What is missing in research about online searching behavior?

Ce qui manque à la recherche sur le comportement en contexte de téléréférence

Raya Fidel
Graduate School of Library and Information Science
University of Washington, FM-30
Seattle, WA 98195, USA

Experiments in online searching behavior have failed to explain the phenomena they were designed to study. An examination of the variables used in research to date finds that the most commonly used independent variables are searcher characteristics. Search-process and search-outcome variables are the most commonly used dependent variables. The search-process variables are inadequate because they are situational and subject to constant change. Significantly, these variables measure the cost of a search rather than the intellectual processes involved in answering a request. While these experiments could support the discovery of the "productive searcher," they cannot uncover searching behavior. Only in-depth analyses of the search process itself can lead to productive research.

Les expériences sur le comportement en contexte de téléréférence n'ont pas réussi à expliquer les phénomènes qu'elles devaient étudier. Un examen des variables révèle que les variables indépendantes les plus utilisées sont des caractéristiques de l'utilisateur. Les variables internes les plus utilisées sont reliées au processus de repérage et aux résultats obtenus. Les variables reliées au processus de repérage sont inadéquates parce qu'elles sont situationnelles et sujettes à changement. Ces variables mesurent le coût de la recherche plutôt que les processus intellectuels requis pour répondre à une question. Ces expériences pourraient mener à la découverte du "chercheur productif" sans pouvoir décrire son comportement. Seule une analyse en profondeur des processus de recherche peut produire une recherche productive dans ce domaine.

Experiments in online retrieval have been conducted for more than two decades. An extensive, well documented literature covering the specific area of online searching behavior reveals that most experiments have produced unexplained results: they could not support the hypotheses they were designed to substantiate. Some have even produced results which contradict common sense.

© The Canadian Journal of Information Science / Volume 12, No. 3/4, 1987 / pp. 54-61.

This past phase of experimentation in the emerging phenomenon of online searching of bibliographic databases contains valuable exploratory ventures. Though their direct contribution to our understanding of searching behavior is limited, the experiments have made an invaluable contribution to understanding the problems involved in studying searching behavior itself.

When an experiment presents unexplained findings, investigators may either accept the findings and modify their knowledge, or they may examine the study design and identify the factors that contributed to unsatisfactory results. In this paper, I take the latter approach. I will show that the experience gained through experiments in searching behavior proves that the *search process itself* must be studied before we can examine the factors affecting this process. Further, I will show that the traditional method of analyzing the search process is inadequate.

Experiments in online searching behavior, like any other kind of experiment, have to be conducted with a clear understanding of which variables to define explicitly — independent, dependent, or controlled — and which to ignore. For example, The National Library of Medicine (NLM) carried out a large-scale experiment to test the effect of various training methods (the independent variable) on search process and search outcome (the dependent variables). Other variables considered to be significant but whose effects were not supposed to be tested (e.g., level of experience, subject specialty, or the database searched) were deliberately controlled by selecting only experienced MEDLINE searchers as study participants. Still other variables (e.g., personality traits of searchers or their cognitive styles) were ignored (Wanger and McDonald, 1980).

Following established quantitative experiment methods, most online studies have used a well-established set of dependent variables: precision, recall, and unit cost measure the quality of the answer set (the search *outcome*); number of commands used, number of search terms used, number of cycles, number of sets viewed, connect time, and speed together measure the search *process* itself. The independent variables are unique to each study. To date, most experiments have focused on the *searcher* and chose various characteristics of searchers as the independent variables. In attempts to uncover the profile of a "good searcher," the variables tested have ranged from experience in searching (Fenichel, 1981), through method of training (Wanger and McDonald, 1980) and cognitive styles (Brindle, 1981: Woelfl, 1984; Saracevic et al., 1987), to personal traits (Bellardo, 1985).

Experiments in online searching behavior, however, are different from other online-retrieval experiments because they are aimed at investigation of the effect of the independent variables on the search process. Thus, search-process variables are unique and central to studies in searching behavior because they are the variables that reflect the actual searching behavior, while search-outcome variables are common to most online-retrieval experiments. The NLM study, for instance, measured whether the average number

of commands or of search terms used by searchers who had received one type of training was significantly different from the number used by searchers with another type of training.

