Reports

of the

Intelligent

Systems

Group

WHERE EXPERT SYSTEMS INTERSECT WITH INFORMATION RETRIEVAL: Issues and Applications

by

Gail Thornburg

May 1986

File No. UIUCDCS-F-86-956

ISG 86-~~15~~ 14

File No. UIUCDCS-F-86-956

WHERE EXPERT SYSTEMS INTERSECT WITH INFORMATION RETRIEVAL:

Issues and Applications

by

Gail Thornburg

May 1986

Intelligent Systems Group
Department of Computer Science
University of Illinois at Urbana-Champaign

ISG 86-15

This research was supported in part by the Office of Naval Research under grant N00014-82-K-0186, in part by the National Science Foundation under grant NSF DCR 84-06801, the Defense Advanced Research Projects Agency under grant N00014-K-85-0878, and the Institute for Scientific Information (ISI), Philadelphia, PA.

ABSTRACT

This paper analyzes several issues related to the intersection of artificial intelligence and information retrieval, especially in the area of expert systems. In particular, we look at problems such as determining choice and form of representations in information, transparent interface as "black box" in information systems, and lack of system self-knowledge (performance brittleness). It is indicated that in the area of expert systems for information retrieval, major concerns are user awareness of the limits of a given system, user modelling in general, and facilities for system self-awareness and explanation. Expert system interfaces to larger conventional systems, designed to serve an end user (patron or information client) directly, have dominated research in expert systems in IR, though there exist numerous other possibilities for AI research in other areas of information retrieval, such as classification, indexing, and reference referral systems.

I. INTRODUCTION

The area of expert systems is a specialization within the larger area of artificial intelligence, which in turn is a specialty within a larger, multidisciplinary context. Likewise information retrieval draws on other disciplines both as problem areas and as conceptual relatives. One problem with viewing "information storage and retrieval" (IS&R) as a possible domain for the development of an expert system is that some form of representation is inherent in any information system (e.g., a citation in an index as a surrogate "representation" for the actual paper). Construction of an expert system in IS&R might involve building a representation of such an information representation, but might also mean building a representation from scratch. There is little work on IS&R expert systems reported in mainstream AI publications, though information scientists have published numerous systems development reports outside the major AI channels.

This paper discusses possible roles of expert systems in information storage and retrieval (IS&R), and explores issues in the development of such systems. It will attempt to cover the main themes in expert systems as related to IS&R, though it will not attempt to identify all systems extant in this fast-growing area. An expert system is defined here as any flexible computer program/system which can solve complex problems at an expert level, and explain itself to the human user. Divisions in systems identified are largely by task or subdomain. This paper will be divided into three main sections:

- Identification of some of the major issues of dealing with expert systems in IS&R;
- A section discussing individual systems developed so far; and
- A section on needs and issues not yet addressed in actual applications.

Since no single knowledge representation "paradigm" prevails, overall comparisons among systems using rules, semantic networks, frames, scripts, logic formalisms, or combinations thereof are inherently inconclusive. Notwithstanding this limitation, this paper will attempt to look at issues and commonalities from a functional viewpoint.

Information storage and retrieval is meant in this paper to encompass bibliographic organization in the broadest sense; that is, it concerns any information organization tasks/problems which can been seen as pertaining to, or the responsibility of, information science or librarianship. Thus, indexing, abstracting, systems design, and other areas are implied along with actual retrieval of factual or reference (citations) information. Because of the large amount of work done in the area of online bibliographic retrieval systems, this area may by default receive disproportionate attention.

A problem which emerges from the IS&R side of the study concerns the definition of expert systems. While it appears that within AI, rough consensus exists on a definition, the literature in information retrieval suggests synonymous use of the terms "expert system", "intelligent front-end", and "user-friendly interface." That is, the designation "expert system" is sometimes applied to systems not built by the process of knowledge acquisition from a human expert in a given domain. In this paper, the systems are necessarily taken at face value, but an attempt has been made to clarify the authors' varying definitions of expertise or intelligence in a system. There may be reason for concern over such artificial synonymity, insofar as it tends to ignore the possibility of intelligence in a machine system not designed as an expert system, or implies that user-friendliness in a system not designed as an expert system always requires machine intelligence.

II. GENERAL PARAMETERS

One well-known rubric[1] holds that appropriate domains for construction of rule-based systems are likely to be: 1) domains where knowledge is diffuse, not codified in any one unified theory (e.g., clinical medicine preferable to physics); 2) those domains where processes can be represented as a set of independent actions; and 3) those in which knowledge can be differentiated from the way in which the knowledge is used. Forsyth [A.2 ,p.14] gives a fairly similar characterization of "suitable/unsuitable" features in potential domains.

"*Suitable*"

Diagnostic

No established theory

Human expertise scarce

Data "noisy"/imprecise

"*Unsuitable*"

Calculative

Magic Formula exists

Human expertise plentiful

Facts are known precisely

These categorizations must of course be judged critically against unique features of any prospective domain, since, for example, the existence of the "calculative" need not mean that the algorithms involved are not quite difficult to manipulate. Or, if human expertise is not especially scarce, it may be diffuse and uncodified.

