

Users' Persistence in Scanning Lists of References

Stephen E. Wiberley, Jr., and
Robert Allen Daugherty

One of the basic information-seeking activities where library users might experience information overload is scanning lists of references. Because much information seeking is discretionary—users may abandon a search without fear of penalty—lists that are too long may influence users to stop searching. This article summarizes both librarians' opinions about the number of references that constitute too many and reports of users' behavior with lists of various lengths. It covers preferences for maximum numbers of references from online searches, online public access catalog postings, and manual indexes. It concludes with implications for practice and research.



Information overload is a familiar term in the 1980s, but little is known about how it affects the behavior of individual information seekers.¹ As J. Michael Brittain has pointed out:

The belief of most practising librarians has been that information *per se* is good; the more the better. There has been little or no concern about users' ability to digest and make use of information, or about the optimum flow of information in terms of work tasks and users' ability to process information. We should not be too critical of library researchers on this front. These problems are formidable indeed, and social scientists, including computer scientists and psychologists, have themselves not been particularly successful in researching into human information processing.²

The effects of information overload become particularly interesting when one considers that much information seeking is discretionary, i.e., the person seeking information may abandon a given search strategy either because the information sought is not vital or because an alternative strategy seems potentially more fruit-

ful. Common experience shows that people do not like to work their way through masses of information. Executive summaries, abstracts, and review essays all stand as evidence that readers are impatient with long presentations. Thus, when confronted with a large number of references, a searcher may use discretion to adopt a new search strategy or to end the search entirely. Designers of information systems need to take this into account.

Is there a maximum number of references beyond which most people will not scan, choosing instead either to abandon entirely a search for information or to adopt an alternate strategy? This is one of the most basic questions a systems designer must ask about information overload. References may include postings on video display terminals and entries in manual indexes. If such a number or limit exists, then designers need to develop systems accordingly. If not, then they should not waste resources offering help that generally is not needed.

A review of the literature indicates that no one has addressed this question sys-

Stephen E. Wiberley, Jr., is Bibliographer for the Social Sciences and Robert Allen Daugherty is Circulation Librarian at the University of Illinois, Chicago, Illinois 60680. The authors express their sincere thanks to Robert W. Karrow, Jr., and Gunnar Knutson for criticism of a draft of this article.

tematically, but some either assume or offer evidence that most people do have a limit on the number of references they will scan. One investigator who assumes that to be true has given a name to the limit. David C. Blair defines the "anticipated futility point" as the "maximum number of retrieved documents that an inquirer would be willing to *begin* browsing through."³ While Blair concentrates on the user's response to documents, his term "anticipated futility point" could be applied to lists of references as well. Others have not given the limit a name, but have offered two kinds of evidence about it: (1) opinion about how many references are too many, and (2) reports of users reactions to particular numbers of references.

This article attempts to summarize and compare these opinions and reports, but makes no claims to cover all the relevant literature. Nevertheless, the summary is extensive enough to bring the question of persistence in discretionary scanning of lists of references to the attention of researchers and practitioners and to provide them with information on which to base further investigations and practice. But first, discussion of discretionary information seeking is helpful.

COMPULSORY AND DISCRETIONARY INFORMATION SEEKING

People seek information in a variety of circumstances and from a variety of sources. In many cases a person risks a penalty for not finding information that is known to be in a source. For example, bank personnel will search long and hard to find a cancelled check supporting the bank's position in a lawsuit. Similarly, a student will assiduously seek required readings that an instructor has stressed as crucial to success in a course. Failure to find these readings may diminish the student's success. Such searches may be called *compulsory information seeking*.

Compulsory information seeking differs from most information seeking done in academic libraries and from much other information seeking that people do. Non-compulsory or *discretionary information seeking* entails searching for information

that is not essential and/or whose source is not known with certainty. For example, a student knows exactly the source of information relevant to his term paper. He decides not to retrieve that information because the paper is already adequately documented. Here information is not essential. Or, an executive must report about her company's salary structure. She knows that needed data are in a computer file, but cannot remember which file. While the overall search for these data is compulsory, the search of individual files is discretionary since the content of each is not fully known. Thus, the executive may use discretion to abandon the search of a given file. Discretionary information seeking also includes cases where both information sought is not essential and the searcher is uncertain about its source.