Let us look more closely at these commonly used variables and analyze the reasons why they have produced unexplained results in studies of online searching behavior. Consider the example of the independent variable called "experience in online searching." At least two large-scale experiments have shown that there is no correlation between this variable and search-process or search-outcome variables (Fenichel, 1981; Howard, 1982). These results obviously contradict common sense because we all know that searchers do learn something through experience. Indeed, investigators always control this variable in their studies because they are convinced that experience does effect online searching behavior.

While the notion that experienced searchers are "no better" than novices is still maintained by some researchers, most agree that the above experiments, though carefully and rigorously designed, proved inadequate to reveal relationships between searching experience and both the search process and its outcome.

Let us examine this independent variable of "searching experience" further. Its measurement presents some problems because experience can be measured in a variety of ways: by the length of time a person has been involved in online searching, by the number of searches performed, or by a combination of both. While this is a technical problem in the design of an experiment, the issue can be resolved because experience is basically measurable.

Next, let us examine search-outcome and search-process — the two sets of dependent variables. Search-outcome variables were developed as performance measurements based on clearly defined objectives, but there is no theoretical basis for the development of search-process variables: they are essentially different from search-outcome variables.

Among the three search-outcome variables, precision and recall traditionally measure the quality of answer sets. Even though their use presents a variety of theoretical and technical problems, they are based on theoretical requirements. As Soergel explains, recall is based on the assumption that information systems should provide complete information, and precision is derived from the requirement that the information retrieved be concise (Soergel, 1985, 109–127). The variable "unit cost" — the cost of each relevant citation — is the newest among outcome variables to be studied, and it emerged when online systems became common. Unit cost actually measures the efficiency of the search process and could be used together with utility-based measurements. While closely related to precision, unit cost is also determined by search effort.

The variables used to express the search process were developed along with the experiments in online searching behavior. Unfortunately, there is

no theoretical requirement to support the selection of these particular variables; they represent the level of effort used to perform a task but they do not describe the search process itself. For example, the variable "the number of search terms used" measures only how many terms were keyed in by a searcher. It does not provide data about the search process — data such as the number of concepts selected, the number of terms selected to represent each concept, or the semantic proximity among the terms that represent each concept.

Indeed, the experiments in online searching behavior showed that, regardless of training, experience, personality traits, or cognitive styles, searchers use the same level of effort in accomplishing their tasks. This result is not altogether surprising, and it certainly is very different from concluding that there are no variations in searching behavior. It is time to re-examine the selection of search-process variables altogether.

To do so, one must first consider the objectives of studies in searching behavior. More specifically, there is a need to raise the following questions: Why are studies of online searching behavior carried out? What knowledge and understanding is to be gained and how will the results be used? Answers to these questions are necessary if research is to be productive.

Although these questions have not yet been answered by the scientific community, one is tempted to assume that the quest for the "productive searcher" guided most previous experiments — specifically on discovering which personal characteristics induce a high level of effort. With this aim in mind, previous experiments did not fail; though they did not discover a profile for the productive searcher, they did single out the factors that are *not* characteristic of the productive searcher. The task is now to test additional factors (i.e., independent variables) until the pertinent ones *are* found.

But even this approach is not as simple as it may seem. First, the dependent variables measure the physical effort invested in a search rather than the intellectual processes. As such, they cannot have value judgment attached to them. Is it "good" to use many commands, many search terms, many cycles, or to view many sets? Further, is it necessarily the case that speedy searchers are better than slower ones? (Remember, speed represents less thinking time.) The most one can say is that given the same answer set, searchers who spent less time in retrieving the set are more productive. However, since it is almost impossible to find two searchers who will retrieve the same answer set to a given request, even this use of the variable "connect time" is not sufficient to judge productivity.

These search-process variables actually measure the cost of a search. Results of experiments using these variables may affect hiring policies in libraries, including the introduction of personality or cognitive tests, but are likely to have very little explanatory value. The cost of a search is very different from the quality of a search. To assess the latter, one needs to

look at the intellectual components of this process. No search-process variable currently used reflects this component; a better understanding of the search process itself would reveal such variables.

Second, the search-process variables are not adequate as dependent variables because they are situational. To test these variables, I asked ten experienced searchers to search a set of two requests, and then, two months later, to re-do the searches. The score of most search-process variables recorded for each individual searcher in the first round was different from the ones recorded in the second. There was no apparent trend; the average variation was 49% and the maximum was 600%. This finding produced a new complication because it showed that the level of effort a searcher is using to perform an online search depends on situational variables, probably such factors as mood, additional commitments, or willingness to perform a particular search at a particular time. Therefore, if one is strongly committed to the discovery of the characteristics of the productive searcher, one should investigate situational variables.