---

[1] R. Davis and J. J. King, "An overview of production systems," in: E. Elcock and D. Michie, eds., *Machine Intelligence 8* (Chichester, England: Ellis Horwood, 1977), pp. 300–332.

Note that a "diagnostic" problem is considered to be any area where there exist several possible answers, and the difficulty lies in selecting the right one or the least improbable ones. What lends interest both to traditional librarianship and information science as potential domains is the great "generality" of knowledge held by specialists in each area.

Two expert systems/AI review articles may be noted as indicators of the growth of interest among information specialists. Cercone and McCalla [A.1] delineate the intellectual roots of AI within linguistics, computer science, electrical engineering, psychology, and philosophy. They identify eight sub-areas in AI:

- Natural language understanding
- Computer vision
- Search/problem-solving/planning
- Theorem proving and logic programming
- Knowledge representation
- Machine learning
- Expert systems
- Miscellaneous efforts including game planing, speech understanding, AI approaches to education,

and so on.

Yaghmai and Maxin [A.5] cover definitions and approaches to knowledge representation in expert systems, summarize current state-of-the-art limitations, and identify issues and trends for the future. They make the interesting prediction [p.303] that in the 1990's expert systems will start to proliferate in areas where there were previously no human experts.

III. ISSUES IN THE CONTEXT OF INFORMATION STORAGE AND RETRIEVAL

A. Capabilities Needed

One could consider intelligent system research to focus both on systems which "behave" intelligently, and on intelligent interfaces to systems with larger unintelligent components. Sparck Jones [B.11], well known for her experimental work with non-Boolean information retrieval methods, takes this view.

One example worth explaining here is that of the large bibliographic utilities which have emerged in the last 10–15 years. DIALOG, BRS (Bibliographic Retrieval Service), and SDC (System Development Corporation) are three of the larger examples of these online vendors of multiple bibliographic (citation)and non-bibliographic (fact-retrieval) databases. As these vendors may offer access to more than a hundred databases, each of which may contain millions or hundreds of thousands of multi-field records, use of inverted index files (i.e. access via alphanumeric arrays with pointers to each record address applicable) has been necessary to avoid impossibly long search times online. One of the main advantages to the Boolean search operations used is the relatively high speed obtained by examining only those portions of the index actually used in a query. Unfortunately Boolean algebra does not allow for partial matching, and is further limited by its inability to weight output by ranked decrease in order of matching criteria. Future improvements on the current systems architecture are still in the realm of speculation. For the moment, many of the "intelligent interfaces" noted by Sparck Jones are directed at improving on the command-based, basically nonfriendly online systems which the major vendors offer today.

Basic capabilities needed for internal knowledge representation in an intelligent information retrieval system include, as Sparck Jones sees it:

1. Classification and concept formation,

2. Summarizing and abstracting,

3. Selection and retrieval filtering, and

4. Planning and modeling.

It should not surprise us that these look familiar, if we accept Weiss and Kulikowski's description of the classificatory nature of knowledge- based expert systems [A.4, p.17]. The Hayes-Roth text observes, indeed that "...workers in this field speculate that the most important by-product of expert systems will be the codification of knowledge" [A.3, pp. 27-28].

Basic needs for such a system might also be described in terms of types of knowledge needed. Pollitt's [B.6] characterization of types to incorporate include the categories:

1. System,

2. Search,

3. Subject knowledge, and

4. Knowledge of the individual user.

One could go a step further than this, and instead of types of knowledge, look at functional types of knowledge representation. Smith and Warner [B.10] describe such a taxonomy of information retrieval representations, based on the represented focus of each category. These authors list representations for:

1. Objects — e.g. documents as objects; or queries, such as representations which link terms with boolean operators, vs. those which act on the query as a vector of terms (possibly weighted);

2. Relationships — of document to a given query (relevance), of term to term (as in linkages of corresponding terms in different controlled vocabularies), of document to document (as in citation analysis);

3. Processes — representations of documents, for instance, vary chiefly in the processes or algorithms (in automatic indexing) used to generate them—each process could be seen as a method for identifying optimal indexing terms;

4. The choice of repositories of knowledge to be represented — not only documents as information sources, but human experts, and moreover collections of data;

5. Knowledge about users — ideally, dynamic representation, as for instance in representing the cumulative experience of users of a system in a form helpful to later users; and

6. Scope and limitations of the system: in other words, how well the system can define itself to the user.

This capability of system self-awareness is an issue Buchanan [B.1] explores from an external viewpoint, in the context of documentation. He notes that system manuals usually do not adequately discuss the appropriate uses of a program, nor define the program's limitations. He stresses the desirability of creating intelligence assistants that carry an awareness of their own problem-solving abilities and can explain them. The rationale for use of online documentation seems to be supported by the experience of the designers of one interactive tutorial LISP program;[2] Winston found that students learn LISP several times faster using the online tutorial than in typical classroom/lab setting, and has further stated that most users seem averse to consulting the printed manuals at all.[3]

---

[2] *Golden Common LISP*, Version 1.00, Gold Hill Computers, Inc., 1983.

