Searchers' Selection of Search Keys: III. Searching Styles

Raya Fidel
Graduate School of Library and Information Science, University of Washington, Seattle, WA 98195

Individual searching style has a primary effect on searching behavior. The case study method provided data about elements of searching styles through: (1) observation of 47 professional searchers performing their job-related searches; and (2) analysis of verbal and search protocols. Statistical associations among a number of variables reveal three dimensions of searching behavior: level of interaction, preference for operational or conceptual moves, and preference for textwords or descriptors. The interactive searcher actively modifies search strategies and uses a relatively large number of search keys (or search terms). The operationalist searcher prefers to employ operational moves and is less concerned with recall than his conceptualist counterpart. The free-text searcher prefers to use textwords, has developed a habit of not consulting a thesaurus, and is more likely to regularly receive practical questions. Findings also indicate that searchers encountered difficulties in achieving satisfactory recall, regardless of their searching style. Future research should focus on mechanisms to improve recall and on factors that affect the development of searching styles.

Introduction

Interest in investigating online searching behavior has declined in recent years because most experiments have failed to provide conclusive results. A common explanation for this failure is that individual searching styles override most measured attributes of searching behavior (Fidel, 1987). Although the term *searching styles* is freely used in the literature about online searching, it is not clear just what the concept embodies. We still need to understand what characteristics of searching behavior constitute a searching style, that is, in what way one individual searcher is different from another, all external conditions being equal.

This article is based upon work supported by the National Science Foundation under Grant number IST 85-09719.

Received January 17, 1990; revised April 30, 1990, June 27, 1990; accepted September 7, 1990.

The study reported here investigated the rules searchers used when they selected search terms (or keys), and the modifications of search strategy (or moves) they made to improve search results. The data collected for the study suggest typical characteristics of searching behavior that identify elements of searching style: Individual searchers differ from one another in their degree of interaction during a search, and in their preference for type of move and for type of search key. The data also point to general characteristics of searching behavior that were common to all searchers: Searchers put more effort into improving recall than in improving precision, but when they did try to improve precision, they usually made moves of one type.

The Method

To investigate the rules employed by searchers for the selection of search keys and the moves they made during a search, 47 professional searchers were observed performing their regular, job-related searches of bibliographic databases (Fidel, 1990). Each searcher was observed for approximately five searches, for an overall total of 281 searches. Two formal models were developed from search protocols, from verbal protocols of thought processes while searching, and from interviews with searchers to determine reasons for their search-key selection:

(1) *The selection routine* which is a decision tree that describes the rules used by searchers in the selection of search keys, descriptors or textwords. This model is beyond the scope of this article; it is described in detail in a previous article (Fidel, 1991a).
(2) *Moves in online searching* is a list of modifications of search strategies that are aimed at improving the result of a search. They are presented in Table 1. The moves are of two types: operational moves which do not change the meaning of a request, and conceptual moves which change the meaning of a request. Further, the moves are divided into three groups, depending on their purpose: moves to re-

 CCC 0002-8231/91/070515-13$04.00

TABLE 1. Moves in online searching.

Operational moves		Conceptual moves	
Moves to reduce the size of a set			
Weight 1	Limit a descriptor to be a major descriptor.	Intersect 1	Intersect a set with a set representing another query component.
Weight 2	Intersect free-text set with a broader descriptor.	Narrow 1	Intersect a descriptor set with a set created by more specific free-text terms.
Weight 3	Limit free-text terms to occur in a predetermined field.	Narrow 2	Qualify descriptors with role indicators.
Weight 4	Require that free-text terms occur closer to one another in the searched text.	Narrow 3	Select a narrower concept.
Weight 5	Limit to documents of a certain form.	Intersect 2	Intersect sets with role indicators.
Negate	Eliminate unwanted elements by using the AND NOT operator.		
Eliminate	Eliminate a term from the formulation.		
Limit 1	Limit to documents written in a particular language.		
Limit 2	Limit to documents published, or indexed, in a particular period of time.		
Limit 3	Limit to documents retrieved from a specific portion of the database.		
Limit 4	Limit to sources that have, or do not have, a certain term in their titles.		
Cut	Submit only part of the retrieved answer set, arbitrarily selected.		
Moves to enlarge the size of a set			
Add 1	Add synonyms and variant spellings.	Expand 1	Enter a broader descriptor or term.
Add 2	Add descriptors as free-text terms.	Expand 2	Group together search terms to broaden the meaning of a set.
Add 3	Add terms occurring in records of relevant citations retrieved.	Expand 3	Group together a descriptor with an equivalent role indicator.
Add 4	Add terms from database's index that have a high number of postings.	Expand 4	Represent a query component explicitly only by qualifying another component with role indicators.
Add 5	Move to a new database.	Exclude	Exclude from a formulation concepts present in most documents in a database.
Include	Group together a descriptor with all the descriptors that are its narrower terms.	Expand 5	Supplement a specific answer set with sets representing broader concepts.
Cancel	Eliminate restrictions previously imposed.		
Moves to increase both precision and recall			
Refine	Find a "better" search key.	Probe 1	Construct an indexing-probe set.
		Probe 2	Use the difference among the number of postings for a search term in various databases to decide how to represent components in each database.

duce the size of a set, to enlarge the size of a set, or moves to improve a set by increasing both precision and recall. A detailed description of the moves is available elsewhere (Fidel, 1985).

In addition to the models, eight variables were defined: some that represent characteristics of searching behavior such as average number of search keys per request, and others that represent characteristics of searchers such as subject specialty. These variables were measured at the searcher level where the data for each searcher were averaged over all the searches she or he performed for the study, so that each searcher was considered a distinct instance (a total of 47 instances).

The eight variables examined were:

(1) *Number of moves.* The average number of moves, i.e., search-strategy modifications, made by a searcher per search.

