# Towards Expert Systems for the Selection of Search Keys*

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

**Intermediary expert systems are designed to mediate between end-users and complex information retrieval systems. However, since most of these expert systems are based on text analysis rather than on models of human searching, they cannot process request-related criteria, such as precision or recall requirements. Analysis of the searching behavior of human intermediaries revealed a routine for the selection of search keys—free-text or controlled vocabulary—along a decision tree. Examples of decision rules demonstrate that although further research is required, these rules can be automated to significantly enhance the adaptability of intermediary expert systems.**

## Introduction

It is believed that end-users will very likely search their own requests online when search processes are simplified or made friendlier. The prevailing approach to providing easier and friendlier user-system communication is to develop "interface" or "intermediary" systems. Indeed, systems such as CITE [1] are already available for public access, and others, such as CONIT [2] or CANSEARCH [3], are being tested in experimental settings.

Through such systems, users are freed from encounters with many peculiarities in databases and search systems, and yet can benefit from a large range of capabilities. In particular, users can enter a request in a loosely structured format, preferably in a natural language, sentence-like expression. An intermediary system processes the request terms, displays information to users, and asks for some sort of feedback. The information displayed may be in the form of a list of subject areas, databases, search keys (i.e., strings of characters that are entered to be searched for occurrence in pre-defined fields), or actual citations from which users are asked to make a selection, possibly in ranked order. Interactions of this nature usually proceed until users terminate the session.

Some intermediary systems are actually helper systems: they support shortcuts in end-user training by providing menu-driven interaction or online help programs. Typically, helper systems require end-users to make most of the decisions during a search process, or they drastically simplify searching by reducing the number of options to a minimum.

Intermediary expert systems, on the other hand, can be quite powerful. Expert systems replicate the performance of an expert in a particular area by incorporating the knowledge of an expert with rules for making inferences on the basis of this knowledge. Systems knowledge may or may not be intended to model actual processes as they would be performed by human experts.

This article illustrates how to model searching behavior of human intermediaries by demonstrating that formal rules for the selection of search keys can be extracted from human experts. These rules cannot be incorporated yet into a knowledge base of an intermediary expert system because they are incomplete and were derived from searching behavior that is limited in its subject area. The work presented here, however, clearly indicates that with more research a complete set of rules can be established. It also provides guidance for future exploration and points to various issues that could be readily considered by designers of intermediary expert systems.

## Modeling the Selection of Search Keys

Search processes consist of three basic intellectual components: (1) definition of query structure; (2) selection of search keys; and (3) evaluation of feedback. We focus here on the second component. In a database that

*This study was supported by the Graduate School of the University of Washington.

Received March 1, 1985; revised May 15, 1985; accepted June 13, 1985

 CCC 0002-8231/86/010037-08$04.00

offers the capability, an intermediary system must examine each term of a request and consider its representation: as a controlled vocabulary key, as a free-text key, or as both. Expert systems vary in the degree of freedom they provide users in this selection: some dutifully search only those search keys designated by a user; some use search keys designated by a user to generate additional search keys; and some automatically generate search keys without user participation.

Intermediary expert systems use mapping algorithms to generate search keys. Whether user support and involvement are high or low, a term may be mapped to a descriptor (a search key from a controlled vocabulary) through an exact or other kinds of match, or it may not be mapped to a descriptor at all. While some systems, such as CANSEARCH [3], use subject-specific approaches, most existing intermediary systems, expert and helper, use mapping algorithms that are based on text characteristics, such as word-occurrence frequency or statistical associations.

The most apparent drawback of text analysis algorithms is their lack of sensitivity to request-specific requirements that cannot be directly deduced from a request statement. Consider, for example, a request about "the attitudes of anorexic students toward themselves during examinations periods." One user is interested only in anorexic students, another wants to get all the information available on the topic but is primarily interested in students and is willing to look at material about student behavior during examinations periods. The system decides, say, to use *anorexic students* as a single search key, but may or may not suggest the term *students* as a search key, following its own algorithm.

