# The Effects of Domain Knowledge on Search Tactic Formulation

**Barbara M. Wildemuth**
*School of Information and Library Science, 100 Manning Hall, CB#3360, University of North Carolina, Chapel Hill, NC 27599-3360. E-mail: wildem@ils.unc.edu*

**A search tactic is a set of search moves that are temporally and semantically related. The current study examined the tactics of medical students searching a factual database in microbiology. The students answered problems and searched the database on three occasions over a 9-month period. Their search moves were analyzed in terms of the changes in search terms used from one cycle to the next, using two different analysis methods.**

**Common patterns were found in the students' search tactics; the most common approach was the specification of a concept, followed by the addition of one or more concepts, gradually narrowing the retrieved set before it was displayed. It was also found that the search tactics changed over time as the students' domain knowledge changed. These results have important implications for designers in developing systems that will support users' preferred ways of formulating searches. In addition, the research methods used (the coding scheme and the two data analysis methods—zero-order state transition matrices and maximal repeating patterns [MRP] analysis) are discussed in terms of their validity in future studies of search tactics.**

## Introduction

When conducting an online search, a searcher enters one or more terms, gets a response from the system, and iteratively modifies the terms until satisfied with the results. Each iteration in the search formulation and reformulation process can be considered a search move. One way to examine these moves is to focus on the search terms selected in each iteration, as suggested by Shute and Smith (1993). For example, one move might consist of the substitution of a narrower term representing the same concept as that searched in the prior move (e.g., the first move might be a search on the term "reference works" and the second move might be a search on the term "encyclopedias").

Received June 4, 2003; revised July 23, 2003; accepted July 23, 2003



A set of moves that are temporally and semantically related can be a search tactic. While significant work has examined the individual moves that searchers make (Bates, 1979, 1987; Fidel, 1985), it is equally important to examine the *sequences* of moves made by searchers in order to understand the cognitive processes they use in formulating and reformulating their searches. The current paper attacks this problem by examining the tactics of medical students searching a factual database in microbiology, focusing on changes in students' tactics as their domain knowledge (i.e., their knowledge in the domain of microbiology) changed over time. In addition, it makes a methodological contribution by comparing two different methods for analyzing sequences of search moves.

## Background

A searcher's domain knowledge is his or her knowledge of the subject area (i.e., domain) that is the focus or topic of the search. The searcher's domain knowledge may affect the process of search strategy formulation and reformulation, as well as retrieval success and the outcomes of the search. Studies of search tactics and the effects of domain knowledge on searching behaviors and outcomes are briefly reviewed here.

The sequencing of search moves into search tactics can be viewed as one component of search quality (Debowski, 2001), and a few studies have examined search tactics by examining the moves of which they are composed. Tolle (1983) examined transitions from one command to the next in the transaction logs of a first-generation online library catalog. One of the most striking results from this research was that an error was most likely to be followed by another error. Qiu (1993) used similar methods to study the fit of Markov models to search tactics in a hypertext system, and found that a second-order model had the best fit to the data. Kiestra, Stokmans, and Kamphuis (1994) stipulated that the minimum "meaningful unit" of search behavior consists of three successive actions, and so analyzed these fragments of the online catalog searches of 15 university students. They discovered 553 different patterns used in the searches, but only 65 were used by more than one study participant. In another study of online library catalog searches, Berger's (1994) findings suggest that users are not as unsuccessful as indicated in previous studies, and that the *series* of commands used in completing a single search provides "the best clues to the users' intentions and searching performance" (p. 1). In a detailed study of medical students' searching, Wildemuth and her colleagues (1992) found that just a few tactics (i.e., sequential combinations of moves) were used most frequently. Vakkari, Pennanen, and Serola's (2003) study of psychology students searching PsychINFO used Bates's (1979, 1987) original specification of tactics, and found that the students used the intersect (adding a term), limit (using field restrictions) and vary (replacing one term with another) tactics most frequently. Jansen, Spink, and Saracevic (2000) examined over 50,000 Web search engine queries and found that only 22% of the queries were modifications of a previous query; using a coarse-grained coding of the search moves to analyze 191 of the searches, they found that the most common user session was "a unique query followed by a request to view the next page of results" (Spink, Jansen, & Ozmultu, 2000, p. 321). Each of these studies examined the sequences of moves made by searchers during the course of a search, but it is difficult to draw general conclusions from their results. Though similar analyses were conducted in each study, each used a different set of search move definitions. In addition, none investigated the effects of domain knowledge on the formulation of those search tactics.

Domain knowledge is the searcher's knowledge of the search subject or topic, and is conceptually distinct from knowledge of searching techniques. In an early study of library catalog use, Bates (1977) distinguished subject familiarity (i.e., "knowledge of a specified academic field") from catalog familiarity (i.e., "knowledge of the structure of the system of Library of Congress [LC] subject headings [subject terms]") (p. 162). Allen (1991) focused on topic knowledge, defined as "factual knowledge" of the search topic (p. 188), while Dimitroff (1992) focused on system knowledge, operationalized as the user's mental model of the retrieval system, made up of eight components (e.g., the contents of the database, knowledge of multiple fields within a record, and Boolean search capability), and Hoelscher and Strube (1999) focused on Web expertise, defined "as a type of media competence" (p. 305). The focus of the current study is on the searcher's knowledge of the domain (i.e., each student's knowledge of microbiology) and how that domain knowledge affects his or her search behaviors and outcomes.

A few prior studies have investigated the relationship between the searcher's domain knowledge and various aspects of the *process of searching*. In a study of the search tactics of graduate students from two social science domains, Hsieh-Yee (1993) found that domain knowledge affected the amount of off-line preparation for the search, the amount of time spent monitoring their searches, and the frequency with which terms were combined. Kiestra, Stokmans, and Kamphuis (1994) used a similar research design to examine the effects of both domain knowledge and search knowledge on the amount of time taken and the number of moves included in online catalog searches. Their results, with regard to the effects of domain knowledge, were inconclusive. In four studies of searching in full-text databases, Marchionini, Dwiggins, Katz, and Lin (1993) compared domain experts with intermediaries (i.e., searching experts). Domain experts tended to focus on the answers to the search questions and had definite expectations for the answer and the context in which it would be found, while the intermediaries focused on the problem statement and query formulation, imagining that they were gathering documents for an end user. Carmel, Crawford, and Chen (1992) found that, while domain experts' and novices' general browsing patterns in a hypertext database did not differ, novices used referential links, abandoned topics, selected "unknown" topics, and examined topics of personal interest more often than experts, and they examined topics related to expert knowledge less often than experts. McDonald and Stevenson (1998) also examined the effects of domain knowledge on hypertext navigation. While knowledgeable subjects, overall, took less time to answer questions and navigated to fewer irrelevant nodes, these effects did not occur for the groups offered spatial maps as navigation aids. Bhavnani (2002) compared novice and expert search behaviors on the Web. Five health care searching experts and five shopping experts conducted searches in their own areas of expertise and in the other domain. The results revealed "the existence of domain-specific search knowledge" (p. 611). In another study, Bhavnani and Bates (2002) used hierarchical goal decomposition to better understand the knowledge required to complete particular search tasks. Because each of these studies included only a small sample, many of the findings could as appropriately be attributed to individual differences as to level of domain knowledge. Thus, the evidence concerning the effects of domain knowledge on searching behaviors is inconclusive.