(2) *Operational moves ratio.* The percentage of operational moves, that is: the total number of operational moves, divided by the total number of moves made by a searcher. This variable indicates whether a searcher prefers to make one type of move over the other. Searchers who score relatively high on this variable are called *operationalist* searchers because they prefer to make operational moves, and those who score relatively low are called *conceptualist* searchers.

(3) *Number of search keys.* The average number of search keys selected by a searcher per search.
(4) *Textwords ratio.* The percentage of textwords selected, that is: the total number of textword keys, divided by the total number of search keys selected by a searcher. This variable reflects the tendency of a searcher in the selection of search keys. Searchers who score relatively high on this variable prefer to use textwords and those who score relatively low prefer descriptors.
(5) *Thesaurus neglect ratio.* The percent of textwords entered *without* consulting a thesaurus, that is: the total number of terms entered by a searcher without consulting a thesaurus, divided by the total number of search keys entered by the searcher.
(6) *Recall tendency.* The percentage of moves made to increase the size of a set, that is: the total number of recall moves, divided by the total number of moves made by a searcher. This variable reflects the degree to which a searcher is usually concerned with improving recall.
(7) *Subject area.* The subject area in which a searcher specializes. This variable had four values: medicine, the sciences, social sciences (including both the social sciences and the humanities), and general (for searchers who habitually search requests in a variety of subjects, as is often the case in public libraries or with independent consultants). This variable was selected to examine whether the subject specialty of a searcher affects his or her searching behavior.
(8) *Environment.* The environment in which a searcher works.This variable had three values: practical environments, theoretical environments, and general environments. A practical environment is a work place in which searchers are usually called upon to search requests that result from immediate and practical problems, for instance, most small or medium-size consulting companies or industries. In contrast, a theoretical environment is an establishment whose users are often involved in research or investigation, for instance, universities or regulatory agencies. Search environments that could not be assigned any of the first two categories were called general environments. This variable was selected to examine whether the nature of the requests habitually searched has an effect on searching behavior.

Correlation tests between these variables, data about the frequency of moves selection, as well as data about the reasons for search-key selection, together clearly point to practices in searching behavior that are common to all searchers and to those practices that are typical of an individual searching style.

Searchers' Selection of Moves

Table 1 lists moves that were used to modify search strategies. The table includes 20 operational moves, that is, moves that do not change the meaning of a request. These are divided into moves to reduce the size of a set (12 moves), moves to enlarge the size of a set (seven moves), and moves to simultaneously increase both precision and recall (one move). Thirteen moves are conceptual moves, that is, moves that do change the meaning of a request. Of these, five are moves to reduce the size of a set, six to enlarge the size of a set, and two are made to increase both precision and recall simultaneously. A full explanation of the moves has been published elsewhere (Fidel, 1985).

Table 2 reports the frequencies with which searchers selected different moves. For each move, it gives the number of times the move was used and the frequency with which it was used in relation to the total number of moves.

Operational or Conceptual Moves

The 47 searchers made a total of 1,244 moves in their 281 searches. Of these, 497 (40%) were conceptual and 747 (60%) were operational.

One operational move, the move to add a database (Add 5), was induced by the availability of databases, while the others were induced by consideration of the request. The move to add a database is often called for when a complete run of a database is split into a number of databases, each covering a different period of time. Searchers selected this move 312 times. If we eliminate this move from the list, the relative proportions of conceptual and operational moves changes: 435 moves (47% of the total) were operational moves and 497 (53%) were conceptual moves. That is: ***No specific type of move was most common: About half of the moves selected by searchers were conceptual moves and the other half were operational.***

While searchers in general do not prefer one type of move to another, individual searchers often do prefer a specific type of move. Of the 47 searchers, 25 (53%) made one type of move more than 70% of the time, and 35 (74%) made one type of move more than 60% of the time. Three searchers selected moves of one type only. Further, Analysis of Variance shows that searchers who score highly on the operational moves ratio are significantly different from those who score lower on this variable ($F(46, 280) = 1.8$, $p < .01$). These findings show that the type of move a searcher prefers is a matter of individual searching style. As will be explained later, operationalist searchers demonstrate other typical characteristics of searching behavior, as do conceptualist searchers.

Precision and Recall Enhancement

Moves to enlarge the size of a set are almost always made to increase recall. In contrast, moves to reduce the size of a set may or may not improve precision; at times they are made simply to reduce the size without eliminating nonrelevant citations, and at other times they actually improve precision. In this article, how-

TABLE 2. Frequency of move selection.

Move	Number of times	(%)	Move	Number of times	(%)	Total
Moves to reduce the size of a set						
Weight 1	35	2.81	Intersect 1	75	6.03	
Weight 2	2	.16	Narrow 1	8	.64	
Weight 3	37	2.97	Narrow 2	6	.48	
Weight 4	18	1.45	Narrow 3	52	4.18	
Weight 5	15	1.20	Intersect 2	4	.32	
Negate	34	2.73				
Eliminate	5	.40				
Limit 1	13	1.04				
Limit 2	63	5.06				
Limit 3	10	.80				
Limit 4	1	.08				
Cut	22	1.77				
Subtotal	255	20.50		145	11.65	400
Moves to enlarge the size of a set						
Add 1	63	5.06	Expand 1	56	4.50	
Add 2	7	.56	Expand 2	81	6.51	
Add 3	43	3.46	Expand 3	2	.16	
Add 4	1	.08	Expand 4	21	1.69	
Add 5	312	25.08	Exclude	4	.32	
Include	15	1.21	Expand 5	145	11.66	
Cancel	27	2.17				
Subtotal	468	37.62		309	24.84	777
Moves to increase both precision and recall						
Refine	25	2.01	Probe 1	37	2.97	
			Probe 2	7	.56	
Subtotal	25	2.01		44	3.54	69
Total	748	60.05		498	39.95	1246

ever, *enlarging a set* is used synonymously with improving recall and *reducing a set* with improving precision.

