of the database content is available in other forms, albeit less convenient ones.

Another motivation expressed was that of expected value to be derived from using the database. Among the values expected were financial benefit, multiple use of results or ability to create a marketable product. Time was identified separately. Time both in the sense of saving time and of meeting an urgent need was expressed as an influence by several searchers, with more interest shown in the post-search survey.

A noticeable change between the pre- and post-search interviews occurred in two of the factors mentioned. Work related use showed a slight decrease in the number of times it was mentioned; on the other hand, use due to database considerations increased. The reason for these slight changes may be the mismatch between the academic searchers and the database, which is general purpose in nature and not oriented to academic research. Also, exposure to online searching broadens one's decision-making base, perhaps allowing the nature of the database to be a factor because more is known about it. This appears to have occurred even when the database was not particularly pertinent to the searcher's needs.

Research Issues

Protocol analysis of think aloud data provides a rich resource for the study of search behavior. Some of the questions that interest us are: what is the cause of particular errors; what are the types of errors; what changes can database producers implement to make full text searching more effective; what is the influence of affect on strategy formulation; in what ways does choice of initial strategy influence subsequent strategy decisions; how does the search goal evolve in reaction to relevance judgements of retrievals; what are the dimensions of relevance judgements for full text retrievals; and what blindnesses develop in searching. We are currently working on analyses that provide information on these areas.

Acknowledgements

The authors wish to acknowledge the assistance of DIALOG Information Services, Information Access Company, Veronica Lane Associates, and the Council on Library Resources. We particularly wish to thank the graduate student monitors, Man Evena Shu, Tina Sutton, and Jen Ping.

NOTES

- (1) Kay Durkin, Janet Egeland, Lorrin R. Garson, and Sheldon W. Terrant, "An Experiment to Study the Online Use of Full-Text Primary Journal Database," in Proceedings of the 4th International Online Information Meeting: 1980 December 9-11, (London, England. Oxford: Learned Information Ltd., 1980), 53-56.
- (2) Lorrin R. Garson and Stanley M. Cohen, User's Manual. Primary Journal Database ACS Full-Text

File. (Washington, DC: American Chemical Society, 1983).

- (3) Sheldon W. Terrant, "Special Report: Publishing Scientific Information Today...And Tomorrow," Chemical and Engineering News, 61 (April 25, 1983) 51-58.
- (4) Sheldon W. Terrant, Lorrin R. Garson, Stanley M. Cohen, and B E. Meyers, "ACS Primary Journal Online Database," in Proceedings of the National Online Meeting: 1983 April: New York, NY. (Medford, NJ: Learned Information Inc., 1983).
- (5) Karen J. Winkler, "Chemical Society Offers Full Texts of Scholarly Journals by Computer," The Chronicle of Higher Education, (June 22, 1983) 23-24.
- (6) Jack Franklin, Michael C. S. Buckingham, J. Westwater, "Biomedical Journals in an Online Full Text Database: A Review of Reaction to ESPL," in Proceedings of the 7th International Online Information Meeting: 1983 December: London, England (Oxford, England: Learned Information, Ltd.; 1983), 407-410.
- (7) Carol Tenopir, "Users and Uses of Full Text Databases," in Proceedings of the International Online Meeting, London, England, December, 1988 (Oxford, England: Learned Information, Ltd., 1988), 263-270.
- (8) Carol Tenopir and Man Evena Shu, "Magazines in Full Text: Uses and Search Strategies," Online Review, (1989. In press).
- (9) Leon A. Jakobovits and Diane Nahl-Jakobovits, "Learning the Library: Taxonomy of Skills and Errors," College and Research Libraries, 48 (May 1987) 203-214.
- (10) Patricia Sullivan and Peggy Seiden, "Educating Online Catalog Users: Assessment of Needs," Library Hi Tech, 3(2) (1985) 11-19.
- (11) Peter Ingwersen and Soren Kaae, User Librarian Negotiations and Information Search Procedures in Public Libraries: Analysis of Verbal Protocols (Denmark: Copenhagen Royal School of Librarianship, 1980), 91-92.
- (12) Nicholas J. Belkin, "Discourse Analysis of Human Information Interaction for Specification of Human-Computer Information Interaction," The Canadian Journal of Information Science, 12 (3/4) (1987) 31-42.
- (13) Belkin, 36-40; Ericsson, K. Anders and Simon, Herbert A. Protocol Analysis: Verbal Reports as Data. (Cambridge: MIT Press, 1984).
- (14) Jakobovits, 207; Diane Nahl-Jakobovits and Leon A. Jakobovits, "Managing the Affective Micro-Information Environment," Research Strategies 3(1) (Winter 1985) 17-28.