Knowledge-Based Search Tactics for an Intelligent Intermediary System

PHILIP J. SMITH, STEVEN J. SHUTE, and DEB GALDES
The Ohio State University
and
MARK H. CHIGNELL
University of Southern California

Research on the nature of knowledge-based systems for bibliographic information retrieval is summarized. Knowledge-based search tactics are then considered in terms of their role in the functioning of a semantically based search system for bibliographic information retrieval, EP-X. This system uses such tactics to actively assist users in defining or refining their topics of interest. It does so by applying these tactics to a knowledge base describing topics in a particular domain and to a database describing the contents of individual documents in terms of these topics. This paper, then, focuses on the two central concepts behind EP-X: semantically based search and knowledge-based search tactics.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*search process*; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*frames and scripts*

General Terms: Human factors

Additional Key Words and Phrases: Bibliographic information retrieval, document retrieval, knowledge-based search tactics, knowledge-based systems, semantically based search

1. INTRODUCTION

In searching bibliographic databases, information seekers often need help in defining or refining their topics of interest [22, 28]. Some searchers need to learn more about the topic area in order to decide exactly what it is they are interested in. Other searchers have a clear idea of their interests but have difficulty expressing this interest (even in natural language to a search intermediary). Still

This work has been supported by contracts from the Chemical Abstracts Service and the Applied Information Technologies Research Center.
Authors' addresses: P. J. Smith, S. J. Shute, and D. Galdes, Department of Industrial and Systems Engineering, Ohio State University, 210 Baker, 1971 Neil Ave., Columbus, OH 43210; M. H. Chignell, Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90007.

others begin with a clearly defined and stated topic but find that it is far too broad (or too narrow), retrieving too many (or too few) document surrogates to be acceptable. They must, therefore, find a way to refine their topic.

Over the past five years, we have been studying ways to assist such information seekers [19–21; 37–42]. In particular, we have been addressing two questions:

(1) How do domain-knowledgeable search intermediaries assist information seekers in defining or refining their topics of interest?
(2) Can a knowledge-based system provide similar forms of assistance?

Such a knowledge-based system could provide valuable assistance to the human search intermediary who is not an expert in the domain of interest to an information requester. It could also move us a step closer to the goal of supporting effective direct end-user searching. In addition, the concepts identified as part of the effort to construct a computerized intermediary could prove valuable in the education of human intermediaries [8].

1.1 Current Search Methods

As discussed above, we have been studying expertise in searching existing bibliographic databases. Hence, it is important to understand the nature of the search task as it currently exists.

The purpose of bibliographic databases is to assist information seekers in identifying published documents that are pertinent to some topic of interest. Such databases, therefore, contain document surrogates describing the contents of individual documents. Each surrogate typically includes a listing of the document author, title, and date of publication, and an abstract and a list of descriptive terms, as shown, for example, in Figure 1. These descriptive terms have been assigned by a human indexer who has read the full document. Below is an example of the contents of such a document surrogate from the *Chemical Abstracts*.

To retrieve document information such as this from a bibliographic database, the information seeker must create a query that consists of keywords connected by logical operators [34]. (Searches based on an author or title can, of course, also be conducted.) Logical operators such as AND, OR, and NOT can be used. Positional operators requiring a certain proximity are also available. (The query RADIOACTIVE (W) WASTES retrieves only documents containing the character string RADIOACTIVE immediately prior to or with the character string WASTES.) Truncation and wild card operators also are often available. For example, the query

TREATMENT AND (MANGANESE-54 OR 13967-71-0) AND WASTEWATER#

would retrieve all document surrogates containing the character string TREATMENT, the character string WASTEWATER followed by a single letter or a blank, and either the character string MANGANESE-54 or the string 13967-71-0.

For most information seekers, constructing effective queries can be a task difficult enough to require assistance. This difficulty arises in part because searching such databases is a prediction task.

```
AN  CA110(12):103729p
TI  Treatment of radioactive waste wash water from nuclear
    installations
AU  Schmack, Peter:  Hiller, Ludmilla
CS  VEB Bergmann-Borsig
LO  Ger. Dem. Rep.
PI  Ger. (East) DD 260201 A3,21 Sep 1988, 4 pp.
AI  Appl. 292297, 9 Jul 1986
CL  C21F9/10A
SC  71-11 (Nuclear Technology)
DT  P
CO  GEXXA8
PY  1988
LA  Ger
AB  In the treatment of radioactive wastewater contg. metal nuclides by
    a 2-step flocking pptn. process with the addn. of an acid, esp. H2S04,
    the pH is set (3 after the lst pptn. step and the wastewater is mixed
    with an Fe(III) salt soln., preferably FeC13 or Fe2(S04)3;  an Fe
    hexacyanoferrate suspension, prepd. from the same Fe salt soln. and
    a K4Fe(CN)6 son., is added to the treated wastewater to sep. Cs
    nuclides. The clear water pumped off is mixed with NaOH (to pH 10-
    12) in a 2nd reaction vessel to ppt. excess Fe salts and the residual
    nuclides.
KW  Radioactive wastewater treatment;  metal nuclide pptn. radioactive

    wastewater
IT  Surfactants
        (anionic, removal of, from radioactive wastewater)
IT  Radioactive waste
        (wastewaters, metal nuclide pptn. from, by iron salts)
IT  7705-08-0, Ferric chloride, properties 10028-22-5, Ferric sulfate
    14038-43-8
        (pptn. by, of metal nuclides from radioactive wastewater)
IT  10045-97-3, Cesium-137, properties 10198-40-0, Cobalt-60,
    properties 13966-31-9, Manganese-54, properties 13967-70-9,
    Cesium-134, properties 13967-71-0, Zirconium-95, properties
    13967-76-5, Niobium-95, properties 13981-38-9, Cobalt-58,
    properties 14391-76-5, properties
        (pptn. of, from radioactive wastewater)
IT  1314-56-3, Phosphorus pentoxide, uses and miscellaneous
        (removal of, from radioactive wastewater)
```

Fig. 1. An example of the contents of a document surrogate from the *Chemical Abstracts*.

First, the searcher must predict how the semantic content of relevant documents will be described in the document surrogates, and then, predict the specific language that will be used to represent the intended meaning. The searcher must generate the appropriate set of synonyms and cognates to capture documents relevant to specific concepts (e.g., POLLUT? OR EFFLUENT# OR WASTE# ··· to retrieve documents relevant to pollution). The searcher must furthermore generate similar search strings for all specific cases of a concept (e.g., BIRD# OR ROBIN# OR PENGUIN# OR ··· to retrieve documents discussing birds).

Second, the searcher must decide how to capture the semantic relationships among concepts by combining the strings for individual concepts into a full query

[23]. The searcher must, for instance, decide whether to use AND rather than LINK or WITH to connect keywords representing different concepts.

Finally, this all assumes that the information seeker has a clear idea of what information is needed. Often, this is not the case, and as the search progresses, the "user's notion of what he requires becomes clearer to himself, and may shift in emphasis" [26, p. 4]. Thus, for these individuals, searching represents a learning task as well as a prediction task.

1.2 Search Expertise and Current Systems

As described above, most commercially available search systems require an exact match between the user's query and the character strings present in the retrieved document surrogate [4]. If, for instance, a user enters the query INFORMATION (W) RETRIEVAL AND (EXPERT (W) SYSTEM# OR KNOWLEDGE (W) BASED), that user will not retrieve a document with an abstract stating: "The quality of information retrieved from alternative designs for expert systems or knowledge-based systems is assessed."