“. . . users often have little or no reason to persist in using systems that are 'unfriendly.' "

It is important for those who design and service information systems to realize that much, perhaps most, information seeking is discretionary. Because of this, users often have little or no reason to persist in using systems that are "unfriendly." Numerous factors impinge on the persistence of users. Most of these factors are beyond the influence of those who create information sources. For example, they cannot shape the personality of the user, increase the time available for the search, ensure that the site will be comfortable, or that the searcher's peers and superiors will encourage persistence. Designers and compilers do, however, have some control over how the manual index or computer system will display a given number of references and assist in their manipulation. The review of the literature that follows concentrates on the maximum length of a list of references that people are willing to scan in online databases, online public access catalogs (OPACs), card catalogs, and printed indexes.

PREFERENCES FOR REFERENCES: ONLINE SEARCHING

End-users and search analysts have shown great diversity and range in their preferences for number of references from an online search. (An online search may be defined as an interactive query by an intermediary or an end-user of one or more bibliographic databases compiled and loaded by one or more vendors. For example, an intermediary or an end-user may search *ERIC* or *Psychological Abstracts* and access those databases through BRS or Dialog or on CD-ROM.)

In an overview of online searching, Arleen N. Somerville raised the issue of finding too many references. She urged intermediaries to "determine the range of total citations acceptable to the user." She asked, "If a search produces 150 references, is that acceptable?"⁴ One hundred-fifty is higher than most other standards. One search service placed the limit per search at 100 references without abstracts, 50 with abstracts.⁵ A second service conducted by Lockheed normally supplied between 70 and 90 citations per search.⁶ A third service, which searched ERIC, held that "when the results appear to be on target and sufficient in quantity (30? 50? 70?) we generally manipulate no further unless we are preparing for a research paper or for someone whom we know wants total recall rather than precision."⁷ Marcia Bates observed that students generally consider 75 or 100 references to be the limit for a search.⁸ Finally, an SDI service for chemists and engineers limited the bibliographies it created each week to 35 references.⁹ In assessing the relevance of such standards to user persistence, it is important to remember that intermediaries who set standards are concerned with the cost of a search as well as information overload.

The limits and ideal numbers discussed thus far were set by intermediaries. While they probably took into account end-users' preferences, they did not report directly what users prefer. Reports about end-users suggest that the preferred limit lies somewhere under 50 references, but

show that the range varies greatly. On the one hand is the account of a search where the end-user apparently wanted all 303 citations retrieved.¹⁰ At the other end of the spectrum, David Raitt states that scientists and engineers with whom he works "are content to find a relatively low number of references (c. 10-15)."¹¹ Like Raitt, Charles Anderson and Ann Weston have found that patrons in their public library prefer 10 to 20 very relevant citations.¹² Reportedly, one chemist would regularly search online, "narrowing his strategy until he had 50-60 references, then printing these offline."¹³ And, in a program that trained end-users on BRS/After Dark, the mean number of references retrieved was 31.5.¹⁴ These are limited data, of course, and none was gathered with the purpose of describing the upper limit of number of references that users prefer to scan. Yet they give some sense of preference and range. Moreover, they remind us that individual circumstances are crucial in determining the number of references a particular person at a particular time is willing to use.

John Edward Evans has compared end-users' reactions to one program where there was no limit to the number of databases searched in response to a query with reactions to a program where searching was limited to one file. Multifile searching typically retrieved 94 citations per search, while single file searching usually yielded 70. End-users seemed to prefer the shorter lists:

Researchers seem to be more interested in a relatively restricted list of subject-specific citations appropriate to their research, not in the accumulation of vast bibliographies of possibly useful sources.¹⁵

Unfortunately, the value of these data is limited by Evans' admittedly unscientific sampling of end-users' opinions. Despite this, the conclusion that 70 references is optimal falls near the median number recommended by search analysts.

