

# INDIVIDUAL VARIABILITY IN ONLINE SEARCHING BEHAVIOR

Raya Fidel  
Graduate School of Library and Information Science  
University of Washington  
Seattle, Washington

## ABSTRACT

Variability in the behavior of searchers who answer test requests may influence results of experiments in information retrieval. Data collected from search transcripts and verbalized protocols of searches for two identical requests performed by professional online searchers show that: (a) searcher frustration can affect test results; (b) searcher variability depends on the request searched; and (c) sources and forms of searcher variability can be identified.

## INTRODUCTION

Beginning with the Cranfield studies [1], numerous experiments have been carried out to discover relationships between design variables, such as the nature of an index language or method of indexing, and subsequent retrieval performance. One recent example is the investigation of the effect of the type of search keys (e.g., controlled descriptors, terms in documents' titles) on information retrieval systems [2].

To determine the effect of design variables on retrieval, researchers typically have compared attributes of answer sets retrieved under one condition with similar attributes of sets retrieved under other conditions. recently, search process attributes have been compared as well. To facilitate these comparisons it has been common to search the same request repeatedly under varying conditions. For example, to test the effect of vocabulary control on information retrieval, one can search the same requests on system A (descriptor searching only) and on system B (free-text searching only) and then compare the sets retrieved for each request-system pair.

Regardless of the degree to which a system in such experiments is automated, someone search a selected set of test requests. This searcher may be the investigator, a professional intermediary, or an end-user. Most experiments, employ more than one searcher so that each searcher can search each request only once. The assumption is that if searchers were to search a test request twice, the second search would benefit from the experience and results of the first search. Having the same person search a request twice, then, introduces an additional -- and undesirable -- variable: the number of times a person has searched a request.

To overcome this unneeded variability, researchers elect a number of searchers to participate in their study, each searching each test request only once. This method is believed to introduce more control over test requests since a request is searched on each system with the same request-related knowledge and experience. Using this method, however, implies that each test request is searched on each system by a different searcher. Variations among the resulting answer sets may, therefore, be determined not only by varying attributes of tested systems but also by the differences among participating searchers. While searcher variability has been known to affect tests results (e.g., [2]), there are no data to reflect the magnitude of its effect or to suggest possible methods to control it.

An exploratory study is being conducted to determine whether searcher variability could be analyzed to explicitly express its effect. The study as a whole focuses on the search process and has three objectives:

1. to develop means to express searcher variability that will be useful for retrieval experiments;
2. to explore whether searcher variability is dependent on the request searched;
3. to identify possible sources and forms of searcher variability.

## STUDY METHOD

Ten experienced intermediaries searched two test requests, verbalizing their thought processes while formulating the search strategy and while searching online. Their search transcripts and answer sets were then analyzed and compared.

These searchers were selected from among the members of the Seattle Area Hospital Library Consortium and are similar to one another in their work environment and experience. All are hospital librarians who graduated from a library school more than four years ago. All but one have been searching for more than two years. Only one searches less than eight requests a week, and half search more than fifteen. Additionally, three of the ten searchers did not attend the initial training provided by the National Library of Medicine (NLM). Two of the searchers have formal education in medical-related subjects, and all but one provide free searching services.

Two requests were submitted to the searchers:  
1. I want to know everything about health fairs.

2. Does health promotion have any effect on the involvement of people in their own health care?

The first request is the simplest of the two: the central concept, health fairs, is concrete and well defined; it is matched exactly with a descriptor in the MeSH vocabulary, and the number of citations posted to the descriptor is reasonable. The second request is somewhat more complex: while the first concept, health promotion, can be exactly matched to a descriptor, it is not well defined; in addition, the second concept, involvement of people in their own health care, is not expressed in a term that can be searched in a straightforward manner.

The searchers were asked to approach these requests as if they were ordinary and regular requests. They were instructed to use the search system with which they were most comfortable and to search a database they would normally search for these requests. All searchers, but one, selected the NLM search system, and all searched the Health Planning and Administration database. Two searchers additionally approached the Medline database.

