DOCUMENT RESUME

ED 229 009 IR 050 162

AUTHOR	Marcus, Richard S.
TITLE	Investigations of Computer-Aided Document Search Strategies.
INSTITUTION	Massachusetts Inst. of Tech., Cambridge. Laboratory for Information and Decision Systems.
SPONS AGENCY	National Science Foundation, Washington, D.C. Div. of Information Science and Technology.
REPORT NO	LIDS-R-1233
PUB DATE	1 Sep 82
GRANT	NSF-IST-80-06516
NOTE	74p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Biomedicine; *Computer Assisted Instruction; Costs; Feasibility Studies; *Information Retrieval; *Man Machine Systems; Medicine; *Online Systems; *Reference Services; Relevance (Information Retrieval)
IDENTIFIERS	End Users; *Search Intermediaries

ABSTRACT

Controlled experiments were conducted with an enhanced experimental intermediary system, CONIT (COnnector for Networked Information Transfer), to test how effective such a system could be in assisting end users in online searching of medical and biomedical literature. A total of 16 end users, none of whom had previously operated CONIT or any of the four bibliographic retrieval systems used in the study, performed searches on 20 different topics with no assistance other than that provided by CONIT itself (except to recover from machine and software problems). The same topics were then searched by human expert intermediaries (librarians) with the end users present. Sometimes CONIT and sometimes the human expert were clearly superior in terms of such parameters as recall and search time. In general, however, end users searching alone with CONIT achieved somewhat higher online recall at the expense of longer session times. It was concluded that advanced experimental intermediary techniques are capable of providing search assistance whose effectiveness at least approximates that of human intermediaries in some contexts. Details of the enhanced CONIT system and its costs are also discussed, as well as the possibilities for even more advanced intermediary systems, including those which perform automatic database selection and simulate human experts. A 60-item bibliography is provided. (Author/ESR)

September 1, 1982

Report LIDS-R-1233

INVESTIGATIONS OF COMPUTER-AIDED DOCUMENT SEARCH STRATEGIES

by

Richard S. Marcus

This material is based upon research supported by the *National Science Foundation Division of Information Science and Technology*, under Grant Number NSF IST-80-06516. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author, and do not necessarily reflect the views of the National Science Foundation.

LABORATORY FOR INFORMATION AND DECISION SYSTEMS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Cambridge, MA 02139

ABSTRACT

Controlled experiments have been conducted with the enhanced experimental intermediary system CONIT (COnnector for Networked Information Transfer) to test how effective such a system could be in assisting end users in searching compared with human expert intermediary search specialists. Some 16 end users, none of whom had previously operated either CONIT or any of the 4 connected retrieval systems, performed searches on 20 different topics using CONIT with no assistance other than that provided by CONIT itself (except to recover from computer/software bugs). These same users also performed searches on the same topics with the help of human expert intermediaries. Sometimes CONIT and sometimes the human expert were clearly superior in terms of such parameters as recall and search time. In general, however, users searching alone with CONIT achieved somewhat higher online recall at the expense of longer session times. We conclude that advanced experimental intermediary techniques are now capable of providing search assistance whose effectiveness at least approximates that of human intermediaries in some contexts. Also discussed in this report are details of the enhanced CONIT system and its costs. In addition, there are discussions of the possibilities for even more advanced intermediary systems, including those which perform automatic database selection and simulate human experts.

ACKNOWLEDGMENTS

This report constitutes the final report for Grant NSF IST-80-06516 entitled "Investigations of Computer-Aided Document Search Strategies". The grant was awarded by the Division of Information Science and Technology of the National Science Foundation and the grant period ran from June 15, 1980, to June 14, 1982. Additional support for the overall project of which these investigations were a part was given by the National Library of Medicine under NIH Grant LM03120.

The author wishes to acknowledge the many contributions of his colleague, Professor J. Francis Reintjes, who was co-principal investigator on the NSF grant. We recognize the important contributions made by the following M.I.T. graduate, undergraduate, and special students who participated in the project: Carolyn Cole, Thomas. J. Deane, Michael G. Feinstein, Felix Samuel Hsu, Victor K. Inada, Mark V. Kane, Esther P. Jaffe, Siu-Wah Victor Kwong, Tuan Anh Luong, Wai-Cheong Dick Man, Richard E. Robbins, Sara B. Weber, and Man-Kam Kenneth Yip. In addition, Dr. Geoffrey E. Bock, consultant, and Ms. Weber are acknowledged for their work in the performance of the experiments described herein; an especial thanks is due to Ms. Weber for her help in preparing this report. Finally, we wish to express our sincere appreciation to the many end users and the several information specialists who participated in our experiments — for reasons of confidentiality we do not give their names.

CONTENTS

		Page
1.	Introduction	1
2.	Enhanced CONIT	3
3.	Framework of Experiments	12
	A. Experimental Objectives and Subjects	12
	B. Experimental Procedures	16
4.	Experimental Results	21
	A. Quantities Measured	21
	B. Analysis of Results	25
5.	Instructional and Learning Considerations	36
6.	Automatic Database Selection	41
7.	Interpretation of Results	44
	A. Potential Experimental Biases	44
	B. Tentative Conclusions	47
8.	The Current Prospects for Intermediary Systems	50
	A. Cost/Benefit Considerations	50
	B. Other Computer Intermediary Systems	54
9.	The Future for Intermediary Systems	58
10.	References	64

Tables

1.	Experimental Users and Topics	14
2.	Experimental Statistical Results	22

1. Introduction

The end user of information provided by online bibliographic retrieval systems usually requires a human information specialist acting as as intermediary searcher to access the databases effectively (see, e.g., Wanger [WANG76]). Research and development activity in the recent past (see, e.g., Marcus and Reintjes [MARC81a]) has attempted to determine the extent to which computerized intermediary systems(1) acting as assistants to the end user could replace the need for human search intermediaries. One such intermediary system is CONIT [MARC81a].

CONIT connects to several different retrieval systems but presents to the user what appears to be a single, common (virtual) system by allowing user requests in a common language. These requests are translated by the intermediary into the appropriate commands for whatever retrieval system is being interrogated. The intermediary itself provides instruction so that even inexperienced end users can operate it. Additional search aids are provided to help the user search the heterogeneous databases.

Experiments with one version of CONIT (identified as CONIT 3) have shown [MARC81a] that it is possible for computer intermediary systems to assist end users, who had no previous experience in operating retrieval systems, to obtain in this way information they needed from dozens of heterogeneous databases on four different systems. In these experiments users found relevant information typically beginning within 20-30 minutes of their

---

(1) Note: We and others (for example, Goldstein [GOLD78]) have at times in the past referred to such a computerized assistant system as an "interface." However, this term can be confused with the term 'computer/human interface' which we take to mean those aspects of an online interactive computer system which are at the boundary between the human and the computer and which are used directly by the human and through which he or she 'sees' the system (e.g., the command and response languages, terminals, etc.). Also, as these assisting systems have incorporated more and more functions, not all of which may be directly apparent to the human user, we have been led to adopt the term "intermediary system" as was used, for example, by Meadow [MEAD79] to reflect this more corporeal, as opposed to superficial, status.

online session; instruction was provided entirely by the intermediary system -- no additional human assistance was necessary (except to handle certain system problems). The few reported studies (see, e.g., [LANC71,72; FENI80,81; SEWE76; RICH81]) of end-user operation -- even for recently improved systems -- record that either the users were not inexperienced searchers, or significant amounts of standard (non-computerized) instruction and help prior to or during the online session was given, or limited search effectiveness was achieved, or the number of databases and systems accessed was limited, or a combination of these factors obtained.

Along with the positive achievements demonstrated through CONIT 3, several questions were raised. For example, although all experimental users were able to retrieve *some* relevant documents, the recall levels were determined to be rather low: from 0.2 down to 0.01. Furthermore, while session times -- and other performance measures -- were acceptable to -- indeed, praised by -- the experimental users, the question was raised as to whether human expert information specialists acting as intermediaries could achieve greater effectiveness in shorter time. Finally, there was the question of whether a computer intermediary system could be made sufficiently effective so that its performance would equal or be superior to that of human experts in all respects.

In this report we describe investigations carried out in order to answer, at least in part, these questions. In this report we shall first describe enhancements made to the experimental CONIT intermediary system. Then we shall describe experiments performed with the enhanced CONIT which include searches done by highly experienced human expert search intermediaries and by inexperienced end users. Next, we interpret the results of these experiments so as to present our current understanding of the answers to the questions. Finally, we discuss intermediary system costs and benefits in order to discuss their current prospects, concluding with a discussion of the possibilities of even more advanced intermediary systems in such areas as automatic database selection and simulation of human expert search performance.

2. Enhanced CONIT

Several design principles that helped contribute to the success of the earlier CONIT systems [MARC82] are:

(1) The heterogeneity of existing systems is replaced by the commonality of the virtual system.

(2) The complexity of current system/user interfaces is replaced by a simpler and easier-to-use interface.

(3) Effective instruction is given by the computer to assist the user.

(4) Relatively few basic retrieval operations, of the many retrieval functions available on existing systems, are provided; but these satisfy most needs of most end users.

(5) Even among the few basic retrieval functions, beginning end users initially are taught still fewer *core* functions; additional capabilities may be taught as needed.

(6) Inexperienced users can take advantage of relatively simple methods for developing search strategies that are effective across heterogeneous databases.

A number of techniques were employed to help effectuate these principles. One technique centered around the idea of a simplified command/argument language with several natural-language features. A second theme was the computer assisted instruction (CAI) built into the system itself to help users learn commands themselves as well as effective search strategy formulation methods. This CAI included a menu oriented approach integrated with a carefully-developed expository basis, a hierarchically structured explanation base, and a contextually sensitive decision algorithm that tailored the instruction and explanation to the particular situation.

Finally, our solution to the problem of effective searching by inexperienced users across databases with heterogeneous indexes is based on a natural-language, free-vocabulary approach to searching. This approach, whose efficacy has been supported by several research studies [OVEH73, LANC71, KEEN73], emphasizes the use of natural-language keyword stems as the basis for searching. In the searching operation itself, the keyword stems are matched against both free- and controlled-vocabulary terms under which documents have been posted. A match is counted if any word or phrase has a word stem corresponding to one of the user-given keyword stems. All searches on a single keyword stem are then combined with the Boolean OR (union) operation. Finally, the separate unions are combined with the Boolean AND (intersection) operation.

In our enhanced intermediary system, which we call CONIT 4, we generally sought to retain these principles and techniques while incorporating additional techniques to assist the user further. However, in one respect we somewhat reduced the restraint imposed by Principle 4: the limitation of retrieval operations to a small core of essential ones. Thus, for example, citation searching was explicitly incorporated within the common language command structure; other functional additions are described below. Nevertheless, many functions of retrieval systems were still excluded from the common language commands (e.g., search by explicitly identified word proximity, or by specifically identified field -- except for basic subject, author, and citation fields; and selection of an arbitrary combination of catalog fields for output presentation -- as, e.g., the ORBIT and MEDLINE [but not DIALOG] systems allow).

On the contrary, the main thrust of the enhancements was to incorporate techniques which aid the user in carrying out his search, but which are *not* generally fully available on existing information retrieval systems. That is, existing systems either do not have a particular technique at all, or do not have it to the degree to which it has been developed for CONIT 4. Of course, even where certain aspects of these techniques may be available on individual systems, the ability to use these techniques in a virtual-system mode across a network of heterogeneous systems is unique to the intermediary system environment as presented by CONIT.

The rationale for the development of the computer intermediary in this fashion is threefold:

(1) The existing principles and techniques of the CONIT-3 intermediary apparently worked quite well.

(2) In particular, the principle of concentrating on the basic core information-retrieval functions seemed highly satisfactory.

(3) Our analysis of the previous experiments indicated that the main problems for users of CONIT 3 were (a) the development of effective search strategies; and (b) further alleviation of 'mechanical' problems like knowing what to do when various system buffers overflow or how to perform certain operations more easily, hopefully in such a way that the user could concentrate on identifying for the intermediary *what* he wanted to accomplish and not have to be concerned with complicated details of *how* to request and/or accomplish it.

New Techniques in CONIT 4

A number of changes were required simply to maintain the capabilities previously reported for CONIT 3. First, a few changes in the MIT MULTICS computer system, on which our experimental CONIT systems have been developed, necessitated a few minor changes in CONIT. MULTICS is a good system for subsystem development — which was, of course, a major factor in our using it as a development tool; changes in MULTICS are generally made so as to avoid or minimize the requirements for its users to make changes.

Second, changes are regularly made in the host retrieval systems and these sometimes require changes in CONIT. For example, when ORBIT began requiring a user security code in its login protocol a few changes had to be made to the CONIT rules that execute this protocol. These changes, and other more complex ones to be detailed below, were made relatively easily primarily by changing or adding rules. We believe this ease of modification supports our decision [MARC79] to develop intermediary software on a production rule basis. A second, more subtle, example is that of a system — in this case NLM MEDLINE — permitting a database explanation to be obtained directly from the EXPLAIN command in all situations as opposed to requiring a FILE command to connect to that database first. This is an example of a kind of change that does not *require* a corresponding CONIT change -- the old method would still work -- but where a change in the intermediary system could make operations more efficient and/or less confusing to the user.

There were also a number of changes to the basic CONIT development procedures and core execution programs to improve ease of development and efficiency. For example, a special MULTICS command procedure was devised so that rule changes made in a 'source' segment would be automatically collected and compiled (via the rule generator program) into the execution table of rules. This command procedure reduced the programmer effort while also enhancing maintenance of documentation. A second example is the improvement in the rule search procedure which speeded up this part of the intermediary system execution significantly. The heart of this improvement is the maintenance of a pointer to the last rule matched (rules are stored as list structures). Unless some special context dictates otherwise, the search for the next rule to be matched and executed starts with this last rule instead, as formerly, at the beginning of the table of rules. To take maximum advantage of the potential for improved efficiency provided by this enhancement, rules are generated and ordered to follow a sequential order for the 'normal' cases while still preserving the minimization of number of rules through the principle of looking at specific rules before more general ones [MARC77].