Another objective for the study of searching behavior, however, is the quest for design principles that will guide the construction of better information systems. With this objective in mind, one studies the effect of searching experience on searching behavior to discover what is affected by experience. Once the knowledge gained through experience in searching is made explicit, it can support the design of information systems and, in addition, it can be integrated into training of novice users.

For example, suppose a study were to find that searchers always familiarize themselves with the peculiarities of the databases they regularly search — the indexing, reliability of abstracts, or with the quality of coverage of particular subject matter — and that they use this knowledge to improve answer sets. This finding would indicate that familiarity with the database searched plays a role in improving search results. Designers of information systems, then, might decide that information about indexing, abstracting, or coverage policies should be easily available to searchers, and that a reduction in the number of databases needed by a searcher might be helpful. Training programs could be designed to teach novice users how to familiarize themselves with databases.

Further, when looking for factors that affect design, one is not interested in all possible elements of the search process that are affected by searching experience, but only in those that contribute to the success of a search. The success of a search, in turn, is measured by search-outcome variables, which as yet have not been correlated with search-process variables (except for some obvious relationships such as the correlation between effort variables and recall (Fenichel, 1981)). As we have seen, to discover which elements of the search process are important to retrieval we need first to understand the search process itself; we need "to understand what is actually

happening at the man-machine interface in online systems" (Fenichel, 1980, 125) — a statement that reflects the state of knowledge as well today as it did in 1980.

How can we find out what is actually happening at the human-machine interface? One method is to guess which elements of the search process are significant and to carry out experiments that test the effect of these variables on search-outcome variables. Katzer and his colleagues, for example, measured the effect of various types of search keys (e.g., descriptors, free-text terms in titles) on search results (Katzer et al., 1982).

While methodologically sound, this approach presents a major problem: it forces searchers to use a search strategy which they may not consider optimal for the request. This problem could have been avoided if the search process had been better understood. For example, a study of online searching styles in which searchers were observed performing their regular, job-related searches found that searchers use the difference between the types of search keys as a central vehicle to improve search results (Fidel, 1984). Because Katzer's experiment confined its searchers to using only one type, it prevented the searchers *a priori* from conducting a successful search.

Another approach to understanding the search process is to analyze the search process itself as performed by actual searches. One such study found that experienced searchers consistently used certain databases to help them to formulate a search strategy, and that they were willing to follow leads that were suggested by the retrieved citations, even though these approaches deviated from concepts included in the original request (Fidel, 1984). Selecting a certain search strategy or following new leads obviously affects search outcome, and therefore are important elements of the search process. Moreover, it seems that these elements are influenced by experience because novices have no preferred databases and may be reluctant to follow new leads.

These are, therefore, "promising" dependent variables for experiments to test the effect of experience on searching behavior, but they may not be the best choice for an experiment to measure the effect of cognitive styles or personality traits. In fact, there is no reason to assume that there is a universal set of search-process variables that would be affected by any independent variable one might choose to test. Here again, an understanding of the search process would facilitate an informed selection of search-process variables likely to be affected by a given independent variable.

An informed selection of search-process variables would, at the same time, determine the variables to be ignored. This is not a trivial matter: the notion that some unidentified variables — ones that have been ignored — override the effect of the independent variables has also been used as a reason for the unexplained results in searching-behavior experiments. While no concerted efforts have been made to identify such variables, some attempts have

been made to show the effect of variables that are regularly ignored. The most prominent variable in this category is "the nature of the request" (Bates, 1987).

Two studies designed to measure the effect of this variable demonstrated that the "difficulty" of a request affects the search process and its outcome. The NLM study, already mentioned, originally set out to measure the effect of training method. From a total of 18 requests, each of the 200 searcher-subjects was assigned three requests to search. Each of the three requests had a different degree of difficulty, and the degree of difficulty had been determined *a priori* by searcher-judges. The results show a consistent correlation between the difficulty of a request and search-outcome variables: average recall ranged from 7.5% achieved for a "difficult" request to 70% achieved for an "easy" one, and precision scores ranged from 10.5% to 94% (Wanger, McDonald, 1980, vi-6).