Vakkari (2002) argued that domain knowledge will affect people's *ability to choose appropriate search terms*, and followed up this argument with an empirical study (Vakkari, Pennanen, & Serola, 2003). In examining psychology students' searches at two points in their development of research proposals, they found that the "clearest change in students' searching was the use of a wider and more specific vocabulary" (p. 459) as students learned more about their research topics. Allen's earlier (1991) findings were parallel: Undergraduate students with more domain knowledge used more search expressions while searching a topic in space exploration in an online library catalog. Most of this effect was attributed to those high-knowledge students who, when they experienced difficulty with the search, introduced new terms into their strategies. Hsieh-Yee (1993) found that library science students used more of their own terms in a search on a library science topic and used more thesaurus terms and more synonyms in the search on a topic in which they had little domain knowledge. Similarly, Shute and Smith (1993) found that personal knowledge in chemistry had an effect on term selection for *Chemical Abstracts* searches. In a study of hypertext searching, sixth-graders were more adept than fourth-graders in their selection of terms and their manipulation of those terms in retrieving relevant encyclopedia articles (Marchionini, 1989). Results from these studies consistently support the proposition that the searcher's level of knowledge in the domain of the search will affect the type and number of terms selected for incorporation in the search strategy.

Several studies have examined the effects of differences in domain knowledge on *retrieval success*. Usually operationalized as search recall, these effects have not been conclusively demonstrated. While Jacobson and Fusani (1992), Maidenberg (1991), and Marchionini (1989) did find that domain knowledge had a positive effect on retrieval success, a variety of studies in a variety of domains (including Allen, 1991; Borgman, Hirsh, Walter, & Gallagher, 1995; Dimitroff & Wolfram, 1995; Saracevic & Kantor, 1988; Wildemuth, de Bliek, Friedman & File, 1995) found no such relationship. Based on these studies, the effects of a searcher's domain knowledge on search results are still in question.

Just two studies have looked beyond retrieval success to the *outcomes of searching*, i.e., the searcher's *use* of the information retrieved. Marchionini, Lin, and Dwiggins (1990) compared the performance of four subject specialists, four search specialists, and 15 library science students in answering factual questions, assisted by searches in a full-text hypertext database. Both types of experts were more successful than the students. In another study, de Bliek and her colleagues (1993, 1994) examined medical students' success in answering clinical problems, assisted by a relevant factual database. Across four assessment occasions (representing different levels of domain knowledge), useful retrieval of information always improved students' ability to solve the clinical problems. Taken together, the results of these studies indicate that database use helps the searcher answer questions or solve problems, whatever the searcher's level of domain knowledge.

In summary, a few studies have examined searchers' tactics. No firm conclusions can be drawn from these studies, because each used a different set of "move" definitions as the basis of the analysis. Additional studies will need to work with a definition of a search move that can be generalized across information retrieval systems, and can validly represent the process of users' search strategy formulation and reformulation. Additional studies have explored the relationship between domain knowledge and searching behaviors, but only a few general conclusions can be drawn. Conflicting results have been reported on the effects of differences in domain knowledge on search tactics and retrieval success. There is, however, consistent support for the propositions that differences in domain knowledge affect searchers' selection of terms and that databases can effectively augment personal knowledge in problem-solving situations.

## Research Problem

The current study addresses two research questions and introduces a third (methodological) issue. First, it examines the tactics actually used by searchers in an attempt to understand how searchers formulate and reformulate their search strategies. Second, it focuses on differences in these tactics that might be attributed to differences in the searcher's domain knowledge. Finally, it evaluates a coding scheme for search moves and compares two different methods for analyzing sequences of search moves in terms of these methods' practicality and validity.

## Method

To address these questions, two cohorts of medical students were asked to solve clinical problems, assisted by searches of a factual database. The students answered problems and searched the database on three occasions over a nine-month period. Their search moves were analyzed in terms of the changes in search terms used from one search cycle to the next. In addition, the longitudinal nature of this study allowed us to examine how changes in the students' knowledge of microbiology affected their search tactics. The search tactics were analyzed in two ways: by examining the individual moves (i.e., zero-order state transitions) and then by examining the maximal repeating patterns that appeared across the searches conducted at each occasion.

### *Study Participants*

A sample of participants for this study was selected from two consecutive entering classes of the University of North Carolina at Chapel Hill School of Medicine. Those who had an advanced degree in science or an undergraduate degree in microbiology were eliminated from the subject pool. A random sample of the remaining students was invited to participate; 34 students from the first cohort and 43 students from the second cohort participated in all three of the assessments analyzed in this study.

### *Research Design and Procedures*

On each of three assessment occasions, the students were asked to answer a set of six clinical problems in microbiology, each problem involving several specific questions. The first occasion occurred just before the students received any instruction in microbiology; the second occasion occurred just after the microbiology course; and the third occasion occurred six months after the end of the course. Each assessment session was divided into two passes through the problem set. On the first pass, the students responded to the problems, based solely on their personal domain knowledge. On the second pass, the students re-

TABLE 1. Definitions of moves (adapted from Shute and Smith, 1993).

| | |
|---|---|
| **Beginning moves** | |
| New concept | Enter term(s) for a concept that was not included in previous cycle. |
| **Moves to reduce the size of the set** | |
| Add concept | Add a concept that is not represented in the previous search cycle, using AND. |
| Combine with AND | Combine two pre-existing concepts, using AND. |
| Narrow term | Replace a term with a narrower term for the same concept. |
| Narrow operator | Replace an operator with a narrower operator. |
| Exclude | Exclude a concept or term, using NOT. |
| **Moves to increase the size of the set** | |
| Delete concept | Delete a concept (that was ANDed) from the previous search cycle. |
| Combine with OR | Add a term to a concept that is already represented in the previous search cycle, using OR. |
| Broaden term | Replace a term with a broader term for the same concept. |
| Broaden operator | Replace an operator with a broader operator. |
| **Move to increase both precision and recall** | |
| Replace term | Replace a term with a sibling/cousin term (i.e., a synonym or closely related term) for the same concept. |
| **Other moves** | |
| Error | Typographical, syntactic, and other types of errors. |
| Repeat | Repeat the same search terms in two consecutive moves. |

sponded to one question from each problem (a question that was *not* answered correctly on the first pass) with the aid of INQUIRER, a database of microbiology facts and concepts (Friedman, de Bliek, Gilmer, Twarog & File, 1992). Almost 1,300 searches were conducted by the 77 students over the three occasions. All the students' searches were captured in transaction logs.

The three assessment occasions in the research design represent different levels of domain knowledge among the participants. In addition to the face validity of this claim, results from an analysis of these students' performance on the first pass through the problems provide empirical support for such an argument. At the first occasion, the students answered only 12.6% of the problems correctly, based on their personal knowledge; at the second occasion, they answered 48.1% correctly; and at the third occasion, they answered 27.3% correctly.

The students' searching expertise represents a different form of knowledge that was likely changing over the three occasions. Expertise in searching databases is clearly distinguishable from domain knowledge (Bates, 1977; Hoelscher & Strube, 1999; Hsieh-Yee, 1993), as noted earlier, and its effects on search behaviors have been studied (Connell, 1995; Kim, 2001a, 2001b; Yuan, 1997). Prior to the first occasion, the students had only one hour of group training on searching this database; they might also have had prior experience with searching other online databases. Between the first and second occasions, most of the students used the database regularly (i.e., at least weekly) in preparation of course assignments. Between the second and third occasions, it is unlikely that any of the students used the database, since it was not expected for their academic or clinical work. The effect of searching expertise was not the focus of the current study, but the possibility that it affected the outcomes of the study will be addressed in the discussion of the study results.