Overall, then, the observed preferences of end-users and the recommendations of search analysts generally fall between 50 to 70 references per search. This implies that search analysts should retrieve no

more than that range, unless the end-user requests more.

PREFERENCES FOR REFERENCES: OPACs

While end-users usually access online bibliographic databases through a search analyst, almost all users of Online Public Access Catalogs (OPACs) interact directly with the systems. End-users' persistence in scanning lists of references—usually called postings in the case of OPACs—could be explored through protocol analysis, transaction log analysis, focused group interviews, or by questionnaires, but thus far no one has completed such a study.

Karen Markey has commented perceptively on user persistence, (she uses the term "perseverance"), but her work has emphasized persistence in entering search statements. Since her work does not address systematically the issue of persistence in scanning postings, its value is limited here. Nevertheless, her findings are well worth reviewing. On the one hand, she reports cases where users displayed all 205 items retrieved and printed 51 of 51, 76 of 76, and 88 of 88 found. On the other hand, she describes instances when the user did not scan or print any postings (0 of 266) or displayed only a fraction of the postings: 32 of more than 999, 20 of 673, 19 of 237, 58 of 205, 34 of 123, and 15 of 35. These data suggest users can be remarkably persistent on occasion but, in general, need to scan roughly 15 to 35 postings to comprehend, to their own satisfaction, what a search tells them.¹⁶

The literature about the design of OPACs indicates a range of practice and opinion regarding persistence in scanning references. For example, one OPAC prompts the user to enter additional information when author searches exceed 20 retrievals.¹⁷ Nancy John has suggested that 30 is the maximum number most end-users will scan.¹⁸ The Books Are For Use Project attempted to limit the lists of references it retrieved in each search to 25 items.¹⁹ Also, OCLC, which in many ways resembles an OPAC, prompts users to reduce postings when a search retrieves more than 50. Depending on the applica-

tion, Markey has made a number of recommendations regarding the introduction of aids to the user who is confronted with long lists of postings. In general:

Access points that result in 200 or more retrieved items could be accompanied by a "suggestive prompt" which asks the searcher if he would like to refine the retrieved items by entering additional search terminology. Likewise, when the searcher scans through displays of 100 or more items, introduced into the dialog could be a suggestive prompt that asks the searcher what nonrelevant terminology continually appears in the output or what terminology has appeared in the online display that is particularly interesting.²⁰

For keyword-in-context displays of subject headings that exceed 50 postings, she recommends the results be categorized by subdivisions.²¹ When author searches or title searches or author/title searches exceed 20 postings (or two screens of brief title displays), they should conclude with a message that reminds users of the system's capability to limit the results, e.g., by year, language, or format.²²

" . . . users seem to prefer to scan a smaller number of references from an OPAC than from an online search."

Overall, the data reported here about user persistence and the recommendations for design of OPACs suggest the number of postings that users will scan at an OPAC is somewhat less than the number of references they want from an online search, although there are exceptions. Several search analysts or end-users have shown a preference for more than 50 references from an online search. While some OPAC users will look at more than 50, most are satisfied with looking at fewer than 35 OPAC postings. Only two recommendations for OPAC design allowed for more than 50 postings to be displayed at one time. Use of an OPAC differs from an online search, and three factors may influence users to prefer shorter lists of references from OPACs than from online searches.

First, an end-user seldom receives assistance in use of an OPAC, but intermediaries often search online databases for end-users. Second, the end-user almost always takes a printout of the results at the conclusion of an online search, but OPAC users often have to write down results of their searches. Users tend to want to write as little as possible, and they can usually find all they are willing to write by scanning 35 or fewer postings. Third, many users search OPACs simply to find a few relevant call numbers of books, while users of online search services are often looking for a bibliography, usually of journal articles, on a subject. The bibliography will be referred to intermittently over a period of time, so the user can tolerate more references than he or she normally likes to scan at one time. In contrast, the call numbers sought from an OPAC may be for known items that can usually be retrieved without calling up a large number of postings with them. Or call numbers may be the result of a subject search where the user wants to browse the shelves for books on a given topic. In subject searches, scanning 15 to 35 or fewer postings often tells the user if the search has yielded relevant material. These differences between online and OPAC searches may explain why users seem to prefer to scan a smaller number of references from an OPAC than from an online search. At the same time, it is important to realize that past experience and opinion include exceptions to these generalizations, and future research may change these conclusions.