In my study of search-process variables mentioned earlier, I also asked experienced searchers to search an easy request and a difficult one. They were asked to search the same requests two months later. Both inter-consistency and intra-consistency scores varied greatly from one request to another. The average consistency with which searchers selected terms in both searches for the same request was 64% for the easy request and 45% for the difficult one. Consistency among searchers in the selection of terms the first time around was 61% and 26%, respectively. The average overlap among the sets retrieved by the same searcher was 68% for the easy request and 26% for the difficult one. Similarly, overlap among sets retrieved by different searchers was 70% and 8%, respectively.

These studies provide strong evidence that the difficulty of a request affects the search process and its outcome. However, this finding *by itself* cannot be directly used in research because the difficulty of the requests was determined by "gut feeling." To be useful, the factors that determine the difficulty of a request must first be identified.

Indeed, a typology of requests has always been considered desirable. Recently, Saracevic and his colleagues developed and used a pre-defined classification of questions. For each of the forty real-life questions that were solicited from volunteers, they noted the structure of the question, its domain, clarity, specificity, complexity, and presupposition (its implied conceptual aspects), as perceived by its originator and by the searchers. Their study found a considerable disagreement in searchers' and originators' classification of requests (Saracevic, 1987, 112-118). Further, none of the requests tested could be singled out as particularly "difficult" because the range of precision and recall for a question varied widely for almost all questions (Saracevic et al., 1987, 85).

These results, however, do not contradict previous findings. They merely point out that the classification used by Saracevic is not helpful in determining the characteristics of difficult questions. The starting point of

a request typology is in the search process; only when the difficulties in searching requests are understood can a typology that is useful for the study of searching behavior be established.

The study of the search process is critical to productive research in online searching behavior. This area of research, however, is in its infancy: it will require techniques that support in-depth analyses, provide for flexibility, and facilitate the gathering of data from a variety of sources. While such techniques are still somewhat new to research in online searching, they have proven fruitful.

References

BATES, Marcia J. 1987. Information: The last variable. In *Information: The Transformation of Society, Proceedings of the 50th Annual Meeting of the American Society for Information Science, Boston, MA, October 4-8, 1987*, ed. Ching-chin Chen, 6-10. Medford, NJ: Learned Information.

BELLARDO, Trudi. 1985. An investigation of online searcher traits and their relationship to search outcome. *Journal of the American Society for Information Science* 36 (July): 241-250.

BRINDLE, Elizabeth Ann. 1981. The Relationship between characteristics of searchers and their behaviors while using an online interactive retrieval system. Ph.D. dissertation, Syracuse University.

FENICHEL, Carol Hansen. 1981. The process of searching online bibliographic databases: A Review of Research. *Library Research* 2: 107-127.

FENICHEL, Carol Hansen. 1981. Online searching: Measures that discriminate among users with different types of experiences. *Journal of the American Society for Information Science* 32 (January): 23-32.

FIDEL, Raya. 1984. Online searching styles: A case-study-based model of searching behavior. *Journal of the American Society for Information Science* 35 (July): 211-221.

HOWARD, Helen. 1982. Measures that discriminate among online searchers with different training and experience. *Online* 6: 315-327.

KATZER, J. et al. 1982. A study of the overlap among document representations. *Information Technology: Research and Development* 2 (April): 261-274.

SARACEVIC, Tefko et al. 1987. *Experiments on the Cognitive Aspects of Information Seeking and Retrieving. Final Report for National Science Foundation.* New Brunswick, NJ: Rutgers — The State University (PB87-157699).

SOERGEL, Dagobert. 1985. *Organizing Information: Principles of Data Base and Retrieval Systems.* Orlando, FL: Academic Press.

WANGER, Judith and Dennis MCDONALD. 1980. *Evaluation of the Online Process.* Santa Monica, CA: Cuadra Associates (PB81-132565).

WOELFL, Nancy Newman. 1984. Individual differences in online search behavior: The effect of learning styles and cognitive abilities on Process and Outcome. Ph.D. dissertation, Case Western Reserve University.

Research Record

TERRENCE BROOKS, EDITOR

Librarianship and Human-Computer Interaction

Raya Fidel

The interdisciplinary area of human-computer interaction emerged as a partnership among researchers from computer science, business administration, psychology, communication, educational technology, and librarianship (i.e., library and information science), among others. It covers a broad spectrum of subjects, ranging from user studies to information storage to information display.

