Online Searching as a Problem-Solving Process

The broad subject of this article is how online searching (as a special form of information retrieval) can usefully be viewed as a problem-solving process. This is not an especially new idea. Several writers have taken this general position including Linda Smith (1976, 1980), Marcia Bates (1979a, 1979b), Don Swanson (1977, 1979, 1986), and the author (Harter 1984a, 1984b, 1986). However, this discussion will focus on a particular aspect of the identification of useful classes of *heuristics* for online searching. These ideas are not only interesting and significant from a theoretical point of view, but also because of their implications for education and training, for how librarians view end user searching of online catalogs and other in-house systems, and for the design of expert systems for online searching.

If online searching can be viewed as a problem-solving process, what is the problem that is being solved? An early modern description of the so-called “information retrieval problem” was published by Don Swanson (1963). However, as a source of intellectual concern, the problem must be nearly as old as the earliest libraries.

What is the information retrieval problem? Assume that a person is working on a scholarly, scientific, or practical project and has need for particular documents or pieces of data. That person decides to consult a library. How can the wanted information be found and retrieved from the library from among billions of pieces of recorded human knowledge that do not relate to the information need? How can one separate the wheat from the chaff without at the same time losing the greater portion of the wheat? A modern approach to this old problem is to use the tool of online information retrieval.
The discussion is arranged as follows. First, the components of a particular online search problem will be identified—i.e., the question, its context, and the goals the search is to achieve.

Second, a general problem-solving approach to information retrieval will be outlined, and a model for attacking information retrieval problems online will be reviewed. The model stresses the difference between rules and heuristics.

Third, the paper concentrates on classes of search heuristics—tactics that might be employed during a search. Two typologies for representing major classes of heuristics will be identified—first by the type of action taken, and second, by the search state that produced the action. A class of heuristics for a particular search state will be explored in detail.

Finally, implications for education and training, end user searching, and the design of expert systems for online searching will be briefly explored.

COMPONENTS OF THE SEARCH PROBLEM

What are the dimensions of a “problem” in online searching? One sometimes sees a problem described simply as a question: “I want survey results reporting attitudes of U.S. citizens toward Russia.” Although every search problem certainly involves a question such as this, it can involve much more as well.

An online search problem has several components:

— the question: a verbal or written statement of a relationship between concepts;
— the context: the environment, background, or setting for the question;
— the retrieval goals: what the end user hopes to achieve in results and costs.

First consider the search question. Several levels of questions can be defined. Robert S. Taylor (1962, 1968) has described a model in which four stages or conceptual levels of a question are identified:

| Q₁    | the visceral need |
| Q₂    | the conscious need |
| Q₃    | the formalized need |
| Q₄    | the compromised need |

Taylor describes the visceral need as the underlying “information need of the end user” which may not be consciously recognized and is probably not expressible linguistically. The conscious need is fuzzy and ambiguous and is intimately connected with the context of the problem. Its linguistic
expression may be rambling. At this stage the end user may talk about the problem (and its context) to his colleagues in the hope of understanding and clarifying these ambiguities.

According to Taylor's model, the formalized need is a precise, unambiguous expression of need. At this stage the question may be qualified by the end user and may not include a discussion of the problem context. Finally, the compromised need is a revised version of the formalized need restated in light of how the end user believes the information retrieval system (whether it be a librarian, a card catalog, a CD-ROM system, an OPAC, an expert system, or a commercial search service) can best address the question. At this stage the question may be a simple list of words or even a single subject heading.

Whether or not Taylor's model is accepted in its entirety, it makes important points regarding at least some search problems. First, it distinguishes between the concepts that make up the problem and the words used to represent them. As the end user (and searcher) moves from Q₁ to Q₄, the world of concepts is largely abandoned. Instead, one deals with terminology—i.e., words, phrases, and other symbols. If the search analyst is not careful, he or she will not accurately grasp the underlying concepts of the information problem. If the problem is presented at the level of Q₄, the search intermediary—human or machine—must presumably try to move to earlier levels. Ideally the intermediary will be able to understand the question at level Q₂ as a colleague of the end user. Note that this more extensive delving into a question is not always necessary; a discriminating searcher will know when it is and is not needed.

