RP976
RP975

# DESIGN QUESTIONS IN THE DEVELOPMENT OF EXPERT SYSTEMS FOR RETRIEVAL ASSISTANCE

*Richard S. Marcus*
*Laboratory for Information and Decision Systems*
*Massachusetts Institute of Technology, Cambridge, MA*

ABSTRACT:

A number of efforts have been ongoing to investigate the prospects for expert computer systems that would match or excel human experts in providing assistance to users of retrieval systems. In this paper we contrast such efforts in terms of the models for retrieval and assistance they subsume and in terms of the techniques for performing effective retrieval and for developing expert assistance systems. Further, we state and attempt to support three premises: (1) to provide a truly comprehensive expert retrieval assistant requires a very extensive knowledge-base development; (2) there are significant questions concerning retrieval models and assistance techniques which need to be resolved in developing such expert systems; and (3) although expert retrieval assistance development is difficult, it shows promise for deepening our understanding of the retrieval process from a basic scientific viewpoint as well as for improving search techniques themselves. In support of these premises we discuss some of our recent experiences in the development of our CONIT experimental retrieval assistance system.

DESCRIPTORS: Expert systems; retrieval assistance; search models; search strategy; search modification operators.

## 1. Introduction

Despite the continuing progress of personal computers in this microcomputer age, a high percentage of bibliographic reference retrieval still requires, and is likely in the forseeable future to continue to require, accessing mainframe-based retrieval systems for their very large databases which are regularly updated and for the wide range of functionality that they maintain.

While these major, centralized retrieval systems are being gradually improved over time -- sometimes in small steps and sometimes in more major ways, it is still also true that because of complexity and heterogeneity, access to and operation of these systems poses difficulties for users, especially inexperienced *end users*.

Thus the development of intermediary assistance systems, as originally conceived and investigated by us and others [MARC81], to overcome these difficulties has continued to gather momentum. [See, e.g., [BORG85], [CASE85], [CRAM85], [LAMB85], [HAWK85], and [HUSH86]. In some cases these assistance systems are actually front ends built into the retrieval systems -- often as menu modes; the front-end systems of this kind may assist users of the given retrieval system but they do not serve a *gateway* function -- that is, assist access to a *multiplicity* of systems. Some self-standing intermediary systems may also connect to only a single system, or even only one or a few databases *within* a single system (e.g., SearchHelper for the IAC databases on Dialog [LISA84] or the newly announced GRATEFUL MED [WOOD86] for the MEDLINE and CATLINE databases on the NLM ELHILL system.

Recent research by us and others has turned to the question of how effective one can make retrieval assistance systems. One expression of this new research is given in the nomenclature: "*expert* retrieval assistance systems." The term "expert" correctly implies a certain association with a range of related activities in the area of artificial intelligence and knowledge-based systems. Unfortunately, there is a great deal of ambiguity and confusion about what the term "expert systems" implies. In fact, there is no precisely defined, commonly accepted understanding of the term. One definition is that the system *perform* as effectively for some application as human experts do. Other definitions emphasize various *characteristics of the structure of* and *techniques used by* the system. Some characteristics that, to the extent they are present individually or in combination, are taken as distinguishing what may properly be called expert systems include: (1) simulation of the manners and methods of the human expert; (2) high level of sophistication and, in particular, one or more of the following; (3) a natural-language interface; (4) a capability to explain what it is doing and why; (5) use of an extensive "knowledge base" with sophisticated inferencing and planning functions; (6) an emphasis on automation or computer assistance in preference to reliance on the intelligence of the human user; and (7) certain software techniques -- e.g., LISP or LISP-like programing languages and production-rule mechanisms.

Our own bias is to emphasize the performance criterion; the characteristics are important to the extent they determine actual performance, or may be critical for future performance enhancements, or, in our application area, shed light on information science considerations. In fact, we have argued that a too literal emphasis on some characteristics can be counterproductive; for example, "straight" natural language may not be as effective as mixed interface modes [cf. (3) above], human "expert" searchers may use non-optimal strategies [cf. (1) above], and mixed-initiative interaction (extensive user control possible) may be superior to automated modes [cf. (6) above].