Besides making changes about the way they do existing operations, the retrieval systems also, of course, are continually making additions to their capabilities. One kind of addition that we regularly reflect into CONIT is in the availability of databases through the retrieval systems. With very few exceptions (e.g., MEDLINE'S SERLINE) CONIT makes *all* databases of the 4 retrieval systems available in the common (virtual) mode. (Note that absolutely all functions and databases of each system are available to the user in the 'pass-thru' mode in which the user talks 'directly' to the retrieval system by stating his request in the command language of the retrieval system.) At the end of the period of experimentation for CONIT 3 there were 97 databases that a user could access in the common command language. Of these 97, 16 were 'duplicates' in the sense of being available on another of the 3 other retrieval systems; therefore, there were 81 unique databases at that time. (This counts each individual file of a database as a separate database; thus, the Chemical Abstracts database with its 3 files, covering different years, was counted as 3 databases.) At the beginning of the experimentation period with CONIT 4 there were 156 total and 126 unique databases. The corresponding figures at the end of the experimentation period were 176 total and 155 unique. (At the time of this writing -- March, 1982 -- the numbers are 226 total and 196 unique databases.)

Various improvements to CONIT itself have been made; they range from relatively minor ones to quite major ones. Our previous experiments indicated the importance of carefully worded instructions to help users understand how to work with the intermediary system. On the basis of analysis of user reactions a number of changes were made in the instructions. For example, we changed the heading of the instructions for improving search precision from 'ways to find better documents' to 'ways to find fewer documents'. A single change like this seems trivial -- it is likely to affect only a small subset of users in a fraction of their usage -- but we strongly believe that the cumulative effect of literally hundreds of such considerations can make the difference between a highly 'friendly' (easy-to-use) human-computer interface and one that confuses and confounds as much as it helps.

Some other changes in the instructions were similarly not very important in themselves, but point to future areas of potential high significance. Thus, for example, certain databases were categorized by what other dictionary databases might be useful in finding search terms for the given databases. The explanations of how to develop search strategies was then specialized so as to provide that information whenever the particular databases were being accessed. As we shall discuss in more detail below, the specialization of assistance according to problem context, of which this is one example, will undoubtedly be an important part of more sophisticated assistance.

A few other relatively minor improvements in CONIT may be listed. The ability to express, and execute, citation searches in the two citation databases has already been mentioned. The CONIT command for this is "FIND CITATION <term>", abbreviated "FC <term>", where <term> is the specification for the citation search formulation. As a second example, an argument, EVERY (abbreviated EV), was added to the SHOW (DOCUMENTS) command. This gets translated so that the retrieval system is requested to output every document in the given set. Finally, a timetable of the regularly scheduled times when the four retrieval systems are supposed to be available was generated and maintained in CONIT. This timetable, or selected portions, is available on request by the user. Thus the command SHOW SCHEDULE WEEKLY (or just SSW) gives the user the full week's schedule for each system whereas the command SHOW SCHEDULE TODAY (SST) gives the current day's schedule. Also, whenever the user requests access to a database, the intermediary system checks the timetable; if the system implementing that database is not scheduled to be up, the user is so told and given the option of trying to connect to it anyway (sometimes systems *are* available at unscheduled times); if the database is available on two systems of which only one is scheduled to be up at the given time, the intermediary selects the system that is scheduled to be up (irrespective of the regular, default database).

The more basic and significant additions to CONIT can be categorized as belonging to four areas: (1) search history and reconstruction; (2) automated keyword/stem searching; (3) individualized database searching; and (4) automatic database selection. Additions in the first three of these areas were used in the formal experiments for CONIT 4 and these will be described in detail below in the immediately following discussion. The automatic database selections additions (ADBS, area 4) were not used in these experiments; the detailed nature of the ADBS and the reasons for their exclusion from the experiments will be discussed in Section 6.

Search History and Reconstruction. Existing retrieval systems have various devices for saving search formulations for use at another time — possibly on a database different from the one for which the search was originally done. Some problems with this feature, especially for inexperienced users, include [1] the necessity to "save" explicitly the search formulation, and [2] the loss of the search formulations when changing databases or systems or when the buffer storage for searches becomes filled -- unless special precautions are taken by the user to avoid such losses.

In CONIT 4 we have alleviated these problems by maintaining a record of all past searches. For each search, CONIT remembers the full search formulation, the database and system in which the full search was run, the number of documents found in the resultant retrieved set and in any component sets formed in creating the resultant set, the set names as given by CONIT and by the retrieval system, and whether the set is currently available in the retrieval system. With one exception, all this information is displayed online at the user's request by the command "SHOW REVIEW" (abbreviated SR). The one exception is the retrieval system set name which is not needed by the user in common (virtual) mode; a user in pass-thru mode can, of course, get this information by directly sending the appropriate set review command for the given retrieval system. If the user then requests any component or compound search formulation to be repeated in any database or set of databases, CONIT refers to the search history and repeats the search, after connecting to the appropriate systems and databases. Note that if a given search is compounded from several other searches, then repeating that search requires that CONIT perform a like number of search and combination operations in each database.

Furthermore, if a set has, for any reason, been dropped from the retrieval system in which it was retrieved, the user need not be concerned or even aware of that fact; if the user requests the intermediary to output from that set or to combine that set with another, CONIT will first regenerate the set and then perform the requested operation. In addition, if the maximum number of sets allowed in current memory by a retrieval system has been reached, CONIT will automatically clear sets from the system if additional searches are requested by the user in that system. The user, too, may clear any sets from the retrieval system with the command "CLEAR".

In displaying excerpts from database indexes for users, some systems associate a short code word, or tag, with each displayed term for ease of reference by the user in search statements. Goldstein has demonstrated [GOLD78] that a computer intermediary can simulate this capability so as to give a user the same convenience in specifying a search for a system that does not have the tag feature. We have extended this type of assistance in CONIT in two ways: [1] users can request what amounts to a Boolean OR (union) search by specifying a range of tags, and [2] the actual search terms (not just the tags) are saved so that the search may be repeated in different databases.

Automated Keyword/Stem Searching.

Previous research (MARC79) has indicated that effective search results can often be obtained by searching on the keyword stems found in a user's natural-language search expression. Earlier versions of CONIT provided instruction to the user in this search methodology; such instructions contributed importantly to the search sucoess of experimental users of the computer intermediary. In the current version of CONIT, we have automated part of the process to make it faster and easier to use.

Thus, for example, the user may request a search on the topic "transplantation rejection". CONIT takes each word in the user's phrase and derives a stemmed form for each, using a stemming algorithm (LOVI71). In this case the stemmed forms "transplant-" and "reject-" are derived. Then CONIT requests truncated searches on each of the stemmed forms in all the indexes that can be searched with a single command in the connected database. The sets retrieved from the individual subsearches are then combined with the Boolean AND operator (intersection) to yield a resultant set. CONIT names the subsearches and the resultant search and reports to the user the number of documents in each set. All this is done automatically without user intervention; the user can then work with any of the named sets.

If any of the subsearches yields null results, CONIT suggests scanning the index terms around the non-responsive search term. If, on the other hand, a truncated subsearch causes a search buffer overflow, CONIT replaces the truncated stem search with an exact-match, full-word search so as to avoid the overflow condition.

Individualized Database Searching. Often, search formulations need to be tailored to the peculiarities of the database being searched. We have begun investigating how to help users individualize their searches with two simple kinds of aids. In one aid, the user is permitted to request author searches in a common format; CONIT then translates this format into the one appropriate for the database being searched -- e.g., correct spacing and punctuation between last name and first initial is supplied by CONIT.

A second aid involves a specialization of the keyword/stem searching scheme mentioned above. Where the database has not been implemented so as to post documents under each word of a multiword controlled-vocabulary index term, the scheme given above might not retrieve a document posted under the given term. For the databases for which this is true -- a small minority of those in the four systems CONIT connects to, but including some of the more important databases in the biomedical area -- CONIT performs a full-phrase search in addition to the keyword searches.

3. Framework of Experiments

A. Experimental Objectives and Subjects

The primary objective of the experimental program was to evaluate the enhanced CONIT intermediary system in terms of its ability to allow end users to satisfy their informational needs by accessing a network of heterogeneous information systems through the computer intermediary. An important aspect of the evaluation of the intermediary computer system was to compare the cost effectiveness of searching as done by end users searching on their own with CONIT with that obtained by the users working with an expert human intermediary search specialist. Our research mandate was to place special emphasis on searching in the medical and biomedical domains.

As we have just indicated, there were two kinds of experimental subjects: the end users and the human intermediary search specialists. There were four human intermediaries. Each was an expert bibliographic computer search specialist from one of three bibliographic search services in the Boston area. Each had several years of experience as a search specialist in a primarily academic or medical/hospital setting -- although industrial and other non-institutional users are regularly served by these services. Three of the four had subject specialties centering in the medical area. One of these three served as the primary experimental human expert searcher and the other three served as human experts for only one end user each.

There were several classes of end users in this current round of experiments. A number of usages were of an informal or demonstrational nature; they provided valuable information and insight on the nature of the experimental intermediary system, but the results lacked the definiteness of carefully controlled experiments. What we shall report on below is a series of controlled experiments with the CONIT 4 system.

For these formal, controlled experiments we sought end users with a current information need to serve as experimental users of the CONIT intermediary system. Notices were posted at MIT and other institutions offering free access to information from computer bibliographic databases in the medial and biomedical areas to those who agreed to participate in our experiments. A number of other users were identified as likely prospects under the same criteria through the aid of the several search services who cooperated in our experiments. A final category of users included those identified by members of our research project itself. These included three undergraduate students who had been accepted as members of the project team but had not yet started any regular project assignments. (As a rule, incoming project members are asked to serve as end users before they know the details of the CONIT system and the experimental procedures so that they may gain some insight from the point of view of the experimental end user.) In each case, potential experimental users were interviewed to determine their informational needs and backgrounds, especially with respect to their previous computer experience.

Several dozen potential users were identified in this way and 16 of them participated in the formal experimental program. Only a few potential users were rejected for not having appropriate topics; the reason for the non-participation of the others was largely a question of incompatibility with project schedules. A list of the search topic titles for the 16 participants is given in Table 1. The search topics are identified by a single letter or a letter/digit combination. Each letter (from A to P) refers to one of the 16 users. Where

TABLE 1. EXPERIMENTAL USERS AND TOPICS

USER/FIELD*		TOPIC
A	DOC/NUT	Immunofluorescent techniques for DNA staining
B	DOC/BIO	Kidney transplantation rejection statistics
C1	PHYS	Use of media techniques in psychotherapy
C2	PHYS	Efficacy of film in education
D	PROF/SOC	Government influence on health care management
E1	DOC/NUT	(1) Modeling of spontaneous motor activity (2) Effect of diet on animal behavior
E2	PDOC/NUT	Role of particular ions in specific excitable tissues
F1	DOC/BIO	Ultrastructure and biochemistry of Allogromia
F2	DOC/BIO	Colchicine in microtubule assembly/disassembly
F3	DOC/BIO	Intermediate Filaments in Erythrocytes
G	DOC/BIO	Lipid metabolism in isolated perfused rat kidneys
H	DOC/BIO	Transglutiminase in receptor-mediated endocytosis
I	PHYS	Kaposi's sarcoma in lymph nodes with allergic complications
J	UG/ENG	Properites of biocompatible polymer materials
K	PROF/ENG	Computer-aided diagnosis of eye diseases
L	UNIV/UIL	University Interactions with Industry
M	DOC/ENG	Environmental Planning Information Systems
N	UG/ENG	Effect of kainic acid on spontaneous motor activity
O	UG/ENG	Beneficial and harmful effects of fasting
P	UG/ENG	Etiology and therapy in stuttering

*USER TYPES: PHYS=physician; DOC=doctoral candidate; PROF=professor; UG=undergraduate; PDOC=postdoctoral; UNIV=University Administrative Staff

FIELD: BIO=biology; NUT=nutrition; ENG=engineering; SOC=sociology; UIL=University-Industry Liaison more than one topic is applicable to a given user the separate topics are indicated by an appended digit (e.g., "C.1 and C.2). Two users searched on two topics each (on separate occasions) and one user searched on three topics. Thus there were 20 different topics searched. Generally, each topic was searched on a separate search session, as will be explained below.

While each search topic (except for topic L) has some direct relevance to the general medical/biomedical area, there is obviously a wide spectrum of fields covered. Various biological and medical topics are included. In many instances there is clearly a multidisciplinary character to the topics. Disciplines other than the medical and biomedical ones include engineering, information technology, politics, education, and administration. The significance of the multidisciplinary nature of many of the topics is discussed below in Section 7A.

The professional/educational status of each user is also indicated in Table 1. The totals in the various categories are: 2 medical doctors; 1 university (non-academic) staff; 2 professors; 1 post-doctoral fellow; 6 graduate students (doctoral candidates); and 4 undergraduate students. None of the end users had himself previously operated either CONIT itself or any of the four retrieval systems accessible through CONIT. Several had either used a human intermediary to help search a bibliographic retrieval system or, in one case, had operated a different retrieval system (the Library of Congress online card catalog system). Practically all 16 had some familiarity with computers and most had some experience with operating computers from terminals. The significance of this kind of experience with computers will be discussed when the experimental results are analyzed.

B. Experimental Procedures

All experiments were carefully controlled and monitored. To the extent possible, the same set of procedures was used for each experimental user and the same set of data points was recorded. Experimental procedures were an extension of those for our CONIT-3 experiments [MARC81a]. At the beginning of each experimental session, an experimental supervisor briefed the end user concerning the nature of the experiment. The oral briefing took about three or four minutes. The experiment supervisor answered any user questions after the briefing and he was available if difficulties arose in the course of the online session. The user was requested to try to resolve any questions or problems on his own if at all possible and to call on the supervisor only if he felt hopelessly lost or if the system was broken. If the user did ask a question that was not related to a system problem, he was asked to go back to the intermediary system for instructional assistance.

The online session itself — that is, the user's interaction with CONIT at the computer terminal — began immediately after the briefing. In this series of experiments the primary terminal used was the DEC LA-120, a 120 CPS printing terminal. [The 120 CPS terminal was made part of the experimental setup after the fifth session; before that a 30CPS printing terminal was used.] A printing terminal was chosen in order to allow users to refer fairly easily to past instructions and retrieval results. Previous experiments with a 30 CPS printing terminal indicated that many users would prefer higher speed; these CONIT-4 experiments tended to support that conclusion.

The experiment supervisor remained in or near the room where the user worked at the terminal in order to provide unobtrusive monitoring of the session and of the user (beyond what was possible from the full record of user commands and of CONIT responses, as maintained on the terminal typescript and a computer "audit" file) and be available for questions or system problems. When in the session room the supervisor would also observe visually any activity of the user not recorded on the computer. The user himself was asked to make written or mental notes of particular problems or other reactions he might have, so as to be able to relate his/her experiences more completely to the supervisor at a post-session debriefing. As was indicated above, the supervisor did not prompt the user or otherwise interfere with the course of the interaction unless computer system bugs or computer communications problems appeared to hinder the continuation of the session. This intervention was required several times in the experiments, primarily to handle communications caused system failures, as we shall discuss below.

