Development and Testing of Expert Systems for Retrieval Assistance

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

ABSTRACT:

Investigations have continued in the development and testing of experimental expert systems, under the generic name CONIT, for assisting all classes of users to access and operate mutliple, heterogeneous online retrieval systems. New experiments have now been run in a more "open" and "realistic" environment — users operating an enhanced CONIT at times and places of their convenience (e.g., their own offices) but with definite monetary constraints. These experiments indicated that, while previous results showing easy learning with some relevant retrieval still obtained, the time and cost pressures did have a negative impact on the effectiveness of the searching. Some of the improvements in the enhanced CONIT showed promise for increasing effectiveness and efficiency in time/cost pressured environments. Major additional improvements were deemed possible and highly desirable in several areas including (1) a true *integration* of menu and command modes, and (2) a more directed, dynamic, and comprehensive capability for planning assistance and search *evaluation* — both before and after searches and contemplated search modifications were run. We have designed and partially implemented a modified CONIT intermediary system to evaluate these additional improvements. The new integrated command/menu interface teaches commands by dynamic example in response to menu selections while allowing users to give command or menu inputs at their own discretion without explicitly changing modes.

1. Introduction

Intermediary assistance systems for the online bibliographic retrieval application have been proliferating in the last few years as the perception has deepened that such systems can, in fact, improve accessibility to and operability of the retrieval systems, especially for the less experienced user. Such intermediary systems as CSIN (Chemical Substances Information Network), SCI-MATE, In-Search (now upgraded to Pro-Search), Data-Ease (IT), Search Master, SearchHelper, PC/Net-Link, EasyNet, etc., have been announced (See, for example, [LISA84]).

Each of these systems has, in general, at least one feature that distinguishes it. Thus, for example, SCI-MATE features an internal DBMS (database management system) into which document records from the retrieval systems can be transformed for local common database creation; In-Search features appealing extensive graphics help; Data-Ease features a technique for automatic database selection from a free-vocabulary topic statement; Search Master offers a microcomputer-based version of CSIN's "scripts" for handling specialized search situations for inexperienced users; and EasyNet allows users to charge their search costs to their credit card accounts.

What have tended to be lacking are adequate objective scientific evaluations of how effective the intermediary assistance systems *are* in any absolute or comparative senses — and how effective they *could be* (allowing for the further development of assistance techniques).

For our own CONIT intermediary assistance system we performed a series of controlled experiments [MARC83b] in which data were collected for such parameters as recall and time. These experiments — which were performed on a so-called "standard" version of CONIT — showed that, in certain contexts, effectiveness in terms of recall for computer intermediaries could average as high as effectiveness achieved by expert human (information specialist) intermediaries, although at the cost of greater session time at the computer terminal.

Some of the features and techniques of standard CONIT to which we attribute its relative success include: (1) a simple, easy-to-use *common command language* (CCL) leading to a *translating, virtual-system* approach to a *network* of heterogeneous systems and databases; (2) extensive, dynamic *computer assisted instruction* (CAI) for teaching users how to use the CCL and how to *develop and modify search strategies* in a heterogeneous database context; (3) selection of only the *most important retrieval functions* to include in the CCL and only the most important core of those to recommend to the users initially; (4) the *automatic handling* of certain procedures such as login protocols, search overflow conditions, and search regeneration and repetition in multiple databases; (5) an automated *Boolean, all-fields, content-word-stem search* operation based on a user's free-vocabulary topic phrase description. These techniques have been supported by a research and development philosophy which incorporates a *modelling* approach to the search process and a table-driven, *production-rule* approach to software development.

In view of the relatively sophisticated nature of the standard CONIT system and its research and development environment, as well as the achievement of results comparable with expert humans in some contexts, it may be justified to categorize standard CONIT as an "expert" system [YAGH84]. However, some of the success of standard CONIT is attributable [MARC83b] to an emphasis on certain rather simple search techniques (see point 5 above) in contrast with the greater emphasis by human experts on more sophisticated searching (e.g., proximity and controlled-vocabulary searching). In addition, human expert intermediaries clearly are much more sophisticated and effective in understanding the user's problem and applying that understanding to the search process in *certain contexts* (this is true despite the fact that CONIT's simple search techniques appear better *on average* than more sophisticated ones). Recognizing these facts, we have sought to investigate the extent to which increasing effectiveness and efficiency could be provided in intermediary retrieval assistance systems through the *selective* incorporation of more sophisticated and, perhaps, more human-like techniques.