By these criteria most current retrieval assistance systems fall far short of the expert level; certainly, there have been few objective, scientific experiments and analyses to test performance effectiveness. We have argued [MARC85] that experiments with a version of our CONIT system demonstrated a level of effectiveness and range of characteristics that place it marginally in the expert category. Our current research is aimed at ascertaining the prospects for a more truly expert level of search assistance.

In the remainder of this paper we attempt to illustrate the extensive development required for expert retrieval assistance systems, some of the issues relevant to such systems, how we have sought to resolve these issues, and prospects for the practical utility and theoretical importance of such developments.

## 2. Requirements for Comprehensive Retrieval Assistance

The CONIT system that we have described [MARC83a, MARC85] as having marginally expert qualities, while possibly being the most comprehensive and sophisticated retrieval assistance system (for its time, at least), still was clearly deficient in many respects as compared with an ideal expert assistant. This was true despite the fact that the program took up well over two megabytes of code and was run in a mainframe system -- the M.I.T. Honeywell-6180 based Multics mainframe time-sharing environment -- that added significantly to the capabilities of the program code itself.

The kinds of enhancements we saw as desirable were of three general categories: (1) more-or-less straightforward extensions of functionality already demonstrated; (2) new functionality of a type similar to, but with significant differences from, that previously handled; and (3) "intelligent" assistance aiming at some significant degree of understanding of users and their problems and a sophisticated capability for planning and evaluating search strategies and tactics.

CONIT provides a virtual-system approach to heterogeneous system access through a common command-language (CCL); requests in the CCL are translated to the appropriate commands for the connected remote retrieval system. However, not all retrieval-system functions and features have been fully incorporated into the CCL. For example, search specifications were not possible in the CONIT CCL in the areas of proximity searching (by word adjacency, number of intervening words, same sentence or field, etc.), important term searching, selected field searching (other than author, citation, and basic subject fields), user-specified truncation and masking, and nesting operations with parentheses. Retrieval set display operations in the CCL included only a few predefined field combinations (e.g., title, standard, and full record) but not the full variety of combinations and formats as allowed, for example, by the NLM and SDC ORBIT systems. Other functions omitted include sorting retrieved document sets by various field aspects, setting up SDI (current awareness), cancellation and individualized addressing for offline printouts, ordering the full text of documents, etc. As we have previously argued, the fact that CONIT was as effective as it was despite handling only a few functions, is an indication that we chose those retrieval functions and their augmentations that are, indeed, most important.

Another dimension of the first category enhancement is the number of heterogeneous retrieval systems accessed. Easily, several dozen major bibliographic retrieval systems are extant as compared with the three CONIT handles (Dialog, NLM, and ORBIT). Furthermore, among these dozens of systems are thousands of databases to be found versus the 300+ on CONIT's three; the typical database has 15 or more fields as contrasted with the two regularly used by CONIT for searching (author and base subject) and three groupings used for display purposes. To have a retrieval assistance system handle all these complexities we estimate it would have to be an order of magnitude bigger than the CONIT system previously mentioned.

The second category would include, as a basis, such specialized retrieval functions as simultaneous searching of multiple databases, displaying the matching terms in retrieved documents, finding common terms among (relevant) documents, and automatically applying these terms, or other terms, listed as synonyms for search terms, as alternate search terms. A second subclass of functions in this category includes those that would, strictly, not be classified as retrieval functions as such but that could *support* users of such functions in related activities. Some examples in this subclass include post retrieval editing operations on document records, inputting document records into one or more databases, modifying document records (e.g., by individual user annotation and indexing), etc.

A third subclass of functions in this category would include those not directly related to retrieval but needed by users of retrieval systems for other purposes. Electronic mail, conferencing, generalized database and word processing operations could be mentioned here. While these functions take us outside the realm of retrieval assistance, integration of diverse operations is the ultimate requirement for optimum computer system utilization.