Examining next the purpose of the moves, data in Table 2 show that 63% of the moves were employed to increase the size of a retrieved set, while only 32% of the moves were made to reduce the size of a set. Although the move Add 5 is often necessitated by the distribution of information among databases, it is always made to improve recall and therefore should be counted here, even though it was eliminated from the comparison of the frequency with which conceptual and operational moves were made. Therefore: ***The number of moves to increase recall was almost double the number of moves to increase precision.***

This observation suggests that low recall is a problem searchers frequently encounter in searching. As will be explained later, additional evidence points to the difficulty in achieving satisfactory recall in the databases currently available.

Table 2 shows that the number of operational moves made to reduce the size of a set is almost double that of the conceptual moves made for the same purpose. On the other hand, a similar comparison among the moves made to enlarge the size of the set reveals different relationships. The number of operational moves is higher, but, after eliminating the move Add 5 from the comparison, the number of conceptual moves is almost double the number of operational moves in this category. Similarly, the number of conceptual moves to increase *both* precision and recall is almost double the number of operational ones. This leads to the observation that: ***While searchers made both operational and conceptual moves to improve recall, they employed mostly operational moves to improve precision or otherwise to reduce the size of a set, and conceptual ones to improve both precision and recall.***

The tendency of searchers to increase precision primarily with operational moves points to a general attribute of searching behavior that is not dominated by individual searching styles. Therefore, to further test this finding, two variables were defined on the *search level*, where each search was considered a distinct instance (a total of 281 instances): (1) operational moves ratio, which is the number of operational moves divided by the total number of moves made during a search; and (2) precision moves, which is the number of moves made to reduce the size of a set in a search. A Pearson Product-Moment Correlation test shows that operational moves ratio is directly related to precision moves ($r(279) = .240$, $p < .01$). That is: ***Precision moves are more likely to be operational than conceptual.***

The observation that conceptual moves were rarely used to improve precision can be explained by the reluctance of searchers to narrow the meaning of a request. All the conceptual moves to reduce the size of a set narrow the meaning of a request in one way or an-

other (Table 1). On the other hand, operational moves, by definition, keep the meaning of a request intact. Thus, while searchers freely broadened the meaning of a request in order to increase recall, they tried to avoid narrowing it when higher precision was sought, and instead preferred to use operational moves.

Generally it must be surmised that when, in the searcher's estimation, results are "off target" and both precision and recall are unsatisfactory, searchers cannot rely on operational moves alone. Most often, a change in the meaning of the request is needed, that is, a conceptual move.

These conclusions suggest that the type of a move a searcher makes depends not only on the searcher's personal searching style, but also on the purpose of the move. A conceptualist searcher, for instance, would probably prefer to make operational moves when she aims at improving precision. As a result, one should not expect a searcher to employ moves of one type only. Searching style, whether operationalist or conceptualist, is not "pure," and thus can be determined only by observing a tendency in the selection of moves over a number of searches.

Individual Range of Moves

The array of moves selected by any individual searcher was rather limited. Of the 33 moves available to searchers, the average number of moves that constituted a searcher's repertoire was 8.32, with median of 8.00 and standard deviation of 3.52. The maximum number of individual moves that a searcher employed was 17, and the minimum was three. ***On the average, each searcher employed about 25% of the moves that were available.***

This result validates the commonly held belief that searchers acquire a repertoire of moves from which they select the pertinent ones when a problem arises. Further, it shows that this repertoire is rather limited even though searchers are likely to be familiar with most of the moves that are possible. There are probably a variety of factors which put limits on this repertoire, such as the difficulty in recalling all possible moves under time pressure. However, because we do not know what these factors are, it would be useful for searchers to have unobtrusive reminders about possible moves, so they can make their selection from all the possible moves. Therefore: ***Search systems should remind searchers of the complete array of moves possible in online searching.***

The Interactive Searcher

The average number of moves a searcher made per search reflects the degree of interaction during a search: the larger the number of moves, the more interactive the searcher. Searchers made an average of 5.04 moves per search, with a median of 4.29 and a standard deviation of 3.74. The minimum average number of moves per search for a searcher was 1.00 and the maximum was 18.78. The relatively large standard deviation among the searchers, and the great difference between the minimum and maximum averages for a searcher, together suggest that the average level of interaction varied greatly from one searcher to another. Further, Analysis of Variance revealed that the average number of moves per search for highly interactive searchers varied significantly from the average for less interactive ones ($F(46, 280) = 4.45$, $p < .01$). That is, each searcher has his or her own typical level of interaction. The following examines what characteristics of searching behavior are typical of searchers who are more interactive than others.

Number of Moves

Statistical analyses show that number of moves is associated with only one variable: number of search keys. A Pearson Product-Moment Correlation test shows that number of moves is directly correlated with number of search keys ($r(45) = .777$, $p < .01$). An obvious explanation for this correlation is that moves are made with search keys, and that entering a search key is a move, in which case the association is trivial. But this explanation is not grounded in findings related to actual searching. Examination of the model Moves in Online Searching (Table 1) shows that of the 33 moves only 12 require the use of search keys for their execution, and those 12 moves constitute only 30% of the moves made in the study. It would seem then that the association between number of search keys and number of moves points to a significant pattern in online searching behavior.

The search key/moves association shows that interactive searchers are characterized by two variables: number of moves and number of search keys. Specifically: ***Interactive searchers, who, on the average, make more moves per search than their colleagues, are likely to use a larger number of search keys than searchers who are less interactive.***

Of equal importance to this discussion is the finding that number of moves does not correlate with any of the other variables.

- Operationalist searchers and conceptualist searchers are interactive to the same degree.
- Interactive searchers use the same proportion of textwords as searchers who are less interactive.
- Interactive searchers neglect to use a thesaurus just as often as their colleagues.
- Interactive searchers are concerned with recall to the same degree as other, less interactive searchers.
- The subject area in which a searcher specializes or the environment in which the searcher works do not affect the searcher's level of interaction.