The remainder of this paper discusses two topics: (1) the augmentation and further testing of standard CONIT to determine whether previous results would hold in more general contexts and (2) the development of intermediary systems incorporating greater sophistication and human-like expertise.

2. Experimental Testing and Evaluation of Enhanced Standard CONIT

One enhancement to standard CONIT which serves both modeling and experimental purposes was to incorporate within CONIT a facility to identify and record all computer-related cost components, both from the remote retrieval systems and the intermediary system. Costs that must be identified include those associated with connect time for the systems and network connectors and with online and offline print charges. This facility permits not only retrospective review of charges but is also the basis for a prospective analysis of future costs for planning purposes. Associated with the cost analysis facility is a new accounting facility which permits individual and group accounts on CONIT and the several remote retrieval systems. Costs for the different accounts are recorded and cumulated dynamically and maximum costs can be set preventing users from accumulating costs beyond set limits.

A new facility, exemplifying the incorporation of intelligent aids in intermediary systems, dynamically selects an appropriate path through the network to a desired database. Rather than simply take a fixed priority list of paths, this path selection algorithm keeps a record of path selection attempts, as well as retrieval system schedules and database availability, and chooses new connector, network, and system links based on current indications of success or failure.

The intermediary system also had to be maintained in the light of various changes to the retrieval systems, including database and protocol changes. (There are now over 300 different databases in the 3 systems — NLM ELHILL, DIALOG, and SDC ORBIT — to which CONIT connects.) In addition, a new condition-handling subsystem was installed which provides a more complete basis for circumventing the effects of error situations, where possible, and — in any case — logging special conditions for later analysis.

With our updated and enhanced CONIT system we have performed new and wider-ranging evaluations of the models and techniques we have developed. Our recent experiments break new ground over our previous experiments in at least four major respects. First, we have gained experience with the new functional capabilities of CONIT. Second, we have switched from the strictly controlled environment in which users operated the computer under our direct observation from terminals in our own laboratory, to an "open" environment in which users engage the system at times and places entirely of their own choosing — generally in their own labs or offices. There is some loss of information in the open environment but this is more than counterbalanced by the greater realism achieved and potential for more extensive user participation in the experiments.

A third experimental variation enables us to obtain additional information: a record of the computer response time and the user response (think) time for each operation — previously, we could not distinguish these two. A fourth, and highly significant variation, is the user involvement with costs. In previous experimental situations the project bore the full cost of all computer charges; in these experiments the full costs are being borne by the user's organization and he is made aware of the amount he has spent so far and the maximum amounts expendable for any one session, for himself in all sessions, and for the organization as a whole. Along with the open environment context, then, these experiments have a much more realistic setting than the previous ones.

The two main organizations participating in the experiments so far are the Hudson River Foundation (HRF) — a New York based environmental research organization, and the M.I.T. Laboratory for Computer Science (LCS) — particularly its medical (clinical decision-making) group. The potential user group at HRF included some 44 persons — primarily biologists and other environmental researchers. The LCS group contained approximately 25 persons, including professors, research staff, graduate and undergraduate students, and associated physicians. The potential users were all given general information about our project and a user's manual (a collection of *some* of the online instructional messages including CONIT access instructions). They were told that their organization was paying for the computer costs and that their initial personal limit was $200 — increases negotiable with their organization.
A summary of results follows. From the HRF group there have been 21 users with a total of 49 computer sessions in a 10-month period. From the LCS group we identified 65 sessions by 11 different users in a 12-month period. Taking both groups as a whole we found that session times ranged from 3 to 74 minutes; typical times were in the range of 20 to 30 minutes. Relevant documents were found in about 60% of the sessions. Typically, 5 to 10 relevant documents were found; however, the figure for absolute recall ranged from 1 to a few cases where extensive bibliographies of 100 to 600 documents were developed.

