Online Searching and the Research Process

Connie Miller and Patricia Tegler

Despite evidence that researchers seek information in ways which are quite different from a logical, linear search strategy model, librarians persist in relating to the information-seeking process as if it were static and product oriented. Even online searching, which is generally considered to be flexible and interactive, is viewed merely as an alternate method of compiling a bibliography. The bibliography is considered a fixed and final product to be measured exclusively according to the limited variables of recall and precision. By having such a restricted view of online searching and its potential benefits to researchers, librarians fail to take full advantage of their role in the academic research community. If librarians wish to be relevant to researchers and to offer valuable services to an important constituency, they must fully understand the organic nature of research and the ways that scholars seek information. They must further understand and facilitate the significant way that online searching can contribute to and enhance the research process.

The view of research as a linear, highly structured, logical process has been challenged by studies which indicate that scholars work in ways which can best be described as cyclical, organic, and intuitive.[1] These illogical and intuitive approaches to research mirror themselves in seemingly random, haphazard approaches to locating pertinent information. Rather than following systematic library search strategies, scholars generally employ less structured methods, such as browsing, consulting with colleagues, or tracing footnotes and bibliographies. Printed indexes, by their very nature, tend to limit creative, cyclical interaction between researchers and information. Online systems, on the other hand, have the potential to facilitate highly interactive seeker-information dialogues, just the type of interchange which is integral and essential to the trial-and-error[2] process involved in research. To date, however, this potential for interaction and its corres-

ponding benefits to the scholar have not been fully realized. Further, the limited ways in which online searches have been evaluated have hindered a full understanding of the organic nature of the online process.

A key to exploiting the potential of the online process for researchers lies in understanding the distinction between seeking information on "topics" and seeking information on "problems." Swanson claims that "creative scientific research does not begin with a 'topic' but with a problem—a researcher must be puzzled, curious, in a sense 'bothered' about something."[3] As scholars research a "problem," the questions they ask and the information they seek shifts and changes. Each new finding alters what follows. "Research," as Maurice Line describes it, "is a process that does not allow for too formal organization."[4] Integral to the loose structure of research is the information-seeking behavior of researchers. Stoan states that "scholars . . . follow

Connie Miller is science librarian at the University of Illinois, Chicago, Illinois 60680. Patricia Tegler is systems librarian at Kirkland and Ellis Law Library, Chicago, Illinois 60601. This paper was presented at the ACRL Fourth National Conference in Baltimore, April 9–12, 1986.

no mechanical procedure of thinking up a topic, doing background reading on it . . . going through the card catalog . . . , consulting indexes for articles," etc.[5] Instead they most frequently locate additional information through the bibliographies of previously identified material. This bibliographic tracking technique closely approximates the actual research process. It is organic and cyclincal and manifests what Swanson describes as the trial-and-error, problem-oriented process of information retrieval.[6]

This trial-and-error process of information seeking is quite different from a highly structured search strategy approach, and unlike it, does not result in a complete, final bibliography. In trial-and-error searching both the process of the search and its products can lead the researcher to alter the original understanding of the problem and may lead to additional sources of information. The traditional methods of evaluating information searches—recall and precision—have completely overlooked this generative, creative aspect of a search. By evaluating the product and not the process, recall and precision limit our understanding of information searches and fail to measure them effectively.

Recall and precision measure a specific retrieved bibliography within the context of a particular database. Recall is the percentage of relevant documents retrieved out of all relevant documents in the database. If all possible relevant items are retrieved from the database or databases, the bibliography has achieved 100 percent recall. Precision is the percentage of relevant documents retrieved of all the documents retrieved. If half of the documents in the bibliography are relevant, its precision rate is 50 percent. Let us assume that for a particular topic there are 200 relevant citations in a database. A search for that topic results in a bibliography of 160 citations. Eighty of the citations are relevant. The search has achieved a recall rate of 40 percent, or 80 of 200, and a precision rate of 50 percent, or 80 of 160.

Both recall and precision depend on relevance, an extremely difficult concept to measure. Harter summarizes two differ-

ent types of relevance which have been identified:

The first type of relevance is "on the topic," which is the kind of relevance that would apply in subject searchs. A document is relevant to a topical query if it is on the subject named by the requester. Relevance in this sense can be judged by an individual or by a community of experts; it is objective and involves public knowledge. The second type of relevance is similar to what Kemp and others have referred to as "pertinence"—it is a subjective, private "creation of new knowledge" by the requester in the context of a personal information need. In this sense, relevance is not a property of a document and a request, but is the property of a document and a requester.[7]

Measurement of recall and precision is generally based on the identification of relevance according to the first, more objective meaning of the term. Search results get high marks for precision when a large percentage of the citations retrieved appear to be "on the topic." The specific information needs of the requester may or may not be met by these highly precise results. Recall and precision, therefore, measure the performance of the database or system. They do not and cannot measure the value of a search to a requester.

The concept of recall itself is a highly questionnable one. In the first place, recall cannot be accurately estimated. Using either definition of relevance it is impossible to determine the total number of documents in a database which are relevant to a given request or requester. Secondly, total recall is rarely desirable or valuable. The retrieval of all relevant documents would frequently constitute too much information, and a surplus can be as problematic as a deficiency. Furthermore, recall can only measure a specific bibliography against a hypothetically relevant portion of a database. It measures the value of the search process itself, therefore, only in terms of a quantifiable product.