Because these existing artifacts require an exact match, many discussions on search expertise focus on issues peculiar to the use of character string searches. In discussing the art of using logical operators, for instance, Oldroyd and Schroder state:

> In online searching, the positional logic capability has made it possible to combine terms in any order in a word, fragment, phrase, link, sentence or citation relationship. The advantages and disadvantages of the strategies that can be employed are reviewed. [27, p. 127]

They go on to discuss the effects of various operators on recall (a measure of the number of relevant documents retrieved) and precision (a measure of the number of irrelevant documents retrieved). As specific examples of techniques to use logical operators more effectively, Harter and Peters list heuristics like "Avoid using NOT in a search strategy: when using it, use it with care; ... Use Venn diagrams to help conceptualize the facets and search logic" [15, p. 417]. Fidel [13, p. 39] suggests rules such as "add truncated free-text terms" if recall needs to be improved.

Vigil suggests that there is more to search expertise than understanding the effects of using logical operators. He contends:

> It is not enough to merely explain to searchers the basic mechanics of searching, Boolean logic, truncation, feedback, etc., but additionally to provide specific plans which efficiently and systematically incorporate these basic mechanics and tactics. [46, p. 286]

Along these lines, Hawkins and Wagers discuss search strategies such as "building blocks, successive fractions, and citation pearl growing" [16, p. 12]. Such considerations lead to rules like "search the most specific aspect first" [16, p. 14].

Many of these search strategies call for viewing retrieved document surrogates to look for cues on how to modify the search strategy. As examples, Harter and Peters suggest heuristics like "Browse some retrieved records for relevancy, missed terms, flaws in strategy, etc.; ... When decreasing specificity by adding terms to facets, modify only one facet at a time" [15, pp. 417–418].

Finally, Bates suggests the use of "idea tactics" to improve searching. Included in her list of such tactics are heuristics such as

> BRAINSTORM Generate many ideas and suspend critical reactions until the ideas are well-formed and can be fully evaluated....
>
> CONSULT Ask a colleague for suggestions or information in dealing with a search. [3, p. 281].

Although it would be possible to build systems that simply suggest such domain-independent tactics or strategies to users when they need help, this has not been the trend. Instead, most computerized intermediary systems have been given substantial knowledge about specific domains. This domain knowledge is then used by the system in assisting the searcher to identify relevant documents.

1.3 Thesaurus Systems

One approach to assisting information seekers is to build in knowledge of synonyms and hierarchical relationships [10, 30, 47]. In its simplest form, such a system requires the user to step through the hierarchy in a top-down fashion. Thompson argues that the presentation of such a hierarchical structure is desirable "because it seems to replicate the structure of human thought processes most closely, thus allowing the simplest, most direct transfer of the man's problem into the structure and vocabulary of the system" [43, p. 372].

Menu-MEDLINE represents one application centered on built-in knowledge of hierarchies [29]. This system makes use of a "top-down stepwise refinement approach to query specification through user concept recognition and selection from menus" [29, p. 1]. Recognizing that searchers often are interested in the combination of multiple concepts (e.g., the PATHOLOGY of LEUKEMIA), Pollitt avoids the time-consuming requirement of multiple top-down searches through the hierarchy by allowing users to request a menu of the qualifiers associated with a particular hierarchy concept. Furthermore, the menu of qualifiers is pruned so that only those qualifiers relevant to a given hierarchy concept are presented.

Pollitt [28] has also described another system built around the hierarchical organization of concepts. This system, CANSEARCH, makes use of a menu hierarchy and rule-based programming techniques, allowing the searcher to step through the hierarchy and select items, and then applies rules to "carry out the task of generating search formulations with extensive concept coordination using both terms and subheadings" [28, p. 138].

Finally, EP-X (Environmental Pollution eXpert) [38, 41] illustrates another approach to providing access to hierarchically organized concepts. (EP-X actually contains more than simple thesaurus-like structures. It also explicitly represents topics. This is discussed later in this paper.) EP-X serves to illustrate several important concepts about thesaurus systems:

(1) The same set of concepts can often be organized into radically different hierarchies based on alternative organizing perspectives. Concepts like "coal mining," "energy production from coal-fired power plants," and "iron ore mining"

KEYWORD LIST

Your keyword list currently consists of the following: **TREATMENT** **RADIOACTIVE WASTES** **NUCLEAR POWER PLANTS**

Fig. 2. Entry of keyword phrases.

INTERPRETATION

62 documents are available on the removal/ treatment of radioactive wastes from nuclear power plants.

Fig. 3. Interpretation of keyword phrases.

have different relationships with each other depending upon whether they are organized in terms of the processes involved (e.g., mining) or the materials involved (e.g., coal). Smith, et al. [45] report that, depending upon the searcher's needs, human search intermediaries make use of different organizing perspectives. On the basis of such empirical data, the initial prototype for EP-X incorporated tangled hierarchies into its knowledge base;

(2) Information seekers may prefer direct entry to some lower level of a hierarchy instead of always moving through the hierarchy in a top-down fashion. On the basis of this notion, EP-X allows users to enter keyword phrases such as STRONTIUM-90 or HEAVY METALS, moving directly to the pertinent section of the hierarchy. Furthermore, consistent with the concept of a thesaurus, each such concept node can be accessed by any of the synonyms (or triggers) associated with it. Thus entry of SR-90 has the same effect as entry of STRONTIUM-90.

(3) Because users of EP-X directly input keyword phrases, the system must also monitor for the entry of ambiguous keyword phrases. When EP-X detects an ambiguous phrase like mercury, it asks whether the searcher is interested in "mercury as a pollutant" or "mercury in a vapor lamp to control biological wastes".

(4) Like Menu-MEDLINE, EP-X has been built with the understanding that users are often interested in the relationships among concepts, rather than in single concepts. Users can, therefore, enter multiple keyword phrases to express such interests as illustrated in Figures 2 and 3.

(5) Unlike Menu-MEDLINE, the hierarchies are not static in character. EP-X uses the full context of the user's entry to prune each hierarchy.

Thus, the SOURCES hierarchy displayed for a search on "SOURCES of ACID RAIN" looks very different from that displayed for a search on "SOURCES of RADIOACTIVE POLLUTANTS."

(6) EP-X "learns" or modifies its context-dependent hierarchies when new documents are indexed and entered into its database. Consider, for example, the effect of entering information on a document that identifies (for the first time) that an already known process such as ion exchange can be used to remove a pollutant such as lead. This entry results in the addition of ion exchange to the hierarchy of concepts indicating processes pertinent to the removal of lead.

As the above discussion illustrates, there are a number of issues that arise when an intermediary system is given thesaurus-type knowledge. Issues arise concerning organizing perspectives, means of access to hierarchy nodes (top-down versus direct activation), focusing of attention (pruning of hierarchies), and learning. Such computerized intermediaries, whether they should be considered "intelligent" thesaurus systems or simply useful thesaurus systems, clarify some of the issues that bona fide knowledge-based systems must deal with. Thesauri contain certain types of knowledge that must be dealt with in designing a knowledge-based system. The issues associated with other types of knowledge are discussed later.

1.4 Partial-Match Systems

Systems using queries that consist of character strings connected by logical operators or that use thesauri to identify appropriate controlled vocabulary terms have a problem: They can miss "many relevant texts whose representations match the query only partially" [4, p. 113]. Over the last three decades, there have been numerous efforts to deal with this problem using term-weighting schemes based on statistical word association [11, 14, 33]. The development of expert systems technology, however, has led to other approaches to coping with the potential relevance of partial matches. These alternatives are now discussed.

1.4.1 *Ruled-Based Systems.* Tong and Shapiro describe a rule-based system, RUBRIC, that "can represent partial relevance" [44, p. 265]. In RUBRIC, knowledge consists of "rule evaluation trees" represented as production rules. A partial tree is shown in Figure 4.

To determine the relevance of a document to a topic such as the "World Series," RUBRIC matches document terms to triggering phrases (leaf nodes) in the tree (e.g., "Cardinals" or "Baseball"). It then uses the associated weightings to compute a measure of the relevance of the document to the topic "World Series."