An analysis of those sessions where no relevant documents were found determined that: (1) about 25% were sessions where the users were just "testing the system" prior to making a serious attempt to find documents on a topic; (2) another 25 percent were where either there *were* no relevant documents for the topic or the user's search strategy was faulty; and (3) the remaining 50% were sessions flawed by system bugs (in almost all of these cases users subsequently found relevant documents in later sessions).

The results of these open-environment experiments obtained to date appear to substantiate basic conclusions from the previous experiments: most users appear to have obtained relevant and useful document references fairly quickly without any instruction other than that given by the intermediary system.

More than 30 different databases from all three systems were searched by the users as a group. (One user searched 11 databases.) Practically all databases searched yielded relevant document references. The broad range of topics and databases searched further supports the conclusion that even among relatively homogeneous groups (and even for single individuals) in the biomedical field there is a need to utilize a wide variety of databases and systems.

However, there were substantial differences from the previous experimental results in terms of the sessions' effectiveness, length, and costs. Average session lengths were less by a factor of about 5 and absolute and fractional recall appear (we did *not* perform the detailed analyses required to determine fractional recall precisely) to be reduced correspondingly.

There appear to be two major factors leading to these correlated differences: cost and motivation. Cost considerations clearly dampen users enthusiasm to persist online until their information needs are fully satisfied. Perhaps even more important, we now see that the group of users from our early experiments were, by and large, highly motivated as volunteers who had important immediate needs for information. On the other hand, while a few of the users in the current round of could be so described, most had less pressing

concerns and were satisfied with less comprehensive results. This is especially true of the HRF users who, in some cases, were responding to a call by the Foundation to try out the system for evaluative purposes. This factor is further highlighted by the fact that, contrary to the HRF situation, several of the LCS users have become quite regular users — in fact, two have become *intensive* users with over 18 sessions each. Another factor tending to reduce session times, though not necessarily recall, is that users have easy access to CONIT whenever they want; therefore, there is less felt need to "cram" in a lot of searching in any one session. The greater computer expertise of the LCS group did seem to lead to quicker learning on average, but there were cases where the "hacker" syndrome (don't read the instructions, just try what seems reasonable) seemed to *retard* learning.

In addition to our own experiments we have begun to permit experiments and demonstrations with CONIT by fellow researchers. So far there have been several users in this category in locations around the country. A number of these research users, including a graduate student at UCLA, are interested in performing comparative analyses of CONIT and other intermediary systems. We intend to broaden the scope of this kind of activity so as to make our own work better known and more beneficial to others while enhancing the opportunities for scientific interaction among researchers.

3. Expert System Development

The open environment experiments have supported our previous conclusions about (1) the utility of the standard CONIT assistance techniques and (2) the desirability of incorporating into the intermediary additional sophisticated intelligence, as well as the calculating and memory powers of the computer, in order to further increase the cost effectiveness of the computer intermediary. One technique we have tested [MARC83b] involves automatic database selection as a consequence of searching the user's free-vocabulary topic expression in a multidisciplinary database in conjunction with a matrix containing weighted database relevancy estimators for classification terms found in the documents retrieved in searching the multidisciplinary database. This technique looks promising but has been too costly for our project environment in terms of maintaining the rather complex software compared with the simpler and reasonably effective aids for assisting database selection in standard CONIT.

Another approach toward these ends is to incorporate elements of our search models into the intermediary assistance system. In an early manifestation of this effort Yip [YIP81] implemented an experimental intermediary system, termed EXPERT-0, which had a rudimentary form of the five stages of our search model. EXPERT-0 was implemented after the style of expert systems of the artificial intelligence genre and facilitated a question-and-answer dialog by which the intermediary system assisted the user in preparing a Boolean topic representation and a search strategy. EXPERT-0 then automatically executed the search strategy, led the user to review the catalog records of documents thus found, and prompted the user to revise the search strategy after reviewing this feedback — particularly in adding or deleting individual search terms and whole concept factors based on relevance considerations.