Taylor's model illustrates how compromises are made and constraints are set on the original information need. It especially stresses the importance of understanding the context of the problem. A search question is a product of an environment—the setting or background from which the question has emerged. An understanding of context will not only help the searcher understand the original problem (as opposed to how it was framed) but also will suggest ways of modifying the search as it progresses if necessary.

Consider the search question posed earlier: "I want survey data reporting attitudes of U.S. citizens toward Russia." The question seems straightforward. However, an interview designed to move beyond the words comprising the question might reveal that the end user is a university professor of political science, that this search problem is part of a larger research project on the evaluation of alternative methods of communicating political information about nondemocracies by the United States and other Western governments to their peoples, that the professor is only interested in research findings, that her actual interests extend beyond Russia to the Soviet Union, Eastern Bloc, and other communist countries
as well as extreme right wing totalitarian regimes of which she names several.

Furthermore, the client is interested only in recent findings (the past few years). She can read only German and English and wants retrieval limited to findings presented in these languages. She is writing a grant proposal to explore these questions further and wants comprehensive results on the specific question (dealing with Russia and the United States). She would also like results to be as comprehensive as possible on the more general question. She thinks that there may be several dozen research articles on the former, and perhaps one or two hundred articles on the latter. She can afford to pay up to $50 on the search if necessary. Having gained this information regarding the problem context, the online searcher will be able to formulate alternative approaches to the problem consistent with the overall purpose of the search.

Besides the question and its context, the end user in the example has stated retrieval goals. They are of two types: goals identifying the comprehensiveness and/or purity of the retrieval results and goals stating cost constraints. Comprehensiveness and purity goals are often stated by desired levels of recall and precision:

- recall: the proportion of relevant documents retrieved (from a universe of documents) by the search
- precision: the proportion of retrieved documents that are relevant to the problem

Recall is a measure of the comprehensiveness of the search while precision is a measure of the purity of the output. Usually if a searcher is acting rationally, attempts to improve recall will degrade precision and vice versa. This point will be addressed more fully later.

In summary, an information retrieval problem is comprised of several parts: the question, its context, and the retrieval goals of the end user. The purpose of an online search of a bibliographic database is to solve the information retrieval problem: to identify documents that discuss the concepts of the question in the relationship indicated consistent with the problem context and the retrieval goals of the end user. Note that a given problem may be insoluble; negotiation and compromise may be necessary before a solution can be found.

PROBLEM-SOLVING AND INFORMATION RETRIEVAL

As was observed earlier, several writers have argued that information retrieval can be usefully regarded as a problem-solving process. Such a view takes online searching to be an iterative, trial and error process in
which the searcher takes steps to move closer and closer to the retrieval goals of the end user.

Within the framework of an overall approach to a solution, the searcher prepares formulations to represent the major concepts and their interrelationships. These formulations can be regarded as hypotheses—statements that the searcher believes may succeed in satisfying retrieval goals. Normally, alternative formulations should be prepared as well since initial formulations (as uncertain hypotheses) may fail (for an elaboration see Harter, 1984).

When preparation has been completed, the searcher connects to the retrieval system and begins testing hypotheses. A formulation (a command or series of commands) is entered and the results are tested by printing a sample of retrieved records. The searcher evaluates these against the question, problem context, and retrieval goals. Based on the number of postings retrieved and on an estimate of recall and precision, the searcher tries to move closer to a solution of the problem by stating and testing a new hypothesis. A second formulation is prepared and put to the system. Again a sample of records is evaluated, and again the searcher considers what might be done to move closer to a solution. The process continues to cycle until, in the searcher's judgment, a solution to the problem has been reached (that is, the question, in its particular context, has been satisfied in terms of recall, precision, and cost goals).