### *Coding the Transaction Logs*

The transaction logs were coded using a coding scheme adapted from Shute and Smith (1993). This coding scheme focuses on the concepts represented in each search and the specific terms used to represent those concepts. The coding scheme, as adapted for this study, is shown in Table 1. While it is much easier and more efficient to code moves with the system's command names (usually such coding can be captured directly from the transaction logs), using a coding scheme based on the concepts and terms incorporated in the search has several distinct advantages. First, it allows us to directly examine the searcher's use of terminology and, thus, an important aspect of the searcher's thinking as the search strategy is formulated and reformulated. Second, this scheme is generalizable across a large number of searching systems. It could be applied to search strategies formulated to address very complex and well-indexed databases, such as MEDLINE or online library catalogs, as well as full-text (and relatively unstructured) databases such as the Web. Third, the granularity of this scheme has advantages for later statistical analysis of the codes. More detailed coding schemes are too fine-grained to make statistical analysis effective (Wildemuth & Moore, 1995), while a coarser scheme would not provide enough detail. In summary, the advantages provided by this coding scheme outweigh the disadvantages associated with the need to manually code each search move.

The application of this coding scheme to one brief search is illustrated in Table 2. In the first search cycle, the search began (necessarily) with a new concept and the two additional terms were coded as *Add concept* moves. The coding then progresses through the search, comparing each search cycle with the previous cycle, and coding any changes. When one term was deleted in the second cycle, a *Delete concept* code was added to the list. Finally, in the third cycle, a new concept was added. The *Display* moves were also coded. The coding was performed independently by

TABLE 2. Example coding of a search.

| Search field | | Search term | | Coding |
|---|---|---|---|---|
| Cycle 1 | | | | |
| SIGNS & SYMPTOMS | for | STIFF | AND | *New concept* |
| SIGNS & SYMPTOMS | for | HEADACHE | AND | *Add concept* |
| SIGNS & SYMPTOMS | for | BACKACHE | | *Add concept* |
| Result: 0 records | | | | |
| Cycle 2 | | | | |
| SIGNS & SYMPTOMS | for | STIFF | AND | |
| SIGNS & SYMPTOMS | for | HEADACHE | | *Delete concept* |
| Result: 1 record (displayed) | | | | *Display* |
| Cycle 3 | | | | |
| SIGNS & SYMPTOMS | for | STIFF | AND | |
| SIGNS & SYMPTOMS | for | HEADACHE | AND | |
| SIGNS & SYMPTOMS | for | MUSCLE | | *Add concept* |
| Result: 2 records (displayed) | | | | *Display* |

two research assistants. The author resolved the few discrepancies that occurred.

### Data Analysis

The searches for each of the three assessment occasions were analyzed separately, so that comparisons across occasions could be made. Two methods of analysis were used to examine the tactics embodied in the search moves: analysis of zero-order state transitions, and analysis of the maximal repeating patterns that occurred on each occasion. Both methods are inductive approaches to the identification of patterns in search strategy formulation (Buttenfield & Reitsma, 2002).

To create a zero-order state transition table, each transition between a move and the subsequent move was tallied. For example, for the search shown in Table 2, the state transition matrix would show a frequency of one for each of the following transitions: *Begin* to *New concept*, *New concept* to *Add concept*, *Add concept* to *Add concept*, *Add concept* to *Delete concept*, *Delete concept* to *Display*, *Display* to *Add concept*, *Add concept* to *Display*, and *Display* to *End*. The frequency of each transition and its proportion in each set of transitions (i.e., on each assessment occasion) were calculated. A graphical representation of the most frequent state transitions (those accounting for at least 1% of all the transitions in each data set) was then created by linking together these frequently occurring individual transitions.

The same data set (i.e., the search moves) was analyzed for maximal repeating patterns. The analysis of maximal repeating patterns (MRPs) is a technique developed by Siochi and Ehrich (1991), intended for use by system developers as a "relatively low-cost evaluation technique for helping the developer locate interface problems" (p. 310). It is an exploratory sequential data analysis technique (Cuomo, 1994; Sanderson & Fisher, 1994) and can appropriately be applied to the analysis of sequences of search moves (i.e., search tactics). A repeating pattern is a particular sequence of moves that repeats (i.e., it occurs in more than one place in the set of search moves analyzed). For example, if the possible moves are represented as A, B, and C, and the full set of moves is AABCBAAB, then the repeating patterns are AA, AB, and AAB. A *maximal* repeating pattern is the longest such sequence that is repeated within the set of moves being analyzed (e.g., AAB in this example). Software developed by Siochi and Ehrich systematically identifies all the maximal repeating patterns within a data set, and reports their positions within the data set.

For the current analysis, each of the three data sets (one corresponding to each assessment occasion) was analyzed with the MRP software. Any partially duplicated strings were then removed from the results, and the most frequently occurring *unique* MRPs were selected for further analysis. This selection process involved a tradeoff between the length of the MRP and the frequency with which it occurred; generally, longer MRPs (some up to 18 moves long) occurred less frequently (usually just two times) while the shorter MRPs (those with only three moves were the shortest considered in this analysis) occurred much more frequently (up to 305 times). The most frequently occurring MRPs of each length were included in the analysis; approximately one-third of the MRPs of each length were included. These frequently occurring MRPs were then grouped into "families" of similar tactics. In addition to reporting the frequency of each search tactic, each was represented in a diagram of a typical tactic in that family (usually a hybrid of the longest and the most frequently occurring tactics within the group).

### Conclusion

There are two sets of results that will address each of the research questions, i.e., there is a state transition network representing the search moves on each assessment occasion and there is a set of MRPs representing the search moves on each assessment occasion. By examining the similarities in these two sets of findings, we will be able to understand and compare the tactics used by medical student searchers at three levels of domain knowledge. By comparing the differences between these two sets of findings, we will be able to address the methodological issues of the practicality and validity of these two analysis methods.

## Results

The results from each occasion are reported separately and include both the state transition analysis and maximal repeating patterns analysis from each occasion. Differences across the three assessment occasions and differences in the results from each analysis method will be addressed in the Discussion section of this article.

### *Results From the First Assessment Occasion, Prior to the Microbiology Course*

The first assessment occasion occurred just before the microbiology course began. At that point, the students had received training on how to use the database, but their personal knowledge of microbiology was very low (as noted earlier, they completed only 12.6% of the problems correctly on the first pass, based solely on their personal knowledge). On this assessment occasion, 365 searches were conducted, made up of 3,772 moves (averaging over 10 moves per search). The most frequent zero-order transitions among those moves are represented in the state transition network shown in Figure 1.

All the searches began with the *New concept* move (an error was the only other possibility). The most common transitions during a search were *New concept* to *Add concept*, *Add concept* to *Add concept* (possibly multiple iterations during the same search), and *Add concept* to *Display*. This set of transitions accounted for just over half the transitions on the first assessment occasion.

The students also broadened their searches by deleting concepts. The *Delete concept* move occurred 623 times across all the searches. It most often occurred just after an *Add concept* move, and was most often followed by an *Add concept* move.

The *Repeat* move can be considered an error (the same search terms were resubmitted with no changes). It did occur somewhat often on this first assessment occasion, but not nearly as often as the moves involving term changes. There was also a very small number of other moves coded as errors on this first occasion (nine *Error* moves). Given this small number of errors overall, it can be argued that these students had little difficulty in using the microbiology database.

It should be noted that all the frequently used moves involved changes in the concepts incorporated in the search; none involved changes in the specific terms used to represent a concept. In other words, substitution or addition of synonyms to a search rarely occurred. This finding held true on all three occasions.