However, when experienced online searchers select search keys, they examine not only the degree of term-descriptor match, but they also consider other factors. Moreover, these additional factors are quite frequently essential to the success of a search. One wonders why the experience of human intermediaries has been neglected by system designers when it is an important source of knowledge.

There may be two explanations. First, online searching behavior was not being investigated and thus could not contribute knowledge. Second, research and development in information storage and retrieval is familiar with text analysis because of the long and established experience in automated indexing. Although methods and approaches to automated indexing vary greatly, most of them rely on text analysis and are thus text-oriented rather then user-oriented. Only recently, in an experiment at the American Petroleum Institute, a first attempt has been made to develop an automated indexing system which models indexing behavior [4].

As a first step toward knowledge engineering in intermediary expert systems, I analyzed the online searching behavior of several human intermediaries and found that online searchers do indeed follow some rules [5]. In their selection of search keys, they utilize informal and sometimes highly intuitive decision rules. Moreover, these rules can be detected, examined, and presented in a formal structure which can be processed by computers. This formal presentation incorporates knowledge of multiple experts and with further research can be used in a knowledge base for intermediary expert systems.

### *The Study Method*

To examine online searching behavior, eight searchers were observed performing their regular, job-related searching [6]. Searchers who have been searching for more than two years were recruited from among information specialists in scientific areas, primarily in the life sciences. They were studied one at a time, and were asked to verbalize their thought processes during their searching to the degree that speaking out loud would not interfere with their performance. These verbalizations, including the creation of search strategy, were recorded. At the end of the observation period (approximately 10 to 15 searches), each searcher was interviewed to reveal and clarify information not accessible to observation.

Data collected for analysis were about one hundred printed search protocols with transcriptions of verbalized thought processes and additional explanations from the final interviews.

Each instance in which searchers had selected a search key was then identified and the reason for the specific choice was explicitly noted. Analysis of the first ten search protocols generated a preliminary list of conditions under which a particular selection was made. For example, a condition for choosing a free-text key is: to enter straightforwardly a specific concept which might not be a trustworthy descriptor.

All the search protocols were then analyzed against this preliminary list of conditions. Each instance in which searchers had selected a search key was listed under the condition to which it applied. Instances whose condition could not be found suggested a new condition to be added to the list. This analysis revealed that most conditions were considered by most searchers, and only a few combinations reflect searchers' individual idiosyncrasies.

The list of conditions for the selection of search keys is presented in Figure 1 in the form of a decision tree. This set of decision rules is called here "the selection routine."

The selection routine specifies conditions which are *necessary* for a searcher to be able to select a particular type of search key. It describes the most commonly selected path, but there might be complications. As such, the selection routine is not deterministic: it cannot always accurately predict the selection of search keys unless other factors and their impact are known. This routine groups together similar conditions so it could be presented in a decision tree, but it is not meant to represent a necessary sequence in the selection process.

A list of the options in search key selection and the conditions necessary for each option is given in Table 1.