How does the searcher know which hypotheses are viable? Of the infinity of possible ways of modifying a given search in progress, which should be selected? The searcher is guided by intuition and logical ability as well as by his/her knowledge of heuristics.

Heuristics (or tactics [Bates, 1979] or moves [Fidel, 1984]) are actions that are taken to approach a goal. By their nature, heuristics are not foolproof; they promise the possibility of success but offer no guarantee. Heuristics are perhaps best understood by contrasting them to rules. Rules are actions known to have a certain effect in all cases. Rules are not uncertain. Except for technical problems such as bugs in the retrieval system, line noise, or a scratched CD-ROM disc, rules in information retrieval always have the same effect.

In online searching, logon protocols are expressed as rules. Also, the effects produced by the various commands on the retrieval system files may usefully be regarded as rules. For example, the statement that the expression A AND B produces a new set C consisting of postings common to sets A AND B is a rule. Boolean AND always works this way. In addition to conceptual rules such as this, the searcher must adhere to the form in which commands are to be entered to the system. These are called syntax rules.

Obviously a searcher (or an intelligent front-end or expert system) must know certain rules to successfully solve a problem online. However,
while knowledge of rules is necessary, it is not a sufficient condition for success in all cases. It is assumed here (an assumption supported by the professional literature) that online searching is not a deterministic activity governed solely by rules; at the heart of the matter are so-called "rules of thumb"—heuristics for successful searching. Experienced searchers tend to possess knowledge of such heuristics and novices tend not to possess them. A thesis of this article is that in the education and training process perhaps far too much attention is given to the acquisition of rules (which are, after all, the easy things to teach) and far too little attention is given to the acquisition of heuristics. The next section of this article describes major classes of heuristics.

A TYPOLOGY OF HEURISTICS

Numerous papers have been published that identify one or more heuristics (hints, tricks, strategies, approaches) for successful online searching. Library and information science has several professional journals devoted largely to these issues and numerous other journals that occasionally publish relevant pieces.

In a study published in Online Review, Harter and Peters (1986) proposed a typology based on heuristics identified in published studies. They located and read every published article in the professional literature that could be found discussing aspects of "how to do" an online search, in whole or in part, looking for suggestions made by the "experts" writing the articles. That the authors were indeed experts is one of the assumptions underlying this work.

These heuristics were then organized into classes. This proposed typology was largely built around the type of heuristic being considered rather than its function. The article identified six classes of heuristics. These were based on:

1. overall philosophy and approach of the searcher
2. language of problem description
3. record and file structures
4. concept formulation and reformulation (Boolean logic)
5. increasing or decreasing recall and precision
6. cost/effectiveness

While this organization is useful as a preliminary approach, it is perhaps not as helpful as it might be to a searcher looking for the correct approach at a given point in an ongoing search. For application to a search in progress, a typology is required that is based on the present state of the search—i.e., the situation in which the searcher currently finds himself.
Except for the item above on increasing or decreasing recall and precision, the searcher is given little guidance by this typology to which heuristic to pursue next at a given time. Thus this organization tends to be pedagogically oriented; it is better suited for classroom instruction than for guidance during an ongoing search.

This article proposes a typology of heuristics based on search states rather than form. The typology is based on the idea that there are several major states that characterize an ongoing solution to a search problem at least one of which is likely to be encountered in nearly every search.

Heuristics Based on Search States

What are the states in which a searcher might find himself? Ideally, the searcher would like to evaluate the search output against the problem statement and its context as well as recall, precision, and cost goals. Sometimes, however, one is unable to reach this point for technical or personal reasons. There are several preliminary states that one might encounter:

— inability to make connection with the host system;
— no response from system, no prompt, keyboard frozen;
— double characters or no characters are displayed when commands are typed;
— system error message is received;
— a command has been given that the searcher realizes is in error and wants to “take back”;
— utter confusion, helplessness, panic

The first several of these states are technical in nature and the last is personal. In general, these states are properly addressed by rules rather than heuristics. These rules may be listed in search system documentation under the heading of “troubleshooting.” For example, if no characters are displayed when commands are typed, the rule to be followed is to toggle the duplex switch to half duplex; and on some host systems the rule for stopping the execution of a command (taking it back) is to send a break or interrupt signal.