In summary, the finding that the number of moves per search was typical for each individual searcher, in

addition to the finding that number of moves does not correlate with most variables, suggests that some searchers are more interactive than others. That is, the level of interaction during a search is a personal attribute; it is an element of searching style.

Number of Search Keys

Because the degree to which a searcher is interactive is also represented by the typical number of search keys per search, it is useful to examine the variable *number of search keys,* that is, the average number of search keys a searcher used in a search.

Searchers selected an average of 13.31 search keys per search, with a median of 9.20 and a standard deviation of 12.80. The minimum average number of search keys per search for a searcher was 2.80 and the maximum was 68.75. The relatively large standard deviation among the searchers, together with the great difference between the minimum and maximum averages for a searcher, indicate that: The average number of search keys per search varied greatly from one searcher to another. Further, Analysis of Variance found this difference to be significant ($F(46, 280) = 4.07, p < .01$). This finding suggests that with experience, each searcher develops a habit about the "reasonable" number of search keys to be used in a "normal" search. The study results show that the average number of search keys per search is typical for a searcher, and that a person's searching style will determine how extensive will be her or his use of terms.

Like the variable number of moves, the variable number of search keys does not associate with most of the variables. There is one exception: unlike its counterpart, number of search keys does correlate with environment.

Analysis of variance shows that number of search keys is associated with the environment in which a searcher works ($F(2, 44) = 5.22$, $p < .01$). Searchers who work in *practical* environments use an average of 6.76 search keys per search, those in *theoretical* environments use an average of 18.56 search keys per search, and those who work in *general* environments use an average of 11.76 search keys. A post-hoc test shows a significant difference between the practical and theoretical environments. Although environment as a variable lacks a rigorous definition, this association suggests that: ***Searchers who are used to answering practical questions use a considerably smaller number of search keys per search than do searchers who habitually answer theoretical requests.***

This conclusion was unexpected, particularly because it was found that searchers who habitually answer theoretical questions do not make more moves than their peers who answer practical questions. One explanation for this discrepancy might be that theoretical requests usually require high recall, and that high-recall requests require a relatively extensive use of search keys. This explanation, however, is not supported by the findings of this study. The findings show that although science requests that are theoretical may require higher recall than practical requests, searchers who frequently made recall moves did not use a relatively large number of search keys, as will be explained later.

Another explanation is that searchers who answer theoretical requests encounter terminological difficulties more frequently than do their colleagues. Concepts in theoretical questions might be less well-defined than those typical of practical questions and therefore might require a relatively large number of search keys to express each concept. It could be assumed then that searchers in theoretical environments develop a habit of representing each concept with a relatively large number of search keys, but are not necessarily more interactive than their colleagues who answer practical questions. It is beyond the purpose of this study to substantiate this explanation; it is provided here as a suggestion for further exploration.

The lack of association between recall tendency and number of search keys is somewhat unsettling. The finding that searchers who frequently make recall moves do not use a larger than average number of search keys, is contradictory to the commonly-held idea that requests that require high recall would also require a relatively large number of search keys. This finding, however, may be explained by the two observations. First, the use of additional search keys is not typical only of attempts to increase recall. Several moves to increase precision require the use of additional search keys, for example, Intersect 1 or Narrow 1. Second, searchers increase recall *either* by using more search keys, *or* by making other moves to increase recall, in particular, moves that do not require the use of additional search keys. One example of such a move is the move to supplement a specific answer set with sets representing broader concepts (Expand 5), which is commonly achieved by eliminating a request concept from the query formulation. This conceptual move, intended to increase the size of a set, was made 19% of the times that recall moves were made (31% of recall moves, ignoring Add 5), and it does not require entering additional search keys. It can happen then, that when searchers make moves to increase recall they do not use additional search keys.

Operationalist and Conceptualist Searchers

The data collected in this study indicate that individual searchers often prefer a specific type of move, that is, the type of moves a searcher prefers is an element of searching style.

Operational Moves Ratio

The variable that measured this element is operational moves ratio, which is defined as the percentage of operational moves made by a searcher. This variable

correlates with four other variables: textwords ratio, thesaurus neglect ratio, recall tendency, and subject area.

Operational moves ratio directly relates to textwords ratio ($r(45) = .434, p < .01$). This correlation suggests that: ***Operationalist searchers prefer to use textwords and conceptualist searchers prefer to use descriptors.***

Similarly, operational moves ratio relates directly to thesaurus neglect ratio ($r(45) = .413, p < .01$). That is: ***Operationalist searchers are more likely to avoid consulting a thesaurus than conceptualist searchers.***

This conclusion agrees with the previous finding because it is plausible to assume that searchers who prefer to use textwords are more likely to neglect consulting a thesaurus than their peers who prefer to use descriptors.

Another variable that distinguishes operationalist from conceptualist searchers is recall tendency, which represents the degree to which a searcher is usually concerned with improving recall. While a Pearson Moment-Product Correlation test shows that the variables operational moves ratio and recall tendency, as defined above, do not significantly correlate ($r(45) = -.186$, NS), a closer examination of the moves to improve recall is warranted here.

The most common move made by the study's searchers is the move to change a database (Add 5). This is an operational move which is made to improve recall, but is often induced by the distribution of information among databases and not by the personal tendency of a searcher. This situation is most apparent when search-system vendors split a complete run of a database into a number of databases, each covering a certain period of time.

To examine the association between the variables operational moves ratio and recall tendency, the two variables were redefined to exclude the instances where searchers changed databases in order to search additional segments of the same database.

