DEFINING SEARCH SUCCESS:

EVALUATION OF SEARCHER PERFORMANCE IN DIGITAL LIBRARIES

by
Barbara M. Wildemuth
Associate Professor, School of Information and Library Science
University of North Carolina at Chapel Hill
100 Manning Hall, CB #3360
Chapel Hill, NC 27599-3360
Phone: 919-962-8072
Fax: 919-962-8071
Email: wildem@ils.unc.edu

It is a common belief--and empirical data provides supporting evidence--that some people are more proficient in performing online searches than others (Borgman, 1989; Fenichel, 1980-81; Saracevic and Kantor, 1988; Trivison et al., 1986). Precision, the proportion of the retrieved documents that are relevant to the query, and recall, the proportion of the relevant documents that were retrieved by a particular query, are the measures generally used to evaluate differences in search performance. In addition, measures of search term overlap and search efficiency are sometimes reported.

Searching for relevant citations in a traditional bibliographic database, however, differs from searching for specific information in a full-text database with regard to the target of the search. Specifically, the target of a search in a factual database is often a highly circumscribed set of facts or concepts, while the target of a search in a bibliographic database is usually a broader set of citations that are relevant to a query. In this way, searching a factual database more closely resembles searching an encyclopedia (Marchionini, 1989) or searching for a known item in a bibliographic database (Lancaster, 1977). This difference in the target of the search may have important implications for the searcher's success in locating the relevant item(s), the terms selected during the process of searching, and the efficiency with which the search can be performed.

Searching for relevant citations in a traditional bibliographic database also differs from searching for relevant nodes in a hypertext/hypermedia database, with regard to the process of the search. In most such systems, it is possible for users to select one or more terms and provide them to some type of search engine to identify a set of relevant nodes. This situation is, in many ways, analogous to a traditional searching situation. However, it is also possible for the searcher to traverse the network, moving from node to node until he or she arrives at a relevant (or interesting) node or identifies a set of relevant nodes. Because hypertext databases support traversal of the network of nodes so directly, they are also more likely to be used for browsing, while traditional databases are more often seen as a tool for identifying items that meet particular pre-specified criteria. An overstatement of this contrast may make it clearer: a hypertext database is primarily a tool for unfocused browsing (i.e., there is no explicit target)

while a traditional database is primarily a tool for targeted searching (the target may be a single known item or a set of items meeting particular criteria).

Finding ways to describe and evaluate the effectiveness of browsing/navigation behavior will be a challenge for designers and researchers of digital libraries. Some preliminary work in this area has already been undertaken. Canter, Rivers, and Storrs (1985) defined the attributes of navigation patterns (e.g., loopiness or spikiness) based on users' routes through the data structure and did some preliminary investigations of these patterns. More recently, the think-aloud protocols of subject-area experts and novices using a hypertext database on the Vietnam War revealed three types of browsing behavior: search-oriented browse, review-browse, and scan-browse (Carmel, Crawford, and Chen, 1992). Other studies have contrasted searcher behaviors in hypertext and other data structures. In a study of high school students searching on a CD-ROM version of an encyclopedia, half of the students were trained in browsing techniques and half were trained to do Boolean searches; no statistically-significant differences were found in the groups' search behaviors (Liebscher and Marchionini, 1988). In another study, users were given either a search question or a browse question to address to either a linear text or a hypertext database; in terms of the accuracy and completeness of the searchers' answers to the questions, level of expertise was found to interact with format (Rada and Murphy, 1992). Even these few studies demonstrate a variety of methodological approaches to the question of evaluating searcher performance in a hypertext database.

In psychometric terms, "searcher proficiency" is a construct exemplified by a domain of observable behaviors (Nunnally, 1978). Prior research has indicated that the construct of searcher proficiency in the context of a traditional database includes such observable behaviors as the precision and recall achieved in the search results (evidence of the searcher's success in retrieving relevant records); the overlap of the terms selected with a prespecified set of "good" terms (evidence of the quality of the searcher's selection of terms); and the time or number of commands used to conduct the search (evidence of the searcher's efficiency). Measures of each of these observable behaviors have been incorporated in past studies of online searching behavior, e.g., Saracevic and Kantor (1988a), and have even been explored in studies comparing traditional searching and hypertext navigation (Godin, Missaoui, and April, 1993). The question is: what types of measures are most appropriate for evaluating searcher or system performance in the digital library context? Digital libraries are hypertext/hypermedia databases that support browsing and searching for particular items/nodes and particular facts. In other words, they incorporate all the variations in the search process and the search targets discussed above.