A good system error message will include an attempt at diagnosis and will state or imply a rule for solving the problem. Thus the error message “unmatched parentheses” implies that the searcher should review the formulation looking for an extra or missing parenthesis, and to reenter it with the correction made. The rule that might be suggested to beginners when they experience a state of confusion or panic is to disconnect from the system, take several deep breaths, review the output, and try to recover before continuing.
In conclusion, although the states listed earlier are legitimate states in which one might find oneself, they are handled with rules. They are also easily automated. Rule-based states are not considered further.

The remainder of this article concentrates on heuristics suggested for states that result from a comparison of system output with the components of the search problem—the question, its context, and the retrieval goals. The following states are among those that might be achieved as a result of this comparison:

1. *Records not related to the subject.* The search has resulted in the retrieval of documents that do not discuss the topic of the search problem; they are entirely off the subject. They bear no resemblance to either the specific question or its context. What is wrong and what should the searcher do?

2. *Unexpected null set.* The search has resulted in an unanticipated empty set. What are the possible explanations of this state, and what should be done?

3. *Need to increase recall.* The search has resulted in some documents relevant to the question but too few; the comprehensiveness of the search is inadequate when measured against the retrieval goals of the end user. What can the searcher do to increase recall?

4. *Need to increase precision.* The search has resulted in relevant documents, but the percentage of nonrelevant documents is unacceptably large. There are too many false drops when measured against the retrieval goals of the end user. What can the searcher do to increase precision?

5. *Need to limit search.* The search has resulted in the retrieval of an acceptable mixture of relevant and nonrelevant documents, but there are too many records to print; the desired recall and precision goals cannot be achieved for the stated cost goal. What can the searcher do to reduce the number of documents retrieved in a way that is consistent with the problem context and retrieval goals?

**STUDY OF THE STATE OF RETRIEVING NONRELATED RECORDS**

Material in this section is drawn heavily from *Online Search Analyst*, a menu-driven diagnostic and tutorial program for IBM PCs and compatible microcomputers, written by and available from the author (Harter, 1987). The program is organized by the five states listed earlier. It provides a detailed discussion of each state, including the identification of several subclasses of heuristics that might be employed for each state. It includes
numerous examples for each heuristic discussed and is presented in a tutorial fashion.

In the discussion that follows, a closer look is taken at the first of these states—a state that is perhaps rarely achieved by experienced searchers but is commonly found in searches conducted by beginners. The searcher has invented a hypothesis designed to retrieve records on the question. A formulation is prepared to test the hypothesis and is entered into the system. A solution set is created, and a sample of records from the set is printed. To the searcher's surprise, the records appear to have nothing whatever to do with either the question or its context; they are not at all on the subject. What are possible explanations of this state, and what are the heuristics for avoiding this problem in the future?

There are six major subclasses of useful heuristics for explaining the problem and for dealing with it or avoiding it in future searches. The problem may be: (1) poor choice of database, (2) misuse of Boolean logic, (3) semantic or language problems, (4) set number errors, (5) truncation errors, or (6) not enough search facets.

First, conceivably the searcher has obtained results not on the subject because he is searching in an inappropriate database. In particular, newspaper and magazine databases such as National Newspaper Index and Magazine Index are highly multidisciplinary. Not only are documents indexed that represent a wide variety of disciplines, but also such databases contain "popular" material—publications that are not technical in their information content.

Perhaps the searcher is looking for a technical subject in a "popular" database. What may be really needed is a technical, scholarly, or scientific database such as MEDLINE, Chemical Abstracts, or Biological Abstracts.