PREFERENCES FOR REFERENCES: MANUAL INDEXES

Machine-readable systems generate references far faster than manual systems and can provide a copy with virtually no effort on the user's part. Manual systems require the user to do much more work. Consequently, users of manual indexes (including card catalogs) prefer to scan even fewer references than users will scan from online searches and in OPAC use. Preferences for references in printed sources range from 7 to 50, but, for the

most part, do not exceed 20. Limits recommended for card catalogs allow for more references than those given for printed indexes.

John Wallace Metcalfe suggested that entries under a given term in a classified or alphabetical catalog be limited to 50. When entries exceed 50, the term should be further specified.²³ In his *Rules for a Dictionary Catalog*, Charles A. Cutter argued that, wherever possible under subject headings, titles should be "broken up into sections containing from half a dozen to a score."²⁴ In comparing people's behavior with references displayed in different media, R. S. Marcus found that "the user who scans 50 references on a CRT may limit himself to 20 or less on a typewriter and 5 or less in a traditional library."²⁵

Designers and critics of printed indexes in books seem to place the outer limit of a reader's patience at fewer than 10 items. While Alec Clifton-Taylor complained that 17 was too many undifferentiated references to appear after a personal name in an index, most do not allow as many.²⁶ The design of the Modern Language Association's CIFT indexing system allows for no more than 10 bibliographic references to appear under a single descriptor.²⁷ Likewise, J. H. Hexter has criticized the index to the *Proceedings in the Parliaments of Elizabeth I: 1558-1581* for failing to limit undifferentiated strings of page numbers to 10 or fewer.²⁸ Bonnie Gratch, Barbara Settel, and Pauline Atherton also recommend that "if an index entry is followed by 10 or more locators, some attempt should be made to subdivide the entry for ease of reference to specific aspects of the subject."²⁹ Finally, G. Norman Knight contended that "about seven references should be the absolute limit" for undifferentiated page numbers under a given entry.³⁰

The apparently lower tolerance for references in printed, back-of-the-book-style indexes compared to card catalogs is understandable. A card catalog user, like an OPAC user, wants to take from a search selected elements of catalog records, for a few items at most. Transcription of such information from the cards to paper is relatively straightforward. On the other hand, flipping back and forth between an

index and the pages in a book is, if not onerous, much more demanding.

IMPLICATIONS FOR RESEARCH AND PRACTICE

Traditionally librarians have been concerned with maximizing the amount of information they provide users. This is reasonable because users often do not find enough information.³¹ But the concern to maximize retrieval, combined with the common assumption that more is better, can lead to excess. Librarians need to be sensitive to the possibility of overloading users with information. In this regard, one of the most interesting aspects of the "Search Evaluation Questionnaire" developed and recommended by the Machine-Assisted-Reference-Services section of RASD is that it asks users if they found enough from their searches, but does not ask if they found too much.³² The former question is very important, but, as data gathered in this article show, there are times when the latter question is equally or more significant.

The possibility of exceeding a user's tolerance for a list of references should be recognized because so much information seeking is discretionary and users may well abandon a search if they are presented with more references than they are willing to scan. Coping with this is a challenge for system designers and librarians. Greater understanding of its nature is a challenge of research. Researchers must better describe and analyze this problem, so that system designers and librarians can address more successfully its practical aspects. The possibilities for research, system design, and librarianship are numerous. Only a few can be mentioned here.

". . . online systems have particular promise for revealing how users respond to information overload."

In research, online systems have particular promise for revealing how users respond to information overload. A computer-based system can keep a log of

each step in the interaction between user and machine. Because this record of interaction is obtained unobtrusively, the data-gathering process does not affect the user's behavior as it might with other methods such as protocol analysis. Analysis of system logs will not tell all, but such analysis combined with questioning of users after they complete a search session may offer great insight into motivation and behavior in coping with information overload.