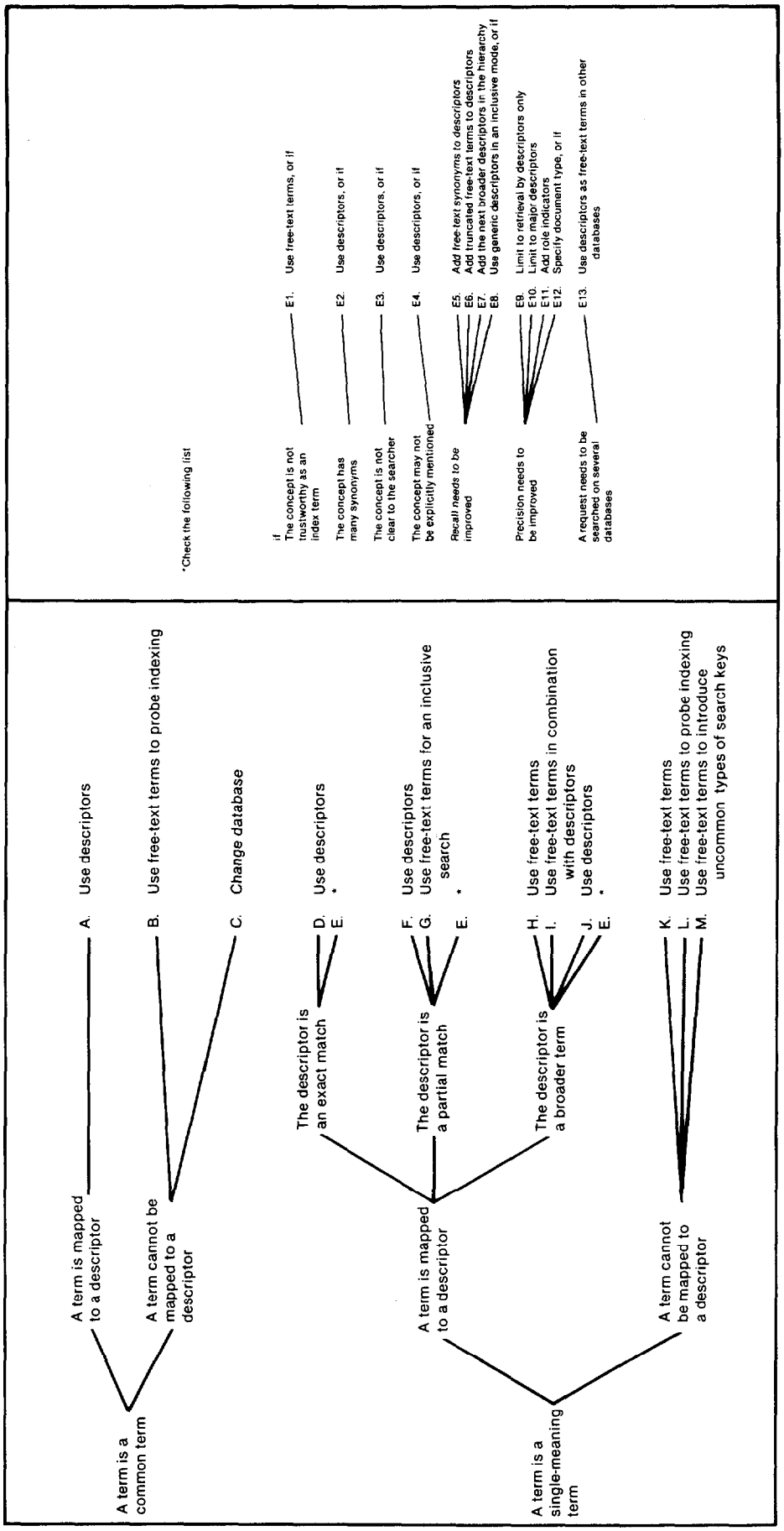


FIG. 1. The Selection Routine.

TABLE 1. A List of Options and the Associated Conditions.

