AN EXPERT SYSTEM FOR DOCUMENT RETRIEVAL

by

Man-Kam Yip

B.S., Massachusetts Institute of Technology

(1979)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

MASTER OF SCIENCE

at the

Massachusetts Institute of Technology

FEBRUARY 1981

© Massachusetts Institute of Technology

Signature of Author ____________________________

Department of Electrical Engineering and Computer Science, December 1980

Certified by ____________________________

Professor Francis Reinitjes, Thesis Supervisor

Accepted by ____________________________

Prof. A.C. Smith, Chairman, Departmental Committee on Graduate Students

ARCHIVES
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

MAY 6 1981

LIBRARIES
AN EXPERT SYSTEM FOR DOCUMENT RETRIEVAL

by

MAN-KAM YIP

Submitted to the Department of Electrical Engineering and Computer Science on January 15, 1981 in partial fulfillment of the requirements for the Degree of Master of Science

ABSTRACT

This thesis describes a computer-based Expert Intermediary System, EXPERT-1, which is developed to aid inexperienced users of bibliographic retrieval services to perform searching efficiently. The approach is to model a human expert's search procedures in terms of search-strategy formulation and explanations. The following types of expertise have been built into EXPERT-1:

1. Ability to assist the user to formulate his search problem;

2. Ability to formulate search strategy based on the concepts and search terms supplied by the user;

3. Ability to assist the user to select appropriate databases to search in;

4. Ability to handle different protocols and command languages; and

5. Ability to assist the user to reformulate search strategy based on partial search results.

The knowledge base of EXPERT-1 is implemented as a production-rule system. Each production rule is of the form IF...THEN...ELSE and represents a meaningful chunk of domain-specific knowledge. Invocation of rules is controlled by a goal-directed forward-chaining mechanism. The clauses of a production rule are written in a Lisp-like notation, i.e., (predicate argument1 ...), and are at sufficiently high level to enhance understandability and writability.

Informal experiments have been performed with EXPERT-1 and preliminary analysis of experimental results raised certain important issues in designing an effective intermediary which can replace human expert searcher.

Thesis Supervisor: J. Francis Reintjes

Title: Professor Emeritus
Acknowledgments

I am deeply indebted to Prof. J. Francis Reintjes, my thesis Supervisor, and Mr. Richard Marcus, the CONIT Project Co-supervisor, for their support and many helpful discussions during this research.

This research is conducted in the Laboratory for Information and Decision Systems at Massachusetts Institute of Technology.
CONTENTS

1. Introduction ........................................................................................................ 1
   1.1 Problems in Bibliographic Information Retrieval ....................................... 1
   1.2 Objectives of The Thesis ............................................................................ 2
   1.3 Overview of the Thesis ............................................................................. 3

2. Overview Of Human Expert Searching .......................................................... 6
   2.1 Roles of User ............................................................................................ 6
   2.2 Roles of Human Expert Searcher ............................................................... 7
   2.3 Expertise of a Human Expert Searcher ....................................................... 8
   2.4 Related Works .......................................................................................... 10

3. Design and Implementation of the Expert System ......................................... 14
   3.1 Design Goals ........................................................................................... 14
   3.2 Knowledge-Based Approach ..................................................................... 15
   3.3 Formalization of Expert Knowledge ........................................................... 16
   3.4 Production Rules as a Representation for Expert Knowledge .................... 23
   3.5 Retrieval, Refinement & Execution of Rules ............................................... 27
   3.6 Summary and Others ............................................................................... 28

4. Example of User/Computer-Based-Expert Interactions ................................. 29
   4.1 An Experimental Session with EXPERT-1 ............................................... 29
   4.2 Summary of Built-in Expertise of EXPERT-1 ............................................ 48
   4.3 Limitations of EXPERT-1 ....................................................................... 48
5. Conclusions .......................................................................................... 50

5.1 Justifications of Design Goals of EXPERT-1 ........................................ 50
5.2 Remaining Issues .................................................................................. 52
5.3 Future Directions ................................................................................. 54

6. References ............................................................................................. 55

Appendix I. DataBase Selection Rules ......................................................... 58

Appendix II. Listing of Production Rules .................................................... 72
1. Introduction

1.1 Problems in Bibliographic Information Retrieval

In the past ten years, many information retrieval systems for science, technology and medicine have been developed at an extremely rapid rate. There are currently more than 500 commercially available bibliographic databases that contain millions of bibliographic citations to the published literature of many subject areas, including engineering, medicine, business, social sciences etc. Some of the databases correspond to online versions of the printed indexes such as Index Medicus, Science Citation Index, the NTIS GRAI and many others. There are also databases such as SSII: that contain summaries of many ongoing research projects in the United States. Documents stored in these databases are typically indexed by both free-vocabulary -- title or abstract words -- and by controlled vocabularies found in various thesauri. Documents are searched by combining various search terms with Boolean operators (AND, OR, NOT).