Some designers of print and computer-based indexes have taken steps to help users cope with too many references or postings. The question, particularly for the designers of machine-driven systems, is whether their aids to users can be made more effective. The studies of OPACs sponsored by CLR have made recommendations in this area.³³ If research determines that a majority of OPAC users do not scan more than 35 postings and online searchers more than 50, designers could take this into account. For example, they could withhold system prompts until these thresholds are exceeded. Limiting the frequency of display, such prompts would presumably make them more noticeable when they did appear and, therefore, more likely to help the user.

Finally, the librarian, whether searching for information or guiding the user in a search, always needs to be sensitive to the possibility of overloading the user. Here, awareness of techniques to deal with the problem is particularly important. Marcia Bates has outlined a number of search tactics that are particularly helpful in coping with information overload.³⁴

As databases become larger and larger and as increasing numbers of them are linked, the possibility increases that those who seek information will encounter lists of references longer than they are willing to scan. Because information overload can influence persons to use their discretion and abandon their searches, it is important for researchers, system designers, and librarians to work together to help users cope. Effective effort to assist users who need help with long lists of references is a significant challenge in our age of information overload.

REFERENCES AND NOTES

1. Joel Rudd and Mary Jo Rudd, "Coping with Information Load: User Strategies and Implications for Librarians," *College & Research Libraries* 47:315-22 (July 1986). Lester Asheim, "Ortega Revisited," *Library Quarterly* 52:215-26 (July 1982).
2. J. Michael Brittain, "Pitfalls of User Research, and Some Neglected Areas," *Social Science Information Studies* 2:144 (July 1982).
3. David C. Blair, "Searching Biases in Large Interactive Document Retrieval Systems," *Journal of the American Society for Information Science* 31:271 (July 1980).
4. Arleen N. Somerville, "The Pre-Search Reference Interview—A Step by Step Guide," *Database* 5:35 (Feb. 1982).
5. Georgia Fox Donati and Martha Moss Kreisel, "Free Online Searching in a Public Library System: An Unscientific Study," *Online* 7:14 (Mar. 1983).
6. Roger K. Summit and Oscar Firschein, *Investigation of the Public Library as a Linking Agent to Major Scientific, Educational, Social, and Environmental Data Bases; Two-Year Interim Report* (Palo Alto, Calif.: Lockheed Research Lab., ERIC Document Reproduction Service, ED 131 857, Oct. 1976).
7. Katherine Clay, "Searching ERIC on DIALOG: The Times They Are a' Changing," *Database* 2:50-1 (Sept. 1979).
8. Marcia Bates, "The Fallacy of the Perfect Thirty-Item Online Search," *RQ* 24:43 (Fall 1984).
9. Ron Kaminecki, "Comparison of Selective Dissemination of Information Systems," *On-Line Review* 1:197 (Sept. 1977).
10. Mary L. Calkins, "Free Text or Controlled Vocabulary? A Case History Step-By-Step Analysis . . . Plus Other Aspects of Search Strategy," *Database* 3:57 (June 1980).
11. David Raitt, [Letter to the Editor], *Online Review* 9:40 (Feb. 1985).
12. Charles Anderson and Ann Weston, "The Costs of Online Searching," *Library Journal* 112:47 (Apr. 1, 1987).
13. Judith S. Haines, "Experiences in Training End-User Searchers," *Online* 6:17 (Nov. 1982).
14. Elaine Trzebiatowski, "End Users Study on BRS/After Dark," *RQ* 23:447 (Summer 1984).
15. John Edward Evans, "Database Selection in an Academic Library: Are Those Big Multi-File Searches Really Necessary," *Online* 4:41 (Apr. 1980).
16. Karen Markey, *Subject Searching in Library Catalogs: Before and after the Introduction of Online Catalogs* (Dublin, Ohio: OCLC Online Computer Library Center, Inc., 1984) p.67-71. For a case where scanning the first 20 postings apparently was enough for the user to comprehend a list of 323 references, see Mary L. Calkins, "Free Text or Controlled Vocabulary?" *Database* 3:58 (June 1980).
17. Karen Markey, *Online Catalog Use: Results of Surveys and Focus Group Interviews in Several Libraries* (Dublin, Ohio: OCLC Online Computer Library Center, Inc., Office of Research, Mar. 31, 1983), p.151.
18. Peter Kidder, "ALA Videoconference, Part 2: Cataloging, Formats and Filing," *American Libraries* 14:465, 478 (July-Aug. 1983).
19. Pauline Atherton, *Books are for Use: Final Report of the Subject Access Project to the Council on Library Resources* (Syracuse, N.Y.: Syracuse Univ. School of Information Studies, Feb. 1978), p.54.
20. Markey, *Subject Searching*, p. 103-4.
21. Karen Markey, *The Process of Subject Searching in the Library Catalog: Final Report of the Subject Access Research Project* (Dublin, Ohio: OCLC Online Computer Library Center, Inc., Office of Research, Feb. 4, 1983), p.8.
22. Markey, *Online Catalog Use*, p.151.
23. John Wallace Metcalfe, *Subject Classifying and Indexing of Libraries and Literature* (New York: Scarecrow, 1959), p.50.
24. Charles A. Cutter, *Rules for a Dictionary Catalog*, 4th ed. The U.S. Bureau of Education Special Report on Public Libraries—Part 2 (Washington, D.C.: Government Printing Office, 1904), p.123.
25. Quoted in David A. Thompson, "Interface Design for an Interactive Information Retrieval System: A Literature Survey and a Research System Description," *Journal of the American Society for Information Science* 22:371 (Nov.-Dec. 1971).
26. Quoted in Robert L. Collison, *Indexes and Indexing* (London: Ernest Benn, 1953), p.72. Similarly, *Magazine Index* on roll microfilm uses subdivisions if there are more than 16 articles listed under a single descriptor. See Mary Ellen Kyger Davis and John F. Riddick, "Readers' Guide to Periodical Literature and Magazine Index: A Comparison," *Reference Services Review* 11:45 (Winter 1983).
27. James D. Anderson, "Contextual Indexing and Faceted Classification for Databases in the Hu-