[3] Patrick Winston, Lecture, at the International Summer Seminar on AI, Dubrovnik, 27 August 1984.

In comparing the capabilities proposed by a number of writers as need for intelligent information retrieval, the consensus seems to emerge as comprising, at minimum, the following performance capabilities:

1. Classification/concept creation;

2. creation of document representations—as in abstracting and summarizing, as well as creating bibliographic citations and access points;

3. building non–document representation of knowledge—from human experts, from collections of data;

4. modelling activities—e.g. user modelling, process modelling (as in choice of a method of indexing), modelling the query formulation;

5. search and retrieval filtering—as in improvements over Boolean term matching, to allow for partial matching; and

6. ability of a system to understand and explain its own limitations.

Through all the theoretical considerations thus far, two main emphases seem to run: the user's awareness, and the self–awareness of the system. As will be seen later, there are costs and tradeoffs connected with each, and the two factors may need to be weighed against each other in the design of a system.

B. Applications

At the most general level, Smith [B.8] mentions three classes of AI applications to information retrieval: reference retrieval, data/fact retrieval, and instructional systems. Presumably Sparck Jones' distinction between intelligent systems and intelligent interfaces to large unintelligent systems could be drawn within each of these three categories.

In another broad grouping, DeJong [B.3] identified four categories of AI applications to IR:

1. human–database interfaces;

2. conceptual indexing;

3. automatic data entry (in which the system can read and "understand" text to be added to a database); and

4. active memory, e.g. involving use of deductive data retrieval, or storing rules that characterize data elements (rather than storing the data elements themselves).

Expert systems could also provide a means of alternative publication. K. P. Jones [B.4] suggests that traditional printed publications which are difficult both to compile and to use, e.g. timetables or chemistry handbooks, might be generated by an expert system which has a report–writing module either as its central function or as a component. As Michie [B.5] notes in his talk on "knowledge refineries," expert system "computers as co–authors" might one day supersede the human writing of manuals. To Michie's comment, one might reply: when that happens, we will probably want "them" to do it better!

More specific examples of expert systems applications which have been suggested for information retrieval include creation of classification schemes for library catalogs, systems for cataloguing (i.e. physical description on an item, plus choice of access points), user–friendly interfaces to large online collections, automatic indexing systems, and query based interfaces to the multi–database systems of various online bibliographic vendors. While all these applications are significant, this paper will focus on the last area, that of online search intermediaries. Descriptions of some typical systems follow.

IV. INDIVIDUAL DESCRIPTIONS OF SYSTEMS

While most of these projects relate to online search "intermediary systems," a few from other areas of information retrieval have been included where useful to illustrate some approach not implemented elsewhere. This listing is by no means exhaustive, but rather an attempt to illustrate major types of intermediary systems developed or under development thus far. Acronyms for the systems will not be defined here unless considered necessary to the reader's understanding.

Several commercial systems have been developed which function as database access software or "gateways," not expert systems in level of performance or in explanatory capabilities, but useful, query-based tools for the novice user of dial-up online document retrieval systems [C.9, C.15]. These may offer features such as command and response translation into more natural language, storage of search strategies and responses, logical multiplexing to permit multiple users (as in Toliver [C.16]), tutorial modules, and other user helps. Unlike front-ends marketed by the online vendor directly, SCI-MATE [C.4] offers a common command language for searching multiple vendors' systems, and adds the feature of a database management system for handling personal databases.

When implemented PLEXUS [C.17] is to be a referral system, i.e., not an interface to a larger remote database, but a source of advice on which reference tools (computerized databases or printed works) may aid in a given search. This system could serve as assistant to the reference librarian or the end user. The prototype system will be restricted to the area of horticulture. In the latest progress report, the authors state that the knowledge base is implemented in PROLOG with the remaining elements (user interface, data base management, production rules) in PASCAL. The system is to be implemented on a SIRIUS1 microcomputer with 850K RAM memory and a 20 Megabyte hard disk.

Trial development of a prototype system in more than one version, i.e. using different hardware and different software shells, has been attempted by only two information projects known to this author. Both are in the area of cataloguing.

The HEADS project [C.1], being carried out at the University of Manchester, uses for one version the ES/P Adviser marketed by Expert Systems International, running on an ACT Sirius 1 or IBM PC. The second version has been rewritten from the first in SAGE version 1.3 (from SPL International), and runs on the Prime 750.

In the ESSCAPE cataloging project, Hjerppe and collaborators [C.5] did a series of test of the commercial tools EMYCIN and EXPERT-EASE for certain cataloging applications. They have recently completed the first version of a corollary system on a related topic: an advisory system for authority control of personal names. Their paper is notable, too, for the useful, concise summary given of the differences between the system versions ESCAPE/EMYCIN and ESSCAPE/Expert-Ease.