These bibliographic databases are of great use to the scientific and medical researchers who need frequent access to the ever-expanding volume of literature in particular areas. However, despite the fact that online access to many of these databases, which are implemented on systems of major commercial vendors (e.g. Lockheed Dialog, SDC Orbit, BRS Retrieval Services), is easily available through various communication networks (e.g. Telenet or Tymnet), the impact of these systems on the scientific community is still limited. Many researchers are either unaware of the existence of these systems or use

1. In this thesis, the term "document" refers to a bibliographic reference about a published paper or report. The bibliographic information usually consists of the following: the authors, title, abstracts, year of publication, source, and index terms which are used to indicate the subject matter of the published paper or report.
2. These are also known as "index terms", "descriptors", or "subject headings".
them in a superficial manner that never fully exploits their capabilities. There are two major reasons for this:

1. The search process often requires a non-trivial amount of effort from the user of such search services. The user not only has to handle the different command languages, protocols, and search features of different retrieval systems and databases, but also has to formulate good search strategies\(^3\) for his particular search topic.

2. The cost of such search services is also quite expensive. The charge for each connect-hour to the vendor varies from $40 to $120. Inexperienced users are most vulnerable since they will spend much time in figuring out how to use the system and browsing hundreds of bibliographies that result from particular searches.

As a consequence, most of the users of these search services will refer their searches to trained search specialists\(^4\) who know how to cope with the differences among retrieval systems and how to formulate effective search strategies. Reliance on such search specialists has obvious drawbacks too. For they charge a non-trivial amount for their services. Moreover, the limited availability of such specialists also creates a hindrance to the widespread use of retrieval services.

1.2 Objectives of The Thesis

The thesis developed here is intended to solve the problems facing inexperienced users of retrieval services by providing an intelligent intermediary that can simulate most of the functions of a human search specialist in such a way that will allow the user to perform the search directly and efficiently. The underlying assumption is that one can perform effective searching by using the existing retrieval systems

---

3. Search strategy consists, in part, of the choice of (1) search terms, (2) search modes for these terms, and (3) boolean combinations of these search terms.

4. In this thesis, the terms "search specialist", "human expert searcher" and "human searcher" are treated as synonymous.
"smartly". It is hoped that such an intermediary can promote the use of retrieval services in the scientific and medical communities by being simple and less costly to use. The so-called knowledge-based approach to construct domain-specific experts that is adopted in this thesis represents the first major effort in applying this Artificial Intelligence technique to solve some of the problems of document retrieval in the field of Information Science.

1.3 Overview of the Thesis

In chapter two, I shall discuss the types of expertise required in a system that can perform most of the functions of a human search specialist during a search process. These include:

1. Ability to assist the user to formulate his search problem;

2. Ability to formulate search strategy based on the concepts and search terms supplied by the user;

3. Ability to assist the user to select appropriate databases to search in;

4. Ability to handle different protocols and command languages; and

5. Ability to assist the user to reformulate search strategy based on partial search results.

I shall compare a number of related works that attempt to solve the "user-interface" problem by different methods.

Chapter three will be devoted to the design and implementation of the expert system. Basically, this is a rule-based system, consisting of three major components:

5. Alternatively, one can assume the existing retrieval systems are "dumb" and proceed to design new retrieval model, e.g., vector processing, cluster analysis etc. [Salton 1968]
1. The **Knowledge Base** implemented by production rules each of which is of the form:
   \[ \text{IF } \text{Premises} \text{ THEN } \text{Actions ELSE Actions} \].

2. The **Global-Database** which saves information about the search process as well as stores the production rules.

3. An **Interpreter** program which controls the invocation of rules and contains functions used by these rules.

Chapter four contains discussion of an experimental session done with EXPERT-1, the current version of expert system. The purpose is to illustrate the main features of the system. The following types of expertise (as discussed in Chapter 3) have been built into the system:

1. Assist the user in characterizing his/her search topic as an intersection of major concepts each of which is represented by a union of search terms.