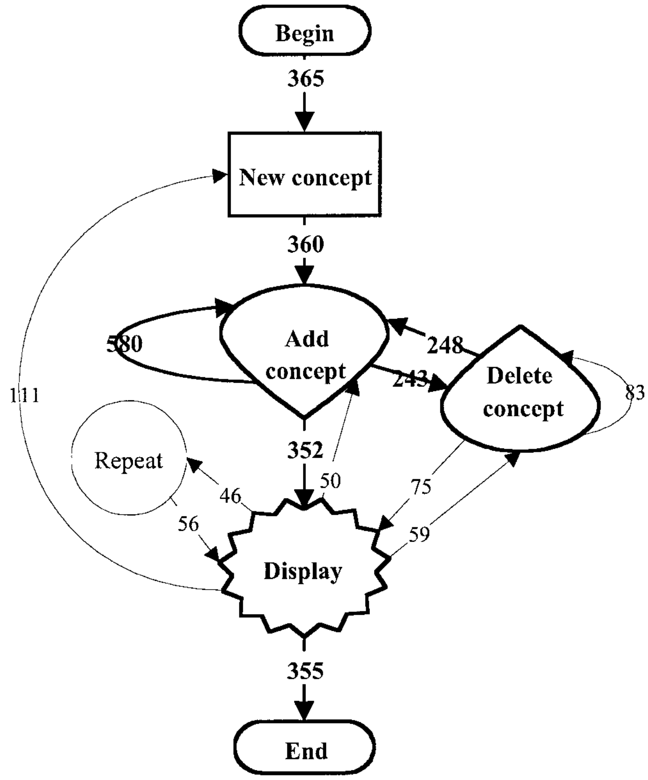


FIG. 1. State transition network of search tactics prior to the microbiology course. (n = 3,772 moves, in 365 searches) (includes all transitions > 1% of total; highlights transitions > 5% of total).

The results from the analysis of the maximal repeating patterns indicated that there were three distinct patterns, or families, of search tactics frequently used on this occasion (see Figure 2). Each of the patterns/families appeared in at least 1% of the searches conducted on the assessment occasion under consideration. The frequency of each of the members of each family is given at the bottom of the relevant column.

The first family of search tactics involved the initiation of a search with a *New concept* move, followed by one or more *Add concept* moves, followed by a *Display* (illustrated in the first column of Figure 2). The individual members of this family, i.e., the unique search tactics, are of varying length. Some encompass entire searches (defined here as all the search moves associated with solving a particular problem), while others are only fragments of a search. As would be expected, the tactics that are shorter (i.e., include fewer moves) occur more frequently. However, it is noteworthy that the longest member of this family (the search tactic shown in the first column of Figure 2; a complete search) occurred 43 times, accounting for almost 12% of all the searches conducted at this assessment occasion. All together, there were 358 searches or search fragments represented in this family of search tactics.

The second family of search tactics discovered among the maximal repeating patterns on the first occasion is characterized by the alternating *Add concept-Delete concept* moves (see the second column of Figure 2). The longest

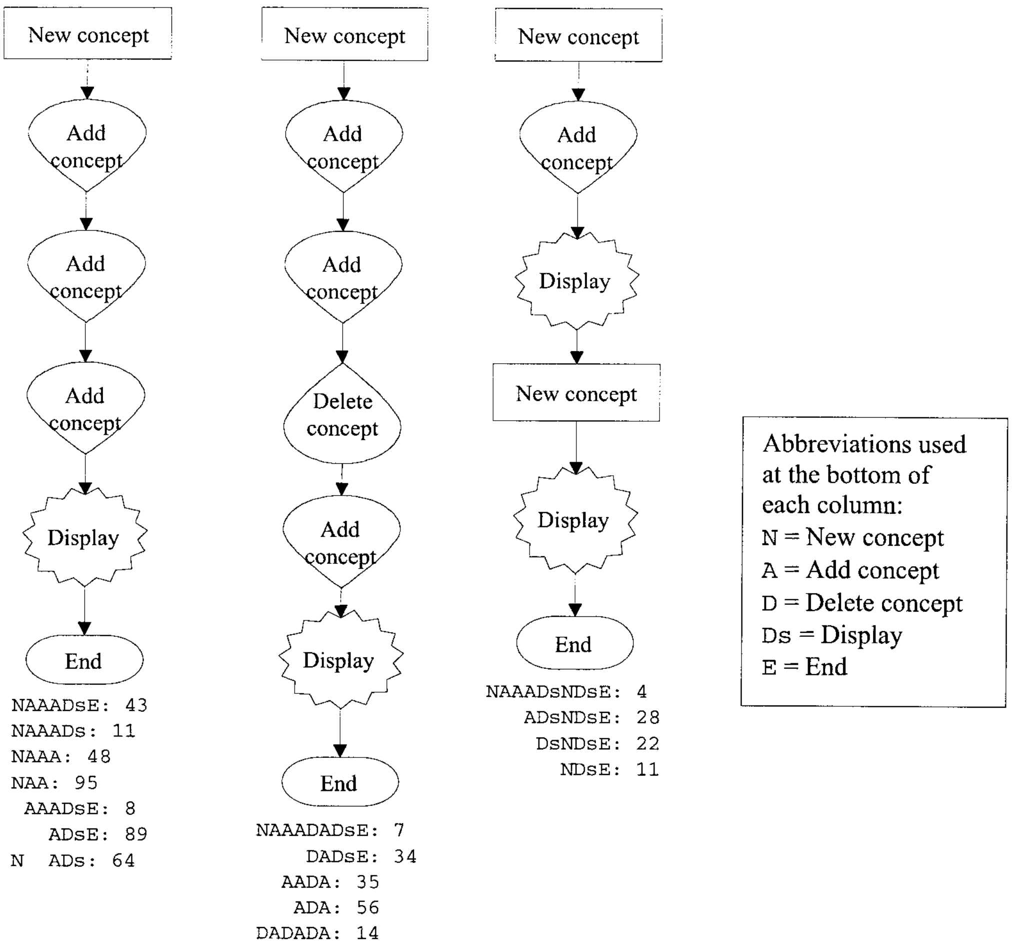


FIG. 2. Maximal repeating patterns prior to the microbiology course. (n = 3,772 moves, in 365 searches).

member of this family occurred only seven times, but several shorter tactics containing both *Add concept* and *Delete concept* moves occurred much more frequently. All together, there were 146 searches or search fragments represented in this family of search tactics.

The third family of tactics occurring on the first assessment occasion included a "restart" of the search (see the third column of Figure 2). In other words, the student began searching, displayed some results, but then began again with a completely new set of terms (i.e., the term contained in the *New concept* move had not appeared in the previous search cycle). This family of search tactics did not occur as frequently as the other two families. In addition, one of the search tactics included in this family was *New concept – Display – End.* Because this tactic did not include earlier searching, it may or may not belong with this family. It occurred 11 times. All together, there were 65 searches or search fragments represented in this family of search tactics.

*Results From the Second Assessment Occasion, Just After the Microbiology Course*

The second assessment occasion occurred just after the microbiology course, when students' knowledge of microbiology was at its highest (they were able to answer 48.1% of the problems correctly on the first pass), and they had gained some facility in using the database. The students conducted 444 searches, encompassing 3,678 moves (averaging about 8 moves per search). The dominant path identified in the state transition analysis of the first occasion became even more dominant at this second occasion (see Figure 3). The *Repeat* move was used so infrequently (only 78 times) that it does not show up on the diagram and, at this second occasion, other types of errors continued to be very infrequent (11 *Error* moves). Students still used the *Delete concept* move to broaden their searches, but this move was much less common.

The results from the analysis of the maximal repeating patterns on the second assessment occasion indicated that there were three distinct patterns, or families, of search tactics frequently used by these students. However, only two of these patterns were the same as those that occurred on the first occasion.

The first family of search tactics was similar to that found on the first occasion: *New concept*, followed by one or more *Add concept* moves, followed by a *Display* (see the first column of Figure 4). All together, this family included 431 searches or search fragments, a proportion similar to its representation on the first occasion.