For the example tree given in Figure 4, Tong and Shapiro illustrate the calculation of relevance for a document containing the words "ball," "baseball," and "championship." The tree shows the leaf nodes representing those words that are present as having a value of 1.0, whereas leaf nodes for words that are not present have a value of 0. Values are then propagated up the tree as follows:

(1) The value of a node leading to a single path is equal to its value multiplied by its weighting. Thus, the node labeled championship has a value of 0.7 (1.0 times 0.7). (Unlabeled branches have a weighting of 1.0);

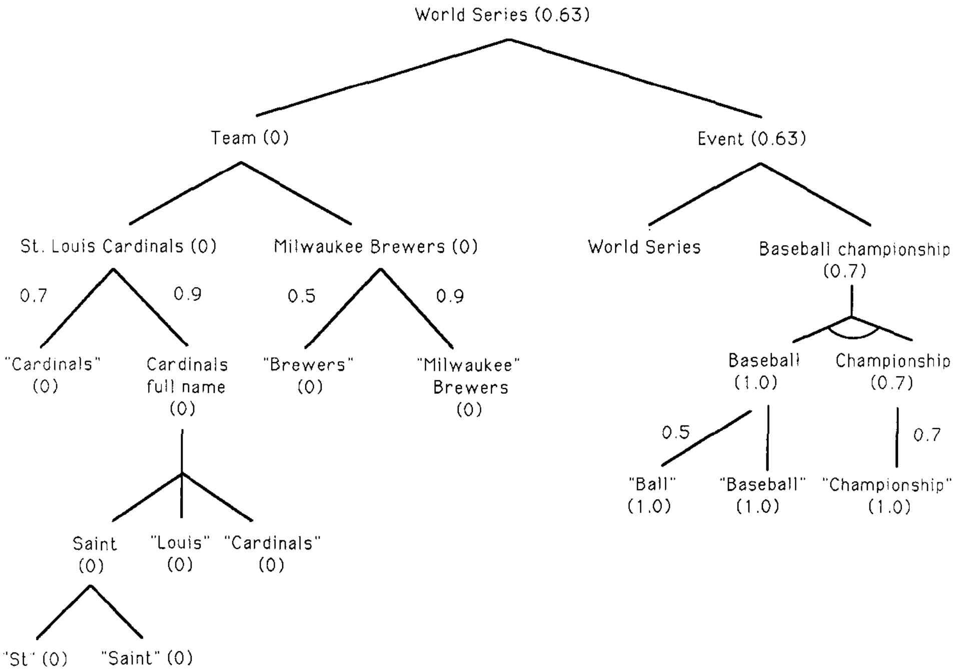

Fig. 4. Rule evaluation tree (after [48]).

(2) The value of a node leading to an AND fork (a set of branches connected by a fan) is equal to the minimum of the values of individual branches. Thus, the node labeled Baseball Championship has a value of 0.7 (the minimum of the values of Baseball (1.0) and Championship (0.7));

(3) The value of a node leading to an OR fork (a set of branches without a connecting fan) is equal to the maximum of the values of individual branches.

RUBRIC, then, deals explicitly with the notion that searching for relevant documents is a prediction task. The words contained in a document are predictors of its relevance to a particular concept or topic. Thus, when the searcher indicates an interest in a subject like the "World Series," RUBRIC can calculate the relevance scores for each of the documents in a database and then present them to the searcher rank ordered according to predicted relevance.

1.4.2 *Spreading Activation in Semantic Nets.* Other systems use different types of control processes to expand the search beyond exact matches to a user's query. In THOMAS, Oddy created a network of associations among concepts or subject terms. In this system "no information is held on the nature of the associations: it is sufficient for this program to know simply that two entities are associated" [26, p. 5]. This network of associations is used to search for related documents.

In THOMAS, the system plays a passive role in exploration. The user is presented with documents retrieved by the initial query, along with associated concepts or subject terms. The user then makes judgments of relevance and indicates which associated terms are important. This new input serves to define the subsequently retrieved set of documents.

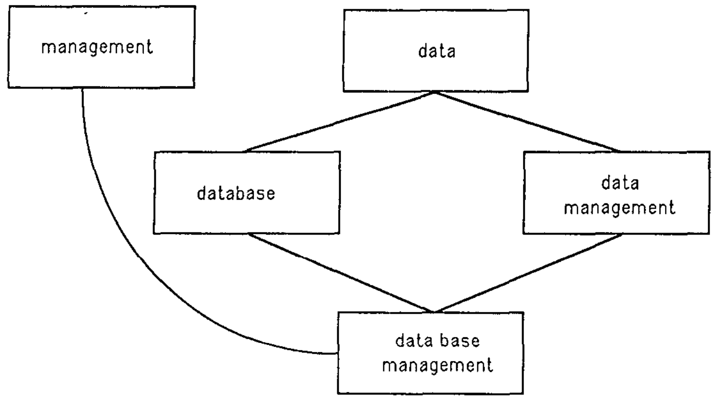

Fig. 5. Partial network of associated concepts (from [39]).

Shoval similarly proposes the use of an unlabeled set of associations among concepts. He proposes a system that is much more active in seeking related concepts and documents, however. Shoval describes a two-stage process based on spreading activation through the network:

> During the 'search' stage relevant knowledge in the semantic network is activated, and search and evaluation rules are applied in order to find appropriate vocabulary terms to represent the user's problem. During the 'suggest' stage those terms are further evaluated, dynamically rank-ordered according to relevancy, and suggested to the user. [36, p. 475]

Figure 5 shows a partial network from Shoval's work. In this example, a user has entered the terms "management" and "data." Spreading activation in the network starting with "data" has led to the concepts "database" and "data management" after one step and "database management" after two steps. Spreading activation from "management" has led to "database management" after one step. In the search stage, spreading activation generates a list of potentially relevant concepts. In the above example, "database management" is identified as potentially relevant. To control the size of the list of relevant concepts, additional rules are also applied to candidates. One such rule is "the metric of 'strength', which is the number of originators, user terms, involved in the selection of a relevant term: the more user terms involved the more promising or important a selected term is assumed to be" [36, p. 479]. The "suggest" stage then further uses measures like the metric of strength to rank order the selected terms for presentation to the user.

Thus, such associative network systems offer an alternative approach to the design of knowledge structures and the design of control processes. Instead of explicitly building in strengths of association or predictiveness, as was done in RUBRIC, these strengths are implicit in the structure of the network and the rules used to act upon the network.

1.5 Explicit Representation

With the currently available commercial systems, interest in a topic is expressed by creating queries for individual concepts and then connecting these queries by logical operators. Thus a person interested in "bioindicators for pesticides" might create a query like

> (BIOINDICAT? OR BIOACCUMUL? OR ACCUMUL? OR BIRD# OR EAGLE# OR HAWK# OR FISH? OR TROUT OR SALMON) AND (PESTICIDE# OR DDT OR MALATHION OR PARATHION).

Similarly, to express interest in a topic like "combined modality therapy for colonic neoplasms" using a thesaurus system like Menu-MEDLINE [29], the searcher could select concepts from the thesaurus hierarchy and combine them with logical operators:

> ALL COLONIC NEOPLASMS AND COMBINED MODALITY THERAPY.

Associative network systems like the one proposed by Shoval do not address the issue of how to better represent a full topic. Rather, they are intended to assist users "in selecting the right vocabulary terms for a database search" [36, p. 475].

Thus, the existing commercial systems, thesaurus systems, and associative networks all rely on the use of logical operators to implicitly represent topics. RUBRIC offers an alternative topic representation, but it is still an implicit representation for two reasons:

(1) The topic (e.g., the World Series) must have been anticipated when rule evaluation trees were built (i.e., a unique tree must exist in the knowledge base for every possible topic);
(2) Weights in the evaluation tree play a role analogous to logical operators in defining a topic for existing commercial systems.