As a second possible explanation of retrieval results totally unrelated to the subject searched, an appropriate database may have been selected, but the command to the search system was put incorrectly. A DIALOG searcher may have forgotten to specify a file and is searching in the default database. Or possibly he made a simple typographical error with the file number.

In an organization that shares passwords, perhaps the searcher was connected to the database that his colleague had been using. In his hurry to get to the matter at hand, he may have failed to notice that he had been connected to the incorrect database.

Two heuristics are suggested:

—Has an appropriate database been selected? In particular, are you looking in a popular file when what is needed is a technical, scholarly, or scientific database?
—It pays to read the information carefully to verify that the correct file is being used.

Another possible explanation of the retrieval of records not on the subject is that the Boolean operators AND, OR, and NOT were used incorrectly. Misuse of Boolean logic is a common reason for retrieval sets that are completely off the subject.

In particular, mistakenly using OR when AND is the correct operator produces a relatively large set consisting almost entirely of nonrelevant records. Although it may seem obvious to many readers, in this author’s experience with students, this is a common error.

\[
A \text{ OR } B \quad \text{is the set of records that are either in set } A \text{ or in set } B \text{ or in both sets}
\]

\[
A \text{ AND } B \quad \text{is the set of records that are present in set } A \text{ and also in set } B
\]

Even if one assumes that a searcher knows which of these operators should be used in a given situation (a questionable assumption), there are still problems. Search systems typically have an implied order of operations to handle potentially confusing combinations of Boolean operators such as in the example: \(A \text{ OR } B \text{ AND } C\). In the DIALOG system, for example, the order of operations is to first carry out all NOT operations then all AND operations, and finally all OR operations from left to right. Exceptions to this order are made possible through parentheses. Hence in the expression: \((A \text{ OR } B) \text{ AND } C\), the parentheses tell the system to do the OR operation first then to do the AND. If the parentheses had been absent as in \(A \text{ OR } B \text{ AND } C\), then \(B \text{ AND } C\) would be done first.

Suppose that one is interested in automation of either information centers or libraries. The expression \(\text{LIBRARIES OR INFORMATION CENTERS AND AUTOMATION}\) would produce a set of documents largely irrelevant to the search problem. The set would consist of records containing both of the terms \textit{information centers} and \textit{automation}. However, these records would be overwhelmed by a much greater number of records containing the word \textit{libraries} in the fields searched. The correct formulation requires the use of parentheses: \((\text{LIBRARIES OR INFORMATION CENTERS}) \text{ AND AUTOMATION}\). The following heuristics are suggested:

—Be sure that you know the order in which Boolean operators are executed in the search system you are using.
—Use parentheses to ensure that certain operations are carried out before others.

Semantic or language problems can also be the reason for retrieving
records which are totally off the subject. There are two kinds of language problems that might result in this outcome: (1) the use of broad, general, or fuzzy terms representing a concept; and (2) the use of terms with multiple meanings to represent a concept. In both cases an unwanted meaning can overwhelm the concept of interest. For example, suppose that several terms are used to represent a concept, and that these terms are combined with Boolean OR as in: RESEARCH or QUEUING THEORY or LINEAR PROGRAMMING. If one term is broad or fuzzy and retrieves many postings (in this case research), then postings from this term will overwhelm postings from the other more specific terms. Although the final retrieval set will contain postings from both the specific descriptors QUEUING THEORY and LINEAR PROGRAMMING, there will be relatively many more postings from the broader term research. This problem can be solved by avoiding generic terms likely to retrieve many postings when it is really a more specific concept that is wanted.

—Use the most specific terms you can find to represent the concepts of a search.
—Avoid generic terms with many postings.

A second possible explanation of a retrieval set that contains records totally unrelated to the search problem is the occurrence of homographs in the database—e.g., words, acronyms, abbreviations—with multiple meanings. Such a term will retrieve records discussing all the concepts named by the term including many that are not on the subject wanted. For example, the term salt will retrieve records discussing various chemical compounds, Salt Lake City, as well as the Strategic Arms Limitation Talks.