The Searcher's Workbench is described by its authors Williams and Preece [C.18] as an intelligent front-end to a model search system. This menu-driven system was designed with tutorial features for novice users, and an escape from the tutorial for experienced searchers. A touch panel is used for input of everything except the search terms themselves.

Taking the search subtask of assisting users with selection of the right controlled vocabulary terms, Shoval [C.12] developed a system using a declarative representation scheme, structured as a semantic network.

The scripts approach is reflected in the CSIN system, described in Horowitz and Bergman [C.6], was characterized as a pre-prototype intelligent terminal designed to streamline the chemical information query process by capturing the results of one search, reformatting them as "acceptable" input to another system, and executing another retrieval. The first prototype "will automate a selected set of predefined multi-system query sequences known as 'scripts'."

Another project reflecting the "scripts" approach to representation also deals with chemical information, in this case *Chemical Abstracts* cited documents on environment pollution. Smith and his collaborators [C.13, C.14] state that this system relies on the encoding of episodic and semantic knowledge in a frame system. The system is notable here as reflecting the perspective of indexers at Chemical Abstracts Service, i.e. the vendor perspective.

Restriction to toy databases has been a limitation for most research in non-Boolean retrieval, in that results may not be generalizable to the large-scale collections of real life. Thus, it may be argued that CITE-NLM [C.3] represents an advance, since this medical natural language interface to an online catalog was able to make use of a large scale database (National Library of Medicine). This system uses a weighted combinatorial search process as a refinement over raw Boolean term matching.

The pilot project INFOS is being encoded as a small production rule system (about 50 rules) dealing with the online search process. INFOS is significant for the audience it addresses: intended use is by small information brokerage companies. The system is being developed by Obermeier and Cooper [C.10] at The Ohio State University.

Marcus' system, EXPERT [C.7], developed out of the earlier CONIT projects at M.I.T. Though also aimed at the end user, EXPERT comes closer to this paper's definition of an expert system as it incorporates, in a menu-driven mode, assistance with database selection, search topic concept formulation, automatic translation of concepts into search statements, and relevance feedback for reformulation of search strategy. This is a production system controlled by goal-directed forward-chaining regime. Though present system requirements are not known, the CONIT-4 version required 1 megabyte of memory in the M.I.T. Multics systems.

Possible uses of machine learning have been little addressed by current systems in information retrieval. One ongoing project is found outside the "intermediary/interface system" group. This concerns automatic thesaurus construction in the domain of medical literature (at the National Library of Medicine). Roy Rada [C.11] and colleagues are trying to build onto MeSH (Medical Subject Headings, a controlled vocabulary of medical terms) to develop a knowledge base for reasoning about document retrieval.

The ADVISE system described by Michalski and Baskin [C.8] is a meta–expert system equipped with multiple forms of knowledge representations, multiple control strategies, and sophisticated learning capabilities. Knowledge may be encoded in the system by means of if/then rules, networks, and relational tables. Nor is the user limited to a single control strategy as is generally the case in current commercial expert system "shells." Learning capabilities can be valuable both the development phase of the system and in its periodic improvment.

Since this meta–expert system constitutes an integrated set of tools for the contruction of expert systems in various domains, there is great potential here for information retrieval applications. Some of this potential is currently being exploited in a current project by this author. The expert system being built deals with one specific area of online information retrieval, choice of database(s) to be searched. This is an area generally underemphasized in current systems research. Though some online vendors have screening files which will tell a searcher the number of postings for a given term (or term combinations) in a given database, use of term postings alone for database choice can be misleading. Types of materials covered, level of specialization, languages indexed, and many other factors may often be as important in choosing the best database(s) for a search as are term postings.

This project will use multiple experts, focusing on databases in life sciences and engineering, and will also attempt to make some observations about the nature of "expertise" in such a "generalist" domain as information retrieval.

The learning capabilities of the ADVISE shell will be exploited, where possible, to compare rules obtained from experts with rules induced by the system from examples. Moreover, the system can use existing rules with new examples to generate new rules, which should be useful for ongoing improvement of the system, even on a search-by-search basis.

V. ISSUES, AND PROBLEMS OF PERSPECTIVE

Issues of representation are several, even within the general categories of what to represent, what form of representation to use, and which are the functions best included in representations. These can be viewed from the system perspective (e.g. indexing), or from the query perspective. Much experimental work in information retrieval has concentrated on development of document representations automatically—e.g., automatic indexing, automatic classification of documents, automatic term clustering. These development efforts are only feasible from the system perspective; two examples are the perspective of the online vendor of multiple databases, and other developers of databases. Though the former may provide enhancements to a database usually developed elsewhere, the vendor is generally limited by prior design decisions made by the original developer.