For the most part, these enhancements to assistance systems in categories one and two can be implemented as either direct extensions of functionality already demonstrated for such systems or along similar lines to already implemented systems. For a truly comprehensive handling of all these enhancements we would imagine at least another order of magnitude increase in size of assistance programs would be required — and we have not even touched the really "intelligent" enhancements yet. Major issues associated with these prospective developments concern the detailed methods of implementation, including hardware and software structure and human factors questions — e.g., can we really scale up two orders of magnitude without major changes in structure and techniques? — and questions of evaluation in terms of analysis of benefits and costs for various classes of users, for different types of problems, and in various contexts.

In the remainder of this paper we consider issues that arise with respect to the third category of desired enhancements: the ones requiring some significantly greater "intelligence" in the program.

## 3. Issues for Expert Retrieval Assistance

In previous papers [MARC81, MARC83, MARC85] we have described some of our preliminary investigations in the development of systems for expert retrieval assistance. In the light of our experience in implementing the first version of such a system we can now discuss some issues that have been brought into sharper focus.

A basic issue concerns what *type of retrieval system and retrieval operations* are to be supported. While the Boolean-based retrieval modality is still predominant in operational systems, research investigators in many cases have pursued a statistical or probabilistic approach to retrieval (see, e.g., [SALT83]). In his design for an expert retrieval assistant, Croft [CROF85] has planned for a system which would switch basic retrieval modes including simple Boolean retrieval and more sophisticated statistical modes. Our approach has been to make the Boolean mode *the* basic retrieval mode. This choice is based on the premise that enhanced (or "smart" or "intelligent") ways of using Boolean retrieval operations will prove superior and that statistical methods will be best as *supportive* techniques rather than basic search modes in expert assistant systems. This hypothesis follows, in turn, from the further premises that it is vital to bring *human intelligence* into play in an *interactive, mixed initiative,* environment and that the human input is best induced when the human can easily perceive the nature of the matching algorithm: a term is or is not present in the Boolean scheme as opposed to the less obvious statistical algorithms where strategy modification may be effectuated by adjusting some numerical parameter(s) without direct and obvious correlation in terms of problem or document description. Because this issue is not yet resolved, it is advantageous that different approaches are being explored in order

to gain evidence for comparative evaluations.

Another issue relates to the type of information to derive from or about the user and his problem. [KORF84] has suggested the value of maintaining a general profile of each user whose elements are employed to modify the search strategy that would otherwise be devised on the basis of the current problem statement itself. We have emphasized an approach concentrating on the problem at hand. The new expert CONIT solicits from the user a problem title, summary description, and certain other problem *qualifications or expectations:* maximum amount of time and money willing to be spent on the search, estimated numbers of relevant documents that exist (the recall base) and that the user already knows about, recall desire, and desired document types. Resource limitations are used to provide warnings when the respective limits are approached. Desired recall and document types influence information given on possibly relevant files. Estimates of the recall base and known relevant documents will be used for (fractional) recall estimates. All of the above qualifications will also be used for more basic system search planning, although the details of those operations have not yet been fully designed. The summary problem description is currently used only to help the user formulate the problem in his own mind and as a resource for the user to find search terms. The problem title may actually be used as a search term.

Clearly, questions that remain for the problem description still cover a wide range: what elements to collect, how to collect and represent them, how to use them and to what effect.

The critical part of the problem description for subject-oriented searches — which, we claim, are at the heart of bibliographic and text-oriented retrieval and distinguish that kind of retrieval from data retrieval — is the description of the *topic*. We have adopted a *Boolean combination of conceptual aspects* as the language for expressing the topic formalization. We have indicated above our rationale for this choice in terms of precision of representation, understandability for the user — especially in evaluating and modifying search strategy, and a felicitous correspondence with most major retrieval systems. We do not believe that any differential relative importance (ranking) of conceptual aspects is usually of prime concern for most users and problems; however, where this *is* the case it can be handled within the Boolean framework by having (a few) alternate topic representations selectively eliminating the less valued aspects. Actually, such selective deletions will be part of the search strategy modification procedures suggested by the expert assistance system. The user is then able to make distinctions based on a differentiated response to different strategies. Note that even this (usually inconsequential, we claim) limitation on a continuous *numerical* ranking of topic aspects does *not* preclude a ranking of search *results*; as we have previously indicated [MARC83a], such ranking can be achieved by differentiating results based on the *nature of the matching algorithm* (title-word match more relevant than descriptor match which is more relevant than abstract word match; exact [phrase] match more relevant than [non-adjacent] stem match; etc.) Eventually, we believe, a full-fledged aspectual ranking within the basic Boolean scheme may provide some second- or third-order benefits, but the introduction of such a complication at this early stage of our expert assistance investigations would be counterproductive.