| OPTION | CONDITIONS |
|---|---|
| | Descriptor Searching |
| Use descriptors | A term is a common term +<br>it is mapped to a descriptor [A].<br>A term is a single meaning term +<br>it is mapped to a descriptor +<br>the descriptor is an exact match [D].<br>the concept has many synonyms [E2].<br>the concept is not clear to the searcher [E3].<br>the concept may not be explicitly mentioned [E4].<br>the descriptor is a partial match [F].<br>the descriptor is a broader term [J]. |
| Add the next broader descriptor in the hierarchy | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>recall needs to be improved [E7]. |
| Use generic descriptors in an inclusive mode | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>recall needs to be improved [E8]. |
| Limit to retrieval by descriptors | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>precision needs to be improved [E9]. |
| Limit to major descriptors | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>precision needs to be improved [E10]. |
| Specify document type | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>precision needs to be improved [E12]. |
| | Free-Text Searching |
| Use free-text terms | A term is a single-meaning +<br>it is mapped to a descriptor +<br>the concept is not "trustworthy" as an index term [E1].<br>the descriptor is a broader term [H].<br>it cannot be mapped to a descriptor [K]. |
| Use free-text terms to probe indexing | A term is a common term +<br>it cannot be mapped to a descriptor [B].<br>A term is a single-meaning term +<br>it cannot be mapped to a descriptor [L]. |
| Use descriptors as free-text terms in other databases | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>a request needs to be searched on several databases [E13]. |
| Use free-text terms for an inclusive search | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>the descriptor is a partial match [G]. |
| Use free-text terms to introduce uncommon types of search keys | A term is a single-meaning term +<br>it cannot be mapped to a descriptor [M]. |
| | Other Combinations |
| Use free-text terms in combination with descriptors | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>the descriptor is a broader term [I]. |
| Add free-text synonyms to descriptors | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>recall needs to be improved [E5]. |
| Add truncated free-text terms to descriptors | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>recall needs to be improved [E6]. |
| Add role indicators | A term is a single-meaning term +<br>it is mapped to a descriptor +<br>precision needs to be improved [E11]. |
| Change database | A term is a common term +<br>it cannot be mapped to a descriptor [C]. |

## The Selection Routine

Before searchers decide how to represent a request term in a query formulation they must answer two central questions: (a) can a term be mapped to a descriptor, and (b) is it a "good" term for free-text retrieval. A searcher maps a term to a descriptor when she/he has decided that a particular descriptor best represents a request term, whether or not there is an exact match between the term and the descriptor.

The second question is a little more complex and requires some explanation. Searchers consider a term to be a "good" term for free-text searching if it: (1) usually occurs in a particular context, (2) is uniquely defined, and (3) is specific in the concept it represents. Such a term will be called here a "single-meaning" term. On the other hand, a term that occurs in *more* than one context will be called a "common" term. For example, in the request about the attitude of anorexic students toward themselves during examinations periods, terms such as *anorexia* and *students* are single-meaning terms. The term *examination,* on the other hand, is a common term; it can occur in a subject-related context ("the best way to take student examinations"), or in a descriptive capacity ("examination of students' responses"), or still further, it can be used very loosely to represent the concept of an inquiry of any kind.

When a term is a common term, searchers do not have much choice in the selection of search keys: if it can be mapped to a descriptor, searchers almost always enter the descriptor as the search key [A] (i.e., option [A] in Figure 1. and Table 1.) because, by definition, it is not desirable to use a common term as a free-text search key.

A common term that cannot be mapped to a descriptor almost always results in unsatisfactory retrieval. Searchers have no choice but to enter a free-text key, preferably in combination with other search keys, in order to retrieve citations, select some relevant ones and review their indexing in an attempt to find descriptors that might possibly be relevant [B]. For example, if the term *examination* cannot be mapped to a descriptor, one can devise a formulation (using the AND operator) that combines the terms *students, anorexia,* and the free-text term *examination.* Reviewing a sample of retrieved citations, one may find that all the relevant citations include the descriptor *Instructional Tests* in their indexing, thus suggesting that this descriptor is an appropriate choice for the representation of the concept *examinations.* Such probing does not always further a search and searchers may then decide to select a different database: one which does allow the common term to be mapped to a descriptor [C].

Single-meaning terms provide more options. If a single-meaning term cannot be mapped to a descriptor, searchers may enter a free-text term as the only search key [K], but they may also probe indexing of relevant citations to make sure that no adequate descriptor is overlooked [L].

In some cases, searchers may use a free-text key to search for a single-meaning term that cannot be mapped to a descriptor in a special way: they require that it occurs in a field other than the common ones, such as the journal title field [M]. Suppose a user is interested only in the psychological aspects of anorexic students, and suppose that the term *psychology* cannot be mapped to a descriptor. Searchers may predict that searching for the occurrence of *psychology* in the text would retrieve a large number of irrelevant citations, and decide instead to retrieve citations to articles whose authors are affiliated with organizations which include the stem *psych* in their titles, or articles that were published in sources whose titles include this stem.