Over the past several years, I (with Charles Friedman, Ruth de Bliek and others) have been involved in a large-scale study (supported by a grant from the National Library of Medicine) that begins to address some of these methodological issues. It is a study of medical students' use of factual databases (Friedman, 1988, 1990) in support of problem solving. In the first phase of the research project, we examined (1) the relationship between students' level of knowledge in a particular biomedical domain, e.g., microbiology, and their searching behaviors and (2) the relationship between their searching behaviors and their success in solving problems with domain-specific database assistance (Wildemuth, 1993; Wildemuth et al., 1995, 1993, 1992; de Bliek et al., 1994; Friedman et al., 1994). In this phase of the research, access to the database

was provided through keyword searching, including the use of Boolean operators. The second phase of the project is exploring the effects of the database access method, i.e., we are comparing student performance on the Boolean-searchable version of the database with their performance using a hypermedia version of the same database (Friedman et al., 1995). The data collected in 1995-1996 will include transaction logs of searches in the hypermedia database, as well as those from searches in the Boolean-searchable version. In addition to this project, I have undertaken studies of end-user searching of an online library catalog (Wildemuth and O'Neill, 1995) and of MEDLINE (Wildemuth and Moore, 1995).

Each of these projects has involved the description and analysis of the sequences of moves that constitute online search strategies. I am currently employing three different methods to analyze sequences of moves employed in searching traditional databases. The search moves were generated by categorizing the user actions (Shute and Smith, 1993) captured in transaction logs. The first analysis method results in state transition matrices, displaying the relative frequency of first-order transitions from one search move to the next. This method has been applied in several prior studies of search behavior (Penniman, 1975; Chapman, 1981; Tolle, 1983a, 1983b; Harris, 1986) and navigation through a hypertext system (Qiu, 1993). The second analysis method involves the identification of maximal repeating patterns (Siochi and Ehrich, 1991) of moves within the search strategies, and further qualitative analysis of those that occur most frequently. To my knowledge, this method has not previously been applied to the analysis of search strategies. The third method is totally qualitative, grouping those sequences of moves that are similar to each other, to identify those that occur most frequently and the circumstances in which they occur (Wildemuth, 1990). I am currently applying these three methods to transaction logs from a longitudinal (three-year) study of medical students' searches of MEDLINE and a set of small factual databases. The substantive results from each method will be integrated to provide a model of end-user searching behaviors and their development over time; the methodological results should provide guidance for the further development of methods for analyzing sequences of search moves in both traditional databases and hypertext databases.

References:

*Borgman, C. L. (1989). All users of information retrieval systems are not created equal: An exploration into individual differences. Information Processing & Management, 25, 237-251.
*Canter, D.; Rivers, R.; Storrs, G. (1985). Characterizing user navigation through complex data structures. Behaviour and Information Technology, 4 (2), 93-102.
*Carmel, E.; Crawford, S.; Chen, H. (1992). Browsing in hypertext: A cognitive study. IEEE Transactions on Systems, Man, and Cybernetics, 22 (5), 865-84.
*Chapman, J. L. (1981). A state transition analysis of online information-seeking behavior. Journal of the American Society for Information Science, 32 (5), 325-33.
*de Bliek, R.; Friedman, C. P.; Wildemuth, B. M.; Martz, J. M.; Twarog, R. G.; File, D. (1994). Information retrieved from a database and the augmentation of personal knowledge. Journal of the American Medical Informatics Association, 1, 328-38.
*Fenichel, C. H. (1980-81). The process of searching online bibliographic databases: A review of research. Library Research, 2, 107-27.

*Friedman, C. P.; de Bliek, R.; Twarog, R.; France, C. L.; Lemmond, G.; File, D. D. (1988). Studying the utilization and effects of a computer-based educational intervention in bacteriology. Annual Conference on Research in Medical Education, 27, 100-105.

*Friedman, C. P.; Twarog, R.; File, D. D.; Youngblood, P.; de Bliek, R. (1990). Computer databases as an educational tool in the basic sciences. Academic Medicine, 65, 15-16.