2. Suggest a ranked list of databases based on the user’s selection of "subject areas" in which the topic falls.

3. Handle all network protocols and login procedures of the retrieval systems.

4. Translate the search strategy implicit in (1) into a sequence of executable search commands that are appropriate for the connected database. The translation mechanism remembers previous search results, thus avoiding the repetition of searches that have been done before.

5. After the user reviews search results, the intermediary suggests to user how to reformulate search strategy; e.g., broadening or narrowing by addition or deletion of search terms etc.

6. Assist user in selecting potentially good search terms from the index terms of documents chosen by the user.

The expert system is especially useful for inexperienced users, and considerably reduces the length of connect time to the retrieval host since the expert system can interact with the host much faster than a human user. In this chapter, I shall also identify some of the limitations of the expert systems.

In the conclusion, I shall summarize the major issues raised in this thesis and suggest further work that may be done along this line of research.
(This page intentionally left blank.)
2. Overview Of Human Expert Searching

In this chapter, I shall give an overview of the process of human expert searching in online document retrieval. The main purpose is to identify the roles of the end user and the human expert intermediary during the search process. Based on this information, I can specify the types of expertise which are needed for the knowledge base of an expert system that can simulate a human expert searcher in the process of document retrieval.

At present, the most effective kind of online document retrieval usually involves the expert searcher working in conjunction with the end user. In the course of this thesis research, I have observed a number of searches involving this team approach. These observations seem to support the following description of the roles of these two participants in the search process.

2.1 Roles of User

Very often, the search is conducted in the presence of the user, i.e., the information requester, who plays two important roles in the search process:

1. Supply the necessary information to the search specialist about the search topic. This may consist of a written problem statement describing the general subject area (e.g. nuclear engineering) of his search, the detailed search topic, and often indications of the major concepts of his topic with a few terms or phrases that express such concepts.

2. Perform relevance judgment on the documents retrieved so that the specialist can have a basis for evaluating and reformulating the search strategies employed.

It is important, however, to note that the user is not expected to have any knowledge about the search

---

6. The term "end user" and "user" will be treated as synonymous, both meaning the one who will utilize the products of the search process.
environment, e.g., system protocols, command languages, search conventions, database characteristics etc.

2.2 Roles of Human Expert Searcher

Because of differences in style and experiences, and differences in problem contexts, human expert searchers do vary in their way of conducting a search process. The following list of functional tasks of an expert searcher is an abstraction of what a typical searcher would perform:

1. **Fact-finding.** Before searching online, the expert searcher usually conducts an interview with the user (typically, five to twenty minutes) during which the searcher gets acquainted with the user's background, information need, and purpose of searching. Often, to make the interview more effective, the expert will have requested the user to prepare a written statement describing his/her information need prior to the interview.

2. **Planning.** Based on the data collected in the initial interview, the searcher begins to plan his/her search. This involves making at least two decisions: (a) deciding which databases are best to search for the given problem statement, and (b) formulating initial search strategy, i.e., sequence of Boolean combinations of search terms and search modes. The latter decision may require the searcher to consult appropriate thesauri for the chosen databases.

3. **Explanation.** After the initial planning, the searcher explains his/her search plan to the user. This enables the user to understand what databases the searcher has chosen and how he/she is going to carry out the search.

4. **Translation.** This task is mechanical. The searcher gets online, connects to the selected database, and translates his search strategy into executable search commands. Notice that the translation varies among retrieval systems because of differences in search conventions and command syntax.

5. **Evaluation & Reformulation.** Normally, the initial search strategy will not be entirely satisfactory. It may fail in recall, i.e., not enough relevant documents retrieved, or it may fail in precision, i.e., many irrelevant documents retrieved. Search strategy reformulation proceeds on the basis of the number of relevant and irrelevant documents found. Some typical evaluation procedures are as follows: If the search strategy retrieves some documents from the database, the searcher will request the user to judge the relevance of some of these documents. If the user identifies some good documents, the searcher can pick out additional search terms from the title, abstracts or index terms. These new search terms may be used in the next iteration of searching. If the initial search strategy retrieves too few documents, the searcher will broaden or loosen the strategy, e.g., by either dropping less important concepts of the problem statement, or adding additional synonyms for each concept. On the other hand, if the search retrieves too many documents, the searcher will narrow or tighten the strategy, e.g., by
eliminating search terms that are too general and do not capture the meaning of the given concepts of the problem statement. This process of evaluation and reformulation is iterated until the user is satisfied with the results.