There is, however, another class of knowledge-based systems that explicitly represent topics. This class of systems uses frames to explicitly represent the relationships among concepts [6].

In discussing CoalSORT, one such system, Monarch and Carbonell suggest:

> Frame-based semantic networks are intelligible and communicable because they contain familiar words and phrases representing thoughts that are conceptually related in familiar, well organized and uniform ways. [25, p. 41].

As an example, they present a frame instantiation for coal liquefaction, which contains the following terms as slots or classes of information:

description:	dissolve solid coal to liquid state through the application of heat, solvent, and catalyst
sub-categories:	catalytic coal liquefaction non-catalytic coal liquefaction
site:	coal liquefaction plant
chemical process:	chemical reaction
also called:	coal dissolution
examples:	SRC-1 EDS CONOCO H-Coal.

Such an organization of concept supports understanding by the computer [35]. In CoalSORT, frames support graphical browsing of the knowledge base in an organized fashion. Another system that is based on frames, PLEXUS, uses them to support natural language input: "There is complete freedom of expression in presenting the problem. It may be a single term, a list of terms, a phrase or a grammatical sentence" [45, p. 103]. Our work on EP-X, a knowledge-based system to retrieve documents on environmental pollution, falls within this tradition [38, 40, 41]. Topics are explicitly represented as frames, which support the interpretation of user entries and the generation of topic refinement suggestions.

2. DEVELOPMENT OF A SEMANTICALLY BASED SEARCH SYSTEM

As part of the initial development of EP-X, we studied the performance of a number of expert search intermediaries. These intermediaries were also experienced indexers at the Chemical Abstracts Service. Hence, in addition to experience in acting as intermediaries, they were very knowledgeable about a particular field of chemistry and about the practices and policies followed at Chemical Abstracts when indexing documents in that field.

Initially, we conducted a number of informal studies of the performance of these search intermediaries. These studies consisted of

(1) Interviews regarding search methods;
(2) Videotapes of intermediaries helping actual information seekers;
(3) Videotapes of intermediaries helping us conduct searches, when we were role-playing to determine how the intermediary would respond under various conditions.

A sample of the resultant data is given below.

Intermediary:	OK. What area are you interested in searching?
Requester:	OK. The title of what I'm doing is "Effects of Acid Rain on Vegetation and Wildlife."
Intermediary:	OK. How broad do you want this to be?
Requester:	Since I plan to use it for a problem-solving report, I want to narrow it down to one solution for acid rain, reducing it.
Intermediary:	How much do you know about acid rain?
Requester:	Not a whole lot really. I know it's caused by sulphur dioxide in the air and nitrogen escaping into the air.
Intermediary:	There are different kinds of precipitation, and I'm putting in a whole bunch of them. You know, things that we don't typically think of when we say acid rain, but could be acid snow or acid hail, depending on how cold it is . . .
Requester:	Yes. I guess they do have it where it's colder too.
Intermediary:	Well, places like Scandinavia and also northern Canada are very badly affected, partly because this acid gets into the snow. So you have the snow pack and then, in the spring, when

everything melts you've got this massive dose coming ... So there's also a location on the surface of the globe that might be of importance ... Would you be interested only in the United States? ...

Intermediary: Now we'll need a whole bunch of words that stand for vegetation ... Would reviews give you enough information?

On the basis of such data, we concluded that there were four critical aspects of the interactions that we had observed:

(1) Communication between an information seeker and a search intermediary involves interpretation at a semantic level. When an information seeker requests documents on the use of cadmium in electrodes to remove heavy metals from wastewaters, for instance, the intermediary knows that the user is interested in cadmium as a component in an electrode rather than as a pollutant.

(2) On detecting an ambiguity, the intermediary asks the information seeker to resolve it, often explicitly pointing out the alternatives that should be considered. For example, an information seeker asking about the accumulation of pesticides in fish might be asked whether the interest is in fish as bioindicators or in the effects of pesticides on fish.

(3) The intermediary actively helps the information requester to explore and refine the topic by making suggestions based on a knowledge of the domain (and, when necessary, based on document descriptions that have been retrieved and displayed). A request for documents on pollution by radioactive substances might, for instance, lead to the suggestion to look for documents on pollution from fallout and nuclear power plants.

(4) The intermediary handles the mechanics of mapping an understanding of the topic of interest into an overall search plan or strategy (including the generation of lower level details like the selection of logical operators). During the example dialogue given earlier, for example, the intermediary entered the computer command "S acid (w) (rain or pptn or hail or snow or fog)" without any discussion of why the With operator was being used.

Our initial version of EP-X focused on the first two aspects of performance, communication at a semantic level, and active assistance in identifying and resolving ambiguities.

2.1 Semantically Based Search

Critical to the performance of these human experts was their knowledge of the domain of interest. This knowledge allowed them to understand the meaning of the information seeker's topic statement (i.e., they could translate the topic statement into some sort of semantic representation). It also allowed them to identify and help clarify ambiguities. For instance, when a person expressed an interest in pesticides, the search intermediary knew that the searcher was interested in chemical compounds used for particular purposes. Similarly, a request for documents on the topic "mercury in natural bodies of water" activated knowledge available to the intermediary and prompted the question: "Are you interested in all occurrences of mercury in natural waters or just the anthropogenic (man-made) ones? We generally think of pollution as just including the anthropogenic ones."

2.1.1 *Knowledge Representation.* In order to achieve similar performance from a computerized intermediary, a representation of the meanings of concepts and topics is needed. In EP-X, we have chosen to represent this knowledge in the form of hierarchically defined semantic primitives and frames [1, 18, 24]. Earlier in this paper we discussed the use of such hierarchies by EP-X. Frames represent a second class of knowledge used by EP-X.

Topics in the field of environmental chemistry were represented by frames such as

> Removal of a *pollutant* in *polluted medium* using a *removal/treatment process.*

Each italicized concept represents a slot that can be filled by a specific instance. An instantiation of this frame, for example, would be the

> "removal of *mercury* in *wastewater* using *ion exchange.*"

Frames, then, indicate the usage of a concept (mercury as a pollutant versus mercury in a mercury vapor lamp used to treat biological wastes). They also indicate the relationships among concepts. In EP-X, the possible slot fillers for a given slot are organized in tangled hierarchies to indicate class relationships.

The other important class of knowledge captures the mapping of words to semantic primitives. A number of different kinds of ambiguities and equivalence classes have been identified and captured as different kinds of triggers. One class of triggers captures direct equivalences (US, USA, and United States all trigger the same semantic primitive). Another captures ambiguities in word sense (Mercury triggers three alternative word senses: Mercury as a chemical substance, Mercury as a planet, and Mercury as a Greek god).

2.1.2 *Database Representation.* The knowledge base in EP-X describes topics pertinent to the field of pollution. The database indicates which documents are relevant to particular topics. Each document is represented as a frame instantiation with an associated list indicating the authors, title, abstract, and so forth.

This representation of documents as frames does not, however, support efficient processing for certain functions performed by EP-X. Consider, for instance, a searcher interested in "methods for removing phosphates from wastewater." To find the relevant documents based on a frame representation, the system would have to look through the entire set of documents, checking to see which ones had the slot for removal methods filled, had phosphates as the slot filler for pollutants, and also had wastewater as the polluted medium.