Search transcripts for each request-searcher pair were analyzed according to the search process variables used by Fenichel [3] and to the searcher consistency measurements suggested by Saracevic [4]. In addition, the overlap between answer sets was determined using measurements developed by Katzer [2]. These variables were selected for the study because they relate to the search process and its outcome. Moreover, they have been used as dependent variables to measure effects of design variables on information systems. They are here measured when design variables are fixed. Therefore, variability in test results should be attributed only to searcher variability.

## RESULTS AND DISCUSSION

Data for the first phase of this exploratory study have just been collected, and their analysis is only preliminary and minimal at this time. The results of the completed study will be reported at the conference. It is useful, however, to show now some of the intriguing findings, even though they are preliminary.

A. Searcher frustration. One fact that stood out at the very beginning of data collection is the deep frustration that the searchers experienced because they were not able to conduct a pre-search interview. While some saw no difficulty in searching the first request, all had trouble searching the second one. At various points during the search, they voiced a strong dissatisfaction with their moves and the results they were able to obtain. All searchers found it important to explain that their performance would have been much better were there a user present to be interrogated. Although this

finding is not surprising -- after all, the importance of a pre-search interview has never been downplayed -- its significance has been neglected.

The ramifications of eliminating the interview from searches performed in experiments should be investigated. One example illustrates how this factor can bias test results. Three searchers felt that they could not complete their searches for the second request without consulting the requester. They decided, therefore, to print trial sets: each represented a different approach to answering the request and included a few citations which were neither the best nor the most typical. At that point, they explained, they would use these sets to help the requester to better define the request. In fact, the number of citations retrieved by each of these searchers was about half of the mean number of citations retrieved for this request.

When analyzing the data, however, one has to consider these sets as answer sets. Ignorant of the purpose of such "answer sets," researchers might mistakenly consider the low precision and low recall that are likely to be measured for such sets as the effect of some design variables.

B. Searcher variability. Data collected in this study also strongly suggest that searcher variability depends on the request searched. Variability among searchers in scores for dependent variables is more pronounced for the second request than for the first one. This observation is supported by three findings as displayed in Table 1: (a) standard deviation scores for all search process variables for the second request are consistently higher than those for the first one; (b) exhaustivity score, which measures searcher consistency, is much higher for the first request than for the second one; (c) overlap among sets retrieved for the first request is dramatically higher than overlap among sets for the second request.

The first seven variables are search process variables. For these, the difference between the ranges of values for each request highly supports the assumption that variability among searchers depends on the request:

1. The values for the number of commands used for the first request ranged from 2 to 19, and from 13 to 36 for the second one. The difference in this range is more dramatic for commands to create a set to represent a single term: from 1 to 3 for the first request in comparison to 1 to 8 for the second.

2. Number of search keys used, whether controlled-vocabulary descriptors or free-text keys, ranged from 1 to 5 for the first request but from 5 to 17 for the second.

3. The difference in the number of cycles (a sequence of commands for (a) entering search

terms, then (b) combining terms, and then (c) printing or displaying) was not extreme: 1-5 for the first request vs. 2-8 for the second.

4. Variability in the number of sets viewed, excluding sets that were printed to be included in the answer set, on the other hand, depended strongly on the request. While searchers viewed between zero and 3 sets for the first request, they viewed between zero and 8 sets for the second.

5. Connect time (in minutes) from beginning to end of online session was also affected by the request. It ranges from 2 to 15 minutes for the first request but from 9 to 32 for the second.

6. The range of speed of searching did not vary greatly between the two requests: from 0.33 commands per minute to 1.75 commands per minute for the first request and a range of a similar magnitude for the second request: 0.97 to 2.71 commands per minutes. Data show, however, that speed of searching the second request is consistently higher than that recorded for the first request, with half of the searchers doubling their speed in searching the second request. The notion that speed of searching depends on the request should be investigated further.

7. Number of citations retrieved ranged from 21 to 48 for the first request and from 16 to 68 for the second, supporting the suggestion that variability in the number of citations retrieved depends on the request.

The remaining variables actually measure variability in searcher behavior.