2.3 Expertise of a Human Expert Searcher

The observation in the previous section indicates a human expert searcher needs at least the following types of expertise in order to conduct a successful search:

1. Ability to identify the search requirement of the user by assisting him/her to formulate the search problem in terms of major concepts and search terms;

2. Ability to formulate search strategy based on the information supplied by the user;

3. Ability to select appropriate databases to search in;

4. Ability to handle different system protocols, search conventions and command syntax of different retrieval systems.

5. Ability to assist the user to evaluate partial search results, and reformulate search strategies based on the evaluation.

Item (4) in the above list is largely mechanical. Database selection - item (3) - is usually not difficult but does require enough knowledge about the subject areas covered by each database. Normally, a few dozen bibliographic databases will cover most of the search topics requested. Item (1), the formulation of the search problem in terms of concepts and keywords, largely depends on the user's understanding of the search topic and ability to decompose it into meaningful concepts. The role of the searcher here is to lead the user into doing such decomposition appropriately.

Items (2) and (5) vary among human searchers. However, many experts tend to prefer controlled
vocabularies over free vocabulary terms\textsuperscript{7} in their searching. And one frequently used search strategy is: first, perform a \textit{union} (i.e. Boolean "OR") of all the search terms for a given concept, and then an \textit{intersection} (i.e. Boolean "AND") of the unions of all concepts of the search statement. Another strategy is the converse of the former: first, do the intersection of terms from different concepts, and then "or" these results with a different set of terms of the concepts. In other words, intersections are performed before the union of terms. This latter strategy, of course, requires, in general, more searches to get all possible term combinations. However, it may lead to some results more quickly since not all synonymous terms will be used if the searcher is satisfied with the result produced by some subset of the combinations of these terms.

In summary, the above types of expertise are necessary for any computer-simulated expert searcher which can model a human expert on the level of functional behavior. It is likely that a human expert searcher could make use of his/her understanding of the professional field(s) in which the search topic resides to assist search strategy formulation and the like, but the author believes that such information could be obtained from the user provided that the computer-based intermediary is able to ask the relevant questions.

\textsuperscript{7} "Free vocabularies" are non-controlled terms, including terms found in the title or abstracts of a document. In general, a term is "free" for a particular database if it is not listed in the printed thesaurus corresponding to that database.
2.4 Related Works

The investigation of computer-based expert to assist document retrieval has been conducted in conjunction with the CONIT project (Connector for Networked Information Transfer) [Marcus & Reintjes 1979]. The CONIT system enables inexperienced users to access heterogeneous retrieval system by translating search requests stated by the user in a simple command language into a sequence of search commands appropriate for the system in which the search is to be performed. It also provides instructions for the user to assist him/her in learning to use the command language as well as constructing search strategies. Experimental results have demonstrated that CONIT can help inexperienced users retrieve relevant documents from multiple heterogeneous databases. However, the retrieval effectiveness of users of the CONIT system, in terms of number of relevant documents retrieved per unit time, is usually not nearly as high as that achievable by a human expert searcher working with an end user. There are two main reasons for the lower effectiveness: (1) although the CONIT command language is simple, it still represents an impediment, especially for casual (i.e. infrequent) users; and, more importantly, (2) the instructions provided by CONIT to aid search strategy formulation do not yet bring inexperienced users to an expert level in the relatively short period of time of a computer session.

Another closely related project is Oddy’s scheme of reference retrieval through man-machine dialogue [Oddy 1977]. Under his scheme, documents are organized in a special database which is a richly connected network of associations between documents, authors and subject names any pair of which can be linked. The retrieval process begins when a user indicates his area of interest to Oddy’s program, say, by a simple subject keyword. The program then responds by a document which bears the closest *textual resemblance* to the user’s term. The degree of resemblance is measured by the involvement ratio which is defined to be the ratio of the number of associated items in the document to the total number of
associated items in the image (i.e. representation) of the user's search request. The user is then asked to perform relevance judgment on this document as well as the associated subject names or authors that belong to the document. Using this information, the program can update its image of the user's request and hence can respond with new documents. This process iterates until the user is satisfied with the documents displayed by the program. Oddy has reported that his prototype program achieved a performance equal to that obtainable from a more conventional online retrieval system even without obliging the user to formulate his own query in the traditional sense.

The major disadvantage of Oddy's program is its requirement for a specially organized database which may become extremely complicated as the number of documents in the collection increases beyond a certain point. Hence, it may not be useful for large-scale databases of the kind now available in operational systems. Another drawback of the program is that: although it keeps an image of the user's search request, it does not make the image explicit to the user who is ignorant of how his requests or responses to the program are processed or made use of. Therefore, the user's judgments may not provide as focused a feedback mechanism for search reformulation as possible.