One variation from our experimental procedure for the experiments with the CONIT-3 system relates to the use of printed instructional materials. For the CONIT-3 experiments we divided the users into two groups: those who were given access to printed instructions before beginning the online session and those who started the session without such aids. We determined in those earlier experiments that there were elements of the printed instructions, basically concerning development of online search strategies, which appeared to be helpful. We incorporated those elements into the *online* instruction for CONIT 4.

We have essentially all of the informational items explicitly requestable by the user through the EXPLAIN command available in printed form bound in the format of a reference manual. For the first five users in the CONIT-4 experiments we made the reference manual available to the user prior to the session itself. This was done after the initial interview with the user when it was decided that he would participate in the experiments, usually one or more days before the experiment itself. The reasoning for doing this was that some users prefer to read printed instructions *before* they access any computer system. Generally, as with the first four experimental users, users make no, or very little, use of such printed instructions prior to operations at the terminal -- typically, five minutes or less of cursory scanning. They have been told that the intermediary system is self-instructional and they are content to follow the CAI mode of learning rather than spend much effort on instructional manuals. As with other norms of learning behavior, however, there are often significant numbers of exceptions to any rule. In this case the fifth user spent over three hours reading the reference manual and gave us a detailed critique of its contents when he came in for his pre-session briefing. His online session did require less CAI instruction time and he did appear to have made somewhat better than average use (from a mechanical viewpoint) of the CONIT facilities.(1)

In any case, we decided subsequent to the fifth user to reduce the number of variables inherent in the experiments by eliminating the offer of the reference manual *prior* to the online session. It was kept available during the session for those who preferred that mode of obtaining instructions. A few users did so, mainly, as a way to re-read an explanation they had previously seen rather than shuffle back through the printed typescript from the terminal or request it again on the terminal. The predominant mode, however, was to re-request the explanation at the terminal.

The online session was terminated when the user issued the STOP command, or, in a few cases, when a system problem prevented continuation. The primary motivation for the user's terminating the session appeared to be his feeling that he was at the point of diminishing returns, as far as getting useful information was concerned. At the end of an online session the supervisor held a debriefing conference with the user. The debriefing

---

(1) Note that this result is not necessarily obvious. The CAI is devised to direct users along those lines most likely to help the user according to his context online. Reading a printed reference manual beforehand could negate some of that potential benefit by establishing less-than-optimal propensities *a priori* and by causing the user to resist or ignore potentially better directions given in CAI online. We have seen a few instances of such behavior in our experiments. Thus, for example, user C, who had spent 35 minutes reading the printed instructional manual, apparently misremembered some of the instructions and tried to do a FIND (search) before he had done a PICK to select a database. A related behavior occurs when a very computer-experienced user attempts to outguess the CAI and proceed on his own intuition without following, or even reading, the instructions available through online CAI.

emphasized (1) any problems encountered by the user; (2) overall impressions of the session and its usefulness; (3) the rationale for the user search strategy formulation; and (4) the relevance judgments by the user on the retrieved documents. The users were asked to judge relevance on a four-point scale: high, medium, low, and none. Usually, one online session was sufficient to satisfy the user's needs and conclude that part of the experiment. However, as we shall detail below, a user would occasionally come back for a second or third online session with CONIT.

A second part of the experiment involved a search on the user's topic by an expert human search intermediary. The intention was, as much as possible, to have the human expert search done as such a search would normally be done by that intermediary. Thus, the user would fill out the regular forms required by the search service. These typically requested information similar to that requested for the corresponding forms for the CONIT session. As for other users of these services, the actual searching was performed at the search service (library) location with the user present to assist the searcher. For our experiments a member of the CONIT project was also present to observe and record the session; the project observer did not actively intrude on the proceedings. Before the search session itself the searcher conducted an interview with the user to go over his previously prepared search statement. In general, the searcher had looked at this statement prior to the interview to assure herself that the topic and the statement did not contain any unusual difficulties but she did not spend any significant amount of time preparing for the search before seeing the user. (The search topic was orally explained to the searcher when the search session was being arranged by the project staff person; in a few cases, as we shall explain, the searcher was picked because she was especially familiar with the topic area.) The pre-session interview, whose average length of time was 24 minutes, involved further delineation of the topic and, in some cases, the development of initial search strategies with help from the user. The user was requested to try to act as if he had not already performed a CONIT search; in particular to avoid initiating use of any search strategy or terms that he had developed or found effective on CONIT. It was understood by all parties that the CONIT project was to pay the regular costs of the search. The searcher was asked to follow the guidelines of the user in terms of how extensive a search was to be performed. After the online session itself the user was debriefed to get his reactions and relevance judgments, as in the CONIT case. Again, with a few exceptions, one online human expert search session sufficed.

Subsequent to the CONIT and human expert online sessions any requested offline printouts from those sessions were collected and presented to the user for his relevance judgments. Based on these various judgments we subsequently performed an analysis to discover search strategies that might improve on the searches of the users and the human intermediaries and to estimate the recall base. Searches based on this analysis, both in databases previously searched and others, were then performed. Catalog output from selected documents retrieved from these searches, along with offline output requested during the online experimental sessions, was then presented to the user for further relevance judgments. In a few cases, subsequent rounds of analyst searches and user relevance judgments were performed in order to complete the analysis of search strategies and the recall base for a given topic.

The specific experimental procedures described above were not fully carried out in every instance. Due to a combination of scheduling difficulties and system problems, the completion of all procedures was prevented in several instances. For 4 of the 20 topics we did not proceed to that portion of the experiment involving human expert searches. In one or two other instances there was insufficient information upon which to complete the search strategy and recall analysis. These gaps in the analysis will be detailed in the description of the experimental results which follows.

4. Experimental Results

A. Quantities Measured

Statistical data derived from the various stages of the experimental analyses are summarized in Table 2. For each topic there are two rows of figures. The first row contains data relevant to results applicable specifically to the end user session. The second row contains data relevant specifically to the human expert session. Average and median figures are listed at the bottom of the table.

The quantities measured include the time consumed in various parts of the experimental sessions. TOTAL time is the time from the start to the end of an online session; i.e., for the user, the time in CONIT from the START command to the STOP command. USABLE time is that portion of total time that was available for productive work; it excludes time wasted due to system bugs (e.g., the time between a line drop and the reconnection to the system and reconstruction of the status before the drop). Total time equals usable time for the human expert search; there were no instances of system error in that mode. EXPLAIN time for the user session is that portion of his usable time that was consumed in requesting and reading CONIT explanations; it is basically that time in issuing the EXPLAIN and HELP commands and looking at their responses. It does *not* include time spent in reviewing explanatory instructional material that the CONIT system automatically gives the user other than for the EXPLAIN and HELP commands (e.g., after the results of a search are presented CONIT may explain to the user how to see information about any documents found by that search). EXPLAIN time for the human expert session means the time spent at the interview prior to the online session itself. SEARCH time is that portion of the online sessions consumed by giving the actual search commands and looking at their responses. By DISPLAY time we mean that portion of the sessions consumed in giving requests for displaying

USER	SESSION TIME FIGURES (MINUTES)					RECALL FIGURES					STRATEGY		DOCS SHOWN			NO. FILES			
	TOTAL	USABLE	SEARCH	DISPLAY	EXPLAIN	REL ON	REL TOT	BASE	%ON	%TOT	SETS	TERMS	N	F	T	F	DB	RF	RDB
A	282	206.6	120	67.2	19.0	8	8	-	-	-	37	30	101.	1.	0.	2.	2.	1.	1.
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B	187.4	167.6	101.6	57.1	3.8	16	181	-	-	-	30	15	13.	4.	19.	5.	5.	1.	1.
	12.2	12.2	5.3	6.9	-	5	6	-	-	-	12	10	31.	1.	0.	1.	1.	1.	1.
C1	392.8	323.9	232.9	65.1	14.9	6	939	-	-	-	67	30	59.	4.	15.	12.	12.	3.	3.
	-	-	-	-	-	?	23	-	-	-	14	29	-	-	-	7.	2.	6.	1.
C2	153.2	132.3	66.4	51.4	7.7	4	29	-	-	-	29	6	12.	9.	27.	4.	4.	1.	1.
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D	92	77.4	30.8	9.3	37.3	?	?	-	-	-	3	3	7.	0.	4.	1.	1.	-	-
	26.8	26.8	19.5	7.3	-	?	?	-	-	-	53	14	0.	0.	81.	1.	1.	-	-
E1	361	314	124.9	174	.5	89	132	4120	2.2	3.2	62	18	238.	6.	204.	4.	4.	3.	3.
	181	181	123.2	57.8	-	99	138	4120	2.4	3.3	97	34	135.	5.	309.	9.	3.	9.	3.
E2	496.4	478.2	243.2	204.3	18.9	83	83	400	20.8	20.8	131	38	87.	46.	617.	1.	1.	1.	1.
	118	118	95.4	22.6	20	19	144	400	4.8	36	107	50	113.	0.	0.	3.	2.	3.	2.
F1	91.9	91.9	35.6	16	16.9	25	28	28	89.3	100	2	2	69.	1.	0.	2.	1.	2.	1.
	37.3	37.3	26.2	11.1	20	6	28	28	21.4	100	34	18	27.	9.	0.	3.	2.	2.	1.
F2	83.1	81.6	60.4	11.7	9.5	19	20	171	11.1	11.7	10	8	35.	0.	0.	2.	2.	1.	1.
	35	35	10.9	24.1	20	11	29	171	6.4	16.9	55	8	27.	24.	0.	6.	3.	5.	3.
F3	72	68.2	45.7	19.3	3.2	19	54	120	15.8	45	30	8	58.	0.	0.	6.	4.	5.	3.
	24	24	19.4	4.6	24	16	66	120	13.3	55	42	12	15.	1.	0.	4.	1.	3.	1.
G.	276.9	268.8	148.5	74.4	34.2	6	6	11	54.5	54	73	15	118.	3.	40.	9.	6.	8.	5.
	76	76	59.6	16.4	15	6	6	11	54.5	54	59	12	65.	4.	5.	8.	2.	7.	2.
H	108.3	70.9	26.6	13.3	28.2	7	12	62	1.3	19.6	17	7	14.	1.	0.	2.	2.	2.	2.
	50	50	42.6	7.4	28	4	51	62	6.5	82.3	39	10	28.	2.	0.	5.	3.	3.	3.
I	126.1	122.3	52.8	53.1	16.4	12	12	15	80.0	80	6	4	51.	24.	0.	1.	1.	1.	1.
	55	55	47.5	7.5	21	8	9	15	53.3	60	33	8	16.	6.	0.	6.	1.	6.	1.
J	206.1	148.3	93	29.7	22.6	35	35	193	18.1	18.1	53	8	58.	4.	21.	8.	6.	5.	4.
	53	53	46.2	6.8	13	16	91	193	8.3	47.2	33	18	30.	1.	0.	5.	2.	4.	2.
K	168.9	167.8	96.8	28.2	34.7	2	223	246	.8	90.7	24	12	46.	0.	20.	7.	4.	1.	1.
	33	33	13.3	19.7	9	22	22	246	8.9	8.9	9	3	29.	24.	0.	3.	3.	3.	2.
L	128	84.6	36	28.2	18.6	6	944	977	.6	96.6	11	4	46.	2.	10.	1.	1.	1.	1.
	77	77	?	?	?	-	33	977	?	3.3	-	-	-	-	-	-	-	6.	6.
M	141.3	135.8	78.8	8.1	48.1	4	11	11	36.4	100	5	10	14.	0.	0.	1.	1.	1.	1.
	-	-	-	-	-	-	?	11	-	-	-	-	-	-	-	0.	0.	-	-
N	204.7	160.4	101.2	38.2	14	3	11	23	13.0	47.8	37	10	23.	11.	0.	5.	5.	5.	5.
	30	30	29	1	22	3	11	23	13.0	47.8	13	21	5.	0.	0.	1.	1.	1.	1.
O	121.3	116.0	46.4	27.7	17.8	23	48	128	17.9	37.5	40	12	25.	0.	84.	4.	3.	2.	1.
	71	71	65.2	5.8	50	8	56	128	6.3	43.8	42	21	29.	0.	0.	4.	4.	4.	4.
P	45	39.1	18.3	5.6	12.3	7	31	1200	.6	2.5	15	6	14.	0.	0.	2.	2.	2.	2.
	9	9	6.1	2.9	45	17	112	1200	1.4	9.3	59	30	10.	0.	10.	7.	7.	5.	5.
CASES																			
Users	20	20	20	20	20	19	19	15	15	15	20	20	20	20	20	20	20	19	19
Experts	16	16	15	15	12	14	16	15	13	14	16	16	15	15	15	16	16	16	16
AVERAGES																			
All Users	173.1	153.6	81.7	50.8	19.1	24.7	137.9	513.7	25.0	44.8	37.5	12.3	54.5	5.8	53.1	3.4	4.	2.7	2.
Cor Users	186.9	162.8	87.9	49.1	18.9	19.7	147.7	513.7	25.2	48.5	34.1	12.3	57.1	3.6	67.3	3.7	4.4	2.4	2.2
All Exp's	55.5	55.5	40.6	13.5	23.9	17.1	51.6	513.7	15.4	40.6	43.8	18.6	37.3	5.1	27.0	2.4	4.6	4.3	2.4
MEDIANS																			
All Users	127.1	119.2	60.4	28.2	17.4	17.5	41.5	128	15.8	41.3	30	9	46.	1.5	7.	2.5	3.0	2.	1.
Cor Users	147.3	134.1	72.6	28.9	17.4	8	31	128	15.8	45	29.5	9	46.	1.	4.	3.5	4.0	2.	1.5
All Exp's	43.7	43.7	29	7.4	20.5	9.5	31	128	8.3	45.5	40.5	16	28.	1.	0.	2.0	4.5	4.	2.

Table 2. Experimental Statistical Results

Key to Table 2

USERS: Each letter stands for a different user; numerical suffixes indicate distinct topics for given user.