8. The first measurement of this sort is exhaustivity, which measures the agreement between search terms selected by searchers. Asymmetric exhaustivity between a pair of searchers,  $E(i,j)$ , is defined as the number of search terms selected in common by two searchers divided by the number of search terms selected by searcher  $i$ . Mean scores of exhaustivity for the two requests (0.61 vs. 0.26) clearly demonstrate that searcher variability was larger for the second request.

Citations retrieved by all searchers were collected together to form a union set of distinct citations retrieved for each request. A total of 73 citations was retrieved for the first request while the second one resulted in 240 distinct citations. Since the mean number of citations retrieved for each request is not highly different from that of the other, the gap in the total numbers can be explained by low overlap.

9. Asymmetric overlap between a pair of searchers,  $E(i,j)$ , is defined as the number of citations retrieved in common by the two searchers divided by the number of citations retrieved by searcher  $i$ . Mean scores for overlap (0.70 vs. 0.08) show again the

increased variability demonstrated in the second request.

C. Sources of variability. The evidence that is provided here needs to be reinforced by a study to test the hypothesis that searcher variability is affected by test requests. The special importance of this exploratory study, however, is in its potential to discover the sources and forms of searcher variability. At this preliminary stage, it is too soon to analyze these aspects systematically. One example, however, illustrates the line of investigation to be taken in the study.

Although the request about health fairs, can be searched straightforwardly, three searchers adopted a broad view; they assumed that even though the user is interested in health fairs, he or she should get material about other activities related to health promotion. While not distinctly different from the rest in scores for the second request, they are easily singled out in their scores for the first request: they used more commands (mean of 15 vs. a mean of 7.5 commands for all searchers); entered more search keys (mean of 6 vs. 3.4); had more cycles (3.3 vs. 2.0); viewed more sets (2.3 vs. 1.2); performed the longest searches (13.7 vs. 7.6); and retrieved the largest sets (44.7 vs. 28.7); their exhaustivity scores are the lowest (0.29 vs. 0.61) but they score relatively high on the individual measurement with one another; and, their overlap scores are remarkably lower than the rest (0.40 vs. 0.70), scoring again highest on the individual measurements among themselves.

These data clearly suggest that one source of searcher variability is the perception of requests by searchers and that one of its forms is the degree of specificity in which searchers perceive requests. These findings also suggest that this variability is more likely significant for requests that seem to be specific and that could be searched straightforwardly.

The implications of such findings to the design of retrieval experiments are far-reaching. One suggestion, for example, is that test requests be simple and straightforward so as to keep searcher variability on a relatively low level and to control for the degree of specificity in which searchers perceive requests. In the next phase of the study, the same searchers will be asked to search the same requests again, about two months after their first session. With hope, that data analysis will uncover additional sources and forms of searcher variability, indicating which sources and forms are situational and which are inherent to differences in searching behavior.

NOTES

1. Cyril W. Cleverdon, Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems. (Cranfield, England: College of Aeronautics, 1962.)

2. J. Katzer, et al., "A Study of Overlap Among Document Representation," Information Technology: Research and Development, 2 (1982) 261.

3. Carol Hansen Fenichel, "Online Searching: Measures that Discriminate among Users with Different Types of Experiences," Journal of the American Society for Information Science, 32 (January 1981) 23.

4. Tefco Saracevic, "Measuring the Degree of Agreement between Searchers," Proceedings of the ASIS Annual Meeting, 21 (1984) 227.

Table 1. Summary of results

	Request 1 Health Fairs		Request 2 Health promotion	
	Mean	Standard Deviation	Mean	Standard Deviation
Commands used	7.5	5.73	23.8	6.99
Search keys used	3.4	2.50	8.7	3.65
Cycles	2.0	1.25	4.3	1.70
Sets viewed	1.2	1.13	3.4	2.11
Connect time	7.6	4.74	15.3	7.94
Speed	1.0	0.42	1.7	0.59
Citations retrieved	28.7	11.40	38.2	18.59
Exhaustivity	0.61	0.27	0.26	0.09
Overlap	0.70	0.23	0.08	0.06