Using these new definitions, operational moves ratio relates inversely to recall tendency ($r(45) = -.405$, $p < .01$). That is: ***Operationalist searchers put less emphasis on recall than do conceptualist searchers.***

Further, the subject area in which a searcher specializes has a significant effect on operational moves ratio ($F(3, 43) = 6.31, p < .01$). The averages are as follows:

Subject area	Operational moves ratio (%)
Medicine	45
Social sciences and the humanities	51
Science and technology	76
General	79

A post-hoc test found a significant difference between general searchers and both medical and social-sciences searchers, as well as between medical and science searchers. That is: ***Science searchers and searchers who have no subject specialty are more likely to make operational moves than their colleagues in other subject areas.***

The large percentage of operational moves among generalist searchers can be explained by the nature of their task. They are called upon to answer requests in a large variety of subjects. Unlike searchers who specialize in one subject area, their knowledge of the subject of a request is usually limited. This limitation prevents them from making conceptual moves because conceptual moves, since they change the meaning of a request, require some subject knowledge. A person who is familiar with the subject of a request is more likely to feel comfortable modifying its meaning for the purpose of a search than a person who has little experience in the subject matter.

While the tendency to make operational moves among generalist searchers is well understood to be inherent in the nature of their searching, finding this tendency among science searchers is puzzling. The significant difference between science and medical searchers could be explained by the average number of databases they use per search. Data show that medical searchers used an average of 1.33 databases per request, while science searchers used an average of 2.64 per request. It is possible that having to deal with a larger diversity of databases and thesauri, science searchers who otherwise would appear to be conceptualist find it overwhelming to manipulate the meaning of a request. Medical searchers who tend to be conceptualist, on the other hand, are more free to follow their personal tendencies because they handle a smaller diversity of databases and thesauri: they typically search MEDLINE with the MeSH vocabulary. Incidentally, the average number of databases per search could also explain the operational tendencies among generalist searchers who used an average of 2.48 databases per request.

Operational moves ratio does not significantly correlate with number of moves, nor with number of search keys, as explained in the previous section. In addition, it does not significantly correlate with environment ($F(2, 44) = 1.24$, NS). ***The environment in which a searcher works has no effect on the searching style of the searcher, whether operationalist or conceptualist.***

The "Free-Text" Searcher

When the study's searchers gave reasons for their search-key selection, they often mentioned a general preference for a certain type of search key. Some searchers said that they preferred to use descriptors, and others explained why the use of textwords was usually beneficial to their searching. It is clear then that some searchers have a strong preference for one type of search key. Textwords ratio, which represents the percentage of textwords entered on the average by a searcher per search, is the variable to measure this tendency among searchers.

Textwords Ratio

Analysis of Variance indicates that the variation among searchers on this variable is significant ($F(46, 280) = 5.16$, $p < .01$). This finding reinforces searchers' perception that a preference for textwords or descriptors is a matter of searching style. Further, textwords ratio is associated with four variables: operational moves ratio, thesaurus neglect ratio, subject area, and for science searchers, environment.

The finding reported in the previous section revealed that textwords ratio is significantly associated with operational moves ratio and points to the conclusion that operationalist searchers prefer to use textwords while conceptualist searchers prefer to use descriptors. Similarly, data show that the variables thesaurus neglect ratio and textwords ratio are directly related ($r(45) = .660$, $p < .01$). This association is trivial, however, because it is obvious that searchers who prefer to use descriptors are more likely to consult a thesaurus (in which they find the descriptors) than searchers who prefer to enter textwords, and because neglecting to consult a thesaurus most frequently leads to entering textwords.

More pertinent are the effects of subject area and environment. Analysis of Variance shows that subject area as a variable correlates with textwords ratio ($F(3, 43) = 13.16$, $p < .01$). On the average:

Subject area	Textwords ratio (%)
Medicine	34
Social sciences and the humanities	39
General	57
Science and technology	76

A post-hoc test shows that the difference lies between science searchers on the one hand, and medicine and social sciences searchers on the other. Textwords ratio for general literature did not differ significantly either from social sciences and humanities or from science. That is: ***Science searchers are more likely to use textwords than their colleagues who specialize in other subject areas.***

At first glance, this finding seems to support common belief. It has been long assumed that searches in the scientific literature do not require the use of controlled vocabulary because the scientific terminology itself is already controlled. This argument, however, is not a valid explanation for this finding because of the difference between science and medical searchers. Medical terminology *is* scientific terminology, and there is no evidence to assume that it is less controlled than other scientific terminologies, no matter how one defines or measures the degree of terminological control. Yet medical searchers used the smallest proportion of textwords while science searchers used the largest proportion. That is, while the fact remains that science searchers use more textwords than other searchers, the degree of control in the science terminology does not explain this phenomenon.

Further, data indicate that the tendency among science searchers to prefer textwords may be *caused* in part by their need to search a relatively large number of databases for each request. Study results reported in a previous paper (Fidel, 1991b) show that having to use a number of databases for a request encourages searchers to use textwords. In addition, these data disclose that searchers refrained from using descriptors when they perceived that the thesaurus or the indexing of a database were of poor quality. Thus, there is enough evidence to indicate that the discrepancy in textwords ratio between science and medical searchers is not inherent to the subject area but due instead to the databases that are available in each subject area and to the quality of their thesauri.

The nature of the environment, across all subject areas, has no significant effect on textwords ratio ($F(2, 44) = .69$, NS). However, Analysis of Variance shows that for those who search the scientific literature, the searcher's environment has a significant effect on this variable ($F(1, 21) = 7.43$, $p < .05$). Science searchers who typically answer requests that address practical problems used textwords 86.84% of the time; those who typically search for theoretical requests used textwords 67.28% of the time. ***Science searchers who typically answer practical questions are more likely to use textwords than science searchers who usually address theoretical problems.***

The finding that environment in general does not affect the textwords ratio, but has an effect within science searching, may imply that the subject area has a larger effect on the selection of search keys than whether the requests searched are of a practical or theoretical nature.

However, it is plausible to speculate that within each subject area, practical questions encourage the use of textwords because they are more likely to include concrete and well-defined terms that are adequate for free-text searching than are theoretical requests. The failure of this study to find such an association for subject areas other than the sciences may be due to deficient sampling: the samples of searchers within other subject areas were small and therefore possibly not representative enough.