The least problematic term is one that is single-meaning and also can be mapped to a descriptor. Such a term can be entered either as a controlled vocabulary, as a free-text key, or as both. Here, searchers are free to deal with other factors when selecting search keys. It is useful to show the conditions under which searchers select free-text keys, and those under which they choose to use descriptors.

*Selection of Free-Text Search Keys*

A single-meaning term can be mapped to a descriptor through an exact match, partial match, or a term might be mapped to a broader descriptor. When a single-meaning term is mapped to a descriptor through an exact match, searchers may use a descriptor [D], or elect to consider a variety of factors [E].

Partial match usually implies mapping a request term to a narrower descriptor, or to a group of narrower descriptors. If suitable, searchers use a free-text key to inclusively search concepts that are not grouped together by the hierarchy of the controlled vocabulary [G]. If, for example, the request term *students* is mapped to descriptors such as *Foreign Students, College Students,* or *Undergraduates,* and a descriptor *Students* does not exist, the free-text key can be used to retrieve information about any type of student. It should be noted that in some search systems use of the free-text key *students* also would retrieve citations that are indexed with descriptors which include the term. This is a source for constant confusion for searchers because the routine changes from one search system to another.

When a single-meaning request term is mapped to a broader descriptor, searchers may prefer to preserve the specificity of the request and use free-text search keys [H]. If they are concerned with the precision of the set to be retrieved, they enter free-text terms in combination (using the AND operator) with the broader descriptor to which it is mapped [I].

Regardless of the degree of match between a single-meaning term and a descriptor, searchers may still prefer to use free-text search keys for three reasons. First, if recall needs to be improved, searchers use both descriptors and free-text terms as search keys [E5]. For a further increase in recall, free-text keys are entered in a truncated form [E6]. Second, if searchers think that a particular descriptor may be assigned inconsistently by indexers, they consider the use of a free-text key to be more trustworthy [E1]. Suppose, for instance, a controlled vocabulary includes the descriptor *Nutrition* and also the descriptor *Diet*—each one to be assigned for distinct representation of a subject. When looking for nutrition-related problems of anorexic students, searchers may find the distinction confusing and thus assume that indexers are likely to be inconsistent in assigning these descriptors. To compensate for indexers' errors, they may use both a descriptor and free-text keys, or only free-text keys.

Lastly, if a request is to be searched on several databases, searchers may map single-meaning terms to descriptors in only a few of the relevant controlled vocabularies. Running the same query formulation against several databases, they then, in fact, search some of the terms as free-text keys in some of the databases [E13].

*Selection of Controlled Vocabulary Search Keys*

The most straightforward use of a descriptor to represent a single-meaning term is when a term is exactly matched with a descriptor and no other apparent constraints exist. However, searchers may elect to enter a request term as a descriptor when it is mapped to a descriptor through a partial match [F], in which case it is mapped to a narrower descriptor, or when the term is mapped to a broader descriptor [J]. Such decisions depend on the nature of the request, and when searchers suspect that precision might not be satisfactory, they may combine free-text terms with a descriptor [I].

Even single-meaning terms may have some attributes that will make them unattractive for free-text searching. Regardless of the degree or nature of a term-descriptor match, searchers most often prefer to enter a descriptor when: (1) a term has many synonyms [E2]; (2) a concept and its use is not clear to a searcher [E3]; or (3) when a concept is likely to be implied rather then explicitly mentioned in the searched text [E4].

The previously mentioned request about anorexic students provides a clear example of the last condition. The request concept *attitudes toward themselves* can indeed be entered and searched as a free-text phrase in most search systems. However, this concept can be expressed, directly or indirectly, in various other phrases, such as "students displayed attacks of self-hatred", or "narcissism level dropped with time." Searchers, then, consider descriptors such as *Self Image,* or *Self Esteem* to be more reliable than free-text terms for searching.