Another related work is Doszkocs's CITE system [Doszkocs 1979]. The CITE system is a user interface for the MEDLINE system. An online user is allowed to enter search queries in English sentences, say, a paragraph. The English sentence query is then processed by the same indexing algorithm used for generating MEDLINE inverted files. Valid search terms are identified after elimination of common stopwords. 8 CITE then essentially performs a logical equivalent of combinatorial search, i.e., the x out of n search, for those valid search terms. Documents are retrieved and they are ranked according to the

---

8. Stopwords are essentially non-content-bearing words such as articles, pronouns, prepositions etc.
frequency of occurrence of search terms in these documents. The ranked documents are then displayed for user to judge their relevance. Mesh headings of those relevant documents are included into the list of valid search terms, thus modifying the original query. Another iteration of combinatorial search is performed until the user is satisfied by the results.

One disadvantage of the CITE system is that it creates an illusion for the user who might think that CITE is a natural language understanding system which really understands his request. This illusion may inhibit users from appreciating the need to modify their search queries to achieve better results. Another defect is: although the combinatorial search and frequency-of-occurrence ranking may be useful in some situations, it is by no means universally applicable in all situations. There are many queries that seem to require specific combinations of search terms to get effective results simply and effectively. In these cases, knowledge of the actual logical combinations of terms used for searching may be of great importance to the user in suggesting search reformulations.

Finally, there are many document retrieval systems or intermediary systems that have been implemented in the past ten years or so for the purpose of increasing search effectiveness and ease of use. Some examples are the SMART [Salton 1968, 1971], BROWSER [Williams 1969], SIRE [McGill et al. 1976], IIDA [Meadow 1977] and the FIRST [Dattola 1979] system. Typical features they provide include: simplified commands, computer-assisted instructions, combinatorial searching, ranked output, and relevance feedback. Yet, all of them suffer from one or more of the disadvantages described above. More importantly, none of the systems mentioned above did ever try to simulate in any comprehensive way human expert performance in terms of strategy formation techniques and explanation capability. It is the latter approach which, I hope to show, will provide the kind of computer assistance superior to that available in the systems cited above.
(This page intentionally left blank)
3. Design and Implementation of the Expert System

3.1 Design Goals

The primary purpose of the thesis is to investigate the concept of an intelligent intermediary system which can model the behavior of a human search specialist (on the level of functional tasks as discussed in section 2.2) and thus enables inexperienced users to perform bibliographic searching efficiently. There are three design goals that guide the development of such a computer-based expert searcher:

1. *Ease of Use.* The user of the expert system should not be required to learn any command language. He/she only needs to say what he wants to search on, and should not be bothered by the details of the search conventions, system protocols, database characteristics etc.

2. *Efficiency.* This can be measured in terms of amount of time (hence, cost) required for the user to obtain the desired documents. The usual recall and precision analysis\(^9\) is, of course, a good measure too.

3. *Maintainability.* The resulting system should be easy to maintain. This demands readability and writability of the programs implementing such a system. Although maintainability is hard to measure, its importance is obvious for any successful software system of non-trivial size.

The above goals are quite general. It is certainly possible that many implementation schemes can satisfy these goals (perhaps, to different degrees). Consequently, I do not intend to argue that the particular approach adopted in this thesis, namely, the *knowledge-based* approach, is the best method to construct computer-based expert systems. There is, however, ample experiences gained from the work of *knowledge engineering* and its parent science -- Artificial Intelligence (AI) in the past ten years [Feigenbaum 1977] to indicate its potential utility for this purpose. Some of the recurring themes that emerge from these works such as *generation-and-test*, *rule-based representation*, and *explicit representation of a rich body of information*.

---

\(^9\) Recall is defined as the ratio of the number of relevant documents retrieved over the total number of relevant documents available. Precision is defined as the ratio of the number of relevant documents retrieved over the total number of documents retrieved.
domain-specific knowledge strongly influence the present work as will be seen in the next few sections.