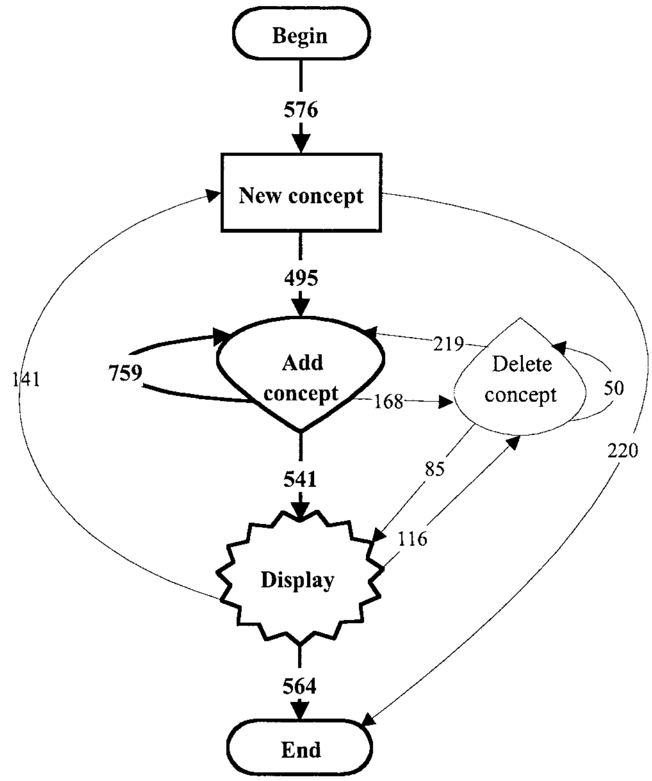


FIG. 3. State transition network of search tactics just after the course. (n = 3,678 transitions, in 444 searches) (includes all transitions > 1% of total; hightlights transitions > 5% of total).

The second family of tactics discovered on this occasion (see the second column of Figure 4) also resembled the second family discovered on the first occasion: it is characterized by the alternation between *Add concept* and *Delete concept* moves. All together, this family included 79 searches or search fragments; it occurred much less frequently at this assessment occasion than at the first occasion.

The third family of tactics discovered on this occasion (see the third column of Figure 4) is quite different from the "restart" pattern discovered at the first occasion. On this second occasion, just after the microbiology course, students often selected one search term (i.e., they made a *New concept* move), displayed the results, and ended the search (this single-move search occurred 81 times). Alternatively, the *New concept – Display* sequence was followed by another *New concept – Display* sequence, i.e., they replaced the original search term with a completely different search term representing a different concept. This family included 125 of the searches or search fragments occurring on the second assessment occasion. It does match the 11 searches at the first occasion (see the third column of Figure 2) that consisted solely of a *New concept* move and *Display*.

### *Results From the Third Assessment Occasion, Six Months After the End of the Microbiology Course*

The third assessment occasion occurred six months after the microbiology course ended. In the interim, the students had not received any additional instruction in microbiology, and had no need to use the database. Their level of personal knowledge had dropped to a midway point between their knowledge on the first occasion and their knowledge on the second occasion (they solved 27.3% of the problems correctly on the first pass). On this assessment occasion, the students conducted 489 searches, encompassing 3,238 moves (averaging about seven moves per search).

On this occasion, the state transition analysis indicated that the dominant path through a search had shifted somewhat (see Figure 5). While still more common than the *Delete concept* move, the *Add concept* move was not as prominent as it had been at the first two occasions. It appears that the decrease in the number of moves per search can be attributed to less iterations of the *Add concept* move. On this third occasion, the students most often moved from their first search move, *New concept,* directly to a *Display*. Even though the students had not used the database for several months, the number of errors (*Repeat* moves or *Error* moves) was very low (63 *Repeat* moves and 15 *Error* moves).

The results from the analysis of the maximal repeating patterns on the third assessment occasion indicated that there were just two distinct patterns, or families, of search tactics frequently used (see Figure 6). These two families

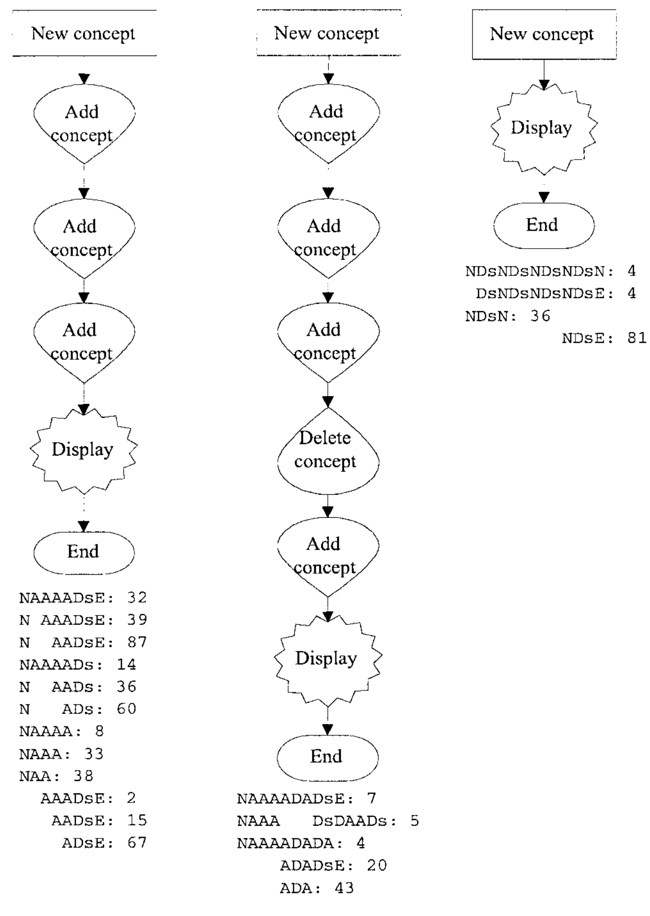


FIG. 4. Maximal repeating patterns just after the course. (n = 3,678 transitions, in 444 searches).

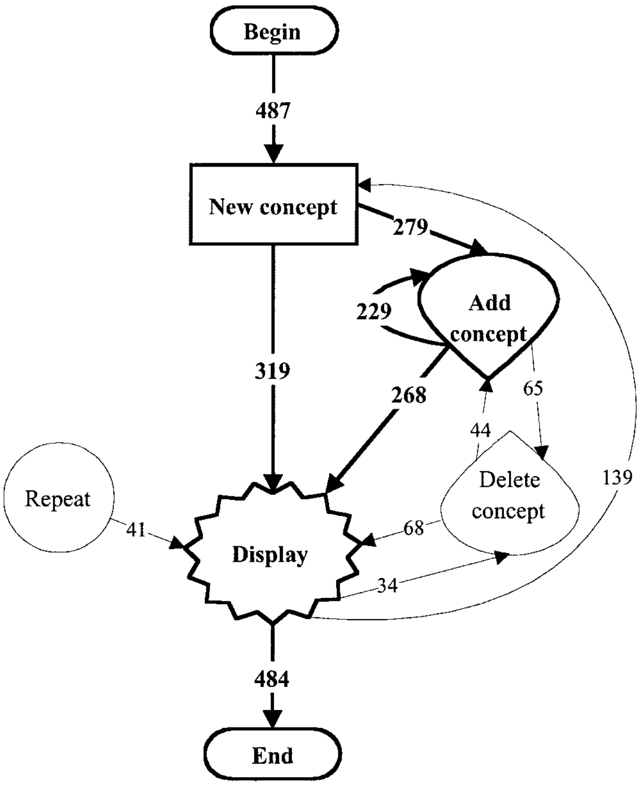


FIG. 5. State transition network of search tactics six months after the course ended. (n = 3,238 transitions, in 489 searches) (includes all transitions > 1% of total; highlights transitions > 5% of total).

represented the first and third families from the previous occasion.

The first family was very similar to that found on both of the previous assessment occasions, including initiation of the search (a *New concept* move), one or more *Add concept* moves, and a *Display* of the results (see the first column of Figure 6). This family included 260 searches and search fragments, a much lower proportion of the total than on the previous two occasions.