SESSION TIME FIGURES (MINUTES)
TOTAL: Time for online session
USABLE: Total time minus time lost due to system bugs
SEARCH: Time spent on search commands
DISPLAY: Time spent on document records
EXPLAIN: (for user) Time spent on instructional commands
(for expert) Time spent at pre-search interview

RECALL FIGURES
REL ON: Number of relevant documents found online
REL TOT: Total number of relevant documents found
BASE: Recall base
% ON: Fractional recall for documents found online
% TOT: Fractional recall for all relevant documents found

STRATEGY
SETS: Number of retrieval sets
TERMS: Number of search terms

DOCUMENTS SHOWN: The number of documents shown by output mode:
N: normal (citation) mode
F: full (whole record) mode
T: title only

NO. FILES:
F: The number of files searched
DB: The number of databases searched
RF: The number of files in which relevant documents found
RDB: The number of databases in which relevant documents found

CASES: Number of cases for which data exists

AVERAGES: Average value of given parameter; shown for:
(1) all end users, (3) all expert sessions,
(2) end users where corresponding expert values are available

MEDIANS: Median value of given parameter

DATA
1. In each pair of rows, first row is for user CONIT session and second row is for expert session.

2. Hyphen (-) indicates data unavailable

3. Question mark (?) indicates data of questionable accuracy

document record information online and in looking at the responses to those requests. Taken together, search and display (and explain, for the user session) time make up the bulk of the usable time spent online. Minor amounts of time are occasionally spent on other operations, e.g., requesting offline prints, etc.

The main parameter by which the effectiveness of the results is measured is the number of relevant documents found. The number of relevant documents identified by reviewing some portion of their catalog records online is labeled REL ON. In addition to those documents found online, there were other relevant documents found through reviewing the offline printouts requested by the user or the expert searcher. The user did not always make relevance judgments on every document in those sets which he or the expert dumped offline. In those cases we extrapolated the number of relevant documents in the total set from the number in the sampled set. The total number of relevant documents found on a topic either online or offline by user or expert is given in the column labeled REL TOT. The estimated number of documents in the recall base for a given topic (i.e., the total number of relevant documents on that topic in all databases) is given in the column labeled BASE. Obviously, the recall base as so designated is the same for the user and for the human expert. The recall base was estimated on the basis of analysis of search results by the user, the human expert, and by project analysts. For the purpose of this summary a document was considered "relevant" if it was marked as being of "high" or "medium" relevance by the user; "low" relevance documents are lumped with "no" relevance documents as being nonrelevant. Fractional recall is then calculated as the percent of the recall base found in searching. The columns for fractional recall for documents found online and for all documents found are labeled %ON and %TOT, respectively.

Several other parameters were calculated. The number of documents for which information was looked at online is broken down by the type of information displayed:

normal (citation), full (the whole catalog record), and title only. The columns labeled N, F, and T contain the numbers of documents displayed in each of those categories in the respective order. The number of retrieved sets found online is given under SETS. The number of search terms employed is given under TERMS.

The number of databases searched (successfully) was a parameter of interest. For the purpose of this calculation we allow for the fact that some "databases" are made up of several separately searchable "files". For example, the Chemical Abstracts database was actually 3 files distinguished from each other by the years of coverage in the corresponding printed abstract journal. In the columns labeled F and DB we list the number of files and databases searched. In the columns labeled RF and RDB we list the number of files and databases searched in which relevant documents were found.

B. Analysis of Results

The first question we might want to answer is how well the users learned to use the CONIT commands. Based on the high success in achieving results to be detailed below, it is clear that users did, in fact, learn to master CONIT commands rather well, at least. For the CONIT–3 experiments we measured such parameters as the amount of time it took to learn the system well enough to perform particular tasks. A review of the results of this round of experiments showed a similar rapidity of learning. There is, it appears, very little difficulty in learning the CONIT basics; we shall discuss in Section 5 below the question of instruction and user learning in more detail.

The first, and perhaps most important general result to observe, is that the CONIT users did rather well in retrieving relevant documents, both in an absolute sense of numbers and in the relative senses in comparison with the human experts and with the total number of relevant documents in the databases — i.e., the recall base. For the 19 topics for which we have data, CONIT users retrieved from 6 to 944 relevant documents, with an average of 148 and a median of 31. For the 16 topics for which we have corresponding data for the human experts, the range is from 6 to 144 total relevant documents found with an average of 52 and a median of 31. (Note that for even numbers of data points we calculate the median by averaging between the two middle points.) If we take the 16 topics for which we have both user and expert data, perhaps a fairer comparison, we see that the relative superiority of the users is somewhat diminished in terms of the average (down to 138) but somewhat enhanced in terms of the median (up to 41.5).

Perhaps an even fairer comparison is made by looking at the individual comparisons between user and expert more closely. Let us make this comparison along a five point scale. At the extremes of the scale we have those situations in which user or expert did much better than the other -- we identify this situation as one in which one person was at least 20 percent better than the other. Sometimes the measurements are identical. The two other cases are the intermediate ones: user or expert did somewhat better, but there is less than a 20 percent difference in the data points. For this way of comparing we see that the expert did much better than the user in 6 cases whereas the user did much better in 5 cases. In 2 cases the expert did somewhat better and in 3 cases the results were identical. By this way of comparison we could say that the experts had a slight advantage over the users for this measurement of search effectiveness.

Let us now consider one measure of search *cost*, namely time spent in performing the search. In an operational environment time could be translated into dollar costs in a number of ways. The time spent online is usually the biggest single factor in determining costs from the retrieval systems. The time spent by the expert is a major factor in determining any cost overlays by a search service over the basic retrieval system costs. The time spent by the user is also, of course, a very real, and perhaps determining cost, as perceived by the user, although it is not often translated directly in monetary terms. As a first comparison along these lines we may look at that parameter we call usable time. For usable time we see that the experts have a dominant advantage over users. Experts did much better — i.e., spent at least 20 percent *less* time — in every one of the 16 cases for which we have a direct comparison.

In terms of cost-effectiveness it might be thought by looking only at the measures of total numbers of relevant documents retrieved and time spent that human experts are significantly superior in that they are, on the average, as good — or almost as good — in recall while they take much less time. However, the usable time spent parameter, as far as person-minutes is concerned, could be doubled to take account of the fact that both user and human intermediary are required to spend this time. Also, the interview prior to the expert online session requires the time of both user and expert searcher. If we calculate the total person-minutes, we see that the human-expert/user team actually spent somewhat *more* than the CONIT user does in terms of person minutes: 159 minutes versus 154 minutes.

In making the computer/human comparison there are a number of other criteria by which users judge the costs of the search. Perhaps the most important of these other criteria are the total (calendar) time required to get the information they seek and the total costs and effort involved. Users were quite consistent in expressing their appreciation for the relative ease of using CONIT compared with the difficulties that are inherent in the human intermediary mode. The perceived difficulties included making a mutually agreeable time for the online session and having to explain their problem in written and oral form to another party. Also, the library location for the online expert session may have been less convenient than our LIDS laboratory location for some of the users. Furthermore, the users' desire to get detailed results as quickly as possible led them to a rather different mode of searching than that employed by the human experts. In particular, users spent considerably more online time looking at the actual catalog records of retrieved document records than did the human experts. Some data that support this observation follow. The 20 users spent an average of 49 minutes looking at document information. The 15 human expert cases for which we have data spent an average of only 13 minutes. The corresponding figure for users for those 15 cases is 51 minutes. This same four-to-one ratio is maintained when medians are considered. The median is 29 minutes for 20 users and 28 minutes for users in the 15 cases for which the expert median was 7 minutes. This result is further supported by the data on the number of documents displayed online by category of catalog information requested.

Our debriefing interviews with the users and the human experts provide some explanation of how this difference came about. As we have previously said, the users highly valued getting complete, usable information as quickly as possible. Document citations and abstracts contain the needed information, or, at least, information sufficient to locate the needed sources. In addition, since costs of the retrieval system operations were being borne by the project, no additional costs, other than for their own time, were perceived by the users. For the human experts, on the other hand, we had something of a conflict. On the one hand they had been instructed by us to conduct searches as much as possible in the manner they normally did. On the other hand, we told them not to worry about any search costs — which were to be borne by the project — and to follow the dictates of the users with respect to search comprehensiveness and style. One of the important functions of the human experts in their normal mode is to keep costs as low as possible while balancing user needs and resources. In doing so these experts will try to keep to a minimum the largest single component of the costs: online time at the terminal. This, in turn, is achieved by keeping the online session moving along as briskly as possible; the expert typically takes only the minimum time necessary to explain to the users what she is doing and why (the experts try to prepare the users in the interview before the online session) and requests the minimum amount of document information for online display — just what is sufficient to assure (with user concurrence) that the searches are giving adequate results and that the offline printouts requested will complete the user's need for information.

In resolving this conflict it appears that the human experts chose some intermediate ground between their regular searching practices and those directly responsive to the 'cost-free' experimental environment and the users' natural desires. Qualitatively, the practices employed in these experiments by the human experts seemed closer to their normal practices (i.e., emphasizing cost reduction) than to what might be interpreted as applicable to a more strict adherence to the 'cost-free', user-directed modality. Another reason for the higher display time by the users is the fact that they operated primarily at a 120 CPS terminal compared with the 30 CPS terminals used by the experts at their search centers and thus were able to achieve more productive results in terms of output for a given period at the terminal.

In any case, we see that the number of relevant documents found and viewed online was much higher, in general, for the user CONIT sessions than for the human expert sessions. In particular, the average of 19 users was 20 relevant documents. The average for 14 human experts was 17 documents whereas in those 14 cases the user average on CONIT was 25 documents. The median figures show an even more striking difference. The median figure for the 14 human expert cases was 9.5 documents while the corresponding 14 CONIT user cases had a median of 17.5 documents (the median was 8 documents for the 19 users). This large superiority for CONIT users is further evidenced in the pairwise comparisons. In 8 cases the CONIT user found and displayed many more (again using our 20% difference criterion) documents online than the human expert and in one case the CONIT user displayed somewhat more documents. In 2 cases identical results for this parameter were noted for user and expert. On the other hand, the human expert displayed many more documents in only 2 cases and somewhat more in 1 case.

Before further analyzing the *differences* between human and expert intermediary searchers, let us first review some statistics which support our previously made statement that *both* end users and human intermediaries did well. The most pertinent parameter may be fractional recall — '% TOT' in Table 1. The average and median figures for this parameter all range in the span of 0.4 to 0.5, which we feel is quite good. In contrast, the figures for the CONIT-3 experiments showed an average fractional recall by the end users of only 0.05 and a median figure of only 0.025. Clearly, the enhanced intermediary techniques of CONIT-4 provide for much greater effectiveness of searching.

A small part of the improved effectiveness in terms of recall could be attributed to the difference in topic comprehensiveness between these two sets of experiments. For the CONIT-3 experiments the average recall base was estimated at 780 and the median was 810. For the CONIT-4 experiments the average was 513 and the median was 128. On the average we would expect the the higher the recall base the lower the fractional recall would be.

It is difficult to find out comparable recall statistics in the general literature. Usually, the recall base is established, if at all, as that found explicitly in a few more-or-less standard searches, as opposed to the numerous, broad searches that we have carried out in estimating this parameter. A figure of 0.4, for example, for recall found in the less comprehensive way could easily be inflated by a factor of from 2 to 4 over the true recall figure. Results from recent experiments at Syracuse University [KATZ81] tend to support our contentions. In these experiments very low overlap of results by expert searchers imply recall values may average as low as 20% or lower. Of course, the actual level of recall achieved depends on may factors including the type of indexing available and the level of relevance or utility demanded in the judgment of "relevant" documents.

With only the relatively meager statistics provided in the various experiments, we do not have sufficient evidence to *prove* that our CONIT users actually achieved higher than 'normal' recall values. Indeed, it might be argued that the experts' recall figures, which are roughly comparable to the users' figures, are probably in the 'normal' range for document reference retrieval. However, we would counter that the users' desire for high recall, the probably higher-than-normal 'free' computer time that was used, and the at least subconscious desire to 'do well' in these experiments (which they knew would be subjected to extensive comparative analysis) all lead to the hypothesis that the experts' recall, along with that of the end users, was probably on the high side in terms of what is regularly accomplished by expert searchers.

Our users did seem to generally want high recall. It may be that their interests, centering on academic and research pursuits, were such as to put them in a category of users seeking higher-than-normal recall. In a number of instances the users were fairly good in predicting beforehand what the recall base would be. In other instances they either could not make any estimate or their estimate was very low. Users were, in general, quite happy with the results of their searching, including the level of recall. In some cases (e.g., user E) we observed that both user and human expert intermediary thought that they had high recall and were content with their searching, but it turned out under analysis that the recall base was actually much higher and the user did, in fact, want the much larger number of documents discovered by the analyst's searching.

Even where users may not want to, or be able to (where the recall base is measured in the hundreds or thousands), see every document, they often want at least to know how many relevant documents there are and be able to view extensive sample titles and/or abstracts so as to insure that they know the general nature of all the relevant literature available and to increase the certainty that they are not missing any important documents. Our experiments lend strong credence to the hypothesis that existing expert searching generally gives only rather vague, generally intuitive, notions of the recall levels achieved. We suggest that one important improvement in current search practices would incorporate more quantitative and definitive recall evaluation procedures (see Section 9 below on future considerations).

Let us now look at other parameters and detailed experimental results by which we can compare and contrast end user and human expert searching. Befitting their much greater searching experience, the human experts were by and large at least superficially more sophisticated, complex, and comprehensive in their searching. From the table we see that the information specialist generated more retrieval sets than the end users — an average of 43.8 compared to 37.5 and a median of 40.5 compared to 30.0. Also, the human expert derived more individual terms for searching: an average of 18.6 compared to 12.3 for the end users and a median of 16 compared to 9 for the end users for this parameter.

In addition, the information specialists regularly took advantage of such precision-enhancing devises as proximity searching, important term searching, and subheadings and other controlled vocabulary searching. Furthermore, these specialists used such recall-enhancing devices as truncation searching and searching on all more specific terms for a given term (e.g., MEDLINE "EXPLOSION" searching). By contrast, the end users made no use of proximity searching or important term searching as these search modes are not available in the CONIT common command language and none of the users was sufficiently knowledgeable in the command languages of the retrieval systems to be able to use these features. In fact, none of the end users ever used *any* retrieval system commands directly in the pass-thru mode via the SEND command.

Controlled vocabulary term searching is possible in a fashion in CONIT common command language mode through a combination of (1) dictionary/thesaurus term lookup, index file browsing (SHOW INDEX), or selection of terms from the document catalog records and (2) the FIND EXACT or FIND TERM commands. A number of end users did make more or less effective use of this mode using terms found in document records or index files (no user made any successful use of the dictionary/thesaurus files); however, the information specialists, with their experience with thesaurus/controlled vocabulary terms, naturally made significantly greater use of this search modality.