In addition to providing search keys for "problematic" terms, controlled vocabularies provide special means to improve precision and recall. When searchers perceive precision and/or recall to be unsatisfactory, they may decide to take advantage of these means and elect to use a descriptor to represent a single-meaning term. Although

these routines are quite well known, it might be helpful to mention them. When searchers decide that recall needs to be improved, they select a controlled vocabulary key because they can add the next broader descriptor in the hierarchy [E7], or use generic concepts in an inclusive mode [E8]. Controlled vocabularies readily suggest broader descriptors, and thus make it convenient to indeed broaden a concept. Moreover, broadening the meaning of a concept is not always possible in free-text searching since the broader concept may be a common term. Inclusive searching—which is quite straightforward in descriptor searching—facilitates retrieval of citations indexed under a descriptor as well as those indexed under its narrower terms.

Lastly, when searchers predict that precision may not be satisfactory, they may select a controlled vocabulary key because they can exercise various ways of adding weights to search keys such as: to limit to retrieval by descriptor only [E9]; to limit a descriptor to be a major descriptor [E10]; to add role indicators [E11]; or to specify document type [E12].

## Discussion

The selection routine clearly shows that the process of selecting search keys as performed by online searchers can be formalized into a decision tree. Moreover, several suggestions for improvements in existing systems are immediately apparent; others will require more research.

First, the pattern of the selection routine illustrates the significance of decisions made during search key selection to the success of a search. This pattern shows that when a term is "not adequate" for searching, i.e., it is a common term and/or it cannot be mapped to a descriptor, only a few options are available for searching. Only six of the twenty-five options in the selection routine are suggested for such terms, and some, such as the use of free-text terms to probe indexing, require a fair amount of creativity on the part of an intermediary. On the other hand, when a term is "good" for searching, i.e., it is a single-meaning term and it can be mapped to a descriptor, intermediaries can look at several options, as presented in the check-list. In other words, only after terminological difficulties or peculiarities in representing request terms have been overcome for searching, can an intermediary consider additional factors that are essential to the success of the search. Therefore, intermediary expert systems must be able to resolve terminological difficulties before they can be equipped to deal with additional factors.

Second, using the selection routine, one can identify flaws in existing intermediary expert systems and at the same time propose methods to overcome these failings. While some flaws can be readily identified and possible remedies suggested already at this time, other issues require additional research before their nature can be clearly defined. To demonstrate the ability of the selection routine to illuminate such issues, a few examples are discussed below.

One of the flaws in existing intermediary expert systems that readily stands out is their inability to distinguish between common and single-meaning terms. This distinction is important because if a term is a common term, experienced searchers almost always select it as a descriptor even if they have to change databases (unless they use it in a trial [B]). Yet, to my knowledge, no existing system, provides safeguards to advise end-users against the use of common terms as free-text search keys. For example, when asked about *"information retrieval,"* CITE suggested *Information Systems, Information Services, Information Theory,* and *Information Centers,* as descriptors, and *retrieval, retrieving,* and *information* as free-text search keys [1]. Experienced searchers normally will avoid using these common terms as free-text keys, though end-users may prefer them because none of the descriptors exactly matches the original concept.

The idea that some terms are not suitable for automated processing is not new. In linguistics, homonymy and polysemy are specific cases of common terms which might be better described as ambiguous terms or terms that belong to more than one semantic domain. These concepts are essential to thesaurus construction. From the information science field, Fugmann, for example, differentiates between "individual" and "general" concepts, the latter being non-lexical concepts which are better searched with controlled vocabulary [7]. In addition, the assumption that some terms carry more information than others is a fundamental premise in automated indexing and abstracting.