A highly significant feature of our expert system development is the *clear separation of the topic representation from the search strategy*. This, in fact, may be considered a prime differentiation between what (the better) existing intermediary assistance systems *can* do and what expert assistance systems *aspire* to do. Thus a good intermediary system will allow a user to create and easily modify a search strategy, but we must attain the expert level before there are capabilities for having the user, with system aid, develop a formalized topic representation and, from that, *several distinct* search strategies with clearly understood (by user *and* system) *rationales* for going from one strategy to another for a given topic and for evaluating the results.

Thus, for example, a conceptual factor included in the topic representation may have a direct correlation to a search component strategy (SEARCH A). However, it may be decided -- tentatively, at least — that a more comprehensive, but not too imprecise strategy (SEARCH B) *may* be achievable by *not* including a search component for that conceptual factor. For expert assistance we believe the modified search strategy should have a standing *separate from* (*not replace*) the original strategy and that both strategies should be associated with the same conceptual representation. In that way the system better "knows" to assist users compare results of different strategies (e.g., look at the "SEARCH B AND NOT SEARCH A" set) and evaluate the searches themselves (e.g., to suggest that the user consider whether it is the absence of the given factor that is causing any irrelevance in SEARCH B — in which case, perhaps, further suggest that the user [and/or systems] seek a more recall-oriented [broader] formulation of the search for the given factor [short of the extreme of elimination entirely, as in SEARCH B]).

While we continue to affirm our initial decision to keep conceptual representation and search strategy separate, we have found it challenging to devise exposition to explain the distinctions where we find them useful and hide the distinctions where they might be more confusing than helpful. A manifestation of this dilemma is observed at the preliminary stage of the assistance session when the system attempts to help the user formulate the conceptual nature of the problem. One would like to make an unrestricted Boolean structure available to the user both for conceptual formulation and search strategy. However, that is likely to be overly complicated and confusing for most end users and problems. In our first expert system [MARC81] we reduced the level of apparent complexity by making the conceptual framework a simple intersection of conceptual factors. However, as simple as that framework was, we found users had difficulties with the very notion of identifying conceptual factors. Prompted by our positive experience with "standard" CONIT, we decided to base the user's initial conceptual formulation on the user's (natural language) keyword phrase which the system elicits from him. Each significant word in the keyword phrase is taken as an initial conceptual factor. Thus, in a way that appears to be natural for users, the initial conceptual formulation is *drawn from him* without any explicit effort. Rather, the system *explains to the user* what it has done in the way of formulating the initial conceptualization.

Another step is to transform the conceptual formulation into an initial search strategy. Here, again, we believe our preliminary attempt at an expert system [MARC81] has led us to a better approach. In that preliminary attempt we suggested that the user try to expand each conceptual factor by adding alternate terms *before the first search*. We found that this took some time, that inexperienced users may not do a good job without

feedback, and that the initial search takes longer and that it is more costly and harder to evaluate. This led to our current scheme which is to take each conceptual factor and perform a direct search on it and intersect the resultant components. This superficially would appear to be a narrow strategy emphasizing precision: however, because the individual searches are done on a truncated stem all-fields algorithm, the result is, on average, actually a good balance between recall and precision, a blend among basic strategies enunciated, for example, by Markey and Atherton [MARK78].

Another decision we have made -- all these decisions are really issues that need to be justified by experimental testing and analysis -- was to recommend that the user run this initial search formulation to get quick feedback rather than try to optimize the strategy *before* the fact.