The multiple meaning problem is most serious with multidisciplinary and popular databases. In Magazine Index, the term mole might refer to any of the following concepts: a spy, a congenital spot on the human body, an insectivore, or one gram molecule. Mole as spy is a popular term, but mole also has various technical or quasitechnical meanings in medicine, gardening and lawn care, biology, and chemistry. The searcher may have overwhelmed the “mole as one gram molecule” postings with postings to spies, congenital spots, gardening problems, or the biology of insectivores. The following heuristics are suggested:

—Watch out for homographs—words with more than one meaning.
—Use controlled vocabularies—thesauri, lists of subject headings, or classification codes—or specific search fields whenever possible to restrict a term to a particular meaning.

A careless set number error or command language error may also explain the lack of relevance of a retrieval set to the concepts of interest. In
particular, the searcher may simply have typed records from an inappropriate retrieval set.

Often beginners follow their planning notes too unthinkingly. Their plans call for displaying records from set 7, so that is the command given even though for various reasons set 7 may not refer to the correct set in the search as it was actually conducted. A simple typo might also explain the result obtained.

—Watch for typos of set numbers.
—Don’t follow your planning notes too slavishly; adjust set numbers in your notes to reflect what has actually taken place in the search.

A common error on DIALOG is to confuse the way in which set numbers are used with the combine and select commands. The command combine 3 AND 7 creates the Boolean intersection of sets 3 and 7. However, the command select 3 and 7 will create a set for the numeral 3, another set for the numeral 7, and combine these using Boolean AND. The result of the latter command will have nothing whatever to do with the previously created sets 3 and 7 or the underlying search problem.

To combine sets 3 and 7 using the select command on DIALOG, type: select S3 AND S7.

—Be careful of confusing a numeral as a search term with a set number.
—Always precede a set number with the letter S when using the select command on DIALOG to perform Boolean operations.

Truncation is another possible explanation for retrieving records totally different from the subject. If a term is truncated too much, words representing different roots will be combined to form a retrieval set. Since many of these terms will not relate to the concept of interest, the result may be records totally off the subject wanted.

For example, searching on the truncated stem LIB will retrieve records containing the words liberty, libation, libby, and libel, as well as the library terms of library, librarian, and librarianship. The stem that one should search is librar.

Searching on the truncated stem US for the concept of “United States of America” will retrieve records containing the terms use, ustinov, and usher, as well as the wanted terms US and USA. Postings from these terms may greatly exceed the number of postings resulting from US and USA.

—Be careful where you truncate. A carelessly truncated stem may result in the retrieval of many records not on the concept wanted.
—Use Boolean OR to combine terms that are alphabetically related rather than truncation when truncation is likely to result in retrieval on unwanted terms.
Finally, it is possible, though not likely, that the reason the searcher has obtained records that seem to be totally unrelated to the subject is that he has not employed enough concepts (or facets) in his formulations. Assume that the “subject” to be represented has three concepts to be intersected: A AND B AND C. If the original formulation is comprised of only two of these facets, then records will not be on the specific subject of the search. If this is the case, a possible solution is to create a search formulation for the third facet and intersect it with the previous two facets.

—Consider forming another facet to intersect with existing facets.

Note, however, that the tendency of most beginning searchers is to overspecify rather than underspecify problems. Underspecification is unlikely to be the source of difficulty, and other possibilities should be considered first.

MOST RATIONAL PATH

The previous section of this article proposed six subcategories of heuristics for a particular search state—i.e., that records retrieved in a search formulation are completely unrelated to the search problem. A total of fifteen heuristics, each representing a class of possible actions, were suggested as possible ways of addressing this retrieval state. Each of the fifteen, or a combination of several, are possible explanations (and solutions) of the state.