On the other hand, the special functional features of CONIT along with its instructional capabilities considerably reduce the potential advantage of the human expert. The automatic keyword/stem truncation all-fields searching done by CONIT actually tends to push the end user searching into a *more* recall-oriented direction than that we observe for most human experts. Explicit 'explosion'-type searching is still the province of the knowledgable searcher, but the increasing tendency to automatic posting under broader terms — for example, the MEDLINE's "pre-explode" — together with CONIT's default search mode, which uses this type of indexing when available, is reducing the need for the explicit knowledge of this powerful tool. Thus, for example, in topic N, the end user was able to perform searches of the pre-explosion MESH terms 'brain' and 'behavior' in MEDLINE; without the pre-explosion feature user N would have fallen far shorter of matching the human expert's performance than she did. In contrast, for topic J the information specialist explicitly made good use of an explosion on term 'polymers' which was not pre-exploded and therefore was not easily available to the end user.

Nevertheless, with their greater expertise it is quite reasonable that the human experts did, in fact, develop considerably more sophisticated and comprehensive-appearing search strategies. Why, then, did not the human experts do much better than the end users in their results? There are several answers to this question. In some cases, of course, the human experts *did* do better. They achieved their results in less online time, even if we look specifically only at the "search time": average of 40.6 minutes for the human experts versus 87.9 minutes for the 15 corresponding end users or a median of 29 minutes for the human experts versus 72.6 minutes for the end users. Part of this relative speediness was due to the precision-enhancing modes mentioned above. Contrarily, in many cases the end users took advantage of the additional time they spent in two ways: (1) they simply took longer to get to the same kind of searches and results and (2) they adopted a more exploratory behavior, including the previously mentioned greater output of document records online, which led in some cases to the discovery of more efficient search strategies and in some cases to greater recall through the searching of additional databases (for example, see user N).

Part of the increased exploratory behavior for the end users could, in some cases, be attributed simply to their greater feeling of the lack of restraint over using 'free' computer time, as we have discussed. However, it is our belief that another factor could be a negative effect of the human experts' greater experience: since they 'know' that certain databases are less likely to be useful, they do not even try them. This appears to be the case, for example, in topics F3, G, J, N, and P.

There are two other reasons which sometimes seem to favor the 'naive' end users. In the first place, the end user obviously knows his problem better than the human intermediary. We have observed that there is sometimes a failure to transmit fully and correctly either the nature of the topic of the other goals and parameters of the search -- for example, the desire for high recall or for rapid access to the search results (i.e., more emphasis on online as opposed to offline output). A second component of this factor is that end users may intuitively appreciate what the best search terms are. It is true that the human intermediaries will generally ask the end users for suggested search terms; however, the ultimate decision on which terms to use and in what order is the responsibility of the intermediary searcher and will not necessarily follow the intuitions that the end users themselves feel freer to follow in their end-user CONIT searching. Related to this problem is the mistake of overspecifying the search statement in an attempt to make sure that all aspects of the topic are covered. The end user may say some particular aspect "must" be covered in searching; this may lead the information specialist intermediary to include that aspect in the search strategy when the other aspects may be sufficient to maintain sufficiently high precision.

Another explanation for these kinds of results has also been mentioned by us and others in previous research. (See, e.g., [DVEH73].) That is, it often appears that fairly simple strategies work rather well. Thus those rather simple keyword strategies that CONIT promotes often seem to work better than the more sophisticated strategies employed by the human experts. In particular, it seems that many times the precision-enhancing devices employed by the experts are counterproductive: the amount they raise precision, if at all, is not worth the additional time and expense involved or the consequent loss in recall.

Some examples of the previously mentioned situations are given below.

In topic I the information specialist tried hard to characterize within the search strategy the immunological, lymphatic, and allergic aspects of the user's interest. In one database, MEDLINE, this raised the precision (over the results from the end user's simple CONIT strategy of the disease term 'Kaposi's Sarcoma' alone) from 19% to 25% (not very spectacular) while reducing recall from 100% to 83% (the recall loss in a second database, EXCERPTA MEDICA, for this strategy would be 70.5% to 58.8%).

For the topics of user E1 there was a gross loss of recall by both end user and information specialist. The end user's main problem was in not recognizing that he had to generalize the 'spontaneous motor action' concept to a term like 'behavior'. For the information specialist it was a problem of overly specific search strategy: the term 'behavior' was recognized as important but it was coordinated with too many other terms; this problem was apparently due to the end user's poor explanation of what was truly critical for his problem and the information specialist's failure to test this through simple, broadened searches. It must be said, however, that to get high recall for this problem one must apparently accept rather low precision (under 10%).

Similarly, for topic E2 the recall failure for both end user and human expert can be attributed to the end user's strong and erroneous bias (not corrected by the information specialist) to incorporate the apparently inappropriate animal-type terms.

For topic F1 the end user had a simpler, faster, and more effective search strategy.

For topic G the information specialist, again following the end user's statements, was overly restrictive in search strategy formulation by insisting on coordinations including the aspect 'kidney'. In this case, the end user himself did *not* make the same mistake.

For topic H the information specialist did successfully find (in the MESH vocabulary) a synonym for 'transglutiminase' (glutymal transpeptidase) and thereby raise recall. The end user, however, taking advantage of the SHOW INDEX function, used searches on stemmed forms more effectively.

For topic P a failure to communicate adequately the wide-ranging nature of the topic led the information specialist to concentrate on certain subsets of the problem (therapy over causes).

5. Instructional and Learning Considerations

As we stated earlier, the users, by and large, learned most of the basic CONIT commands fairly quickly and well. In the CONIT-3 experiments we noted the amount of time that it took users from the time they started their online sessions until they first successfully used key CONIT commands: PICK (to select and connect to a database); FIND (to perform a search); and SHOW (to output specified catalog information from retrieved document references). The average time in minutes to achieve these critical search junctures for the 6 CONIT-3 users was 9, 19 and 25 minutes, respectively. For six of the CONIT-4 users (C1, F1, F2, G, H, and I) the corresponding figures are comparable and, in two of the three cases, actually somewhat lower: namely, 11, 17, and 22 minutes, respectively.

While there are too few cases and too much variability in individual cases to draw statistically significant conclusions, these figures do support our intuitive conclusions derived from analyzing the experiments that CONIT-4 users appeared to learn system commands approximately as quickly as the earlier CONIT-3 users had. This appears true despite the somewhat more complicated nature of CONIT-4 and the added online instruction given to the users in the early part of their engagement with CONIT in the nature of search strategy formulation suggestions. Countering these factors that would potentially increase users' time consumption, we have noted that modifications to instructional dialog and more automated CONIT system functionality (for example, automatic phrase searching) could be expected to lead to faster learning and effective use, especially in the latter stages of the critical search junctures.

In fact, the statistics given above do point in that direction: namely, the six CONIT-4 users averaged 2 minutes *more* to achieve their first PICK but 2 minutes *less* to achieve their first FIND and 3 minutes *less* to achieve their first SHOW as compared with the CONIT-3 users. Again, however, we must point out that there is insufficient data to confirm this kind of hypothesis at the present time.

While these results appear relatively promising in the absolute and comparative senses — how well CONIT-4 users learned and how well they learned compared to CONIT-3 users — we did observe, on the other hand, many difficulties in CONIT learning and use. Users made various mistakes of commission, omission, and other misunderstandings. For example, user E typed 'sd 5' which gave him the SHOW DATABASES results for area 5 when he actually wanted to show document 5 (i.e., 's d 5' — need a space between 's' and 'd'). A second example is that user J typed 's 4,7,9' in an attempt to look at 3 different non-consecutively numbered documents with one command — not yet possible in the CONIT common command language. As a third example, user E attempted to use the FIND command to combine two sets instead of the COMBINE command. Two other examples, derived from the session of user N: she typed 's 5' instead of 's d5' in an attempt to see document 5 and she used both database name *and* number in the REPEAT command. We have previously cited the case of user C who tried to FIND (search) before PICKing a database in which to search.

In general, users were able to recover from these primarily syntactic mistakes with relatively little trouble; the error messages or other responses from the intermediary system were usually sufficient to prompt the users to remember or look up the correct format. In a number of cases it is apparent that a somewhat more comprehensive or sophisticated intermediary system should avoid the problem entirely or relieve it greatly. Thus, it is planned to allow a sequence of non-consecutive document numbers in a SHOW command. This will require a *sequence* of several commands for the translation to those systems which do not allow such a functionality for a single command. Similarly, the FIND command *should* allow set numbers as arguments.

In some cases the mistake was more of a semantic one. Thus user K typed 'f computer aided diagnosis of eye diseases' in contradiction to the CONIT instructions which said only *keywords* should be used in search statements. As before, a more sophisticated intermediary should *automatically* recognize this situation and simply ignore the common, non-content words like "of".

Users were less than optimally efficient and effective in their use of CONIT in that they failed to take advantage of existing features of the intermediary system. Thus, a few users did primarily single word searches and subsequent intersection searches when the phrase-type search was available to them. Similarly, searching multiple databases with the repeat command could have speeded search operations for some users. In other cases it was clear that showing the full document record for relevance feedback purposes should have assisted users in search strategy formulation. Similarly, the SHOW INDEX command could have been used to find alternate search terms in a number of cases in which it was not so used.

CONIT-4 experimental users spent considerably more time at the terminal than did CONIT-3 users. The average and median times for CONIT-4 users were 173.1 and 127.1 minutes, respectively, versus 86 and 70 minutes, respectively, for these same parameters for CONIT-3 users. Even if we deduct the 'system-problem' times for CONIT-4 users, their session times are still much longer: an average of 153.6 minutes and a median of 119.2 minutes. In fact, in 5 cases (as we detail in Table 2) the users spent two online sessions to complete the searches on their topics. Also, three users came back to perform searches on a second topic and one of those three searched a third topic.

One apparent reason for this additional time is the greater success in finding relevant documents afforded by the enhanced CONIT-4 intermediary system. We have previously cited the much higher absolute recall figures for CONIT-4 users compared to CONIT-3 users. The multi-topic usages all resulted from users specific, unsolicited requests to act again as experimental users. In these cases the users felt that their initial search sessions had been successful and they had additional topics that they expected would meet with similar successful results using CONIT.

In any case, these extended and multiple usages gave us an opportunity to more closely compare the learning curve for users. While we do not have definitive results on this factor, our observations and the accounts of the users themselves provide some indications of the nature of the learning curve. While the initial period of extensive CAI in a CONIT session (i.e., the first general explanations of the interactive dialog and the basic commands followed by explanations of basic search strategy formulation issues and the specific commands EXPLAIN, PICK, FIND, and SHOW) usually take only 10–20 minutes, it is observed that a period of learning consolidation through use may be required before the user feels comfortable with the operations. Thus, the user may express this level of achievement toward the end of a one-hour or two-hour session. Users also express and show a greater sense of confidence at second and subsequent online sessions. Some indication of this learning curve is found in the EXPLAIN times of the multi-topic users. User C spent 35 minutes reading the printed offline manual before the first online session on his first topic (C1). In that session he spent 11.7 minutes of EXPLAIN time. He had a second session on this topic three weeks later in which he spent only 3.2 minutes of EXPLAIN time. Other statistics suggestions user C's learning curve are the times to first PICK, FIND, AND SHOW which were 15, 19, and 23 minutes, respectively, for session one and only 2, 10, and 12 minutes, respectively, for session two. Four weeks later user C had the online session for his second topic. No additional offline reading of the instructional manual was made and the EXPLAIN time was 7.7 minutes (perhaps showing a small fallback on the learning curve due to forgetting.

User E, as we have mentioned, spent 3 hours analyzing the printed instructional manual before the online session for his first topic. Since he felt that he had learned essentially all he needed to know for operating CONIT through the printed medium, he spent only 0.5 minutes of EXPLAIN time in his first session. For the online session for his second topic (E2) which occurred approximately one year after his first session, user E did not review the printed manual any further; his online EXPLAIN time was 18.9 minutes.

User F showed a steady reduction in her EXPLAIN times in her three sessions, all of which were in the experimental mode in which no prior use of the printed manual was allowed. Her online EXPLAIN times were 16.9, 9.5, and 3.2 minutes, respectively; the second session was 2 months after the first and the third was 1 month after the second.

Our observations on learning successes and difficulties support conclusions on the computer/human interface techniques we have stated on previous occasions and summarized in Section 2 above. Namely, the interface should, for optimum ease of use, be well designed for simplicity and have sufficient CAI capabilities. We feel that our choice of simplified command/argument language with natural language features is justified. The need for a *variety* of instructional techniques has been supported, if *all* classes of users are to be satisfied. Some techniques which we see as valuable but which we have *not* fully integrated into the computer system include:

(1) an example search session that the user could reproduce which would demonstrate system operations;

(2) printed instruction manuals and brochures of various levels of comprehensiveness;

(3) pre-online (standard) instructional classes;

(4) online human consultant help (via computer messaging) on demand;

(5) bi-modal response language and instructional levels (for the experienced as well as the naive user mode currently in CONIT);

(6) more sophisticated, dynamic, and individualized error and misuse detection and instruction;

(7) even faster terminals and telecommunications for higher information retrieval system to user throughput.

6. Automatic Database Selection

In addition to the other enhanced techniques described in Section 2 we have investigated the concept of using an existing multi-disciplinary database together with a "classification mapping" scheme to select databases automatically. The technique being studied (described in greater detail by Deane [DEAN80] and Marcus [MARC80]) involves several steps. First, the user describes his topic by a natural-language phrase. The CONIT language form for this is "FIND DATABASES <phrase>" (abbreviated "FDB <phrase>") where <phrase> is the natural language phrase. The computer intermediary then performs an automated keyword/stem search (see above) using this phrase in a multidisciplinary database having documents which have been indexed according to a classification scheme, as well as by other subject indexes. Next, the classification codes for some of the documents retrieved by the search are extracted from the appropriate catalog fields for those documents. These codes are then looked up in a "classification map" which identifies databases relevant to given classifications: the more relevant, the higher the assigned "weight". Databases are ranked by the (weighted) number of times they are found by the map for the codes of the documents retrieved under the search topic.