File selection, (including getting information about the files for selection purposes) running the search, and displaying document information for any of the resultant search sets are all controlled in the ASSIST mode by selection from menu options. The user can, at any point, avoid menus by using commands, which have been illustrated by the ASSIST mode as it executes them based on the user-selected options. If the command is one that seeks information about how to use the system, it is performed and the ASSIST mode with menus is maintained. If the command is a "basic retrieval" command, (e.g., construct or run a search or display results), expert CONIT assumes the user is ready to use command mode and the menus are suppressed. The user can get back to an appropriate menu in the ASSIST mode by giving the HELP command. Also, system detected errors in commands will evoke suggestions to the user to request particular help. A major question is how effective this integration of menu and command modes will be.

Search modification can also be guided by ASSIST explanations and menus. The first choice is whether the user wants recall-enhancing (broadening) or precision enhancing (narrowing) suggestions. For example, if broadening is chosen, the following menu of specific options are presented:

1. Add alternate (synonymous) terms for some searches.
2. Drop some factors (use fewer ANDs).
3. Relax combination requirements (to change ANDs to ORs) -- this means retrieving documents from EITHER one OR another search instead of requiring the documents be in both searches.
4. Replace some search terms with better terms.
5. Replace phrase, exact, or long terms with word or shorter truncated stem searches.
6. Identify other files to search in.
7. Do an entirely new search or combination of searches.

Once the user selects a given broadening technique he is given explanation, including examples, of when, why, and how that technique can be applied. If the user decides to employ this technique, he is then prompted for the search(es) to which he wants to apply the techniques. Next, specific factors are elicited from the user -- which terms to drop, add, or otherwise operate on. After all necessary parameters of the technique are elicited the ASSIST mode issues the appropriate CONSTRUCT commands to generate the desired modified searches and search strategy. Next the user is given the option to run the modified strategy or perform additional modifications.

As we have indicated before, the modified searches are generated *in addition to* (do *not* replace) existing searches so that nothing is lost; the old searches can still be used and comparisons between old and new are also possible. In particular, if a change is made to some *component* of a search strategy, not only is a new component generated, but also new searches are created for any search which *uses* that component. As in other situations, operations performed at ASSIST level are accomplished through commands and so are directly accessible at the command level. Thus, if a modification to search S5 generates a search S12, the command

USE S12 for S5

will create new searches for those searches in which S5 is a component (or sub-component) search.

Modified searches are associated with the same rep ("rep" is our generic name for the structural unit in the conceptual representation formulation) as for the original search; thus the system "knows" what search variations go with any conceptual aspect. Our modification assistance is, so far, aimed at assisting users modify the searches, not the conceptual representation. There are potentially subtle effects among the interrelationships between reps and searches for which we have not yet devised fully comprehensive explanatory and assistance modalities. Thus the broadening techniques listed above are described in terms of search modifications. There are corresponding modifications at the representational level but we have not yet decided when and how to broach their potential application to the user. For example, as we have mentioned before, a user might want to drop a search factor (see broadening technique 2) while keeping the same conceptual representation. On the other hand it may be that on rethinking the problem, it is, in fact, the *conceptual factor* which should be dropped. How to handle these potentially subtle and complicated distinctions in the assistance mode is one important research issue for expert Boolean assistance systems.

We have begun to see the need for defining certain *operators* that can express some of the kinds of modifications desired. Thus we have defined STEMMING, UNSTEMMING, TRUNCATION, UNTRUNCATION, and various PROXIMITY operators by which existing searches or search combinations can be modified in designated ways. Other modifications seem best performed as direct consequences of the CONSTRUCT command which creates Boolean combinations of existing searches. The desired range and scope of operators and operations for search modification is another important research issue.

Other major research issues have to do with search and search modification evaluation, including estimation of recall and cost parameters. We are now completing detail designs for including these operations within the expert assistance system. Once that is accomplished the high-level planning operations by which we can make information retrieval a truly scientific, quantifiable decision-making process can be tackled. (Planning proposals are included in [CROF85] and [BRAJ85]).

In summary, we have discussed a number of design questions that are posed by development of expert retrieval assistance systems. These questions indicate the extensive nature of the development required to achieve computer expertise in this area; however, they also point to hope for significant advances in the theory and practice of information retrieval.