3.2 Knowledge-Based Approach

One effective method to construct a computer-based expert system is the knowledge-based approach familiar in the AI literature. Many expert consultation systems, e.g. MYCIN [Shortliffe 1976], INTERNIST [Pople 1977], DENDRAL [Buchanan & Feigenbaum 1978] etc., have been constructed along this line. The knowledge-based expert systems differs from the early successful AI consultation systems such as MACSYMA [Martin & Fateman 1971] or BRANDAIDE [Little 1972] in that they often embody a rich body of domain-specific knowledge which is represented explicitly in a large sets of rules. Explicit representation has two advantages: (a) it enables the system to reason about its own knowledge, thus allowing efficient retrieval and invocation of knowledge; and (b) it enables the system to explain its behavior. The author is strongly influenced by many ideas that have been used in constructing such knowledge-based systems. From the AI perspective, this thesis can be viewed as an application of these ideas to the field of Information Retrieval which is still largely unaffected by developments in AI (see [Smith 1976]).

According to the knowledge-based methodology, the design of an expert system can be decomposed into three major tasks:

1. Formalization of the knowledge employed by an expert searcher in conducting a search process;

2. Finding a suitable representation scheme to encode and organize such knowledge; and

3. Providing means to control the use of the encoded knowledge.

In the rest of this chapter, I shall discuss how each of the above tasks is handled.
3.3 Formalization of Expert Knowledge

In section 2.3, I have identified different types of expertise that are employed by a human expert searcher in search process. In this section, I shall discuss how the expertise can be formalized.

3.3.1 Conceptualization of Search Problem

The most important prerequisite to a successful search is that the searcher understands the search problem of the user. In order to do this, the searcher requires the user to specify his/her problem precisely. This amounts to identifying the major concepts of the search problem, and the logical relationships between these concepts. For example, in a search for documents about applications of Artificial Intelligence techniques in medical diagnosis, the important concepts could be (1) "Artificial Intelligence", and (2) "Medical Diagnosis". A potentially good document is one that contains both of these concepts.

Since documents are indexed by both free and controlled vocabularies which vary across databases, it is important for the searcher to identify as many synonyms, jargon, and descriptive terms that can express a desired concept as possible. Expert searchers are trained to identify such terms (perhaps, with the aid of a thesaurus). However, in many situations, the searcher has to rely on the expertise of the user to suggest and evaluate various technical terms as search terms for the concepts. Using the above example, the searcher may not know that "problem solving", "reasoning", "heuristics" etc. are terms which are frequently used in AI, and thus are useful as search terms for the concept "Artificial Intelligence".

---

10 "Medical Diagnosis" could also be considered as two concepts, namely, "medicine" and "diagnosis". It is not clear to me what is an appropriate definition for "concept". In this thesis, I take "major concepts" to mean words or phrases or both that serve to identify specific knowledge being sought within the total domain of knowledge.
The conceptualization task is difficult for a computer-based expert because it cannot truly "understand" the search problem (given the present state of natural understanding capability of machines). What it can do instead is to guide the user into doing such decomposition of the search problem and rely on the user's conceptualization of the problem in terms of boolean combination of concepts. There are two specific aids that the computer-based expert can offer in this process:

1. Show inexperienced users how to decompose search problems into meaningful concepts by examples. It is advisable to explain to the user how these concepts will be used by the system,

2. Provide a representation language in which the user can express the relationships among the concepts of the search problem and their respective descriptive terms; this will reduce the "mental" load of the user and allow the computer-based expert to reason about the structure of the problem.

In the current implementation of the expert system, the representation language is extremely simple; it has only two meta-symbols -- and and or. In particular, it allows concepts to be related only by the boolean "AND" so that only documents that contain all of these concepts will be retrieved. The reason for this simple representation is that it can be easily understood by most inexperienced users; thus, much of the ambiguity of the "and" and "or" in our ordinary language in avoided. Moreover, the boolean "and" of all concepts does capture the semantics of a large class of search problems. Whether the restrictiveness of the representation language presents an aid or obstacle to users can be shown by performing experiments of diverse search topics with the expert system, and analyzing the results. The expert system can be easily extended to allow any boolean combinations of concepts if such a feature proves to be desirable.
3.3.2 Formulation of Search Strategy

As I have noted in section 2.3, human expert searchers vary in their method of search strategy formulation. Experiments have shown that the use of "free term" searching with word-stem truncation\textsuperscript{11} increases recall and can be an effective initial strategy [Marcus & Reintjes 1977,1979]. Controlled vocabulary searching, however, is effective in increasing precision of the search. In the design of the computer-based expert, I try to capture the advantages of both types of searching by performing truncated searches for the word-stems of search terms initially supplied by the user and of terms later selected by the user from the titles or abstracts of the documents retrieved during the search, whereas exact searches are performed for the index terms selected by the user from the documents.