From the query perspective, one may adopt the perspective of the expert or the end user. An example of the former, representation of expertise could be developed via functional analysis of the expert's tasks. Deerwester [B.2] takes this approach, in attempting to develop a conceptual description, the Retrieval Expert Model, with the reference librarian as the archetype. He postulates that this expert begins with a set of general search strategies, and constructs other needs-specific strategies by trying and modifying existing search strategies [pp. 58–59].

Taking the user perspective here means developing concepts about users in order to better predict their needs. How this "user" knowledge is to be obtained, or itself categorized, is a major issue. Studies have been done using sterotypes as a mechanism for modelling, e.g. by Rich [B.7]. Another study of users employed schema theory, based on characterisation of research paradigms in a given scientific subfield as "procedural scripts" [B.12]. The idea was that the scripts which a scientist follows, a system might learn.

Note that efforts to model users may obscure somewhat a related issue: whether to maximize operations performed for the user automatically and invisibly, or to stress more interaction with the user in the hope of making clear the subtler system constraints. As so many of the online intermediary systems developed to date assume users are novices,the trade-offs must be considered carefully.

One thorny problem in information retrieval by subject involves the trade-offs between the rigidity of controlled vocabulary (thesaurus) approaches and the enormous imprecision of natural language or "free text" searching. Full-text databases (meaning those which include not only citation but also the complete referenced paper), though often hailed by their developers as the answer to free-text subject search difficulties, have not thus far been proven to increase the quality of the subject search. Natural language appears to be too rich and too contextual not to need structuring for effective subject searching of a literature. In fact, a recent report on one full-text litigation database reported a surprisingly low 20% recall (relevant references retrieved/known relevant references in a database) rate on the average.[4]

Problems related to the lack of any complete English grammar in natural language systems emerge as a concern in this context. In fact, DeJong wonders whether it is actually better to have a natural language "understander" in a system, if the end user would have benefitted from the process of inputting the query in some formal language which would have required clarification of the request in the user's mind [B.3]. Though indeed the user might then have a better awareness of the constraints of Boolean searching, this could make the intermediary system seem more cumbersome and less useful. Moreover, as to the benefits DeJong suggests, one might just as easily argue the reverse: that the formal aspects of search strategy formulation distract the user from the mental clarification of the topic. After all, no one is arguing that the

[4] Blair, D. C., and M. E. Maron, "An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System," Commun. of the ACM, 28, 3 (Mar. 85): 289-99.

remote online vendors who constructed these rigid search statement formats were modelling some idealized user's mental processes.

The online intermediary systems developed to date have generally been oriented to serve the end consumer. That is, systems developed so far are aimed at replacement of the expert, almost to the complete exclusion of "intelligent assistant" systems useful to the library/information professional in a more direct sense. This is not to underrate the value of any user-friendly or tutorial system which can free an information specialist of some of the routine instructional tasks, but to wonder about the potential for "intelligent assistants" as advisors to information experts. At this state of library and information science development, maybe truly expert information systems which can interact directly with all sorts of novice end users are really not yet feasible. If not, system designers may need to address varieties of expert systems to segments of user populations, with separate systems to assist information experts.

Considering the many possible applications noted earlier in this paper, relatively little work has been done outside the area of online search systems. And too often, these end-user systems have had more of the "interface" and less of "expertise" in their functions. The process of conceptualizing an information need involves identification of the essential facets of the question, development of a search statement suitably specific, but not so narrowly worded as to result in retrieval of null sets, choice of a database or databases likely to cover the topic, execution of the search, refinement of the search strategy based on scrutiny of part of the search results for relevance feedback, and evaluation of the results after the search is done. Using the various search languages requires detailed knowledge of the commands and options for the systems involved, but only to a lesser degree, judgement. Formulating and revising a search strategy for a given information request may be more complex, though not inevitably. The area of "choice of appropriate database" has so far been implemented only in rather limited guidance, with choice often left to the user. Relevance judgments as to search results have been left to the user. All in all, expert intermediary systems in information retrieval have scarcely exhausted the problems of the domain.

Why should this be so? There seem to be several factors which tend to limit developments in intermediary systems. One problem, the all–or–nothing drawbacks of Boolean searching, would presumably not be modifiable except by the host vendors of systems. Machine learning techniques such as conceptual clustering [D.1] might be useful in studying the inverted index of an online vendor's system, as for example to produce more useful categorizations of relative levels of subject coverage by databases. This might be less than feasible outside the host system, unless perhaps used for small incremental improvements [as by ADVISE] to an intermediary system via search–by–search analysis of postings, or storage for periodic analysis of postings. Financial incentives for such large–scale alteration by a vendor of its search logic are moreover difficult to envision. The future use of associative memory in large scale integration has been suggested by one writer [D.2] as offering a possible alternative to the present systems, though financial incentives for change remain questionable for such enormous systems. Perhaps—assuming technical feasibility—for such a transformation to take place, the current competitive climate among vendors would have to be altered by the emergence of a new vendor offering such capabilities, even on a small scale trial database.