The second family of tactics on this occasion (in the second column of Figure 6) corresponds to the third family of tactics from the second assessment occasion (in the third column of Figure 4). It is characterized by the selection of a single search concept (a *New concept* move) followed by a *Display* of the results; then, either a new concept replaced the previous search or the search was concluded. On this occasion, this family was completely dominated by the single-term search, i.e., the tactic *New concept – Display – End*, which occurred 111 times. The remaining members of this family occurred only 80 times; however, it should be noted that some of these tactics consisted of up to four *New concept – Display* fragments in sequence.

### *Results Summary*

In summary, there were common patterns in the tactics used by students in formulating their searches. The most common pattern was the specification of a concept, followed by the addition of one or more concepts before a display of the results. While this pattern was the most common and occurred on all three occasions, other patterns did occur and these patterns changed over time as the students' domain knowledge changed.

## Discussion of the Effects of Domain Knowledge on Search Tactics

The results of this study provide a clear picture of differences in search tactics (as well as other search behaviors) based on differences in domain knowledge. When domain knowledge was very low, there were more moves per search (10 on the first occasion, versus 8 and 7 at the second and third occasions, respectively). The higher number of moves was most likely due to the number of changes the students needed to make in their search strategies in order to retrieve appropriate records. This result corresponds to Allen's (1991) finding that more search terms were used when knowledgeable users experienced retrieval difficulties. The medical students were at their lowest level of domain

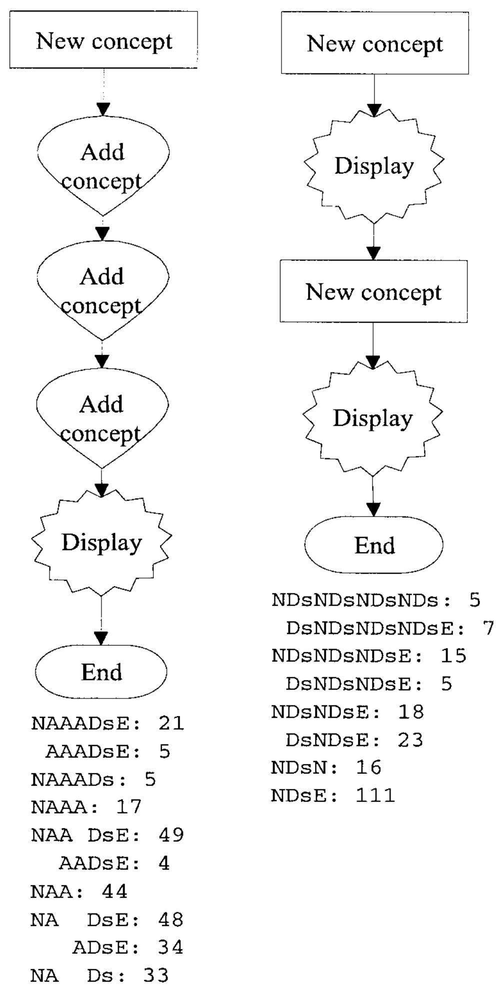


FIG. 6. Maximal repeating patterns six months after the course ended. (n = 3,238 transitions, in 489 searches).

knowledge when they made the most moves, so it may be that their inability to initially choose appropriate terms required them to generate additional search terms in order to retrieve records that would help them solve the clinical problems assigned.

The sequences of moves in students' search tactics also changed as domain knowledge changed. However, to be interpreted most effectively, students' problem-solving performance on the second pass (i.e., with database assistance) must also be taken into account (see de Bliek et al., 1994, for detailed performance data from the first cohort). At the first occasion, student domain knowledge was low; database assistance on the second pass through the problems allowed them to greatly improve their performance. At the second occasion, domain knowledge was at its peak; database assistance still improved performance, but not as much as at the first occasion. At the third occasion, domain knowledge had decreased to a point midway between the first and second occasions; database assistance allowed students to increase their second-pass scores more than on *either* of the other occasions. It is most likely that this unexpected increase at the third occasion can be attributed to the students' ability to recognize relevant information in the database, even though they could not recall it on the first pass through the problems.

In light of this performance data, the changes in search tactics make sense. On the first occasion, students used several moves to add and delete concepts from their search strategies in an attempt to retrieve relevant records. Their low knowledge gave them only the most basic ability to make relevance judgments and to use the information retrieved. In addition, the *Repeat* move was used often, probably reflecting the students' lack of facility with the database. On the second occasion, students were quite practiced at using the database and had high levels of domain knowledge. They used this knowledge to incorporate multiple concepts in their searches (explaining the high proportion of *Add concept* moves), but made fewer changes to their searches (combinations of *Add concept* and *Delete concept* moves). On the third occasion, their domain knowledge (based on recall of microbiology information) was low, but they were quite able to recognize relevant information in the database (as demonstrated by high second-pass scores). The results of this study indicate that they were also able to select a small number of useful concepts to incorporate in their search strategies. Much of the time, they were able to select a single term and then proceed to a display of the retrieved set; next most frequent was the tactic of adding one or two concepts to the initial search concept. Only rarely was a concept deleted from a search. At the third occasion, some errors (in the form of *Repeat* moves) did show up, but these were probably due to the students' lack of database practice in the intervening six months.

One must question whether the changes in search tactics discussed above are attributable to changes in domain knowledge or to changes in searching expertise. Two findings support the conclusion that these changes should be attributed primarily to changes in domain knowledge. First, the changes in domain knowledge were dramatic, as measured by the students' scores on the first pass through the problems at each occasion. Second, students' searching expertise appeared to be relatively high at the first occasion and, while it did improve by the second occasion (in terms of a drop in the number of *Repeat* moves), the number of other errors was relatively stable. Future studies should be concerned with both of these types of knowledge and their separate and combined effects on search behaviors (as exemplified in earlier studies by Bates, 1977; Hoelscher & Strube, 1999; Hsieh-Yee, 1993; Kiestra, Stokmans, & Kamphuis, 1994).

## Discussion of Methods for Analyzing Search Tactics

The research approach used in the current study makes two unique contributions to our understanding of the methods that can be applied to the analysis of search tactics. First, Shute and Smith's (1993) definitions of search moves were slightly adapted and applied to the search transaction logs—the first time that this coding scheme has been applied in a large-scale study of search moves and tactics. Second, two methods for analyzing the sequences of search moves, in an attempt to understand them as search tactics, were applied to these data. The strengths and weaknesses of the coding scheme and of each data analysis approach will be discussed here.

### *Coding of Search Moves*

For many search systems, transaction logs can be used to capture the details of each query and how it is formulated and reformulated within a particular search session. In many cases, these transaction logs can then be automatically simplified into a representation of the sequence of search moves—a representation that can be further mined for two types of information about the search: the sequence of *search commands* used in conducting the search and the *search terms* used in the search. Unfortunately, the commands are of little utility because they are extremely system-dependent and the set of available commands does not correspond well to the logic used by people in formulating a search strategy (as represented, for example, in the moves identified by Bates [1979, 1987] or Fidel [1985]). The search terms are of more interest but are difficult to analyze semantically using automatic methods. For example, Jansen, Spink, and Saracevic (2000) resorted to counting the frequencies of changes in Web search terms, but were unable to automatically categorize the types of changes made. Unless the search terms are all drawn from a controlled vocabulary with a pre-existing syndetic structure, the relationships between terms cannot be easily and automatically interpreted.