We label this technique MSCM -- for multidisciplinary search with classification mapping. In our current implementation of MSCM, the multidisciplinary database used is NTIS (National Technical Information Services) with its COSATI/WGA/GRTA Subject Category (classification) codes. A search in NTIS on "hearing impairment" for example, finds documents classified under the heading "6J". This category code stands for the area "BIOLOGICAL AND MEDICAL SCIENCES, Industrial (Occupation) Medicine" for which the classification map ranks such databases as MEDLINE and EXCERPTA MEDICA as likely to be highly relevant. In a variation of this technique, the user is asked to estimate the relevance of the retrieved documents; this ranking then modifies the weights by which the classification codes influence the estimated relevance of the databases.

It is the document-searching feature that we hypothesize may make the MSCM technique superior in some situations to alternate methods [MARC79, WILL77, ANTO79] which are based on the use of posting information on individual words and terms. The reasoning behind this hypothesis comes from three observations: [1] the searching may, itself, give the desired retrieval results directly from the multidisciplinary database, or indirectly, through repeating the search in other databases; [2] a given free-form problem statement phrase (e.g., "hearing impairment") is not likely to be found in that exact form in any reasonably sized set of terms -- or if it is, that phrase is unlikely to be used for indexing by all databases of interest; and [3] individual words tend to be used too diffusely or ambiguously to provide precise database indications. Thus in our previous example, "impairment" and "hearing" are used widely in many different databases in many areas, but the correlation of the pair of words in relevant documents points clearly to the medical area.

In the CONIT-4 experiments described above we did not use the MSCM automatic database selector but rather used the simpler computer-assisted techniques of CONIT-3 which we have described in previous papers (e.g., [MARC81a]). These simpler techniques proved generally satisfactory in our current experiments. As we have previously indicated, users fairly quickly and easily found most of the most-relevant databases. The biggest single problem seemed to be the failure of users -- and, sometimes, of the human experts -- to use the more general databases like NTIS, CDI, Science Abstracts, etc.

On the other hand, we believe that the newer automatic database selection techniques will become increasingly more attractive as

(1) more databases are added to the computer network, making the simpler schemes less tractable;

(2) these automatic selection schemes are more completely integrated with the other aspects of the intermediary system so that the advantages of search strategy formulation assistance in the multidisciplinary database are better incorporated into the other intermediary system functions; and

(3) more automatic techniques (see Section 9) are incorporated into the intermediary so that even more naive users can effectively do searching and so that searching for all users can be more highly automated at the user's discretion.

7. Interpretation of Results

In some sense, the single most conclusive parameter by which we can consider the relative merits of computer and human intermediary assistance is the evaluation given by the users themselves. Users regularly stated that they preferred the computer intermediary approach to that of the human intermediary. In view of the fact that several objective costs/benefits parameters seem to show that the human experts are still about as good, or even better, than end users working alone with the CONIT intermediary, we must examine these attitudes carefully.

A. Potential Experimental Biases

Clearly, a part of our users' positive attitude toward CONIT stems from the underlying preference for the user search mode in which they can avoid what they perceive as the awkwardnesses of the human intermediary mode: namely, the need to make an appointment with the information specialist at a mutually agreeable time and place; to attempt to explain in detail to the information specialist, who is not a subject expert, the nature of one's problem; and the inability to guide the search directly without more, possibly difficult, interpersonal communication with the information specialist during the search session. Our users are apparently even willing to trade off some loss of their perceived online session time cost, or even recall loss, to achieve the benefits of the user search mode. Also, as we have seen this evaluation may be reflective to a considerable extent of the differences in search modes induced by the 'cost-free' experimental environment. Thus users are clearly less likely to praise their own increased output of relevant document references found online if they are forced to pay -- or, at least, appreciate the cost -- of the additional online time.

In addition, we should consider whether the experimental procedures themselves have biased the user group toward being one which is more likely to be predisposed to CONIT use. One possibility is that our users might come from a class whose academic level and experience lends itself more to successful usage than the general class of potential users. The typical CONIT user, a doctoral student at MIT, can be expected to have a reasonably good aptitude for learning the CONIT system. It is our feeling, however -- based on the kind of skills necessary, the nature of the CONIT instructions, and some experience with users of lesser academic and computer-experience background -- that this is *not* a crucial factor.

Another factor that may be more critical is the nature of the experiment itself. Users who volunteered knew that they were trying out a new information system and most had some idea that there would be some analysis of how well they did with it (although our stated purpose was that we were analyzing how well the new *CONIT system* did and what *its* shortcomings were). The volunteers, then, had to be prepared to meet this challenge. Also, it became known in the course of the experiments that the results of their own searching were to be compared with those of the information specialists. It might be expected, then, that the users might draw some personal satisfaction in being able to declare a greater, rather than lesser, degree of satisfaction with their own searching. We do not doubt that these psychological considerations influenced users in their judgments to some extent. One might anticipate that in the general population there would be potential users who would be less likely to use and praise a computer intermediary system. Nevertheless, we believe that our users' judgments are likely to be reflective of what we might expect from a large part of the potential user population.

Another potential bias in our experiments relates to the almost exclusive concentration of search topics to the medical and biomedical areas. However, as we have previously indicated, the terms "medical" and "biomedical" were taken in the generic sense in which medical and biomedical themes can be intermixed with aspects of other disciplines. In our experiments, these other aspects included those from engineering, sociology, psychology, education, government, industry, computer applications, and planning. In addition, over 20 different databases from 3 different information retrieval system (4 systems if NLM and SUNY MEDLINE are counted separately) were accessed. In these databases was a spectrum of controlled vocabulary and free vocabulary modes of indexing. Thus, while these experiments had a medical/biomedical concentration, we believe that there was a sufficient variety of topics, databases, indexing, and retrieval systems so that there is some evidence that our results and conclusions are of general significance.

Similarly, we do not believe that our concentration on one particular information specialist caused a significant bias in the results. The searching of the other three information specialists did not seem, as far as we could determine, to differ markedly — as a group or individually — from the primary information specialist.

One other possible bias relates to the fact that, except in three cases, the human expert searching was started only after the CONIT searching was completed. (In only one case — user L — was the human expert search *completed* before the CONIT search began; in the other two cases — users C1 and D — there was some preliminary, partial searching done by the human expert before the CONIT search.) Again, despite the attempts we made to have the searching in the two modes be independent of each other, we might expect that the human expert searching would benefit, at least in some subtle ways, from the prior end user CONIT searching. This effect might tend to cancel the previously mentioned one of users performing or judging so as to enhance the evaluation of their own searching.

Our own belief is that these various potential biasing effects, either singly or as a group, did not affect the basic conclusions to be drawn from our experiments. Of course, we also believe that much more experimentation and analysis needs to be accomplished before all of these factors can be properly and definitively accounted for. In these experiments our tentative conclusions (or their inverses) could serve as experimental hypotheses.

B. Tentative Conclusions

In light of this variety of experience from our experiments we have arrived at some initial, tentative conclusions. The main one is that, in fact, the approach manifested in CONIT of a computer using natural English phrases with an all-fields keyword/stem truncated Boolean intersection search derived from those phrases, is an effective search approach across a wide variety of subject topics, disciplines, databases, and retrieval systems. While this conclusion appears especially true for inexperienced end users in their initial (as opposed to more refined) searching, in light of the more effective searching done by our end users compared with the human experts in some cases, we would further postulate that even expert human intermediaries could also benefit at times by a greater reliance on this simple mode of searching. This appears particularly true in two respects: (1) this simpler searching should be emphasized more in the initial for a search; and (2) this searching modality could allow some information specialists to search with some degree of assurance of effectiveness databases that they now avoid.

In regard to the first point, it may be observed that the simpler searching is generally more efficient and less costly in terms of human and computer time. If it does not work as effectively as desired, it can usually be modified fairly easily and quickly so that there is little loss even in those cases where it may be less than fully adequate. Also, we suggest that initial searching in this simpler mode more consistently allows for a more complete evaluation of the comprehensiveness of the searching and for a more effective base for planning refinements and modifications to the search strategy. Of course, it must also be stated that the simpler mode is not *always* less costly; for example, if a truncated search nears or surpasses the limitations of the size and number of sets the retrieval system can handle, it may take some additional time to recoup from this inadequacy of the retrieval system, even if done automatically by the intermediary system, as CONIT generally does.

In regard to the second point, we may note that some of our end users were particularly noted as potential experimental subjects by the computerized search services *because* other topics were multidisciplinary. It was clear that in these and other cases, information specialists will refuse to use, or will be less likely to use, databases and systems for which they do not feel expert. From the results of these experiments it appears that information specialists are shortchanging themselves and their clients in that they could probably get at least fairly good results most of the time by employing on these databases simple search techniques that do not require an expert knowledge of the database or its indexing policy. Of course, it may be somewhat more difficult for an information specialist to access a database which is implemented in a system for which she does not have easy (or any) access, or for which she does not have full and up-to-date knowledge of the command language. This last point suggests that there are currently situations in which even expert human information specialists could take advantage of a computer intermediary system such as CONIT.

While a common simplified general approach to searching most topics and databases has been shown in our experiments to work well -- at least for initial searching in many, if not most, cases -- it is, nevertheless, apparent that there are many situations in which a specialized approach taking account of topic or database peculiarities is required for fully complete and/or optimum performance.

While finding appropriate search terminology through feedback from full catalog record output of documents found through the simplified initial searching is apparently an excellent method for search strategy reformulation in most cases, the speed of the information specialist in some cases shows that an analysis of the controlled vocabulary for appropriate terms can *sometimes* be a more efficient and less costly approach. (How a user or an intermediary -- computer or human -- could *predict a priori* for which cases, or classes of cases, one approach is likely to be more efficient that the other for a given kind of user, is a question for future research.) To help a user make such an analysis of the controlled vocabulary CONIT suggests that certain thesaurus or dictionary databases be used. The particular databases suggested depend on which document database is being searched. Thus, for databases in the medical area, the NLM MESH and the BIOSIS BIOTERMS databases are recommended to the user.

In the experiments we found few attempts to follow these CONIT suggestions. We believe that the suggestions could be made more persuasive by, for example, showing examples of how to perform such searching of these databases and possible positive results of so doing. In addition, an advanced intermediary could demonstrate actual use of thesaurus databases for the particular user problem by automatically searching them for terms employed by the user and suggesting particular matching, synonymous, and other related terms so found to the user as potential search terms. Along with the potential terms themselves, the intermediary could lead the user to instructions on how to get additional information from the thesaurus databases.

We note that the kind of intermediary assistance described above, both in effect in current CONIT and projected for the future, is topic and database specific in one sense, but yet quite general in another. Namely, for example, the same dictionary/thesaurus searching principles, instructions, and implementation methods are applicable, in general, to the various topics and special databases. Furthermore we speculate that most of those situations which appear to require special treatment for certain databases are not the result of essential differences in the topics themselves but rather reflect differences in the *implementation* of certain databases.

Thus, for example, variations in the way certain databases are implemented on different systems include (1) whether or not a multiword index term is posted under each of its individual words (on MEDLINE and some other NLM ELHILL databases it is *not*; on most others it is); (2) whether and how more specific terms in a hierarchical classification scheme are automatically posted under the more general heading (MEDLINE does do this, but how to take advantage of this feature depends on the case — e.g., 'explode' versus 'pre-explode'; most other databases do not); (3) whether terms are automatically posted to certain near synonyms or new terms which were not in existence at the time the document was originally indexed (MEDLINE does more and better along these lines than many other systems and databases); (4) how many different indexed fields there are and which ones are searched in different search modes (MEDLINE has a very restricted default search mode in this respect but it does have an all-fields search mode); and (5) whether and how one can request specialized searching such as proximity, or "important-term" searching. The current CONIT handles these variations to a certain extent; a more sophisticated intermediary would handle them even better, for example, by attempting to develop more specialized search strategies for particular situations.

8. The Current Prospects of Intermediary Systems

A. Cost/Benefit Considerations

Effective end user operation of multiple, heterogeneous computerized bibliographic information retrieval systems has been demonstrated through our experiments with the CONIT-4 computer intermediary system. In addition, user preference for such a mode of access vis-a-vis more traditional human expert intermediary modes has been found in many cases. However, as we have pointed out, our experimental users did not have to weigh the costs in their considerations. We now seek to shed some light on these economic considerations by looking at costs for CONIT and CONIT-like systems.

Costs may be divided into several categories; the two main categories are development costs and operations costs. Operations costs, in turn, may be divided between overhead costs and incremental costs to run individual users. If there were a large user population, as we predict there would be based on our experimental users' reactions (assuming reasonable economics), the developmental and overhead costs would tend to be small compared to the direct incremental costs. Let us then concentrate, at first, on these incremental costs. For an intermediary system, we may divide these costs into costs for the host retrieval systems and for the intermediary system. The major component of retrieval system cost is generally a charge per connect hour which varies with the database. (Other costs from retrieval systems include possible yearly or monthly [minimum] subscription fees, charges for offline printing of portions of documents' catalog records, and, for a few databases, charges for online printing of this same kind of information.) Typical retrieval system connect-hour costs run from $22/hour to $100/hour. If connection to the retrieval system is made via a communications network, an additional fee of approximately $5/hour to $10/hour is imposed.

In estimating the direct incremental costs of the intermediary itself it is worthwhile to consider the costs to the project of running CONIT on MULTICS. The costs of running an *experimental* system like CONIT should be taken as an upper bound to the costs of running an *operational* system that would, presumably, be designed for high efficiency and minimum costs and not carry the burden of procedures designed specifically for research considerations. MULTICS costs are assessed by M.I.T. so as to recover close to 100% of the costs of maintaining this large multiprocessor, time-sharing system (up to 110 simultaneous online users). Typical costs for running CONIT-4 in prime time (9AM to 5PM weekdays) range between about $5 to $15/hour. The cost is variable because MULTICS costs for prime time interactive usage have two parts: one is a fixed $2/hour connect time charge and a second is based on the amount of central processing unit (CPU) resources used; this latter cost is presently $3/CPU-minute.

That up to 4 minutes of a very large CPU processor can be used by CONIT in one clock hour is an indication of the large process requirements in this experimental intermediary system. The CONIT-4 system takes up slightly over one million bytes of storage (258 pages of 1032 36-bit [4 9-bit bytes] words). Unlike the earlier CONIT systems, the CONIT-4 system consumes sufficient CPU resources so that when the MULTICS system is heavily loaded (for example, over about 80 simultaneous and active users) there can be a noticeable degradation in response and throughput times in passing the results of information from the remote retrieval systems. For example, under the worst conditions observed, we have seen throughput reduced from its theoretical maximum of 120 CPS to about 60 CPS.