ACKNOWLEDGMENTS:

The research described in this paper was supported by the National Library of Medicine under Grant LM-03210 and by the National Science Foundation Division of Information Science and Technology under grant IST 8414485. The author also acknowledges the many vital contributions of his project co-workers including the following students: Man-Kam Kenneth Yip, William Tiger Lee, Steven Harvey Schwartz, Monica R. Gerber, Ricardo A. Cardenas, Hing-Fai Louis Chong, Aleks Gollu, and Man-Wah Colina Yip.

REFERENCES:

[BORG85] Borgman, Christine L., Donald Case and Charles T. Meadow. "Incorporating Users' Information Seeking Styles into the Design of an Information Retrieval Interface." *Proceedings of the 48th Annual Meeting of the American Society for Information Science.* 22:324-330; October, 1985.

[BRAJ85] Brajnik, Giorgio, Giovanni Guida, Carlo Tasso. "An Expert Interface for Effective Man-Machine Interaction." in *Cooperative Interactive Systems*, L. Bolc (Ed.), Springer Verlag, 1985

[CASE85] Case, Donald, Christine L. Borgman, and Charles T. Meadow. "Information-Seeking in the Energy Research Field: The DOE/OAK Project. *Proceedings of the 48th Annual Meeting of the American Society for Information Science.* 22:331-336; October, 1985.

[CRAW85] Crawford, R.G. and H. S. Becker. "Toward the Development of Interfaces for Untrained Users." *Proceedings of the 48th Annual Meeting of the American Society for Information Science.* 22:236-239; October 1985.

[CROF85] Croft, W. Bruce. "An expert assistant for a document retrieval system." *Proceedings of RIAO 85,* Grenoble, France. March 1985.

[HAWK85] Hawkins, Donald T. and Louise R. Levy "Front End Software for Online Database Searching Part 1: Definitions, System Features, and Evaluation; Part 2: The Marketplace; and Part 3: Product Selection Chart and Bibliography." *ONLINE* 9(6):30-37; 10(1):33-40 and 10(3):49-58. November, 1985 and January and May, 1986.

[HUSH86] Hushon, Judith M. "How Micro-CSIN, a New Gateway to Online Systems, Stacks Up." *Proceedings of the Seventh National Online Meeting.* 203-210. New York; May, 1986.

[KORF84] Korfhage, R. R. "Query Enhancement by User Profiles." in *Research and Development in Information Retrieval* Cambridge University Press. 1984.

[LAMB85] Lamb, M.R., E. W. Auster, and E. R. Westel. "A Friendly Front End for Bibliographic Retrieval: The Implementation of a Flexible Interface." *Proceedings of the 48th Annual Meeting of the American Society for Information Science.* 22:229-235; October 1985.

[LISA84] Lisanti, Suzana. "The On-line Search." *BYTE.* 9(13):215-230; December, 1984.

[MARC81b] Marcus, R.S. "An Automated Expert Assistant for Information Retrieval". *Proceedings of the 44th ASIS Annual Meeting.* 18:270-273; October, 1981.

[MARC83a] Marcus, R.S. "Computer-Assisted Search Planning and Evaluation." *Proceedings of the 46th Annual Meeting of the American Society for Information Science.* 20:19-21; October, 1983.

[MARC83b] Marcus, R.S. "An Experimental Comparison of the Effectiveness of Computers and Humans as Search Intermediaries." *Journal of the American Society for Information Science* 34 (6):381-404. November, 1983.

[MARC85] Marcus, R.S. "Development and Testing of Expert Systems for Retrieval Assistance". *Proceedings of the 48th ASIS Annual Meeting.* 22:289-292; October, 1985.

[WOOD86] Woodsmall, Rose Marie; and Anderson, John E. "NLM Introduces GRATEFUL MED: A User-assisted Interface to MEDLINE and CATLINE". *The NLM Technical Bulletin.* No. 202:1-8; February, 1986.

[YAGH84] Yaghmai, N.S. and Maxin, J.A. "Expert Systems: A Tutorial." *Journal of the American Society for Information Science.* 35(5):297-305; September, 1984.

[YIP81] Yip, Man-Kam. An Expert System for Document Retrieval. Master of Science Thesis in Electrical Engineering and Computer Science, Massachusetts Institute of Technology. Cambridge, MA. February, 1981.