As regards the boolean combination of the search terms, I choose to perform the union of all the search terms for each of the concepts and then the intersection of these unions to yield the final result. This method, which I think is simple for the user to understand, avoids the difficulty in keeping track of all combinations of search terms as required if one performs intersection of terms from different concepts before the unions. There is a practical disadvantage for this method, however. Since the retrieval systems usually limit the storage space allocatable to a given user, performing the union of search terms first can often create storage overflow whereas the other method does not. One solution to this problem is to encode both of these translation mechanisms into the computer-based expert explicitly so that it can select the appropriate translation based on the search results.

\textsuperscript{11} The truncated mode involves searching on word-stems of full words or phrases.
3.3.3 Database Selection

The problem of automating the selection of appropriate databases to search for a given search topic has been addressed by a number of researchers in the field of Information Retrieval, e.g., [Williams & Preece 1975], [SDC 1978], [Marcus & Reintjes 1979], [Deane 1980]. However, none of these projects has tried to simulate the way in which a human expert searcher makes this decision. Based on my interview with information specialists, I conclude they use simple decision rules of the form (perhaps, to a first order approximation):

\[
\text{IF} \ \text{<topic belongs to subject area A>}
\]

\[
\text{THEN} \ \text{<conclude database_1 with certainty factor f_1>}
\]

\[
\ldots
\]

\[
\text{<conclude database_n with certainty factor f_n>}
\]

To capture this "human" style of decision making, I use a much simplified version of the MYCIN production rules to encode this type of knowledge for a few dozen databases supported by the SDC ORBIT and SUNY/NLM MEDLINE systems (see Appendix-I for a complete listing of these rules). The database selection scheme works as follows:

1. The expert system asks the user to select one of the eight broad subject areas to which his/her search topic belongs. Subject areas correspond to major professional disciplines such as Medicine, Engineering, Social Sciences etc.

2. Once the subject area is known, the user is asked to select one or more fields from the subject area that are most closely related to the search problem. For example, if the user selects Medicine as the subject area, various fields such as Drug Uses, Health Planning, Toxicology, Cancer, General Medicine etc. will be shown to the user for further selection.

---

12. One reason may be that people do not believe that human "expert" decision on this matter is optimal. But, the human decision making process on database selection is simple, fast and economical.
3. Based on the field(s) that have been selected by the user, decision rules of the above IF...THEN... type will be run to obtain the certainty factors of various databases that are relevant to the fields. The certainty factors (CF) recommended by different rules are combined by a simple averaging method. To be more precise, if there are n decision rules that matched the selected fields, and the total CF for database_k is CF-total, then the net CF of database_k is CF-total divided by n. There is one exception to this averaging method, namely, when a database has been assigned a CF of 1.0 (i.e. 100% certainty) by a single decision rule, its net CF will be 1.0 independent of value of CF-total or n. This exception is reasonable since it corresponds to the intuitive meaning of "100% certainty".

The current database selection rules do not have negative recommendation, namely, concluding that certain databases will not be useful for a given field. However, the current scheme can be easily extended to handle this type of inference. Moreover, the current rules only make use of the field(s) related to the search topic. Other sources of information such as types of documents desired (e.g. reports, books, dissertations), names of relevant journals (e.g. New England Journal of Medicine, Yale Journal of Medicine), nature of the information requested (e.g. statistics, federal funding of projects) can all be used in refining the decision process.

As it stands now, the simplicity, naturalness and economy of the database selection rules make the current selection scheme a serious rival to the various database selectors mentioned in the beginning of this subsection.

3.3.4 Translation of Search Strategy into Executable Search Commands

The differences in system protocols and syntax of search commands among retrieval systems present a serious difficulty to inexperienced users who want to access multiple databases belonging to different systems. Even within a single retrieval system, there is reason to believe that the user should not be required to learn a particular command language before he/she can perform a search. This consideration is especially important for the class of users who are only interested in getting the helpful documents fast,
and have no intention (or patience) to learn any command language. In fact, this is one of the reasons that some users prefer to perform searching via a human intermediary.

There is a second reason that the requirement to learn a command language in order to use a retrieval system is undesirable, namely, for an inexperienced user to perform the search directly, he/she has to worry about the formulation of search strategy and the translation of the strategy simultaneously. This not only increases the length of time spent online, but often results in a poor search because of either ineffective search strategy or poor translation or both. The problem is in fact a common source of difficulty for any operator-based systems, e.g., MACSYMA.