Three variables do not correlate with textwords ratio: number of moves, number of search keys, and recall tendency. Previous findings show that interactive searchers, those who make a relatively large number of moves and use a relatively high number of search keys, are not likely to use more textwords than their peers who are less interactive.

Similarly, recall tendency does not significantly relate to textwords ratio ($r(45) = .104$, NS). That is: ***Searchers who are usually more concerned with improving recall than their peers, are likely to prefer textwords to the same degree as do their peers.***

This conclusion indicates although some researchers and practitioners believe that the use of textwords increases recall, in real-life searching this is not actually done.

Thesaurus Neglect Ratio

Thesaurus neglect ratio measures the percent of textwords entered *without* consulting a thesaurus. Because it is clear that searchers who prefer textwords are more likely to neglect to use a thesaurus than searchers who prefer descriptors, this variable is relevant to the description of the "free-text" searcher. Similar to textwords ratio, Analysis of Variance indicates that a tendency to avoid thesaurus consultation is a personal trait contributing to a person's searching style: the difference between searchers who used a thesaurus frequently and those who regularly neglected to consult it is significant ($F(46, 280) = 9.10, p < .01$).

Findings in previous sections show that thesaurus neglect ratio and operational moves ratio are directly related, which led to the conclusion that operationalist searchers are more likely to neglect consulting a thesaurus than conceptualist searchers. In addition, the subject area being searched has a significant effect on the frequency with which a thesaurus is avoided ($F(3, 43) = 3.51, p < .05$). The average frequencies for entering search keys without consulting a thesaurus for each subject area are revealing:

Subject area	Thesaurus neglect ratio (%)
Medicine	0
Social sciences and the humanities	13
General	29
Science and technology	32

Therefore: ***Science searchers are more likely to enter textwords without consulting a thesaurus than searchers who specialize in other subject areas.***

This conclusion concurs with a previous finding: Science searchers are more likely to use textwords than their colleagues. This finding also shows the effect that a database with broad coverage and with a thesaurus of high quality can have on searching behavior. Being accustomed to searching MEDLINE with its MeSH vocabulary, the medical librarians who participated in the study never entered a search key without checking a thesaurus first.

Further, generalist searchers (i.e., those who habitually search several subject areas) entered a significantly larger number of search keys without consulting a thesaurus than did their peers in the social sciences and medicine. This phenomenon can be explained by the fact that generalists search a relatively large number of distinct databases throughout their searching practice. It is plausible to assume, then, that generalists search the largest number of distinct databases even though this factor was not measured in this study. They cannot familiarize themselves with the thesauri of the many databases they search and they are, therefore, more likely to refrain from using a thesaurus.

The Concern with Recall

The frequency with which searchers selected moves (Table 2) provides a first indication that searchers put more effort into improving recall than in improving precision. These data show that the number of moves to increase recall was almost double the number of moves to increase precision.

Additional evidence that searchers generally put much effort into improving recall is provided by other data collected in this study that pertain to the selection of search keys. A detailed presentation of these data is available in the final report (Fidel, 1990), but a short summary of the pertinent findings is provided here.

An examination of the search keys selected by searchers showed that 70% of the time searchers selected the most straightforward key: When a term was mapped to descriptors through exact match they entered the descriptors, but they used textwords when a term was not matched exactly to a descriptor, or when searchers did not know if the term matched. Setting these common options in search-key selection aside, 17% of the options resulted in improved recall. The rest of the options were selected for a diversity of reasons. ***Over half of the times searchers selected an option that was not straightforward, they did so to enhance recall.*** Simply put, if the search was not straightforward, quite frequently searchers encountered, or expected to encounter, low recall.

Additional evidence of the concern with recall is provided by the reasons searchers gave to explain their selection of search keys. These reasons were divided into three categories: Reasons that relate to request characteristics, those that relate to database characteristics, and reasons that reflect general beliefs held by a searcher. Examination of the frequency with which reasons in each category were given showed that: ***Among the request-related reasons, the need to enhance recall was the most frequent reason (35%) for the selection of a certain type of search key.***

This finding is an additional indication that unsatisfactory recall is a problem searchers face frequently.

Recall Tendency

Recall tendency, i.e., the percentage of recall moves made by a searcher across all searches, can be used to test whether the concern with recall is a characteristic typical of a particular searching style. Data presented in the previous sections illustrate that conceptualist searchers are more concerned with recall than are operationalist searchers, but there is no difference in recall tendency between searchers who are more interactive

and their less interactive peers, as there is between searchers who prefer textwords and those who prefer descriptors.

Further, various findings indicate that style of searching does not affect the concern with recall. First, Analysis of Variance shows that the average percentage of recall moves per search does not vary significantly among searchers ($F(46, 280) = 1.39$, NS). Second, the frequency of move selection shows that operationalist searchers were concerned with recall as well: The most commonly made move to increase recall is an operational move, the move to change a database (Add 5). It is clear then that while conceptualist searchers are usually the most concerned with recall, all searchers encounter problems with achieving satisfactory recall.

Subject area and environment affect recall tendency only in combination. Analysis of Variance found that recall tendency is significantly affected neither by subject area ($F(3, 43) = .52$, NS), nor by environment across subject areas ($F(2, 44) = 2.83$, NS). The same analysis for environments within each subject, however, revealed that within the sciences, environment significantly affects recall tendency ($F(1, 21) = 7.29, p < .05$). Science searchers in theoretical environments made recall moves 74.51% of the time, while those in practical environments made such moves 54.94% of the time. ***Science searchers who work in theoretical environments are more likely to be concerned with recall than their colleagues in practical environments.***

This finding is not surprising, because it is commonly believed that theoretical requests usually require higher recall than practical ones. While this association between environment and the tendency to employ recall moves was found only in science searching, it is plausible to assume that it may hold for other subject areas as well. As explained earlier, deficiencies in the composition of the study sample prevents the validation of this assumption.