Another likely reason for limits in progress of online intermediary systems is the difficulty of developing adequate representations for some judgemental tasks carried out by online searchers. Knowing, in a climate of scarce funds, which database will yield the best results for a given query seems to be a form of judgement learned only by considerable experience with many databases. The relatively recent emergence of large bibliographic retrieval systems, and the everchanging character of economically competitive systems, suggest that expert searchers have not yet codified this form of expertise, at least not completely. Still, one could argue that the very process of developing expert systems representations could aid in this codification process.

End users also judge relevance of online search results in expert intermediary systems. At this point in time, that sounds appropriate, as library and information science has yet to find a really satisfactory algorithm to predict "relevance." Perhaps, however, an end-user system could query users as to factors involved in their judgements of relevance (e.g., was the material unrelated in fact? too old? a study the user already knew about?), and store these comments, and periodically use machine learning capabilities to group these judgements in useful classes. A Meta-expert system such as ADVISE [C.8] has adjustable evaluation functions, so that an expert evaluating user comments could run trials weighting various factors as to importance.

In the environment where end user systems predominate, a more troublesome question has been raised: whether end users really *want* to do their own searching [D.3, D.4]. Nor has a taxonomy of difficulty of search types been developed, to allow a system to refer a user to a human search intermediary where a search is difficult or inappropriate for the system involved. Again, this assumes that the system could be endowed with satisfactory knowledge of its own limitations. As evaluations of current search systems tend to indicate that novice searchers using search intermediary systems may be unreliable judges to the comprehensiveness/quality of their search results, generic searches which can be expected to be difficult/fruitless must be made known to the system if novices' searches are not to deteriorate sharply, and perhaps undetectably.

Yet are there in fact recognizable class of queries? And wouldn't any useful taxonomy have to consider difficulty of numerous subcomponents of the pre-online and online phases of the search? Relatively few systems have dealt with much of the pre-online phase, aspects such as concept formulation and development of a query statement. It would appear that better modelling of the reference interview process (between information specialist and patron) will be necessary to improve future systems' performance in pre-search stages.

Note also that only one system (INFOS, C.10] was identified which was categorically designed for information professionals outside libraries. This too was a search-aid system. As an information industry outside library institutions has already emerged in this country, the beginning of diversification of expert information systems is surely due.

VI. CONCLUSION

The evolution of artificial intelligence has been marked by reconceptualizations of the nature of intelligence. What once was considered the "acid test" of "intelligent" machine behavior, once surpassed, has tended to be superseded by new models of intelligent behavior, and newly rigorous tests.

We have identified several interesting areas for information retrieval applications to expert systems and AI in general. Among these, we suggest the following areas for future research: expert systems in IR areas outside online intermediary systems, application of conceptual clustering and other machine learning techniques to cataloguing (e.g. large conventional computerized library catalogs), or to classification and indexing schemes.

The generalist aspects of expert performance in information storage and retrieval domains may be in the short run a challenge, and in the long run a likely benefit to expert systems builders. In the beginnings sketched above, information scientists may find the promise of creative new tools, for innovative approaches to classic questions.

ACKNOWLEDGEMENT

The author wishes to thank Professors R. S. Michalski and Linda C. Smith, both of the University of Illinois, for invaluable advice and comments. This research was supported in part by the Office of Naval Research under grant N00014-82-K-0186, by the National Science Foundation under grant NSF DCR 84-06801, by the Defense Advanced Research Projects Agency under grant N00014-K-85-0878, and by the Institute for Scientific Information (ISI), Philadelphia, PA.

REFERENCES

Background

1. Cercone, Nick, and Gordon McCalla. "Artificial Intelligence: Underlying Assumptions and Basic Objectives." JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE [JASIS] 35, 5 (Sept. 1984): 280–90.
2. Forsyth, Richard (ed.). EXPERT SYSTEMS; PRINCIPLES AND CASE STUDIES. New York: Chapman & Hall, 1984.
3. Hayes–Roth, Frederick, Douglas R. Lenat, and Donald Waterman. BUILDING EXPERT SYSTEMS. Reading, Mass.: Addison–Wesley, 1983.
4. Weiss, Sholom M., and Casimir Kulikowski. A PRACTICAL GUIDE TO DESIGNING EXPERT SYSTEMS. Totowa, N. J.: Rowman & Allanheld, 1984.
5. Yaghmai, N. S., and J. A. Maxin. "Expert Systems: A tutorial." JASIS 35,5 (Sept. 1984): 297–305.