Because of these weaknesses in automatic methods of coding search moves, and in spite of the effort required, the transaction logs in the current study were coded manually. The coding scheme developed by Shute and Smith (1993) has three strengths that led to its adoption for this study. First, it focuses on the searcher's manipulation of the search terms, in terms of the semantic relationships among those terms. They proposed that their scheme describes "knowledge-based" search moves, and so it is very appropriate to studies that are interested in the application of a user's domain knowledge to the formulation and reformulation of search strategies. This coding scheme's second strength is its generalizability across search systems. Regardless of the specific command language used in a search system, the search moves can be represented accurately in this coding scheme. In addition to being applied to searches conducted on the factual database in microbiology that is the focus of the current study, this scheme has also been applied to searches in large bibliographic databases, such as MEDLINE (Wildemuth & Moore, 1995) and *Chemical Abstracts* (Shute & Smith, 1993). The third strength of this coding scheme is its granularity; it is neither too detailed nor too coarse. Prior taxonomies of search moves (Bates, 1979, 1987; Fidel, 1985) were quite detailed. This level of detail provides a complete view of all the actions taken by the searcher in formulating and reformulating a search strategy. However, it is so fine-grained that patterns do not emerge during analysis (Wildemuth & Moore, 1995). At the other extreme are coding schemes that are based on the search system's limited set of commands (e.g., Tolle, 1983); the analysis resulting from these coding schemes is not detailed enough to be useful to system designers. Thus, its emphasis on search term meanings and semantic relationships, its generalizability across search systems, and its level of granularity all recommend this coding scheme for continued use in research on search tactics.

An alternative scheme has recently been proposed by Rieh and Xie (2001). Like the Shute and Smith (1993) scheme, it is of a moderate level of granularity. There are 10 possible codes, organized in three facets: content (specification, generalization, replacement with synonym, parallel movement), format (term variations, operator usage, error correction), and resource (general resource, special resource, site url). It has been applied to a set of 183 Web search sessions, where Rieh and Xie (2001) found that their content facet accounted for 80% of the moves. This emphasis on the content facet provides further support for the validity of Shute and Smith's (1993) coding scheme, which also focuses on the content or semantics of the search tactics (rather than syntactic, logical, or resource-based facets). Researchers are encouraged to continue to focus on the content or meaning of search terms as they conduct additional studies of search moves and tactics.

### *Methods of Data Analysis*

Once search moves have been coded, they need to be further analyzed to understand how individual moves are sequenced as search tactics. Two methods of analysis were used in the current study: state transition matrices and maximal repeating patterns (MRP) analysis. While each of these is based on tabulating the frequencies with which particular transitions occurred, they are both essentially qualitative. They are intended to identify those sequences of moves on which the researcher should focus further efforts at interpretation.

In many ways, these two methods are similar. They both require that the sequence of moves be slightly manipulated, so that the rest of the analysis can be automated. In both cases, their output must be further interpreted to draw any conclusions about the process of search strategy formulation. However, each approach has strengths and weaknesses in terms of its validity.

State transition matrices have been used in a number of studies of search moves, based on transaction log data, beginning with Tolle's (1983) work on online catalogs. The biggest advantage of this method is that every transition between two moves is taken into account when calculating the zero-order state transition matrix. However, it should be noted that, when drawing conclusions, it is most likely that low-frequency transitions will be ignored, as they have been in the current study. The most important threat to the validity of this method is that it views the search tactics as small fragments, allowing the researcher to piece them together only after the matrix is created. In other words, it is quite possible that a diagram representing the most frequent zero-order transitions may illustrate a pattern that *never* occurred in any individual search.

The analysis of maximal repeating patterns (Siochi & Ehrich, 1991) addresses this validity problem. This analysis looks for the longest sequences of moves within a search, and then tabulates the frequency of the occurrence of those sequences. Thus, the outcome of this analysis is potentially a more valid characterization of search tactics and strategies (which, at least in the current study, encompass 7-10 moves) than the zero-order state transitions, which focus on one transition (i.e., two moves) at a time. The weakness of this approach to search tactics analysis is that sequences that do not occur frequently will be ignored when drawing conclusions; however, this problem is not noticeably larger than the comparable problem with state transition matrices.

Methods for the analysis of maximal repeating patterns were originally developed for application in the field of human-computer interaction. Here, it is argued that they are also very applicable to the analysis of search tactics. Nevertheless, MRP analysis reveals only *what* the searcher did; it does not explain *why* the searcher formulated the search strategy in a particular way (Borgman, Hirsh, & Hiller, 1996). Researchers are urged to use MRP analysis in their future studies of search strategy formulation and reformulation and, if possible, augment it with other data collection techniques (such as the think aloud and think after protocols investigated by Branch [2000]) to gain a fuller understanding of the process of search strategy formulation and reformulation.

## Conclusion

A database search is made up of a series of moves, temporally and semantically related as search tactics. Yet we have little understanding of how people approach this task. The current research examined a large number of searches conducted when the searchers had varying levels of personal knowledge about the topic/domain. An analysis of the search tactics used at these varying levels of knowledge indicates that the most common pattern used across all three occasions (i.e., across varying levels of domain knowledge) was the gradual narrowing of the retrieved set through iterative addition of search concepts. While this pattern did occur frequently at every occasion, there were also differences in the search tactics used at each occasion. Thus, it was concluded that domain knowledge does affect search behaviors. Low domain knowledge was associated with less efficient selection of concepts to include in the search and with more errors being made in the reformulations of the search tactics.

These results have important implications for system designers. Current information retrieval systems provide little or no support for searchers' formulation and reformulation of search strategies. If anything, some access to the database's controlled vocabulary might be available. However, this does little to guide the searcher in developing an effective search strategy. If other studies also find that gradual narrowing of the retrieved set is the most "natural" tactic for searchers to adopt, there are many graphical ways to support searchers' visualization and management of the search formulation and reformulation process. Once we have a better understanding of how people conduct searches and can begin to develop a theory describing search strategy formulation and reformulation, then we can begin to redesign our information retrieval systems to better support those who use them.

Though this study has important implications for design, it also has some limitations that cannot be ignored. First, it was conducted with one set of searchers (medical students) in one domain (microbiology). It is certainly possible that different groups of searchers would behave differently. It is also possible that the structure of the domain itself may have an influence on searchers' tactics. Second, the influence of the coding method on the results cannot be ignored. Other possible definitions of moves have been proposed (e.g., Bates, 1979, 1987; Fidel, 1985). Shute and Smith's (1993) coding scheme was selected here because it focuses on the selection of search terms. In addition, it has the advantage of being applicable to any information retrieval system in which search term selection is part of the searching process. Its generalizability should make these results comparable to the results from other studies using this coding scheme, across domains and groups of searchers.

Overcoming these study limitations can only be successful with additional studies of searching behaviors and, in particular, search tactics. We need to know more about how people formulate their searches and about the influences of their personal knowledge on those search tactics. The automatic capturing of transaction logs makes such studies possible, but manual coding of the individual moves is necessary in order to derive meaning from those logs. Only through further studies of search moves in different domains and with different sets of searchers can our knowledge of these behaviors be increased.

## Acknowledgments

This work would not have been possible without the contributions of Charles P. Friedman, principal investigator of the "Information and Cognition in Medical Education" project; the numerous graduate research assistants who coded the search moves; and the student participants. The comments of the anonymous reviewers made significant contributions to the clarity of this paper. Funding was provided by the National Library of Medicine (grant 5R01 LM04843).

## References

Allen, B. (1991). Topic knowledge and online catalog search formulation. Library Quarterly, 61(2), 188–213.

Bates, M.J. (1977). Factors affect subject catalog search success. Journal of the American Society for Information Science, 28(3), 161–169.

Bates, M.J. (1979). Information search tactics. Journal of the American Society for Information Science, 30(4), 205–214.

Bates, M.J. (1987). How to use information search tactics online. Online, 11(3), 47–54.

Berger, M.G. (1994). Information-seeking in the online bibliographic system: An exploratory study. Ph.D. dissertation, University of California, Berkeley. UMI Order No. 9504745.