In summary, while most of the study's searchers extended more effort to improve recall than to improve precision, conceptualist searchers and those who usually answer scientific requests that are theoretical are more likely to be concerned with recall than their colleagues.

Conclusions

The findings reported here illuminate elements of searching behavior that constitute individual searching style, and those that are common to all searchers. While a pronounced concern for recall was common to all searchers, three characteristics of individual searching styles were uncovered:

- the level of interaction during a search, which is measured by the typical number of moves per search and by the average number of search keys per search;
- the preference for type of move, operational or conceptual, which is measured by the relative frequency with which a searcher selects each type of move;
- the preference for type of search key, textwords or descriptors, which is measured by the relative frequency with which a searcher selects each type of search key.

The Interactive Searcher

Data analysis shows that regardless of their preference for textwords or descriptors, the subject matter, the environment in which they search, and whether they are operationalist or conceptualist searchers, some searchers routinely put more effort into their searches than others: Interactive searchers make more moves and enter more search keys than their peers who are less interactive.

This finding is highly relevant to studies of online searching behavior. Search interaction as a variable has played an important role in such studies, particularly in experiments. Such experiments have attempted to determine the effect of various searcher and searching characteristics on searching behavior. Various characteristics have been selected, such as level of experience (Fenichel, 1981; Howard, 1982), cognitive attributes (Brindle, 1981; Woelfle, 1984), personality traits (Bellardo, 1985), and the type of request (Saracevic & Kantor, 1988b). A group of variables, usually called search-effort variables, has been commonly used to assess the search process. First introduced by Fenichel (1981), this group includes variables such as number of commands, number of search keys, number of cycles (a sequence of commands from those used to enter search keys to those used to view results), number of sets viewed, number of search modifications, and connect time. These variables clearly relate to the level of interaction as defined here: An interactive searcher would score higher on search-effort variables than would a less interactive searcher.

Further, although rarely stated explicitly, it seems that the search-effort variables carry a value judgment: the higher a searcher scores, the "better" she is. For example, Fenichel measured these variables, among others, to test the hypothesis that experienced searchers would score differently than novices. She could not substantiate the hypothesis but found instead "unexpectedly large individual differences in search behavior..." (Fenichel, 1981).

The present study leads to two major conclusions which partly explain experimental results that were unexpected:

- High level of interaction is not always desirable, and value judgments should not be attached to this characteristic;
- Effort variables, as commonly defined, are not adequate to represent the search process.

The level of interaction is not a matter of quality. While at times a search may require a high level of interaction to achieve satisfactory results, high interaction

is not always due to quality requirements. For example, a high level of interaction might be caused by insufficient preparation. Another factor that may affect interaction was revealed by the study's searchers who indicated that they preferred to start a search with textwords and then look for descriptors (Fidel, 1991b). Such searchers are likely to be more interactive than others and this difference in the level of interaction does not necessarily reflect a difference in the "quality" of their searches. The level of interaction is, therefore, an attribute of searching style but there is no evidence to show that interactive searchers are "better" than their colleagues who are less interactive.

At present we do not know what makes one searcher more interactive than another, and hopefully future research will address this issue. This study's results, however, suggest that "being interactive" is an *inherent* characteristic of a person that is unlikely to be changed by experience, training, subject area, environment, or by similar variables that are of interest to researchers. It is inadvisable, therefore, to use variables that represent level of interaction to measure changes in searching behavior. For example, if future research reveals that the level of interaction is determined by a certain cognitive characteristic of a searcher, one should not expect experienced searchers to consistently demonstrate a level of interaction that is significantly different from that of novices (unless one is willing to assume that experience in online searching changes that cognitive characteristic).

This shows that the inability of experiments in online searching behavior to provide conclusive results with respect to the search process is partly caused by the poor choice of variables to represent the search process.

Operationalist and Conceptualist Searchers

Another aspect of searching style is preference for a certain type of move, operational or conceptual. Operational moves ratio measured the degree to which a searcher was operationalist, as determined by the moves he made. The results show that operationalist searchers:

- use textwords more frequently;
- are more likely to avoid consulting a thesaurus; and
- are likely to make fewer recall moves than conceptualist searchers.

Although only 25 of the 47 searchers in the study exhibited a strong commitment to one type of move, operationalist searchers differed from their conceptualist peers in their preference for type of search key, their habits relating to thesaurus neglect, and their concern for recall. These findings agree with the detailed description of searching behavior typical of operationalist and conceptualist searchers which was published earlier (Fidel, 1984).

The "Free-Text" Searcher

Textwords ratio as a variable measured the degree to which a searcher preferred to use textwords. Results reported here show that a profile of the searchers who use textwords more often than other searchers can now be constructed. Such searchers are likely to have these characteristics:

- they will be operationalist searchers;
- if, as science searchers, they usually answer practical requests, they will use still more textwords than science searchers who answer theoretical requests;
- they will have developed a habit of entering search keys without consulting a thesaurus.

In addition, searchers who prefer to enter textwords *do not* enter more search keys than those who prefer descriptors, nor are they more interactive than their counterparts.

The nature of the "free-text" searcher as described here raises the question: Is the preference of textwords an inherent attribute determined by factors such as cognitive style or personality traits? Answering this question is significant for research in online searching behavior.

The results of this study cannot provide a definite answer to this question, but they do offer some suggestions. On the one hand, the results show that inherent attributes have some effect on habitual preference in the selection of search keys: it was found that operationalist searchers prefer to use textwords. On the other hand, the results also show that the tendency to prefer textwords is encouraged by the realities of searching: by subject area, environment, and the availability and quality of thesauri.

The conclusion that preference in search-key selection is in part determined by factors external to a searcher's personal traits is supported by another finding. Only 20% of the reasons for selecting a search key stemmed from habitual searching behavior (Fidel, 1991b). That is, the selection of search keys is usually determined by the specific requirements and constraints of a search; the effect of inherent searching behavior on this selection is less extensive.