- manities," in *Information Choices and Policies: Proceedings of the American Society for Information Science, 42nd Annual Meeting*, eds. Roy D. Tally and Ronald R. Deultgen (White Plains, N.Y.: Knowledge Industry, 1979), p.200.
28. J. H. Hexter, "The Birth of Modern Freedom," *TLS* 164:54 (Jan. 21, 1983).
29. Bonnie Gratch, Barbara Settel, and Pauline Atherton, "Characteristics of Book Indexes for Subject Retrieval in the Humanities and Social Sciences," in Pauline Atherton Cochrane, *Redesign of Catalogs and Indexes for Improved Online Subject Access: Selected Papers of Pauline A. Cochrane* (Phoenix, Ariz.: Oryx Pr., 1985), p.459.
30. G. Norman Knight, "Editing of Indexes and their Preparation for Press," in *Training in Indexing: A Course of the Society of Indexers*, ed. G. Norman Knight (Cambridge, Mass.: MIT Pr., 1969), p.171.
31. Markey, *Online Catalog Use*, p.74.
32. Richard W. Blood, "Evaluation of Online Searches," *RQ* 22:266-77 (Spring 1983).
33. Markey, *Subject Searching*, p.134-6.
34. Marcia Bates, "Information Search Tactics," *Journal of the American Society for Information Science* 30:211-13 (July 1979).

**IN FORTHCOMING ISSUES OF
COLLEGE & RESEARCH LIBRARIES**

Foreign Library Materials Prices: A Report
by Frederick C. Lynden

Charting a Career Path for the Information Professions
by Leslie M. Kong and R. A. H. Goodfellow

The Learning Resource Center's Role in the Community College System
by Doris Dale

Evaluating the Smaller Library Collection: The Conspectus Approach
by Larry R. Oberg

Conspectus: An Aid to Collaborative Collection Development and a Source of Practical Internal Benefits

by Anthony W. Ferguson, Joan Grant, and Joel S. Rutstein

Academic Library-State Library Relationships: The Pennsylvania Needs Assessment
by Charles Townley