Non-prime-time cost on MULTICS is reduced in steps to a low of half of that for prime time on weekends or a flat $4/hour in the midnight to 9AM shift. The main overhead costs for keeping the CONIT-4 system online are disc storage charges; at $0.01 per page per day this amounts to about $100 per month for the executable code. That figure can be roughly doubled when you include, as you must for maintenance purposes, the source code; it can be quadrupled or more when allowing for the various files that may be kept for research and development purposes.

Taking $50/hour as a reasonable average for retrieval system direct incremental costs, we see that our experimental CONIT intermediary system typically adds an additional 10% to 30% in costs when the retrieval system is actually connected. To make a meaningful evaluation in terms of costs and benefits we need to take into consideration all of the other costs such as human effort and time, as well as the perceived benefits. There are two main existing modes of operation for an end user against which we can compare potential computer intermediary use. One is the performance of the searches on the retrieval systems by the end user himself or herself. Wanger [WANG76] has stated that only 7% of searches are done by the end user; therefore, it seems evident that most end users do not have the incentive or time to ever learn how to access and use even one retrieval system. For them a computer intermediary is currently necessary if they want to make direct use of these systems without human intermediaries. Some small fraction of end users have the needed time, competence, incentives, and other resources to learn how to use existing retrieval systems; even for these, it seems clear that they could learn how to use a computer intermediary like CONIT much more quickly and, at least until they become proficient in one or more existing retrieval systems, would likely spend less session time, and therefore less money, searching via the computer intermediary.

In the mode where a user would work with a human intermediary information specialist we should consider the various benefits, previously described, that most end users find in being able to do their own searching. Still, even neglecting possible personal time end users might save in avoiding the human intermediary, we need to compare the out-of-pocket costs to the end user in the two modes. As our experimental experience has indicated, the CONIT connect-time per hour charges would be somewhat larger working through a computer intermediary. The main unresolved question is how the total online times would compare for the two modes. In some instances we have seen how the superior subject knowledge and search intuition) of the end user and the special functional capabilities and search philosophy of CONIT could actually reduce times and costs over typical human intermediary operation. Clearly, however, our experimental results, in general, showed a much larger time use by the end users. This larger time, as we have indicated, needs to be considered from several viewpoints. First, users were specifically encouraged to ignore cost considerations; users for whom payment by the minute was a factor (either personal payment or funding from some supporting institution) would presumably be more frugal in their online time. (Whether this kind of concern would inhibit their searching success is a valid question). Second, and in some sense a specific derivative of the first point, we have noted that our end users were consciously outputting more document information online than the human experts.

Clearly, there is a great deal of uncertainty in the precise evaluation of cost/benefit considerations for computer intermediary systems of the degree of sophistication already demonstrated in experimental systems. However, it appears that a whole class of additional usage by end users is now economically possible by those end users who prefer to do their own searching. It is also likely that a fraction of the usages now performed for end users by human expert information specialists could be more effectively done by computer intermediary systems, even for those end users who now find human intermediaries acceptable. Also, some information specialist usage could now be enhanced by computer-like intermediary systems. A more precise determination of the costs/benefits picture awaits additional development and research.

B. Other Computer Intermediary Systems

Testifying to the perceived advantages of computer intermediary systems, a number of such systems designed specifically to aid information retrieval have been proposed in the last few years and many of these have actually been implemented. We shall review several of these developments in this section so as to present some perspective on the status of the field as a whole.

Before discussing other intermediary systems as such, it is worthwhile to point out that assistance for retrieval operations is possible through systems other than intermediary systems *per se*. Of course, the individual retrieval systems themselves provide assistance through the informative messages given users, especially in the "LONG" or "VERBOSE" modes. They also, in general, allow for user-initiated requests for explanations of various topics (for example, "HELP" or "EXPLAIN" commands) and some of them have special instructional or training modes. Instructional modes may be incorporated directly into the retrieval system or they may be self-standing systems in their own right. In the latter category we have the

MEDLEARN system [TILL82] which is designed to assist in learning to use the MEDLINE system and the TRAINER system [CARU78] which emulates searching on several systems.

Meadow's IIDA [MEAD79] was a true intermediary system that assisted users in learning and operating several databases from one particular retrieval system (DIALOG). IIDA, an experimental system, used the basic software and hardware of the CONIT MULTICS system. It kept the command language of DIALOG but experimented with some relatively advanced and sophisticated techniques for detecting errant search behavior by users — for example, "thrashing" (e.g., many scattered searches without looking at results) and "dwelling" (e.g., too many similar combine commands with similar results). IIDA was originally conceived as a means to assist users who were already trained, or being trained, by more traditional methods. However, toward the end of the project some success was achieved [IIDA80] in having practicing engineers learn how to search solely by using IIDA. This result is a further indication of the potential power of intermediary assistance systems.

The "User Cordial" intermediary system developed by Goldstein [GOLD78] at the National Library of Medicine provided access to the CATLINE database (book catalog) of the NLM MEDLINE system. This intermediary, a minicomputer based system, gave an early demonstration of the concept of replacing one computer/human interface with an easier-to-use one. In this case the interface was a simple menu-driven one. In contrast to CONIT, the relatively simpler context of this intermediary system — single database, relatively few functions — the straightforward menu approach may be just as good as the mixed command/argument approach. Techniques developed in the User Cordial intermediary system are being further refined in the Integrated Library System [GOLD81] at NLM which, however, is not an intermediary system as such.

Another intermediary system which intends to provide an online catalog access was one devised by Fayen [FAYE82] for the Dartmouth College Library. A development system, based on a Terak 8510/a microcomputer and written under UCSD Pascal, allows the user to search Dartmouth catalog files which are stored and searched on the Bibliographic Retrieval Services (BRS) main computer in Latham, NY. The system assists the user in preparing his search which is run on the regular BRS software and command language. Plans for an operational system will bring the BRS software into the Dartmouth library's own DEC PDP-11/70. In this case, then, the intermediary is being integrated with the retrieval system and will remain as simply a front end or interface module to it.

The Searcher's Workbench of Preece and Williams [PREE80] is notable as an early example of a microcomputer based intermediary system. It featured a touch panel terminal with an incorporated Exidy Sorceror microcomputer in association with the main intermediary computer, a microcomputer by Alpha Microsystems with code written in AlphaPascal. As part of a plan for integrating user assistance techniques [WILL80] it proposed to incorporate the Vocabulary Switching software of Battelle [NIEH79] which is a system that provides a cross reference to the index vocabularies of a number of databases.

Several other microcomputer based intermediary systems have now been developed for operational use.

The OL'SAM (On-Line Search Assistance Machine) developed at the Franklin Institute Research Laboratory by Toliver [FIRL81] who also was responsible for much of the IIDA system software development. It is designed to run on a NorthStar Horizon 2 microcomputer with software written in UCSD Pascal. It features a common command language and a capability for multiplexing two users over a single telephone line and modem to the chosen retrieval system. It permits storing of search strategy for (rapid) transmittal to the retrieval systems and allows saving of output from the retrieval systems in named files.

Phil Williams at the University of Manchester developed his very compact USERKIT as an intermediary system [WILL81] USERKIT is built directly on the Z80 microprocessor so it is its own microcomputer. Despite its small size and low cost it can perform a number of kinds of assistance for the user including storing command sequences which can be evoked by abbreviated user command and transmitted to the remote retrieval system thus facilitating login and searching operations.

Another microcomputer based intermediary system was developed as a so-called "pre-prototype" intermediary system by the Computer Corporation of America (CCA) [HORO80] for the Chemical Substances Information Network (CSIN) of the Environmental Protection Agency. CCA has marketed a version of this system under the name CAST (Chemical Automated Search Terminal). CAST features automatic login; search strategy preparation, storage, and transmittal; and saving and editing retrieval system output. CSIN itself now has the most extensive operational intermediary system in a prototype version. It runs on a DEC VAX 11/780 system under a UNIX operating system. It has two modes of operation: DIRECT and SCRIPT. The DIRECT mode is essentially a copy of the pre-prototype system as embodied in CAST. The SCRIPT mode [BERG81] allows end users to prepare extensive search strategies in specific chemically-related topics (for example, chemical substance identifiers, toxic effects, and chemical manufacturing related information) through combining system-given query lists and user-generated lists with such specific search parameters as databases to search, dates of documents, and document type.

Several other intermediary system and related developments and plans have been announced in the literature. A partial list includes those by Petrie [DIAN81], Dayton [DAYT80] Neilson [NEIL79], Jamieson [JAMI79], Oddy [ODDY77] Shoval [SHOV81] Crystal [CRYS82], IXO Co. [MORG82], Pollitt [POLL81], Rosenberg, [ROSE81], and Burnham [BURN82]

63

One other system is worthy of mentioning here: the Network Access Machine (NAM) of the National Bureau of Standards [ROSE75], a minicomputer-based system designed to expand user-created macrocode to produce the interactive dialog necessary to access and operate resources on heterogeneous network host computers. The NAM system was specialized for the document retrieval application [TREU82] and one version of it has been used for that purpose in conjunction with a more generalized "Technical Management Information System" (TMIS) at the Lawrence Livermore National Laboratory [HAMP79]. NAM as a general purpose macro-translation system leads us to note that the intermediary approach has been tried for assisting access to and searching of non-bibliographic databases as well as for assistance of functions entirely different from retrieval. An analysis of this broader area goes beyond the scope of this report. Some discussion of attempts along this line may be found in [MARC81a].

All of the activity mentioned in this section lends support to our claim that intermediary assistance systems are now widely recognized as a major factor in improving user access to and operation of computer facilities. The variety of these efforts also indicates that there is no single, clear, definitive approach to take in this broad area. A great disparity exists in the set of functions to be performed by the intermediary and, consequently, on the preferred hardware and software organization of the intermediary. We shall discuss this further in the next section. It does seem evident that no current system matches CONIT in its ability to help inexperienced end users, especially in general searching in the broad spectrum of networks of multidisciplinary bibliographic databases.

9. The Future for Intermediary Systems

The current achievements and immediate prospects for intermediary systems have been shown to be great; the future possibilities for intermediary systems are much greater. We have previously [MARC81a, MARC82] looked into the future for intermediary systems. Some of *that* has now been accomplished, as described in this report. Most of the rest of what we previously discussed still beckons brightly. In this section we shall recapitulate these future-looking prospects and add some new prospects that our research has uncovered.

Reliability is a continuing concern. As we have noted, there were numerous system bugs and problems, the most common being network telecommunications problems. An experimental supervisor usually was able to overcome these problems for the user but they did impose interruptions and learning difficulties. In this regard we look forward to faster and more reliable computer-to-computer telecommunications through X.25 type network connections recently installed and now being expanded in MIT MULTICS and other computers.

The *comprehensiveness* of intermediary systems is an area in which we can look for much expansion. Increased comprehensiveness within the bibliographic information retrieval area can come through extension to additional retrieval systems and databases as well as added retrieval functionality in the common or virtual mode. We have noted that CONIT provided the important, basic retrieval functions in a common language mode but not many of the more special functions (e.g., proximity searching). Some other systems — e.g., CSIN — have developed common modes for handling some of these functions. We are just starting to see how some intermediary system research and development is considering how to broaden the scope of application beyond retrieval *per se* to such related applications as text editing.

The *configuration* of intermediary system software and hardware remains an open area for future research. We have seen intermediary systems located in facilities ranging from time-sharing systems on large mainframe computer complexes to tiny microprocessor-based systems built into intelligent terminals. How to distribute the various components of intermediary system hardware and software is a critical issue.

Perhaps the most exciting prospects for the future lie in the area of *modalities of assistance*. CONIT-4 and its direct ancestors require the user to learn a command language albeit a basically very simple one with much computer assisted instruction to make it relatively easy to learn and operate effectively. Still, there is required at least a few minutes to learn the basic commands. It is likely that some users, especially very computer-inexperienced and casual users, would prefer a computer intermediary that acted more like a human intermediary and would "speak" to them in a more natural, English-like language and "understand" and act on their needs as a human intermediary would do. We have carried out an initial attempt to explore this assistance modality. In two recent papers [MARC81b and YIP81] we have described what we have named the EXPERT CONIT (or simply, EXPERT) system which is designed to simulate a human expert while following the precepts of the so-called "expert" systems of the artificial intelligence genre with their knowledge bases and attention to rule-based programming languages.

EXPERT communicates with the user by asking questions, no command language as such is required. After a few preliminaries at the beginning of the session, EXPERT elicits from the user a problem statement in the form of a conceptual formalization in which the user's topic is expressed as an intersection of concepts or aspects and each aspect is represented as a union of terms. Then EXPERT gets the user to select a broad subject classification of his topic plus one or more specific subject classifications. Based on these selections the intermediary system ranks the databases as to their prospective relevance to the search topic. The user then selects one of these databases for searching.

EXPERT then connects to the selected database on a suitable system and proceeds to translate the concept formulation of the topic into a series of actual search statements. Each term under each concept from the initial concept table yields an all-fields, keyword/stem truncated search followed by a Boolean intersection of the component word/stem searches (i.e., the same kind of search done by CONIT-4 on a user given phrase). Next the union of all these term searches for a given concept is requested. Finally, an intersection of all of the concept union searches is performed. The results of each term search is reported individually to the user as it is received. EXPERT compares the number in the resultant set against the number given in the preliminary search goal statement by the user and gives suggestions on how to broaden or narrow the search if the resultant number is much smaller or larger, respectively, than the user's goal.

In order to provide more information from which the actual search reformulation can take place, EXPERT feeds back additional information on some of the actual documents found. First the titles and authors of the first ten documents are displayed. Then title, author, abstract, source, and controlled vocabulary index terms are shown for each document from the list of ten that the user selected as particularly relevant to his search topic. The user is next directed to select for each document displayed any free-text terms from title or abstract and any controlled terms that he believes may be additional good search terms. EXPERT tags each term with a number so that the user can select terms easily — no spelling of terms required (similar to the SDC/ORBIT PRINT SELECT feature).

The user is then asked if he wants to see more documents (in batches of 10) or proceed on to the search strategy reformulation process. If he decides to go on to reformulation, the various options, and reasons for choosing them (in terms of whether they are narrowing or broadening), are presented to the user. These options are essentially: adding, deleting, or replacing individual search terms or whole concepts. First the user is given the option of adding each of the terms found in the previous feedback process as a search term to one or more of the concepts. Then the user is given the more general replace, add, and/or delete options mentioned above.