Consequently, EP-X also has information about document contents attached to each of the nodes in the concept hierarchies as illustrated in Figure 6. With this representation, the system can simply collect relevant document identifiers from the appropriate hierarchies in order to determine the relevant documents. A request for documents on "nitrates in natural bodies of water," for instance,

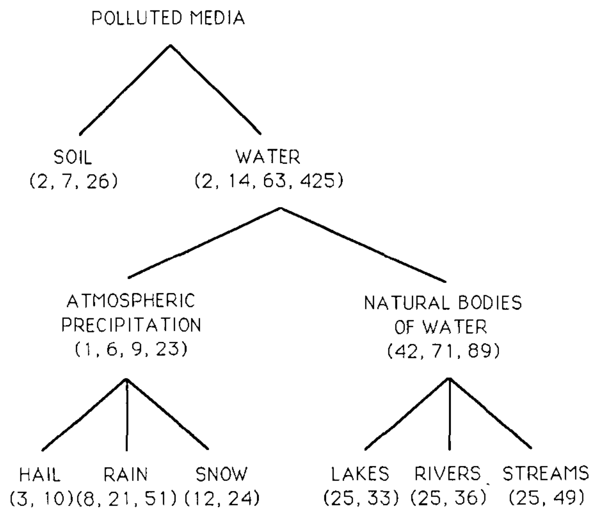

Fig. 6. Partial hierarchy showing document identifiers attached to relevant nodes.

would be processed by

(1) Forming a set that is the union of all document identifiers for NATURAL BODIES OF WATER and all its children (documents 42, 71, 89, 25, 33, 36, and 49 in the example shown in Figure 6);
(2) Forming a similar set for the documents associated with NITRATES and its children;
(3) Finding the intersection of these two sets.

This representation would be highly inefficient, however, for answering a question like "What additional slot could I fill to narrow the current topic?"

To generate a suggestion such as "Would you like to restrict your current topic (the effects of acid rain) to those documents dealing only with the US?," EP-X needs to

(1) Identify the set of documents pertinent to the current topic by using the document identifiers attached to hierarchy nodes;
(2) Look at the frames for all of these documents to see what additional slots are filled by them.

To accomplish this second step using the hierarchy representation (rather than the documents represented as frames) would require searching all of the nodes of all of the hierarchies associated with all of the slots that were not part of the current topic. Consequently, EP-X stores knowledge of document contents both as frames and as lists of identifiers attached to hierarchy nodes.

2.1.3 *Functional Description.* EP-X uses its knowledge base and database to interpret user entries, identify ambiguities, and retrieve document information.

KEYWORD LIST

Your keyword list currently consists of the following: POLLUTION CADMIUM OHIO MICHIGAN

Fig. 7. Keyword entry screen.

INTERPRETATION

94 documents are available on cadmium or its compounds as pollutants in Michigan or Ohio.

Fig. 8. Interpretation screen.

Users enter lists of keyword phrases (ultimately, they should be able to enter complete natural language phrases) as shown in Figure 7. EP-X uses its knowledge base to identify the possible meanings of the user's entry and, if there is no ambiguity, it retrieves the pertinent document information from its database (see Figure 8). In retrieving documents, the hierarchically defined concepts (e.g., Ohio) are used to retrieve all documents discussing the general concept and all documents discussing specific cases of that concept (e.g., specific locations in Ohio). Thus, by combining thesaurus-like functions (discussed earlier in this paper) with frame-based processes, EP-X can communicate with the user in terms of topics.

In addition, to interpret user entries in terms of their meanings, EP-X monitors for and responds to ambiguities. Entry of the keyword phrase "precipitation," for instance, causes EP-X to inquire whether the user is interested in precipitation such as rain, snow, and so forth, or precipitation as a waste treatment method to remove pollutants. Figure 9 illustrates a response to another type of ambiguity (one that suggests the user has an inappropriate mental model of the system's functioning).

2.1.4 *Semantically Based Search: Summary.* As illustrated above, EP-X uses its knowledge of the field of environmental chemistry to communicate with the user about meanings of words and topics. EP-X also assists the user in detecting and resolving ambiguities. These two functions illustrate our attempt to capture the first two critical aspects of human-to-human interaction that we discussed earlier. There are, of course, subtleties that we have not captured in our knowledge

KEYWORD LIST

Your keyword list currently consists of the following :

ACCUMULATION
RADIOACTIVE SUBSTANCES
FISH
FLOUNDER
ATLANTIC OCEAN

CLARIFY SCOPE OF FISH

It is unclear what you mean when you enter both FISH and FLOUNDER as keywords. Are you interested in documents discussing FISH in general (including documents dealing with specific kinds of fish), or only in documents discussing FLOUNDER?

Delete the keyword FISH if you are interested only in FLOUNDER. Otherwise, delete the keyword FLOUNDER (thus indicating that you are interested in documents discussing FISH in general, including documents dealing with specific kinds of FISH). If neither of these alternatives describes your interest, you can delete both terms from your keyword list.

Do you want to:

1 DELETE FISH from your keyword list
2 DELETE FLOUNDER from your keyword list
3 DELETE both FISH and FLOUNDER from your keyword list

Fig. 9. Response to an ambiguity resulting from an inaccurate mental model.

representations such as knowledge of the user's background or intended uses for the desired information. Other researchers are exploring some of these other aspects of human-to-human interaction [5, 7, 9, 32, 45]. The third critical aspect of human-to-human interactions that we listed above focused on active assistance in exploring and refining a topic. Our research on this topic is described below.

2.2 Knowledge-Based Search Tactics

In our initial studies, we frequently observed domain-knowledgeable intermediaries generating suggestions for refining a topic. For instance, an information seeker expressing an interest in "the effects of acid rain" might be told that acid snow, acid hail, and acid fog are other forms of precipitation of possible interest. We hypothesized that such suggestions for topic refinement could be modeled as search tactics that could be applied to EP-X's knowledge base and database. This hypothesis led to two research questions:

(1) What tactics are necessary to account for the refinement suggestions of human search intermediaries?
(2) Can these tactics be implemented in a computational model?

These questions are explored below.

2.2.1 *An Empirical Study.* Bates identifies a variety of information search tactics used by intermediaries. One example is the tactic PARALLEL, "to make the search broad (or broader) by including synonyms or otherwise conceptually parallel terms" [2, p. 208].

We seek to go a step further in modeling this type of intermediary performance. We propose, for example, to model the processes that generate the appropriate "conceptually parallel terms" for a particular topic. Imagine a user interested in the use of precipitative methods to remove heavy metals from wastewater. To broaden this topic, the intermediary could list concepts "parallel" to precipitation. There are hundreds of such concepts (other waste removal methods), however. An intelligent intermediary would not suggest all of these, but rather would suggest only those that are relevant to the current context or topic (the removal of heavy metals from wastewater). We seek, then, to identify the *knowledge-based search tactics* that allow human intermediaries to generate such intelligent suggestions.

As part of this search, we have conducted an empirical study of the refinement suggestions made by a human intermediary. We videotaped the interactions of information seekers with an expert intermediary (an expert in the field of environmental chemistry and in the indexing practices at *Chemical Abstracts*, as well as an experienced intermediary).

In this empirical study, we analyzed the discourses between the expert intermediary and 17 real information seekers as they conducted actual on-line searches of the *Chemical Abstracts*. The information seekers were undergraduates, graduate students, postdoctoral fellows, and faculty at The Ohio State University. All of them were looking for information on topics related to environmental chemistry to meet some real personal need.

Prior to analyzing the discourses for evidence of particular search tactics, we identified a number of coding categories [12]. These categories were identified on the basis of a theoretical analysis of the knowledge structures used in EP-X and on an analysis of the task of on-line searching of a traditional bibliographic search system using logical operators to link character strings.

We define knowledge-based search tactics that to be operations result in a change in the topic of interest as defined in EP-X's knowledge base. Before applying a particular tactic, for instance, the topic of interest might be "phosphates in the rivers of Pennsylvania from agricultural runoff." After application, the topic (the suggested alternative topic) might be "phosphates, nitrites, or nitrates in the rivers of Pennsylvania from agricultural runoff." (In this example, additional slot fillers for an already active slot have been suggested.) Thus, we used our knowledge structures as the basis for a taxonomy identifying different types of search tactics.