As in any partnership, each member in this area of research is expected to make a contribution and to benefit from the results of the collective endeavor. Librarianship has already collected some benefits. For instance, research in automated indexing is performed most often in computer science departments, and research on bibliographic databases utilizes various database management systems. It is not clear, however, what librarianship has contributed to this partnership because most researchers ignore the experience gained in our field.

Nevertheless, librarianship is the only discipline that deals with the organization of knowledge. Long before computers came into use, libraries developed their own databases, such as books in classified order on shelves or card catalogs, to name only two. The practice of developing these databases fed theoretical investigations and was nourished by them.

This long experience resulted in a substantive body of knowledge in the area of information storage and retrieval, but it is yet to be disseminated to other disciplines. Today, researchers in computer science and business administration work on database design in complete isolation from the rich experience gained by librarianship in the building of bibliographic databases. A major barrier to the transfer of our knowledge to other disciplines—though not the only one—is the language we use: principles of design for bibliographic databases are stated in terms that are relevant only to this kind of databases.

There are some exceptions. Classification research has always been concerned with general principles, and Soergel adapted a database approach in his book about organizing information.[1] To make our theory accessible to database designers, however, we need to expand our territory—to generalize more of our findings and principles to fit a generalized database approach.

Using this approach, this writer analyzed the Anglo-American Cataloging Rules (AACR2) because, in database terminology, they are rules for data collection.[2] Clearly, every database needs its own AACR2; that is, a set of rules for data collection. A generalized version of the AACR2 would, therefore, be relevant to data collection for any database.

Analysis of the rules revealed that they fall into several general categories. For example, a rule to determine the publication date when two or more dates are found in an item is a rule about establishing an element in the bibliographic record. It belongs, therefore, to the establishing elements category. This category of data collection is found in all other databases. Similarly, the rule that requires the date to be recorded after the publisher's name is a format rule—it determines the place of the date in the bibliographic record.

Once the AACR2 rules were divided into general categories, it was possible to apply principles from a generalized database approach to examine the structure of the AACR2—in particular, those that concern the relationships among the categories. One of these principles requires that rules for establishing elements appear before rules for the actual format in which information is recorded. The generalized approach revealed that this practice is not always followed by the AACR2, and at the same time, suggested a preferred organization of rules.

Database designers can benefit from such analysis because the rich experience gathered in descriptive cataloging is available to them, and the general categories identified by the rules in AACR2 can now be applied to the creation of rules necessary for their data collection.

The AACR2 is only one example of knowledge that can be transferred to a more general level. With a general approach that focuses on data, or information, rather than on documents, librarianship is ready to make a substantial contribution to research in human-computer interaction.

Other specialties within librarianship, such as thesaurus construction or online searching behavior, also should be examined from a generalized database approach to both assist database design and to further the analysis of these topics within librarianship.

References

1. Soergel, D.: *Organizing Information: Principles of Data Base and Retrieval Systems* (Orlando, Fl.: Academic Pr., 1985).

2. Fidel, R., and Crandall, M.: "The AACR2 as a Design Schema for Bibliographic Databases," submitted for publication.

• • • • •

Besides columns by the column editor, readers are encouraged to submit articles of 1,000–1,500 words on topics related to research to Terrence Books, Graduate School of Library and Information Science, University of Washington, 133 Suzzallo Library, FM-30, Seattle, WA 98195. The database of dissertations in progress can also be accessed through the column editor. Results of a computer search on specific topics or a book catalog of the entire database (about 220–225) records) are available on request.

Raya Fidel is Assistant Professor, University of Washington Graduate School of Library and Information Science, Seattle, WA 98195.

"The Visible College"

ADELE M. FASICK, EDITOR

The Browsing Collection: A Lab for Library Science Students

Douglas G. Campbell and Barbara J. Fahey

What activities involve a student in the average material selection or collection development course? He/she writes a hypothetical selection policy for a hypothetical collection. The requisite number of reviews are read and materials are selected. An order (generally in the form of a bibliography) is prepared and turned in to the instructor. Feedback in the form of patron satisfaction, censorship problems, or budget difficulties come solely from the instructor. There are probably other activities that innovative library science instructors devise to make the student's experience more inspiring and more real, but the student never

Douglas G. Campbell is Director of Library Services, Northwestern Michigan College, Traverse City, MI 49684, and Barbara J. Fahey is Technical Services Systems Librarian and Assistant Professor, University of Wisconsin Oshkosh, Oshkosh, WI 54901.