To be an important part of research, however, the process of an online search and the products of that search must move beyond the restrictions of recall and precision. Hawkins, Bates, Vigil, and others have described heuristic techniques, like title and descriptor scanning, citation

pearl growing, the "notting" out of previous sets, and "interactive scanning," which help a searcher make more effective use of online file capabilities.[8] But the best techniques can still be limited by a topical, recall-and-precision dominated approach. Consider, for example, a search of the ERIC database for citations on end-user searching. Since no thesaurus term *end-user* currently exists, a searcher could develop a group of synonymous terms using the thesaurus or begin with a free-text approach and locate synonyms by printing several citations in the title-descriptor format. Terms like *online-systems, information-retrieval,* or *information-seeking* could be combined with *training, user-satisfaction,* or *surveys* to result in high recall. Limiting the combinations to precise user groups, e.g., *college-faculty* or *health-personnel,* could increase precision. A systematic process of notting out already examined sets could eliminate needless duplication. Logically, this technique of framing an information request in terms of statically defined synonyms could not be faulted.

Titles and descriptors function, however, as more than static synonyms. Words which name or describe books or articles act as signposts, embodying conceptual approaches to research or indicating directions taken. In this fuller sense, they function as powerful disseminators of information whose importance lies less with their potential for becoming part of online search logic and more with their potential to reshape a research question. A logical combination of topical synonyms for the term *end-users* is directed toward the development of a final product: a high recall and/or high precision bibliography. Interacting with words as signposts begins the generative process of reformulating an original information request. This generative process depends upon a willing suspension of logic, a leap, that is into the illogical or intuitive world of pertinence.[9] Harter calls an illogically interactive search a "problem-oriented" online inquiry which is "an integral part of science itself."[10]

How could the search for end-user searching citations have been different? "Full Service Document Delivery: Our Likely Future" is one title that a free text search of ERIC would retrieve. This citation appears minimally relevant at best; none of the logical synonym combinations listed above would retrieve it. By suspending logic, however, and responding to the words in the title, a researcher intent on developing an end-user searching program in an academic library might alter direction considerably, including in the development of the program an online document ordering option.

Whether logically or illogically conducted, an online search results in a bibliographic product. Like the words in titles or descriptors, the importance of this product lies less with its potential to supply information on a topic and more with its potential to alter the direction of research. Each document from the bibliography, while potentially of direct use for its content, also performs the "indirect function . . . of stimulating a reformulation of [a] request."[11] In addition, each document is a primary source, the bibliography of which provides an entry for the researcher into the literature of the field. Because *"the primary literature indexes itself,* and does so with greater comprehensiveness, better analytics, and greater precision than does the secondary literature,"[12] whether the bibliography achieves high recall or consists of citations precisely relevant to the researcher's topic is immaterial. The bibliographic product of an online search functions as a gateway into a citation network and, thus, participates in the cyclical, organic process of research.

Viewed statically or as ends in themselves, neither the process nor the product of an online search can be anything but ancillary to research. A logical search based on heuristic techniques with a high recall or high precision bibliographic product will, almost be default, result in articles or books through which a researcher can gain entry into a citation network. The best bibliographic product, however, the one which includes illogically relevant (pertinent) as well as logically relevant citations, will only result from a search in which information disseminated (through words in titles and descriptors) during the process contributes to a reformulation of

the researcher's request. Illogically relevant or pertinent sources will open up citation networks of their own. This cyclical, organic type of search operates according to what Abraham Kaplan calls its own "logic-in-use, [an] internal logic . . . [which,] as it germinates and develops . . . dictates the sources sought out at each stage along the way."[13] By responding to the document delivery concept, the researcher developing the end-user searching program opened a whole new avenue of literature to explore. Because "logic-in-use" is essential to an organic, research-related online search, Harter predicts the inevitability of end-user searching.[14] While his prediction is undeniably correct, the idea that online searches integral to the research process must operate according to a "logic-in-use" seems related less to who performs a search than to how well a search is performed. If librarians who perform online searches for researchers or who offer online training to end-users fail to understand and to communicate the vital and essential implications of the cyclical, organic, illogical nature of research for the process and products of an online search, they risk becoming irrelevant.

REFERENCES

1. Stephen K. Stoan discusses, in some detail, research on the research process in his article, "Research and Library Skills: An Analysis and Interpretation," *College & Research Libraries* 45:99–109 (1984).
2. Don R. Swanson, "Information Retrieval as a Trial-and-Error Process," *Library Quarterly* 47:128–48 (1977); and "Libraries and the Growth of Knowledge," *Library Quarterly* 50:112–34 (1980).
3. Swanson, "Information Retrieval," p.138.
4. Stoan, "Research and Library Skills," p.102.
5. Ibid., p.102.
6. Swanson, "Information Retrieval," p.138–39.
7. Stephen P. Harter, "Scientific Inquiry: A Model for Online Searching," *Journal of the American Society for Information Science* 35:114 (1984).
8. Marcia J. Bates, "The Fallacy of the Perfect Thirty-Item Online Search," *RQ* 24:13–20 (1984); Peter J. Vigil, "The Psychology of Online Searching," *Journal of the American Society for Information Science* 34:281–87 (1983); and Donald T. Hawkins and Robert Wagers, "Online Bibliographic Search Strategy Development," *Online* 6:12–19 (1982).
9. D. A. Kemp, "Relevance, Pertinence and Information System Development," *Information Storage and Retrieval* 10:37–47 (1974). Kemp's article includes an excellent discussion and definition of pertinence.
10. Harter, "Scientific Inquiry," p.114.
11. Swanson, "Information Retrieval," p.138.
12. Stoan, "Research and Library Skills," p.103.
13. Ibid., p.102, as derived from Abraham Kaplan, *The Conduct of Inquiry: Methodology for Behavioral Science* (San Francisco: Chandler, 1964), p.3–11.
14. Harter, "Scientific Inquiry," p.114.