Our analysis of EP-X's knowledge structures identified 19 such tactics or types of transformations that could, in principle, be applied to generate a suggestion. One such transformation for broadening a topic is "deleting a slot." An example of the application of this tactic would be the transformation of the topic "*acid rain*" from *power plants* in *Europe*" to the suggested topic "*acid rain* from *power plants*." Another transformation, this one for narrowing a topic, is to "add a slot." An example of this would be transforming "removal of *metals* from *wastewater*" into "removal of *metals* from *wastewater* using *chelation or ion exchange*."

The intermediary could generate such transformations either (1) spontaneously (i.e., without any cues from retrieved document information); or (2) in response to some cue in retrieved document information. Note that such a taxonomy does not identify the cognitive processes responsible for triggering and generating specific transformations. It simply characterizes them in terms of their end results.

Advanced graduate students in environmental chemistry were paid to code the discourses in terms of these topic transformations. In the 17 searches, 361 such knowledge-based transformations were generated by the intermediary. Of these, 289 were generated spontaneously by the intermediary. The four most common tactics were

Delete a slot	(36 occurrences)
Add a slot	(81 occurrences)
Delete a slot filler and replace it with a more specific example of the same concept	(54 occurrences)
Add additional slot fillers to an already active slot	(69 occurrences)

In addition to the use of such knowledge-based search tactics (or some equivalent process) to suggest new topics, the intermediary also used domain knowledge to increase the thoroughness of the search on a particular topic. These transformations consisted of adding synonyms and specific cases of a broader concept that was already part of the search. Transformations of this kind occurred 32 times, 30 of them spontaneously generated by the intermediary.

Finally, there was a third class of transformations that did not rely (at least not directly) on the use of domain knowledge. These transformations consisted of changing search restrictions (restricting the languages of retrieved documents, the years of publication, etc.), changing logical operators, or restricting the searched field (e.g., title only versus title plus abstract). Instances of such transformations totaled 28, 1, and 7, respectively, for the 17 searches.

KEYWORD LIST

Your keyword list currently consists of
the following:
BIOINDICATION
PESTICIDES
MOLLUSKS

INTERPRETATION

18 documents are available on the use of
mollusks as bioindicators for pesticides.

SUGGESTIONS FOR
BROADENING

104 documents are available on the use of
clams, fish, fungi, insects or mosses as
bioindicators for pesticides. Thus, you will add
86 documents to your set if you broaden
mollusks to include these other bioindicators
for pesticides.

Do you want to:
1 BROADEN your topic as suggested above.

Fig. 10. Knowledge-based use of the tactic PARALLEL.

The results of this analysis of our empirical study are very striking. This intermediary

(1) Makes extensive use of domain knowledge to suggest topic refinements and to increase such thoroughness;
(2) Generates most knowledge-based suggestions spontaneously, rather than relying on a prompt from a retrieved document;
(3) Relies very little on changes in logical operators (e.g., changing AND to WITH) to refine a query.

Some of these knowledge-based suggestions can be modeled by fairly simple processes acting on EP-X's knowledge structures. Others, however, require fairly rich processing. A good example of this latter case is the transformation of "control of acid rain in the United States" to "prevention of nitrogen and sulfur oxides as air pollutants in the United States."

2.2.2 *Developing a Computational Model.* We have modeled the performance of this intermediary in terms of the application of knowledge-based search tactics to the knowledge base and database in EP-X. For a number of the tactics we have identified, no additional knowledge has to be added to EP-X's frame system

KEYWORD LIST

Your keyword list currently consists of the following: **PREVENTION** **ACID RAIN** **NORTH AMERICA**

INTERPRETATION

1123 documents are available on the prevention of acid rain within NORTH AMERICA.

SUGGESTIONS FOR NARROWING

Acid rain is produced by 16 specific sources in NORTH AMERICA. If you add the keyword SOURCES to your keyword list you will limit your topic to 623 documents discussing specific sources. In addition, the 16 specific sources will be displayed in the hierarchies window, along with a listing of the number of relevant documents from each particular source. You can then further narrow your search to specific sources if you want to. **Do you want to:** **1 NARROW your topic as suggested above.**

Fig. 11. Narrowing by adding a new slot.

with its hierarchically defined semantic primitives. Figures 10, 11, and 12 illustrate the suggestions generated by two such tactics (adding slot fillers to an already active slot and adding a new slot). The tactic illustrated in Figure 10 involves using the current context (bioindicators for pesticides) to identify potentially relevant organisms from a list of thousands. Figure 11 illustrates narrowing the topic by identifying a new slot (SOURCES) that can be used effectively to focus attention on a subset of the currently retrieved document set. Figure 12 shows the pruned sources hierarchy indicating the sources of acid rain in North America that are discussed in its database.

There are other tactics, however, that appear to require knowledge not currently contained in EP-X. The best example is the transformation of "control of acid rain in the United States" to "prevention of nitrogen and sulfur oxides as air pollutants in the United States." This transformation is based on

(1) Knowledge of the transformations that a pollutant can go through as it moves from its source through the environment;

HIERARCHY DISPLAY

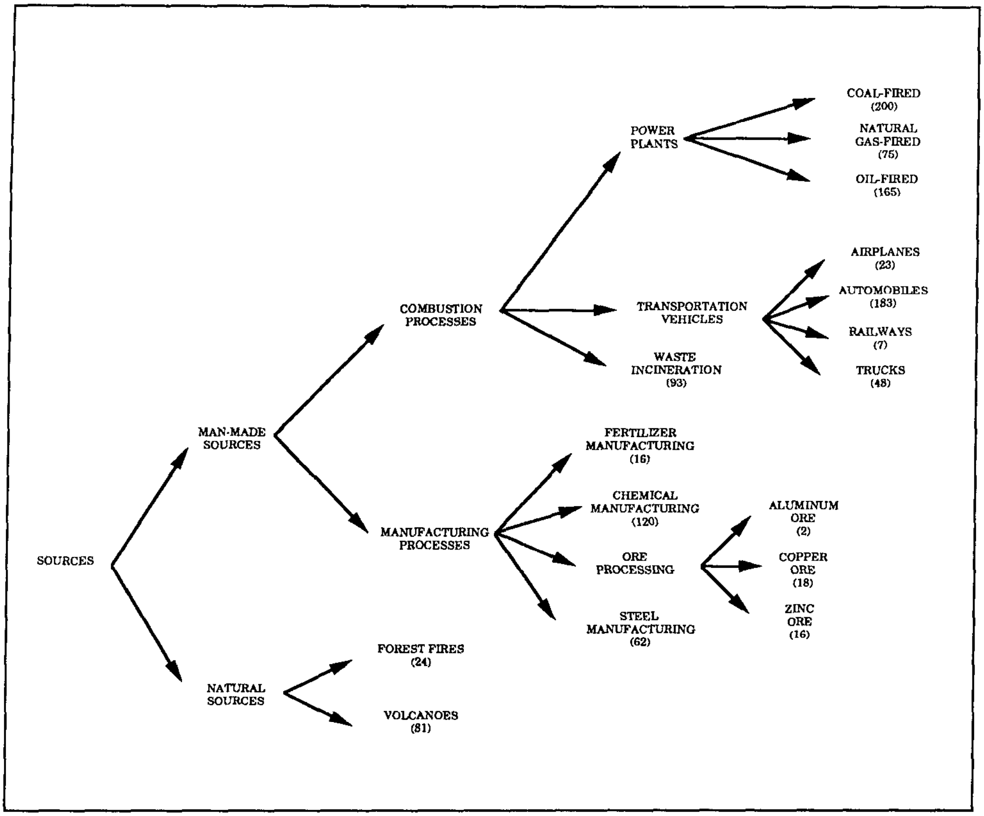

Fig. 12. Hierarchy display corresponding to the suggestion in Figure 11.