As we reported in [MARC81b], there appeared to be important potential for enhanced assistance in aspects of EXPERT-0 but a major deficiency in this preliminary implementation was a lack of integration of the expert modes with the "standard" CONIT modes. In particular, we concluded that for a truly effective intermediary assistant one needed not only the relatively few, albeit highly automated, modes of EXPERT-0 but also the many modes and functions provided by CONIT — including the ability for the user to direct or *initiate* activity (e.g., through the command mode) as well as for the computer system to direct or control operations. Thus we have striven to design, implement and evaluate an enhanced CONIT that would integrate and extend the computer-directed, expert-styled, formalized planning and evaluative, menu-oriented features of EXPERT-0 with the more extensive user-directed, command-oriented, informally-tutorial features of standard CONIT.

The basic design for such a *mixed-initiative* system was described in [MARC83a]. Toward this objective a new *search cataloging* facility was installed to permit search statements (and search results when executed) across the network of retrieval systems to be saved in individual and group "catalogs". This facility will permit testing and evaluation of the concept of the (possibly long-term) development of search strategies and their subsequent utilization within individual or group usage scenarios.

Various data structures devised for the catalog system provide a basis for integrating the standard search structures with the new problem representation and evaluation structures. A meta level was added by which users could in command mode (1) construct topic representations; (2) generate search strategies from these representations; (3) execute the search strategy; (4) modify the representations and/or the search strategies; (5) maintain the appropriate correspondence of *current* search strategies and representations while remembering the past associations (no system that we know of maintains this potentially critical distinction between conceptual representations and actual searches); and (6) run modified searches while automatically remembering and reusing previously run component searches, when possible. In addition, we are in the midst of adding a meta-meta level, labeled "ASSIST," which assists users with a question-and-answer menu-oriented mode in performing appropriate construction, generation and execution operations. One unique feature of ASSIST is the explication for the user of the commands that are implied by his menu selection actions and answers to questions so as to help the user understand what, in fact, is being done for him and ease the way to user-directed command operations if and when he chooses to take such initiatives.

A preliminary version of this interface mode was programmed and tested; it proved unsatisfactory because it required users to switch into a specific command mode to issue commands and it gave users the standard CONIT detailed instructional information about commands when they were executed by the ASSIST mode. Both of these "features" proved confusing to users. In a newer design we allow commands to be executed *without* transferring to a different mode; the computer system distinguishes a command response from ordinary menu selection or argument fill-in. In addition, command-level CAI (computer assisted instruction) is suppressed, except for certain error conditions. The user is automatically transferred to command mode when he issues a non-instructional command. Command mode is now thought of as an experienced user mode and provides little in the way of CAI (except by explicit user request). Whenever the user issues the HELP command, the intermediary system returns to the ASSIST mode and provides a menu appropriate to the user's current context with

respect to his search session.

We are currently completing the design and implementation of the context-sensitive help mode. It not only will allow much more directed instruction on how to specify commands but also will allow the intermediary to tailor its instructions and suggestions explicitly and particularly to the user's individual situation, based on his problem statement and the results achieved so far in the session. We hope to begin testing of this new enhanced expert intermediary system this year.

ACKNOWLEDGMENTS:

The research described in this paper was supported by the National Library of Medicine under Grant LM-03210, by the National Science Foundation Division of Information Science and Technology under grants IST 8201842 and 8414485, and the Hudson River Foundation under grant 4/84X-4. The author also acknowledges the many vital contributions of his project co-workers; in particular, computer science theses were written by Michael Gordon Feinstein, Edward Barry Kutin, William Tiger Lee, Steven Harvey Schwartz, Sara Beth Weber, and Robert Michael Walmsley (all have written Bachelor's theses at M.I.T. and Mr. Lee has also completed his Master's thesis). Ricardo A. Cardenas, Hing-Fai Louis Chong, Aleks Gollu, and Man-Wah Colina Yip are also currently working on the project.

REFERENCES:

[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.

[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.