Bhavnani, S.K. (2002). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. CHI 2002, Conference on Human Factors in Computing Systems (Minneapolis, April 20–25, 2002): Extended Abstracts, 610–611.

Bhavnani, S.K., & Bates, M.J. (2002). Separating the knowledge layers: Cognitive analysis of search knowledge through hierarchical goal decompositions. ASIST 2002: Proceedings of the 65th ASIST Annual Meeting, Volume 39 (Philadelphia, November 18–21, 2002), 204–213.

Borgman, C.L., Hirsh, S.G., & Hiller, J. (1996). Rethinking online monitoring methods for information retrieval systems: From search product to search process. Journal of the American Society for Information Science, 47(7), 568–583.

Borgman, C.L., Hirsh, S.G., Walter, V.A., & Gallagher, A.L. (1995). Children's searching behavior on browsing and keyword online catalogs: The Science Library Catalog Project. Journal of the American Society for Information Science, 46(9), 663–684.

Branch, J.L. (2000). Investigating the information-seeking processes of adolescents: The value of using think alouds and think afters. Library & Information Science Research, 22(4), 371–392.

Buttenfield, B.P., & Reitsma, R.F. (2002). Loglinear and multidimensional scaling models of digital library navigation. International Journal of Human-Computer Studies, 57(2), 101–119.

Carmel, E., Crawford, S., & Chen, H. (1992). Browsing in hypertext: A cognitive study. IEEE Transactions on Systems, Man, and Cybernetics, 22(5), 865–884.

Connell, T.H. (1995). Subject searching in online catalogs: Metaknowledge used by experienced searchers. Journal of the American Society for Information Science, 46(7), 506–518.

Cuomo, D.L. (1994). Understanding the applicability of sequential data-analysis techniques for analyzing usability data. Behaviour & Information Technology, 13(1-2), 171–182.

de Bliek, R., Friedman, C.P., Wildemuth, B.M., Martz, J.M., Twarog, R.G., & File, D. (1994). Information retrieved from a database and the augmentation of personal knowledge. Journal of the American Medical Informatics Association, 1, 328–338.

de Bliek, R., Friedman, C.P., Wildemuth, B.M., Martz, J.M., File, D., Twarog, R.G., Reich, G.M., & Hoekstra, L. (1993). Database access and problem solving in the basic sciences. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, A Conference of the American Medical Informatics Association: Patient-Centered Computing (Washington, DC, October 30–November 3, 1993), 678–682.

Debowski, S. (2001). Wrong way: Go back! An exploration of novice search behaviours while conducting an information search. The Electronic Library, 19(6), 371–382.

Dimitroff, A. (1992). Mental models theory and search outcome in a bibliographic retrieval system. Library & Information Science Research, 14(2), 141–156.

Dimitroff, A., & Wolfram, D. (1995). Searcher response in a hypertext-based bibliographic information retrieval system. Journal of the American Society for Information Science, 46(1), 22–29.

Fidel, R. (1985). Moves in online searching. Online Review, 9(1), 61–74.

Friedman, C.P., de Bliek, R., Gilmer, J.S., Twarog, R.G., & File, D.D. (1992). Influence of a computer database and problem exercises on students' knowledge of bacteriology. Academic Medicine, 67(5), 332–337.

Hoelscher, C., & Strube, G. (1999). Searching on the Web: Two types of expertise. Proceedings of SIGIR '99: 22nd International Conference on Research and Development in Information Retrieval (University of California, Berkeley, August 1999), 305–306.

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. Journal of the American Society for Information Science, 44(3), 161–174.

Jacobson, T., & Fusani, D. (1992). Computer, system, and subject knowledge in novice searching of a full-text, multifile database. Library & Information Science Research, 14(1), 97–106.

Jansen, B.J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. Information Processing & Management, 36(2), 207–228.

Kiestra, M.D., Stokmans, M.J.W., & Kamphuis, J. (1994). End-users searching the online catalogue: The influence of domain and system knowledge on search patterns. The Electronic Library, 12(6), 335–343.

Kim, K.-S. (2001a). Implications of user characteristics in information seeking on the World Wide Web. International Journal of Human-Computer Interaction, 13(3), 323–340.

Kim, K.-S. (2001b). Information seeking on the Web: Effects of user and task variables. Library & Information Science Research, 23(3), 233–255.

Maidenberg, E. (1991). Expertise in online information retrieval. Doctoral dissertation, University of California, Los Angeles. UMI Order No. 92-06013.

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. Journal of the American Society for Information Science, 40(1), 54–66.

Marchionini, G., Dwiggins, S., Katz, A., & Lin, X. (1993). Information seeking in full-text end-user-oriented search systems: The roles of domain and search expertise. Library & Information Science Research, 15(1), 35–69.

Marchionini, G., Lin, X., & Dwiggins, S. (1990). Effects of search and subject expertise on information seeking in a hypertext environment. Proceedings of the 53rd Annual Meeting of the American Society for Information Science, Volume 27 (Toronto, Ontario, November 4–8, 1990), 129–142.

McDonald, S., & Stevenson, R.J. (1998). Navigation in hyperspace: An evaluation of the effects of navigational tools and subject matter expertise on browsing and information retrieval in hypertext. Interacting with Computers, 10(2), 129–142.

Qiu, L. (1993). Markov models of search state patterns in a hypertext information retrieval system. Journal of the American Society for Information Science, 44(7), 413–427.

Rieh, S.Y., & Xie, H.I. (2001). Patterns and sequences of multiple query reformulations in Web searching: A preliminary study. ASIST 2001: Proceedings of the 61st ASIST Annual Meeting, Volume 38 (Washington, DC, November 3–8, 2001), 246–255.

Sanderson, P.M., & Fisher, C. (1994). Exploratory sequential data analysis: Foundations. Human-Computer Interaction, 9(3/4), 251–317.

Saracevic, T., & Kantor, P. (1988). A study of information seeking and retrieving. III. Searchers, searches, and overlap. Journal of the American Society for Information Science, 39(3), 197–216.

Shute, S.J., & Smith, P.J. (1993). Knowledge-based search tactics. Information Processing & Management, 29(1), 29–45.

Siochi, A.C., & Ehrich, R.W. (1991). Computer analysis of user interfaces based on repetition in transcripts of user sessions. ACM Transaction on Information Systems, 9(4), 309–335.

Spink, A., Jansen, B.J., & Ozmultu, H.C. (2000). Use of query reformulation and relevance feedback by Excite users. Internet Research: Electronic Networking Applications and Policy, 10(4), 317–328.

Tolle, J.E. (1983). Transactional log analysis: Online catalogs. Proceedings of the Sixth Annual International ACM SIGIR Conference, 147–160.

Vakkari, P. (2002). Subject knowledge, source of terms, and term selection in query expansion: An analytic study. Advances in Information Retrieval: 24th BCS-IRSG European Colloquium on IR Research, Proceedings (Glasgow, UK, March 25–27, 2002), 110–123.

Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal case study. Information Processing & Management, 39(3), 445–463.

Wildemuth, B.M., de Bliek, R., Friedman, C.P., & File, D.D. (1995). Medical students' personal knowledge, searching proficiency, and database use in problem solving. Journal of the American Society for Information Science, 46, 590–607.

Wildemuth, B.M., de Bliek, R., He, S., & Friedman, C.P. (1992). Search moves made by novice end users. ASIS '92 Proceedings, Volume 29 (Pittsburgh, PA, October 26–29, 1992), 154–161.

Wildemuth, B.M., & Morre, M.E. (1995). End-user search behaviors and their relationship to search effectiveness. Bulletin of the Medical Library Association, 83(3): 294–304.

Yuan, W. (1997). End-user searching behavior in information retieval: A longitudinal study. Journal of the American Society for Information Science, 48(3), 218–234.