(2) Recognition that control can occur (in principle) at any point as the pollutant moves through the environment. In particular, control can occur before the pollutant is created, after it has been created but before it enters the environment, or after it enters the environment.

In the case of this transformation, the intermediary has recognized that nitrogen and sulfur oxides are precursors that develop into acid rain and that it is much easier to control these precursors than to control or reduce the levels of acids in rain directly.

3. CONCLUSIONS

Our work on EP-X, as well as the work by other research groups reviewed in this paper, serves to make a number of interesting points:

(1) There are a variety of functions that a computerized intermediary could play. These include suggesting search strategies, identifying synonyms for a concept of interest, supporting natural language input, and generating suggestions for related topics;

(2) There are significant differences among the various prototypes that researchers are currently exploring, in part owing to the fact that there is considerable diversity in the expertise that different intermediaries have. Our work on EP-X has relied on experts with considerable domain knowledge, whereas others have focused on intermediaries who are generalists that help with the development of search strategies based on domain-independent skills;

(3) The alternative approaches discussed in this paper vary significantly in terms of the kinds of knowledge represented in the prototypes. Some systems focus solely on knowledge for better defining individual concepts. Other systems include an actual knowledge of topics. Systems also differ in terms of the amount of structure imposed on the knowledge base. Some rely on implicit organization of concepts (represented as associative networks), whereas others make very explicit the types of relationships that exist among concepts (thesauri and frame-based systems);

(4) Alternative systems also differ in terms of their control processes. Associative networks rely on processes like spreading activation. Production rule systems make use of metrics for calculating the predicted relevance of terms. Frame systems make use of a variety of inference processes to identify the user's topic and to generate suggestions for topic refinement. Our latest work suggests that some of these processes actually require additional domain-specific knowledge to generate suggestions.

This diversity is to be expected since the field is still trying to determine the range of functions that computerized intermediaries can play. It is premature to decide what method or collection of methods can support large-scale applications in a fruitful manner.

3.1 Future Research

There are a number of research directions suggested by our work on EP-X and by work on other prototype systems. Some have to do with knowledge engineering questions, some are concerned with interface design, and some have to do with computational efficiency.

3.1.1 *Knowledge Acquisition.* The frame system that is the core of EP-X is a relatively simple structure. To create the necessary knowledge structures for a domain, relevant frames, slots, hierarchically defined slot fillers, and triggers must be identified. This is a sizeable but fairly straightforward task for any new domain. Furthermore, the basic control processes that operate on these classes of knowledge are domain independent.

It is important to realize, though, that much of the knowledge base used by EP-X is actually implicit in the database of document surrogates, instead of being stored explicitly in the knowledge base. When a new document is added to the database, EP-X may acquire new knowledge (other than the fact that this particular document talks about some specific topic). The frame system that defines EP-X's knowledge base only knows about things like the fact that removal processes can be used to remove pollutants from specific media and that chelation is a removal process. It does not know that chelation, instead of filtration, can be used to remove mercury from wastewater. This knowledge is contained in the database. In this sense, EP-X learns new knowledge each time a new document is added to its database. Knowledge acquisition is thus an implicit by-product of the indexing of documents. The issue of how to efficiently acquire such knowledge is critical to the success of all the prototypes discussed in this paper. Research on knowledge-based indexing systems represents one approach to coping with this problem [17].

Our most recent study, reported above, indicates that additional domain-specific knowledge is also necessary to provide certain functions. This knowledge appears to consist of temporal and causal relationships (such as the transformation of a pollutant into a new form once it enters a new medium). Just as the slots must be specified for the relevant frames, such temporal and causal relationships must be identified for each new domain. The fact that this extra knowledge appears to consist of these specific types of relationships, however, makes it clear what to look for in a new domain.

It is likely that additional empirical studies will identify an even greater variety of knowledge in use by human intermediaries. Our own work still leaves open, for instance, the question of how particular tactics are selected for application to a given topic.

3.1.2 *Interface Design.* Communicating with a computer is not the same as communicating with a person. Computers have abilities (e.g., rapid generation of graphical displays) that human intermediaries do not have, and vice versa. It is still very much an open question as to how the various prototypes developed to date can most effectively communicate with users. In CoalSORT, a graphical interface was proposed, whereas in PLEXUS natural language input was accepted. Neither those approaches nor the approaches to interface design in systems like EP-X or CANSEARCH have been adequately tested yet.

3.1.3 *Computational Efficiency.* Even given solutions to the knowledge-engineering issues and interface design questions, the computational requirements for large-scale implementations will be enormous. Consider, for example, the issue of providing context-sensitive hierarchy displays similar to the one shown in Figure 12. A full hierarchy for a slot such as "pollutants" will have well over 5000 nodes in it, with hundreds of thousands of document identifiers attached to these nodes. Using brute-force methods to prune such a hierarchy (in real time) for a topic like "pollutants from the energy industry" would be prohibitive. Thus, it will be necessary to develop computational methods to ease this processing burden (such as preprocessing to store knowledge of which pollutants are produced by the energy industry).

3.1.4 *Summary.* EP-X and other prototypes such as PLEXUS, CoalSORT, RUBRIC, and CANSEARCH have served to identify many of the important issues and potential solutions for developing computerized intermediaries for the retrieval of bibliographic information. Considerable research remains to be done, though, to resolve issues pertaining to knowledge engineering, interface design, and computational methods. One clear result of this line of research, though, is that the field of information retrieval has become highly interdisciplinary. Such research issues call for expertise in cognitive psychology, artificial intelligence, computer science, linguistics, and human factors engineering, as well as expertise in information retrieval itself.