Which of these fifteen subclasses of heuristics should be employed in a given search in which this state is experienced? Should they all be considered, one by one, in checklist form? Or, more likely, is it possible that they will not work equally well in a given situation? In that case the searcher would need to carefully consider each heuristic in the context of the particular search being conducted to identify the one most likely to produce the desired retrieval results. It is the nature of heuristics, as compared to rules, that what works sometimes will not work at other times.

Moreover, given that a particular heuristic has been selected to pursue, many different formulations might be employed to operationalize it. Suppose that a searcher wants to increase precision and decides to implement the heuristic: Consider forming another facet to intersect with existing facets. Once the decision to implement this heuristic has been made, there are an infinity of ways to go about it involving choices of terms, proximity operators, fields, truncation, Boolean operators, etc. Thus the problem is not solved merely by selecting a heuristic to pursue; the searcher must then operationalize it by formulating system commands. In general, there are an infinity of possible formulations.
Each formulation would result in a different level of recall, precision, and cost if implemented. Imagine that the universe of possible formulations can be viewed as a three dimensional space where each point in the space is the retrieval result associated with a given formulation. As the searcher moves in this space, he attempts to approach as nearly as possible the recall, precision, and cost goals of the end user.

Some formulations are clearly better than others in the sense that better recall can be obtained for equivalent levels of precision and cost, or that higher precision can be achieved at an equivalent level of recall and cost.

One can hypothesize the existence of what might be called "the most rational path" (Harter, 1986, p. 199) through the infinity of search formulations derived from the various heuristics that might be employed. Visualize the searcher moving through this three-dimensional space. The most rational path is a three-dimensional surface which maximizes precision for given levels of recall and cost. Actions taken by an expert searcher will presumably approximate movements along this surface while actions taken by a novice may more closely resemble random movement through the space.

If cost is kept fixed or ignored, the three-dimensional surface becomes a two-dimensional plane, and the most rational path can be defined as the set of formulations that maximize recall for a given level of precision. The selection and operationalization of a heuristic is then simplified but potentially is still enormously complex.

It should be emphasized that these ideas are hypothetical and have not been subjected to empirical testing. They represent one view of the difference between an "expert" searcher and a nonexpert. However, little is known about how experts go about selecting which heuristics to pursue in a given search, or how to operationalize them when they have been selected. Indeed, little is known regarding which heuristics tend to be superior to others in general (that is, on the average), or even if such an idea makes sense. The idea of a "most rational path" must therefore be regarded as hypothetical.

Clearly many search problems are so simple that the ideas presented in this article are not applicable. A high school or college student who wants a dozen or two nontechnical references on AIDS for a three-page paper probably does not even need an online search. But if one is conducted, the searcher will need to employ few if any heuristics. A "fast batch" approach—an approach using few if any heuristics—will probably work quite well. It is not known what proportion of searches in the real world fall into this "simple" category, although it is obvious that a great many of them can be characterized in this way.
IMPLICATIONS

Training programs for online searching tend to stress learning the retrieval language—that is, to learn the syntax rules and other rules for logging on, inputting commands to the search system, and so on. But it could be argued that rules are relatively easy to teach and at the same time are less significant. (A significant difference between training and education may be that the former deals mainly with rules and the latter with heuristics.) Perhaps too much attention in online searching is paid to training concerns and too little is given to evaluative, conceptual, or artistic aspects of online information retrieval, especially the acquisition and effective use of heuristics. The diagnostic and tutorial software package Online Search Analyst was developed as an initial step in this direction.

End user search systems can be viewed in terms of the analysis just presented. Some end users search Knowledge Index and BRS After Dark systems that are now several years old. Others employ DIALOG, BRS, or other systems directly. Still others are beginning to search CD-ROM and videodisc systems physically located within the library.

By most accounts, end user searching is reported to be enormously successful in achieving user satisfaction. End users are reporting high levels of enthusiasm and success. How are these reports to be evaluated in light of the model of online searching as a problem-solving process, especially since end users typically are not only unfamiliar with search heuristics but also of the basic concepts and rules of online searching (syntax rules, Boolean logic, etc.). How is this possible if online searching is truly a problem-solving process as represented in this discussion?