Theoretical

1. Buchanan, B. G. "Issues of Representation in Conveying the Scope and Limitations of Intelligent Assistant Programs." MACHINE INTELLIGENCE 9. Edited by J.E. Hayes, Donald Michie, and L. I. Mikulich. Chichester, England: Ellis Horwood Ltd., 1979.
2. Deerwester, Scott C. THE RETRIEVAL EXPERT MODEL OF INFORMATION RETRIEVAL. [Ph.D. Thesis, Purdue University.] Aug. 1984.
3. DeJong, Gerald. "Artificial Intelligence Implications for Information Retrieval." RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL. SIXTH ANNUAL ACM SIGIR CONFERENCE, Vol. 17, No. 4, Summer 1983. Jennifer J. Kuehn,Ed. [Preprint.]
4. Jones, Kevin P. "The Effects of Expert and Allied Systems in Information Handling: Some Scenarios." ASLIB PROCEEDINGS 36, 5 (May 84): 213–17.
5. Michie, Donald. "Automating the Synthesis of Expert Knowledge." ASLIB PROCEEDINGS 36, 9 (Sept. 84): 337–43.
6. Pollitt, A. S. "An Expert System as an Online Search Intermediary." FIFTH INTERNATIONAL ONLINE INFORMATION MEETING, London, 8–10 Dec. 81, pp. 25–32. Oxford: Learned Information Ltd., 1982.
7. Rich, E. "User modelling via sterotypes." COGNITIVE SCIENCE 3, 4 (Oct.-Dec. 1979): 329–54.
8. Smith, Linda C. "Artificial Intelligence Applications in Information Systems." ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY Vol. 15 (1980): 67–105. American Society for Information Science. White Plains, N. Y.: Knowledge Industry Publications, 1980.
9. Smith, Linda C. "Representation Issues in Information Retrieval System Design." ACM SIGIR FORUM 16, 1 (Summer 1981): 100–105. [PROC. OF THE 4TH INTL. CONFERENCE ON INFORMATION STORAGE AND RETRIEVAL, Ed. by Carolyn J. Crouch.]

10. Smith, Linda C., and Amy J. Warner. "A Taxonomy of Representations in Information Retrieval System Design." JOURNAL OF INFORMATION SCIENCE 8 (April 1984): 113–21.

11. Sparck Jones, Karen. "Proposals for R & D in Intelligent Knowledge Based Systems (IKBS)." JOURNAL OF INFORMATION SCIENCE 8, 4 (1984): 139–47.

12. Vaughn, W. S., Jr., and Anne S. Mavor. "Simulation of a Schema Theory-Based Knowledge Delivery System for Scientists." [Report, prepared for NSF, under Grant # IST 7904896.] May 1981.

Reports

1. Black, W. J., P. Hargreaves, and P. B. Mayes. "HEADS: A Cataloguing Advisory System." University of Manchester Institute of Science and Technology, Manchester, U. K., [1985].

2. Croft, W. Bruce. "An Expert Assistant for a Document Retrieval System." Proceedings, RIAO 85; Recherche Informations Assistee par Ordinateur (Grenoble, France, 18–20 March 1985). Organized by Centre de Hautes Etudes Internationnale d'Informatique Documentair (C.E.D.) and Institut d'Informatique et de Mathematiques Appliquees de Grenoble (I.M.A.G.).

3. Doszkocs, T. E. "CITE NLM: Natural language searching in an online catalog." INFORMATION TECHNOLOGY AND LIBRARIES 2, 4 (Dec. 1983): 364–380.

4. Garfield, Eugene. "Introducing SCI-MATE: a menu-driven microcomputer software package for online and offline retrieval." CURRENT CONTENTS 14 (1983): 5–15.

5. Hjerppe, R., B. Olander, and K. Marklund. "Project ESSCAPE—Expert Systems for Simple Choice of Access Points for Entries: Applications of Artificial Intelligence in Cataloging." Linkoping University, Dept. of Computer and Information Science. Report # LiU-LIBLAB-R-1985:2. June 1985.

6. Horowitz, Aren J., and Rita F. Bergman. "Improved Productivity in Multi-System Information Retrieval through the CSIN Intelligent Terminal." PROCEEDINGS, SECOND NATIONAL ONLINE MEETING, 24–26 March 1981. Comp. by M.E. Williams and T. H. Hogan. Medford, N. J.: Learned Information, Inc., 1981.

7. Marcus, R. S. "An automated expert assistant for information retrieval." THE INFORMATION COMMUNITY: AN ALLIANCE FOR PROGRESS. PROCEEDINGS OF THE 44 TH ASIS ANNUAL MEETING, 1981. Vol. 18, pp. 270–73. White Plains, N.Y.: Knowledge Industry Publications, Inc., 1981.

8. Michalski, R. S., and A. B. Baskin. "Integrating Multiple Knowledge Representations and Learning Capabilities in an Expert System: The ADVISE system." PROCEEDINGS OF THE EIGHTH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, V. 1. Held 8–12 Aug. 1983, Karlsruhe, West Germany. [Preprint.]

9. National Federation of Abstracting and Indexing Services. "NFAIS to sponsor EASYNET Service." [Press release.] Philadelphia, PA., Aug. 1984.

10. Obermeier, Klaus K., and Linda E. Cooper. "Information Network Facility Organizing System—an Expert System for Information Retrieval." 1984: CHALLENGES TO AN INFORMATION SOCIETY, 1984 ASIS Proceedings, pp. 95–98.