Control over common terms should require relatively modest modifications in intermediary expert systems. First, we have to devise a working definition of what constitutes a common term. For this purpose, it would be useful to test the hypothesis that terms which occur with high frequency in a database are also common terms. Suppose, however, that a system selects to define a common term by, say, a consensus among three knowledgeable searchers who are highly experienced with a database. These searchers can then check each term in a thesaurus, including those that represent an entry or part of an entry and those that occur in a descriptor or in a lead-in entry, and determine which terms are common. Common terms can then be coded so they are not displayed or used as free-text search keys.

This method requires additional effort to identify common terms that do not occur in a thesaurus or in other semantic networks. Some shortcuts can be devised, however, such as the use of a number of thesauri. On the other hand, other methods to identify common terms may apply to all terms in a database whether or not they are listed in a thesaurus. For instance, a test can be conducted to discover whether a correlation exists between the frequency in which a word occurs in a database and its adequacy for free-text searching. If a well-selected

sample of databases proves that common terms occur with high frequency in those databases and vice versa, then common terms can be singled out by frequency counts.

Once a common term is defined, an intermediary system can interact with users. Suppose a common term cannot be mapped to a descriptor. By giving messages to users, a system can interrogate them to determine whether to select another database or whether to enter the term as a free-text search key to probe indexing. This interaction may reveal request-related requirements that are not reflected in a request statement, e.g., that the term is central to the request or that it always should be associated with another term.

A second example of a flaw in existing intermediary expert systems is their failure to suggest the use of free-text terms for inclusive search [G]. As explained earlier, if the descriptor *Students* does not exist and a term is then mapped to descriptors such as *University Students* or *Gifted Students* through partial match, the free-text term *students* can be entered to search for any kind of student. This function can be easily automated. However, one should be careful because some terms are not suitable to be entered as free-text keys for inclusive searching.

The term *attitudes* is a case in point. If an exact match does not exist, the term *attitudes* can be mapped through a partial match to descriptors such as *Employee Attitudes, Mother Attitudes,* or *Negative Attitudes.* In searching for material about attitudes of students toward themselves, a user may find the last descriptor relevant, but the first two are a source for unwanted retrieval. To search the term *attitude* as a free-text key would magnify the problem. In this case, the user is better advised to scan all the descriptors to which the term is mapped and to select the relevant ones.

Thus, we can designate for each word in a multi-words descriptor whether it can be automatically searched as a free-text key or whether it should be displayed to users when a partial match occurs. At this time, we do not have any scale based on systematic investigations that can determine which terms are suitable for inclusive searching in a free-text mode. We can, however, adapt a pragmatic approach and determine the status of each term by consulting with experienced online searchers. In the future, research in terminology may provide more rigorous criteria.

A third example is the inadequacy of intermediary expert systems for term analysis processes. Various statistical approaches could be used to determine which attributes of terms are significant for searching. For instance, the "trustworthiness" of a descriptor can be measured by the degree of consistency with which it is assigned. An extrapolation of the measurement of term consistency suggested by King & Bryant [8] could be used to measure trustworthiness of descriptors. A test database could be indexed by several indexers. A measurement for trustworthiness could then be determined by the ratio between the number of documents to which a descriptor has been assigned by all indexers and the total number of documents to which it has been assigned by any number of indexers (possibly weighted by the number of indexers selecting to assign a descriptor for each document). Here again, each descriptor that is not trustworthy can be flagged. During the search process, a system can then use free-text terms whenever it encounters a single-meaning term which is mapped to such a descriptor. The system may also convey its action to users.

As these examples show, parts of the selection routine can be automated quite easily and with existing techniques. Other parts of the selection routine, such as, when searchers use a descriptor for a single-meaning term because they do not fully understand the concept it represents, may prove to be unsuitable for the design of intermediary expert systems.