REFERENCES

1. AIKENS, J. Prototypical knowledge for expert systems. *Artif. Intell. 20* (1983), 163–210.
2. BATES, M. Information search tactics. *J. Am. Soc. Inf. Sci.* (1979), 205–214.
3. BATES, M. Idea tactics. *IEEE Trans. Prof. Commun. 23*, 2 (1980).
4. BELKIN, N., AND CROFT, W. B. Retrieval techniques. *Ann. Rev. Inf. Sci. Tech. 22* (1987), 109–145.
5. BELKIN, N., SEEGER, T., AND WERSIG, G. Distributed expert problem treatment as a model for information system analysis and design *J. Inf. Sci. 5* (1983), 153–167.
6. BISWAS, G., SUBRAMANIAN, V., AND BEZDEK, J. A knowledge based system approach to document retrieval. In *Proceedings of the Second Conference on Artificial Intelligence Applications* (Miami Beach, Fla., Dec. 11-13). IEEE Computer Society Press, 1985, pp. 455–460.
7. BRANJNIK, G., GUIDA, G., AND TASSO, C. User modeling in intelligent information retrieval. *Inf. Process. Manage. 23*, 4 (1987), 305–320.
8. BRITTAIN, M. Implications for LIS education of recent developments in expert systems. *Inf. Process. Manage. 23*, 2 (1987), 139–152.
9. BROOKS, H. M., DANIELS, P. J., AND BELKIN, N. J. Problem descriptions and user models: Developing an intelligent interface for document retrieval systems. In *Informatics 8: Advances in Information Retrieval* (Oxford, England, Apr. 16–17, 1985). Aslib, London, 1985, pp. 191–214.
10. DEFUDE, B. Knowledge-based systems versus thesaurus: An architecture problem about expert systems design. In *Research and Development in Information Retrieval.* Cambridge University Press, New York, 1984, pp. 267–280.
11. DUMAIS, S. T., FURNAS, G. W., LANDAUER, T. K., DEERWESTER, S., AND HARSHMAN, R. Using latent semantic analysis to improve access to textual information. In *CHI '88 Conference Proceedings* (May 15–19, 1988, Washington, D.C.). ACM, New York, 1988, pp. 281–285.
12. ERICSSON, K. A., AND SIMON, H. A. *Protocol Analysis: Verbal Reports as Data.* MIT Press, Cambridge, Mass., 1984.
13. FIDEL, R. Towards expert systems for the selection of search keys. *J. Am. Soc. Inf. Sci. 37*, 1 (1986), 37–44.
14. GIULIANO, V. E. Analog networks for word association. *IEEE Trans. Military Electronics 7* (1963), 221–234.
15. HARTER, S. P., AND PETERS, A. Heuristics for online information retrieval: A typology and preliminary listing. *Online Rev. 9*, 5 (1985), 407–424.
16. HAWKINS, D. T., AND WAGERS, R. Online bibliographic search strategy development. *Online* (1982) 12–19.
17. HUMPHREY, S., AND KAPOOR, A. The Med IndEx system: Research on interactive knowledge-based indexing of the medical literature. NLM Tech. Rep. #LHNCBC 88-1, National Library of Medicine, Bethesda, Md., 1988.
18. KOLODNER, J. Indexing and retrieval strategies for natural language fact retrieval. *ACM Trans. Database Syst. 8*, 3 (Sept. 1983), 434–464.
19. KRAWCZAK, D., SHUTE, S. J., AND SMITH, P. J. EP-X: A demonstration of semantically-based search of bibliographic databases. In *Proceedings of the 10th Annual International Conference on Research Development in Information Retrieval* (New Orleans, La., June 3–5). ACM, New York, 1987, pp. 263–271.
20. KRAWCZAK, D., SHUTE, S. J., AND SMITH, P. J. The use of metaphors in learning to use a knowledge-based system. In the *2d International Conference on Human-Computer Interaction* (Honolulu, Hawaii, Aug. 10–14). International Commission on Human Aspects in Computing, 1987, p. 114.
21. KRAWCZAK, D., SMITH, P. J., SHUTE, S. J., AND CHIGNELL, M. H. EP-X: A knowledge-based system to aid in searches of the environmental pollution literature. In *Proceedings of the 2d International Conference on Artificial Intelligence Applications* (Miami Beach, Fla., Dec. 11–13). IEEE Computer Society Press, 1985, pp. 552–557.
22. MARCUS, R. An experimental comparison of the effectiveness of computers and humans as search intermediaries. *J. Am. Soc. Inf. Sci. 34*, 6 (1983), 381–404.
23. MAUDLIN, M., CARBONELL, J., AND THOMASON, R. Beyond the keyword barrier: Knowledge-based information retrieval. In *Proceedings of the 29th Annual Conference of the National Federation of Abstracting and Information Services* (Arlington, Va., March 1). Elsevier, 1987, pp. 103–118.

24. MINSKY, M. A framework for representing knowledge. In *The Psychology of Computer Vision.* P. H. Winston, Ed. McGraw-Hill, New York, 1975.
25. MONARCH, I., AND CARBONELL, J. CoalSORT: A knowledge-based interface. *IEEE Expert,* (1987) 39–53.
26. Oddy, R. Information retrieval through man-machine dialogue. *J. Doc. 33* (1977), 1–14.
27. OLDROYD, B. K., AND SCHRODER, J. J. Study of strategies used in online searching: 2. Positional logic—An example of the importance of selecting the right Boolean operator. *Online Rev. 6,* 2 (1982), 127–133.
28. POLLITT, S. CANSEARCH: An expert systems approach to document retrieval. *Inf. Process. Manage. 23,* 2 (1987), 119–138.
29. POLLITT, S. Menu-MEDLINE: A prototype knowledge-based intermediary system for end user searching of references to the biomedical literature. Interim Technical Report for the National Library of Medicine, Bethesda, Md. July 14, 1987.
30. RADA, R., AND MARTIN, B. K. Augmenting thesauri for information systems. *ACM Trans. Off. Inf. Syst. 5,* 4 (1987), 378–392.
31. RAU, L. F. Knowledge organization and access in a conceptual information system. *Inf. Process. Manage. 23,* 4 (1987), 269–283.
32. RICH, E. User modeling via stereotypes. *Cognitive Sci. 3,* 4 (1979), 329–354.
33. SALTON, G. *Automatic Information Organization and Retrieval.* McGraw-Hill, New York, 1968.
34. SALTON, G., AND MCGILL, M. P. *Introduction to Modern Information Retrieval.* McGraw-Hill, New York, 1983.
35. SCHANK, R., AND RIESBECK, C. *Inside Computer Understanding.* Lawrence Erlbaum Associates, Hillsdale, N.J., 1981.
36. SHOVAL, P. Principles, procedures and rules in an expert system for information retrieval. *Inf. Process. Manage. 21,* 6 (1985), 475–487.
37. SMITH, P. J., AND CHIGNELL, M. Development of an expert system to aid in searches of the *Chemical Abstracts.* In *Proceedings of the 47th ASIS Annual Meeting* (Philadelphia, Pa., Oct. 21–25). Knowledge Industry Publications, White Plains, 1984, pp. 99–102.
38. SMITH, P. J., CHIGNELL, M., AND KRAWCZAK, D. A. Development of a knowledge-based bibliographic information retrieval system. In *Proceedings of the 1984 IEEE Conference on Systems, Man and Cybernetics* (Halifax, Nova Scotia, Oct. 10–12). IEEE Systems, Man and Cybernetics Society, 1984, pp. 222–225.
39. SMITH, P. J., KRAWCZAK, D., SHUTE, S. J., AND CHIGNELL, M. H. Cognitive engineering issues in the design of a knowledge-based information retrieval system. In *Proceedings of the Human Factors Society—29th Annual Meeting* (Baltimore, Md., Sept. 29–Oct. 3). Human Factors Society, 1985, pp. 362–366.
40. SMITH, P. J., KRAWCZAK, D., SHUTE, S. J., AND CHIGNELL, M. H. Bibliographic information retrieval systems: Increasing cognitive compatibility. *Inf. Serv. Use 7* (1987), 95–102.
41. SMITH, P. J., SHUTE, S. J., CHIGNELL, M. H., AND KRAWCZAK, D. Bibliographic information retrieval: Developing semantically-based search systems. In *Advances in Man-Machine Systems Research,* Vol. 5, W. Rouse, Ed. JAI Press, Greenwich, Conn., 1989, pp. 93–117.
42. SMITH, P. J., KRAWCZAK, D., SHUTE, S. J., CHIGNELL, M. H., AND SATER, M. Bibliographic information retrieval systems: Using AI techniques to improve cognitive compatibility. In *Proceedings of the 1986 IEEE Conference on Systems, Man and Cybernetics* (Atlanta, Ga., Oct. 14–17). IEEE Systems, Man and Cybernetics Society, 1986, pp. 395–397.
43. THOMPSON, D. Interface design for an interactive information retrieval system: A literature survey and a research system description. *J. Am. Soc. Inf. Sci.* (1971), 361–373.
44. TONG, R. M., AND SHAPIRO, D. G. Experimental investigations of uncertainty in a rule-based system for information retrieval. *Int. J. Man-Mach. Stud. 22* (1985), 265–282.
45. VICKERY, A., AND BROOKS, H. M. PLEXUS—the expert system for referral. *Inf. Process. Manage. 23,* 2 (1987), 99–117.
46. VIGIL, P. J. The psychology of online searching. *J. Am. Soc. Inf. Sci. 34,* 4 (1983), 281–287.
47. WALL, R. A. Intelligent indexing and retrieval: A man-machine partnership. *Inf. Process. Manage. 16* (1980), 73–90.

Received March 1989; revised April 1989; accepted April 1989