11. Rada, Roy. "Machine Learning for Information Retrieval; A Methodological Approach to Automatic Thesaurus Construction." Washington, D. C.: Natoinal Library of Medicine, Sept 4, 1985 [draft].

12. Shoval, Perez (Stangel). AN EXPERT CONSULTATION SYSTEM FOR A RETRIEVAL DATA-BASE WITH A SEMANTIC-NETWORK OF CONCEPTS. Ph.D. Thesis, University of Pittsburgh, Graduate School of Business Administration, 1981.

13. Smith, Philip J., Deborah A. Krawszak, and Steven Shute. "Development of an Expert System to aid in Searches of the CHEMICAL ABSTRACTS." SIGART NEWSLETTER no. 92 (April 1985): 124-25.

14. Smith, Philip J., and Mark Chignell. "Development of an Expert System to aid in Searches of the CHEMICAL ABSTRACTS." 1984: CHALLENGES TO AN INFORMATION SOCIETY, 1984 ASIS Proceedings, pp. 99-102.

15. Tenopir, Carol. "Database Access Software." LIBRARY JOURNAL 109, 6 (Oct. 1, 1984): 1828-29.

16. Toliver, David. "OL'SAM: An Intelligent Front-End for Bibliographic Information Retrieval." INFORMATION TECHNOLOGY AND LIBRARIES 1, 4 (Dec. 1982): 317-26.

17. Vickery, A., H. M. Brooks, B. A. Robinson, and B. C. Vickery. "Expert System for Referral: Progress Report No. 1." University of London, Central Information Service, Jan. 1985.

18. Williams, M.E., and S. E. Preece. "Software for the Searcher's Workbench." COMMUNICATING INFORMATION: PROCEEDINGS OF THE 43RD ASIS ANNUAL MEETING. White Plains, N. Y.: Knowledge Industry Publications, 1980. Pp. 403-405.

Selected

1. Michalski, Ryszard S., and Robert E. Stepp. "Learning from Observation: Conceptual Clustering." Chapter 11 in: MACHINE LEARNING: AN ARTIFICIAL INTELLIGENCE APPROACH. Edited by R. S. Michalski, J. G. Carbonell, and T. M. Mitchell. Palo Alto, CA: Tioga Publishing, 1983. Pp. 331-363.

2. Levine, Gwen. "Alternative Database Structures for Bibliographic Information Retrieval Systems—Their Past and Their Future." NATIONAL ONLINE MEETING PROCEEDINGS-1981, pp. 331-339. Medford, N.J.: Learned Information, Inc., 1981.

3. Tatalias, Jean. "Attitudes and Expectations of Potential End User Online Searchers." NATIONAL ONLINE MEETING PROCEEDINGS—1985, pp. 457-62. Medford, N.J.: Learned Information, Inc., 1985.

4. Bodtke-Roberts, Alice. "Faculty End-User Searching of BIOSIS." NATIONAL ONLINE MEETING PROCEEDINGS—1983, pp. 45-56. Medford, N. J.: Learned Information, Inc., 1983.

BIBLIOGRAPHIC DATA SHEET	1. Report No. UIUCDCS-F-86-956	2.	3. Recipient's Accession No.
4. Title and Subtitle Where Expert Systems Intersect with Information Retrieval: Issues and Applications			5. Report Date May 1986
			6.
7. Author(s) Gail Thornburg			8. Performing Organization Rept. No.
9. Performing Organization Name and Address Department of Computer Science University of Illinois Urbana, IL 61801			10. Project/Task/Work Unit No.
			11. Contract/Grant No. NSF DCR 84-06801 N00014-82-K-0186 N00014-85-K-0878
12. Sponsoring Organization Name and Address National Science Foundation, Washington, D.C. Office of Naval Research, Arlington, VA Defense Advanced Research Project Agency, Arlington, VA			13. Type of Report & Period Covered
			14.

15. Supplementary Notes

16. Abstracts

This paper analyzes several issues related to the intersection of artificial intelligence and information retrieval, especially in the area of expert systems. In particular, we look at problems such as determining choice and form of representations in information, transparent interface as "black box" in information systems, and lack of system self-knowledge (performance brittleness). It is indicated that in the area of expert systems for information retrieval, major concerns are user awareness of the limits of a given system, user modelling in general, and facilities for system self-awareness and explanation. Expert system interfaces to larger conventional systems, designed to serve an end user (patron or information client) directly, have dominated research in expert systems in IR, though there exist numerous other possibilities for AI research in other areas of information retrieval, such as classification, indexing, and reference referral systems.

17. Key Words and Document Analysis. 17a. Descriptors

Information retrieval
Machine learning
Expert systems

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 29
	20. Security Class (This Page) UNCLASSIFIED	22. Price

FORM NTIS-35 (10-70) USCOMM-DC 40329-P71