There are yet further decisions which can usefully be implemented after additional research is performed. Consider a situation in which a single-meaning term is mapped to a broader descriptor. There are three main options as shown in the decision tree (Figure 1.) at the points [H], [I], and [J]. Now, suppose the term *anorexia nervosa* is mapped to the descriptor *Appetite Disorders.* An intermediary system may decide to enter *anorexia nervosa* as a free-text search key and possibly retrieve documents in which the subject is only mentioned, rather then discussed. Or, it can combine this free-text key with the descriptor *Appetite Disorders,* using the AND operator, to retrieve articles that indeed discuss the disorder but may miss relevant documents that were not indexed under the broader descriptor. Or, the system can select to enter the descriptor, in which case relevant documents might still be missed while documents discussing appetite disorders other then anorexia nervosa probably will be retrieved.

No option is better than any other; it all depends on the nature of a request. In other words, one option is required for one type of request and another is required for another type. We do not have yet a general typology of requests that we can use to support the selection of the best option. Further research in online searching behavior, however, can provide criteria to be used in automated systems.

Statistical approaches may suggest some help. For example, one may statistically analyze user satisfaction rate with each of the options. Thus, even though we do not know explicitly which type of request requires which option, we can implicitly detect which type is most common among a particular group of users by the option they find to be most satisfactory. We can then design an intermediary expert system that first will always try the most commonly satisfactory option, then ask for user's reaction, and then utilize the next option if the first failed to produce acceptable results.

A much more promising approach suggests that online

searching behavior be investigated to reveal under which conditions each of the options is selected. We may find, for example, that when a term that is not central to a request is mapped to a broader descriptor—and a user is primarily concerned with precision—searchers decide to combine a free-text search key with a descriptor. Transferring this condition to automated systems will enable a system to decide when and how to interrogate users about request requirements that are relevant to the search process. More specifically, a system may proceed searching independently until it encounters a problematic term, such as one that is mapped to a broader descriptor, and then ask users for a specific kind of feedback. In summary, it is not difficult to envision an intermediary expert system which would: (1) identify situations in which a request statement is sufficient, and conversely those in which additional request criteria are needed for the search process to succeed, (2) list the relevant criteria so that users can provide the pertinent data, and (3) act upon data received to improve search results. Quite a powerful system!

These few examples clearly demonstrate the benefits that could be gained from the study of searching behavior of human intermediaries, and from utilizing the experience of human intermediaries in the design of intermediary expert systems.

The selection routine presented here is not sufficient to develop adaptive algorithms. Many issues, such as the nature of single-meaning and common terms or the conditions for the selection of a broader descriptor, need to be further investigated and rigorously defined. This routine identifies problematic points in the search process and provides guidelines for research into searching behavior that is relevant for automated systems. On the one hand, systems can interrogate users on request parameters—a subject users know best. On the other hand, they can select the most appropriate search keys—a decision casual end-users are not well enough informed to make. Based on searchers' experience, intermediary expert systems can become experts indeed.

## Acknowledgment

The author wishes to thank Dagobert Soergel and Irene L. Travis for their helpful and insightful review of this article.

## References

1. Doszkocs, T. E. "Automatic vocabulary mapping in online searching." *International Classification.* 10(2):78-83; 1983.
2. Marcus, R. S. "An experimental comparison of the effectiveness of computers and humans as search intermediaries." *Journal of the American Society for Information Science.* 34(6):381-404; 1983.
3. Pollitt, A. S. "A 'front-end' system: An expert system as an online search intermediary." *Aslib Proceedings.* 36(5):229-234; 1984.
4. Brenner, E. H.; Lucey, J. H.; Martinez, C. L.; Meleka, A. 1984. "API's machine aided indexing project." *Science Technology Libraries.* 5(1):49-62; 1984.
5. Fidel, R. "Online searching styles: A case-study-based model of searching behavior." *Journal of the American Society for Information Science.* 35(4):211-221; 1984.
6. Fidel, R. "The case study method: A case study." *Library and Information Science Research.* 6(3):273-288; 1984.
7. Fugmann, R. "The complementarity of natural language and indexing languages." *International Classification.* 9(3):140-144; 1982.
8. King, D. W.; Bryant, E. C. *The evaluation of information services and products.* Washington, D.C.: Information Resources Press; 1971, p. 138.