*Friedman, C. P.; Wildemuth, B. M.; de Bliek, R.; Twarog, R. G.; File, D. D. (1994). Database searching proficiency and problem solving proficiency in a biomedical domain. Teaching and Learning in Medicine, 6, 168-74.

*Friedman, C. P.; Wildemuth, B. M.; Gant, S. P.; Muriuki, M.; File, D. D.; Downs, S. M.; de Bliek, R. (1995). A comparison of Boolean and hypertext access to a basic science database," poster presented at the 1995 Research in Medical Education meeting.

*Godin, R.; Missaoui, R.; April, A. (1993). Experimental comparison of navigation in a Galois lattice with conventional information retrieval methods. International Journal of Man-Machine Studies, 38, 747-67.

*Harris, M. (1986). Sequence analysis of moves in online searching. Canadian Journal of Information Science, 11, 35-56.

*Lancaster, F. W. (1977). Studies of catalog use. In F. W. Lancaster, The Measurement and Evaluation of Library Services (Chapter 2). Washington, DC: Information Resources Press.

*Liebscher, P., & Marchionini, G. (1988). Browse and analytical search strategies in a full-text CD-ROM encyclopedia. School Library Media Quarterly, 16 (4), 223-33.

*Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. Journal of the American Society for Information Science, 40, 54-66.

*Nunnally, J. C. (1978). Validity. In J. C. Nunnally, Psychometric Theory (Chapter 3, 2nd ed.). New York: McGraw-Hill.

*Penniman, W. D. A stochastic process analysis of on-line user behavior. (1975). Information Revolution: Proceedings of the 38th ASIS Meeting (pp. 147-48).

*Qiu, L. (1993). Markov models of search state patterns in a hypertext information retrieval system. Journal of the American Society for Information Science, 44 (7), 413-27.

*Rada, R.; Murphy, C. (1992). Searching versus browsing in hypertext. Hypermedia, 4 (1), 1-30.

*Saracevic, T.; Kantor, P. (1988). A study of information seeking and retrieving, II: Users, questions, and effectiveness. Journal of the American Society for Information Science, 39, 177-96.

*Shute, S. J.; Smith, P. J. (1993). Knowledge-based search tactics. Information Processing & Management, 29, 29-46.

*Siochi, A. C.; Ehrich, R. W. (1991). Computer analysis of user interfaces based on repetition in transcripts of user sessions. ACM Transactions on Information Systems, 9, 309-35.

*Tolle, J. E. (1983a). Current Utilization of Online Catalogs: Transaction Log Analysis. Dublin, OH: OCLC.

*Tolle, J. E. Transactional log analysis: Online catalogs. (1983b). Proceedings of the Sixth Annual International ACM SIGIR Conference (pp. 147-60).

*Trivison, D.; Chamis, A. Y.; Saracevic, T.; Kantor, P. (1986). Effectiveness and efficiency of searchers in online searching: Preliminary results from a study of information seeking and retrieving. Proceedings of the Annual Meeting of the American Society for Information Science, 23, 341-349.

*Wildemuth, B. M. (1990). A method for inducing process models from qualitative data. Library & Information Science Research, 12, 329-40.

*Wildemuth, B. M. (1993). Post-positivist research: two examples of methodological pluralism. Library Quarterly, 63, 450-68.

*Wildemuth, B. M.; de Bliek, R.; Friedman, C. P. (1993). Measures of searcher performance: A psychometric evaluation. Information Processing & Management, 29, 533-50.

*Wildemuth, B. M.; de Bliek, R.; Friedman, C. P.; File, D. D. (1995). Medical students' personal knowledge, searching proficiency, and database use in problem solving. Journal of the American Society for Information Science, 46, 590-607.

*Wildemuth, B. M.; de Bliek, R.; He, S.; Friedman, C. P. (1992). Search moves made by novice end users. Proceedings of ASIS '92, Pittsburgh, PA, October 26-29, 1992, 29, 154-61.

*Wildemuth, B. M.; Moore, M. E. (1995). End-user search behaviors and their relationship to search effectiveness. Bulletin of the Medical Library Association, 83, 294-304.

*Wildemuth, B. M.; O'Neill, A. L. (1995). The 'known' in known-item searches: A pilot study," College & Research Libraries, 56, 265-81.