One likely explanation of the reported successes of end user searching is that the recall goals of the individuals using these systems are not especially stringent—that goals of high precision are the norm. A second possibility is that the end user with comprehensive search goals lacks the knowledge to assess the extent to which recall goals are met. (Note that cost goals are typically not a consideration in end user searching supported by a library since connect time charges do not exist or are absorbed by the sponsoring library.)

If the end user only wants a few records, then good success can often be achieved by end users employing controlled vocabularies. In this case the end user could probably get by equally well with a print index. Alternatively, if a comprehensive search is wanted, then the end user is not likely to achieve success for a variety of reasons. This result should be distinguished from a result in which the end user thinks, erroneously, that he has achieved comprehensive retrieval results. An end user who is naïve about the extent of the literature on a subject may well conclude that he has found
“everything” on a topic when in fact only a small portion of the available literature has been retrieved. The long-term effects of such cases are bound to be negative.

Obvious questions that are raised are:

1. Should libraries provide instruction in rules and heuristics for end users? If so, what should be the content of such instruction?
2. What are the implications of providing a complex and sophisticated search system and making it publicly available but not offering instruction in it? Is this ethically defensible?

Finally, some comments are in order regarding the development of expert systems to do information retrieval. Such systems must, presumably, be able to employ heuristics just as human experts do, at least for complex or comprehensive problems. However, systems do not yet exist that can make such independent decisions as:

—The present file is too “popular” for this particular search.
—Boolean operators were used improperly.
—This search term is too generic for the information need.
—This term is too common for use in representing the information needed.
—This term suffers from multiple meaning—steps must be taken to disambiguate it.
—A set number was used improperly.
—A term was truncated improperly.
—Another facet must be created and intersected to increase the precision of the search.

By and large, these decisions must still be made by human beings. Future expert systems in online information retrieval may make such decisions based on ongoing information provided by the end user in an interactive mode. But to do this effectively for complex and comprehensive searches, system designers will need to know much more about how experts search—the thought processes leading to the choice of one heuristic over another, and how particular heuristics are best operationalized. As noted, the current state of knowledge regarding these questions is extremely limited.

Some work in this area has been done. For example, Raya Fidel (1984) has studied how the expert searcher selects terms—e.g., free text vocabulary, controlled vocabulary elements, etc.—by analyzing the online searching behavior of several human intermediaries. Fidel identified several heuristics (she calls them “options”) for selecting search terms, and reduced them to a set of rules formalized into a decision tree. She is
ONLINE SEARCHING

optimistic about the design of expert systems for online searching through the study of the searching behavior of expert human intermediaries.

Although this author shares Fidel's view of the value that could result from a careful study of human experts, there is no optimism regarding the eventual design of an online searching machine that will do as well as or better than human expert intermediaries. Proponents of mechanical translation of languages in the 1950s and 1960s also felt that rules could be proposed that would eventually result in algorithms that would prepare expert translations of texts from one language—e.g., Russian—to another language—e.g., English. The goal of this research was to develop expert systems to do translation of languages. However, systems capable of producing expert translations of general texts without human intervention have never been developed.

Some of the problems inherent in doing online searching are the same problems that eventually caused extraordinary difficulties in creating computer programs to do mechanical translation of languages. These include: the importance of the problem context, the distinction between meaning and the representation of meaning in words and phrases as well as the computer's inability to recognize meaning as such, the existence of homographs in natural language, and the enormous complexity of language, especially at the semantic and pragmatic levels. These are extremely difficult problem areas. Only time will tell whether they are insurmountable.

In the meantime, educational and training programs for librarians as well as end users should be examined with the goal of evaluating the attention given to the study of online search heuristics. The study of heuristics rather than rules addresses the truly difficult and challenging aspects of online searching, and it is precisely here that training and educational programs for novice search specialists and end users, as well as commercial intermediary search systems, fall short.

REFERENCES