But preference in the selection of search keys is characteristic of a person's searching style. It is plausible to assume, therefore, that searching conditions help searchers to crystallize their searching styles. When measuring searching performance, studies of online searching behavior should consider the effect of variables such as the subject specialty and environment of a searcher or the number and quality of databases the searcher habitually searches.

The Concern with Recall

Recall, which measures the completeness of relevant information retrieved, is of special concern in information science research. Concern over recall stems from

the findings of experiments in online retrieval: Most experiments have resulted in relatively low recall scores. For example, in a study of information seeking and retrieving by Saracevic and Kantor (1988a), precision was 57% for all searches but recall was only 22%. As the authors explain, these ratios agree with the results of other studies.

The searchers who participated in the present study frequently attempted to increase recall:

- The number of moves to increase recall was almost double the number of moves to increase precision;
- Among the request-related reasons for the selection of search keys, the need to enhance recall was the most frequent reason given; and
- Among the options in search-key selection that were not straightforward, over half were selected to enhance recall.

The low recall scores obtained in experiments often have been explained by the assumption that searchers in general do not consider recall to be an important factor, or that they prefer to avoid the extra effort that is presumably required to increase recall. In contrast, the findings of this study show that searchers *do* consider recall to be an important factor when they select search keys, and when they modify search strategies.

The discrepancy between the findings of this study and the low recall scores obtained in online searching experiments can be attributed partially to the methods used. In contrast to experiments where searching is performed under artificial conditions, this study examined searchers answering real-life requests submitted by users to whom the searchers are accountable. It is possible, therefore, that the searching observed in this study was guided by a level of recall-consciousness that is higher than the one exhibited in experiments carried out to study online searching behavior. It is difficult, however, to substantiate this explanation because this study measured *concern with recall,* while the others measured *recall performance.* Thus the results of this study cannot be compared with those of experiments, but this observation should be considered in future online searching experiments.

Another possibility is that current bibliographic databases simply do not allow for high recall, regardless of a searcher's experience or searching style.

In any case, the issue of recall in online retrieval needs further study. Designers of both databases and intermediary expert systems should pay special attention to means to improve recall and provide tools that support searchers' attempts to enhance recall. As a beginning, the moves that searchers employ to increase recall can guide system designers in upgrading existing systems. Based on the findings of this study, if the following features were added to search systems and databases, they would help searchers to improve recall:

- an expanded list of synonyms for each term (to support Add 1),
- a clear and easy-to-understand separation between commands for entering textwords and those for entering descriptors (to support Add 2),
- a list of terms which occur in records selected by a user, ranked by the frequency of occurrence (to support Add 3),
- easy access to a database's index which designates the number of postings for each term, descriptors as well as textwords (to support Add 4),
- an automatic execution of inclusive searches— "exploding," or "cascading"— (to support Include),
- an easy mechanism to obtain a clear display of the search history (to support Cancel),
- an easy mechanism to display the hierarchical relationships of a descriptor at any point in the search (to support Expand 1), and
- an index language that includes role indicators (to support Expand 2 and Expand 3).

Further, since the need to improve recall has been expressed consistently by users and researchers, it is time to begin further explorations to discover new ways to improve the recall of retrieved sets.

In summary, this study demonstrates that searching is not such an "imprecise art" as it may seem (Saracevic & Kantor, 1988b); it does exhibit lawful behavior. The study uncovered reasons for the selection of search keys, a typology of search modifications guided by the reasons for these modifications, and a few individual characteristics of searchers that are part of a person's searching style. Moreover, the study points to the manner in which searching style affects searching behavior, and suggests that some characteristics are inherent and others are acquired through professional experience. It is the task of future research to study the factors involved in developing a particular searching style.

Acknowledgments

The study team on this project consisted of Nancy Phelps, Michael Crandall, Cynthia Altick Cunningham, and Kathleen McCrory. Their contribution to the study was substantial and highly appreciated. The online searchers who participated in this study were extremely cooperative spirits. Their collaboration made the study possible. My special thanks to Dagobert Soergel and Elaine Svenonius who reviewed earlier versions of this report. Their insightful comments and suggestions provided an inestimable contribution.

References

Bellardo, T. (1985). An investigation of online searcher traits and their relationship to search outcomes. *Journal of the American Society for Information Science, 36,* 241–250.

Brindle, E. A. (1981). *The relationship between characteristics of searchers and their behaviors while using an online interactive retrieval system.* Ph.D. dissertation, Syracuse University.

Fenichel, C. H. (1981). Online searching: measures that discriminate among users with different types of experiences. *Journal of the American Society for Information Science, 32,* 23–32.

Fidel, R. (1984). Online searching styles: a case-study-based model of searching behavior. *Journal of the American Society for Information Science, 35,* 211–221.

Fidel, R. (1985). Moves in online searching. *Online Review, 9,* 61–74.

Fidel, R. (1987). What is missing in research about online searching behavior? *The Canadian Journal of Information Science, 12,* 54–61.

Fidel, R. (1990). *Extracting knowledge for intermediary expert systems: the selection of search keys.* Syracuse, NY: ERIC.

Fidel, R. (1991a). Searchers' selection of search keys: I. The Selection Routine. *Journal of the American Society for Information Science, 7,* 490–500.

Fidel, R. (1991b). Searchers' selection of search keys: II. Controlled vocabulary or free-text searching. *Journal of the American Society for Information Science, 7,* 501–514.

Howard, H. (1982). Measures that discriminate among online searchers with different training and experience. *Online, 6,* 315–327.

Saracevic, T., & Kantor, P. (1988a). A study of information seeking and retrieving. II. Users, questions, and effectiveness. *Journal of the American Society for Information Science, 39,* 177–196.

Saracevic, T., & Kantor, P. (1988b). A study of information seeking and retrieving. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science, 39,* 197–216.

Woelfl, N. N. (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.