When the user is satisfied that the concept table has been properly modified, EXPERT goes through the process of translating and rerunning the searches. Two differences from the first search iteration are: (1) for greater precision controlled vocabulary index searches are done on an exact phrase search basis instead of a truncated keyword/stem basis; (2) term searches previously made are remembered and used (i.e., they are *not* repeated), thus speeding up the search process considerably.

The feedback and reformulation processes are iterated until the user is satisfied with the results.

While we have run a few successful demonstrations with project members acting as end users, additional work would need to be carried out to extend the functionality and raise the reliability of EXPERT to the point where extensive testing with bona fide users would be possible. However, our analysis of the potential for this kind of computer expert assistant intermediary system suggests the following.

On the one hand, it does seem quite likely that a system of this type could be highly effective, especially for casual, inexperienced users. EXPERT also has features (computer assisted topic formulation and reformulation with speeded searching of the incremental search strategy) that could well serve expert users also. On the other hand, as we have seen in these experiments, the performance of human experts is often not optimally effective; therefore, the design strategy of simply trying to simulate what human experts do is not necessarily the best one. Also, the computerized expert does not lend itself easily to the great variety of detailed intelligent control of search strategy formulation and reformulation possible from the human end user. In addition, any intermediary, whether computerized or not, needs to pay more attention to evaluate how cost effective a search has been and estimate the cost effectiveness of possible additional modifications to the search strategy.

In conclusion, what appears to be needed for the further improvement of the effectiveness of assistance systems for the bibliographic retrieval application is the integration of the more completely computer-directed techniques as found in systems like EXPERT with the more human-controlled techniques as found in intermediary systems like CONIT. In such an integrated system it would be possible to shift modes of operation (under system or user control) as the context (user and topic) demanded. Exactly how such an integration and mode control can be achieved appears to represent the frontier in this area of research. It might be noted that research in automated or computer-directed assistance modes has been carried out by other investigators (see, e.g., [SALT80, DOSZ80, SMIT80, WALK81). Also, we should point out that further development of models and techniques for online searching will undoubtedly be incorporated into future intermediary systems (see, e.g., [ATHE79, BATE79, CALK80, CONG80, FENI80, FENI81, JAHO74, MARK78, OLDR77].

In assessing the potential of intermediary systems it should be noted that the goals presenting themselves for these activities are, in effect, "moving targets". There are several reasons for this. First of all, the retrieval systems are rapidly evolving. Not only does this mean that the intermediary system must be able to adapt to relatively minor (though critical) changes, like changes in protocols, but it must adapt to entirely new functionalities. The retrieval systems themselves adopt the successful techniques developed by the intermediary systems; when *all* retrieval systems do so, this may allow the intermediary systems to drop some functions. However, the tendency to new intermediary and retrieval system functions seems to be leading intermediary systems to greater rather than lesser requirements on the whole.

Finally, we must recognize the vital need for continued testing and analysis of intermediary systems in the context of the retrieval application and the basic information transfer process for which they serve. Our experiments are among the relatively few in the intermediary area and we have pointed out that much more experimentation and analysis is required before the conclusions we have drawn from them can be verified with the desired quantifiable statistical precision.

10. References

[ANTO79] Antony, Arthur. "The Database Index." *Database* 2(4):28–33, December, 1979.

[ATHE79] Atherton, Pauline. *Online Searching of ERIC: Impact of Free Text or Controlled Vocabulary Searching on the Design of the ERIC Data Base.* Syracuse University ERIC Clearinghouse on Information Resources Report, 1979.

[BATE79] Bates, Marcia J. "Information Search Tactics". *Journal of the American Society for Information Science.* 30(4):205–214; July, 1979.

[BERG81] Bergman, Rita. "Beyond SDI: On-line Retrieval Scripts in the Chemical Substances Information Network." *Proceedings of the American Society of Information Science 44th Annual Meeting.* 18:276–278; October, 1981.

[BURN82] Burnam, Judy K; Farrell, Eileen; and Levine, Emil. "Enhancing Online Searching and other Library Operations Using an Intelligent Floppy Disk System." *Proceedings of the 1982 National Online Meeting.* Medford, NJ: Learned Information, Inc. pp41-46; March 30, 1982.

[CALK80] Calkins, Mary L. "Free Text or Controlled Vocabulary? A Case History Step-By-Step Analysis Plus Other Aspects of Search Strategy." *Database* 3(2):53–67, June, 1980.

[CARU78] Caruso, Elaine. "Hands on online: bringing it home." *Online Review* 2(3):251–267; 1978.

[CONG80] Conger, Lucinda D. "Multiple System Searching: A Searcher's Guide to Making Use of the Real Differences Between Systems." *Online* 4(2):10–21; April, 1980.

[CRYS82] Crystal, Maurice I.; Jakobson, Gabriel E. "FRED, A Front End for Databases." *Online* 6(5):27–30; September, 1982.

[DAYT80] Dayton, D.L.; Lundeen, J.W.; and Pollock, J.J. "Automated Techniques for Online Search Guidance: A Review". *Proceedings of the Fourth International Online Meeting.* Learned Press. London. December, 1980.

[DEAN80] Deane, Thomas J. "Automatic Database Selection through Multidisciplinary Searching with Classification Mapping." Master of Science Thesis in Electrical Engineering and Computer Science, Massachusetts Institute of Technology. August, 1980.

[DIAN81] "Microcomputer Assisted Online Information Retrieval." *Euronet DIANE News* 25:4; December, 1981.

[DOSZ80] Doszkocs, Tamas; Rapp, Barbara A.; and Schoolman, Harold M. "Automated Information Retrieval in Science and Technology." *Science* 208:25–30; April 4, 1980.

[FAYE82] Fayen, Emily Gallup; Cochran, Martha. "A New User Interface for the Dartmouth On-line Catalog." *Proceedings of the 1982 National Online Meeting* Medford, NJ: Learned Information, Inc. 87–97; March 30, 1982.

[FENI80] Fenichel, Carol Hansen. "The Process of Searching Online Bibliographic Databases: A Review of Research." *Library Research* 2:107–127; Summer, 1980–81.

[FENI81] Fenichel, Carol Hansen. "Online Searching: Measures that Discriminate among Users with Different Types of Experiences." *Journal of the American Society for Information Science* 32(1):23–32. January, 1981.

[FIRL81] OL'SAM User Manual. Franklin Institute Research Laboratory; Philadelphia, PA. May 20, 1981.

[GOLD78] Goldstein, Charles M.; Ford, William H. "The user-cordial interface". *Online Review*. 2(3):269–275; September, 1978.

[GOLD81] Goldstein, Charles M; Payne, Elizabeth A.; and Freiburger, Gary A. "The INtegrated Library System Online Public Catalog." *Proceedings of the 44th ASIS Annual Meeting*. 18:279–281. October, 1981.

[HAMP79] Hampel, Viktor E; Gallo, L.E.; Kamin, R.A.; Kopytoff, V; and McGrogan, S. First DOE/LLL Workshop on the DOE/STOR Technical Management Information System (TMIS). University of California Lawrence Livermore National Laboratory. Report CONF-791258. 6 Dec. 1979.

[HORO80] Horowitz, Aren J; Eastlake, Donald E, III; and Low, David. "An Intelligent Terminal for Access to Heterogeneous Chemical Information Systems." *Proceedings of the 43rd ASIS Annual Meeting*. 17:83–85. October, 1980.

[IIDA80] Individualized Instruction for Data Access (IIDA). Final Report. Drexel University School of Library and Information Science and Franklin Institute Research Laboratories. June, 1980.

[JAHO74] Jahoda, Gerald. "Reference Question analysis and Search Strategy Development by Man and Machine". *Journal of the American Society Information Science*. 25(3):139–144; May, 1974.

[JAMI79] Jamieson, S.H. "The Economic Implementation of Experimental Retrieval Techniques on a Very Large Scale Using an Intelligent Terminal." *Association of Computing Machinery SIGIR Forum*. 14(2):43–51. 1979.

[KATZ81] Katzer, Jeffrey; McGill, Michael; Jessica, Judith A.; Frakes, William; and Das Gupta, Padmini. A Study of the Impact of Representations in Information Retrieval Systems. Syracuse University School of Information Studies Report. May, 1981.

[KEEN73] Keen, M.E. "The Aberystwyth Index Language Test." *Journal of Documentation*. 29(4):1–35. March, 1973.

[LANC71] Lancaster, F.W. An Evaluation of EARS (Epilepsy Abstracts Retrieval System) and Factors Governing its Effectiveness. University of Illinois Graduate School of Library Science Report. October, 1971.

[LANC72] Lancaster, F. Wilfrid. Evaluation of On-line Searching in MEDLARS (AIM-TWX) by Biomedical Practitioners. University of Illinois Graduate School of Library Science Occasional Paper 101. February, 1972.

[LOVI71] Lovins, Julie Beth. "Error Evaluation for Stemming Algorithms as Clustering Algorithms." *Journal of the American Society for Information Science.* 22(1):28–40; January, 1971.

[MARC77] Marcus, Richard S.; Reintjes, J. Francis. Computer Interfaces for User Access to Heterogeneous Information-Retrieval Systems. M.I.T. Electronic Systems Laboratory Report ESL-R-739; April, 1977. (NTIS:PB269126, ERIC:ED142180).*

[MARC79] Marcus, Richard S.; Reintjes, J. Francis. Experiments and Analysis on a Computer Interface to an Information-Retrieval Network. M.I.T. Laboratory for Information and Decision Systems Report LIDS-R-900; April, 1979. (ERIC Order No. ED 142180)

[MARC80] Marcus, R. S. "Search Aids in a Retrieval Network". *Proceedings of the 43rd ASIS Annual Meeting.* 17:394–396. October, 1980.

[MARC81a] Marcus, R.S.; Reintjes, J.F. "A Translating Computer Interface for End-User Operation of Heterogeneous Retrieval Systems; Part I; Design; Part II: Evaluations". *Journal of the American Society for Information Science.* 32(4):387–303;304–317. July, 1981.

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

[MARC82] Marcus, R.S. "User Assistance in Bibliographic Retrieval Networks Through a Computer Intermediary." *IEEE Transactions of Systems, Man, and Cybernetics.* SMC-12(2):116–133; March, 1982.

[MARK78] Markey, Karen; Atherton, Pauline. ONTAP: Online Training and Practice Manual for ERIC Data Base Searchers. Syracuse University ERIC Clearinghouse on Information Resources Report, June, 1978.

[MEAD79] Meadow, Charles T. "The Computer as a Search Intermediary". *Online.* 3(3):54–49; July, 1979.

[MORG82] Morgan, Chris. "A Revolution in Your Pocket." *BYTE.* 7(4):6–18; April, 1982.

[NEIL79] Neilson, W.J.; Townsend, V.S. "Alberta Computerized Enquiry System: A Feasibility Study". *Proceedings of the Seventh Annual Canadian Conference on Information Science.* pp54–66, Banff, Alberta, Canada, May, 1979.

[NIEH79] Niehoff, R. T.; Kwasny, S. "The role of automated subject switching in a distributed information network". *Online Review.* 3(2):181–194; June, 1979.

[ODDY77] Oddy, R. N. "Information Retrieval through Man-Machine Dialogues". *Journal of Documentation.* 33(1):1–14; March, 1977.

[OLDR77] Oldroyd, Betty K.; Citroen, Charles L. "Study of strategies used in online searching". *Online Review.* 1(4):295–310; December, 1977.

[OVEH73] Overhage, C. F. J.; Reintjes, J. F. "Project Intrex: A General Review". *Information Storage and Retrieval.* 10(5):157–188; 1974.

[POLL81] Pollitt, A.S. "An Expert System as an Online Search Intermediary." *Proceedings of the 1981 International Online Meeting.* 25–32. London. December 7, 1981.

[PREE80] Preece, Scott E.; Williams, Martha E. "Software for the Searcher's Workbench." *Proceedings of the 43rd ASIS Annual Meeting.* 17:403–405. October, 1980.

[RICH81] Richardson, Robert. "End-User Online Searching in a High-Technology Engineering Environment." *Online.* 5(4):44–57; October, 1981.

[ROSE75] Rosenthal, Robert. "Accessing Online Network Resources with a Network Access Machine". *IEEE Intercon Conference Record.* Session 25/3; April, 1975.

[ROSE81] Rosenberg, Steven. "Expert Systems and the Design of Powerful User Interfaces." *Proceedings of the 44th ASIS Annual Meeting.* 18:285–287; October, 1981.

[SALT80] Salton, Gerard. "Automatic Information Retrieval." *Computer* 13(9):41–56; September, 1980.

[SEWE76] Sewell, W.; Bevan, A. "Nonmediated use of MEDLINE and Toxline by pathologists and pharmacists." *Bulletin of the Medical Library Association.* 64:382–391; 1976.

[SHOV81] Shoval, Peretz. "Expert/Consultation System for a Retrieval Data-Base with Semantic Network of Concepts." *ACM SIGIR FORUM.* 16(1):145–149; Summer, 1981.

[SMIT80] Smith, Linda C. "Artificial Intelligence Applications in Information Systems." In: M.E. Williams, Ed. Annual Review of Information Science and Technology. White Plains, NY: Knowledge Industry Publications; 1980: 15:67–105.

[TILL82] Tilley, Carolyn B. "The Use of MEDLEARN." *Proceedings of the 1982 National Online Meeting.* Medford, NJ; Learned Information, Inc. 539–543; March 30, 1982.

[TREU82] Treu, Siegfried. "Uniformity in user-computer interaction languages: a compromise solution." *International Journal of Man-Machine Studies.* 16(2):183–210; February, 1982.

[WALK81] Walker, Donald E. The Organization and Use of Information: Contributions of Information Science, Computational Linguistics, and Artificial Intelligence. *Journal of the American Society for Information Science.* 32(5):347–363; September, 1981.

[WANG76] Wanger, J; Cuadra, C.A.; and Fishburn, M. Impact of On-line Retrieval Services: A Survey of Users, 1974–75. Systems Development Corporation Report. Santa Monica, CA. 1976.

[WILL77] Williams, Martha E.; Preece, Scott E. "Data Base Selector for Network Use". *Proceedings of the American Society for Information Science 40th Annual Meeting.* 14:(34); September, 1977.

[WILL80] Williams, Martha E.; Preece, Scott E. "Elements of a Distributed Transparent Information Retrieval System." *Proceedings of the 1980 Annual Meeting of the American Society of Information Science.* 17; October, 1980.

[WILL81] Williams, Phil. "A New Device to Simplify Online Searching and Reduce Costs." *Proceedings of the Second National ONLINE Meeting,* Learned Information, Medford, NJ. 503–514; March 24, 1981.

[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.