In this thesis, I attempt to transfer the entire translation "load" from the user to the computer-based expert which embodies sufficient knowledge to translate any given search strategy into a sequence of executable search commands appropriate for particular retrieval systems. In the current implementation, the expert can "talk" to two systems -- SDC ORBIT and NLM/SUNY MEDLINE. The expert can be easily extended to handle Lockheed DIALOG. The translation mechanism is also explicitly represented in the conventional situation-action production rules. This will be discussed in more detail in section 3.4.

3.3.5 Evaluation & Reformulation of Search Strategy

Successive refinements to the initial search strategy is almost always necessary in order to achieve good search results. Evaluation of a previous search strategy is often done by the user whereas the searcher performs the reformulation based on the result of the evaluation. This is an important stage of the search process during which the user can contribute substantially to the success of the search by identifying (a) relevant documents from the retrieved set; and (b) good search terms from the relevant documents to augment the original search strategy. Secondly, the user's evaluation provides a measure (or better,
estimation) for the recall and precision of the previous search. Low recall indicates broadening of the search strategy is required while low precision indicates narrowing is desired.

Reformulation of search strategy thus consists of two parts:

1. Modify the Concept Table, which represents the original search problem, by either adding additional search terms under each concept or deleting search terms that are too general to convey the meaning of a particular concept.

2. Modify the search requirement by adding new concept(s) to the Concept Table or deleting old concepts which are too specific.

The computer-based expert can make different suggestions to the user -- either to broaden or narrow the search -- and ways to achieve them based on the estimated recall and precision obtained from the user's evaluation. Finally, it should be emphasized that the role of the computer-based expert during this evaluation and reformulation is not to make the decision of either narrowing or broadening of the search but rather to suggest what refinement is likely to be useful and how the refinement can be carried out.

3.3.6 Summary of Formulation of Expert Knowledge

In this section, I have suggested various types of expertise that are employed by the human expert searcher, and ways to encode them in some sort of rule-based representations. I shall now discuss more fully the syntax and semantics of the production rules which constitute the knowledge-base of the expert system.
3.4 Production Rules as a Representation for Expert Knowledge

The effectiveness of production rules as a representation for expert knowledge has been discussed by a number of researchers in AI e.g. [Davis & King 1975], [Shortliffe & Buchanan 1977], and [Feigenbaum 1977]. The advantages of a rule-based representation can be summarized in the following terms: modularity, naturalness and extendability. Each rule usually encodes a modular chunk of domain-specific knowledge that is comprehensible to a human expert in that domain. That makes the task of explanation of the behavior of the system simpler. It also enables the program to reason about its own knowledge which is explicitly represented [Davis 1980]. Interactions between any single rule and the rest of the knowledge are minimized, thus increasing the extendability and maintainability of the system.\(^{13}\) In the rest of this section, I shall discussion the syntax and interpretation of the production rules used in the current system, and show examples how they can be used to encode expert knowledge.

3.4.1 Syntax of Production Rules

The production rules used in the current system is a slight extension to the conventional \(\langle\text{situation}\rangle \rightarrow \langle\text{action}\rangle\) pair. Each rules is of the form

\[
\text{(defrule } \langle\text{rule-name}\rangle
\begin{align*}
&\langle\text{premise}_1\rangle \\
&\ldots \\
&\langle\text{premise}_n\rangle \\
&\rightarrow \\
&\langle\text{then}_1\rangle \\
&\ldots \\
&\langle\text{then}_n\rangle \\
\text{ELSE } &\langle\text{else}_1\rangle \\
&\ldots
\end{align*}
\]

---

13. The terms "extendability", "maintainability" and "modularity" are hard to define precisely, and difficult to measure. They roughly mean how easy it is to change or extend the system's software.
<else_n>)
where the ELSE parts are optional. <Premise> should have the form ⟨function⟩ {<parm>} {<val>}. {...} indicates that this argument is optional. (or ⟨premise⟩ ⟨premise⟩) is also allowed. The ⟨premise⟩ must be evaluated to either true or false. ⟨then⟩ should have the form ⟨action⟩ {<arg1>} {<arg2>}. ⟨else⟩ should be like ⟨then⟩. The "⇒" arrow indicates that "then" clauses follow. "ELSE" indicates the "else" clauses follow.

Intuitively, the rule means the following: if the premises are evaluated to be true, then the actions specified in the "then" clauses are executed. If the premises are evaluated to be false, then either the "else" clauses are executed if they are present, or the rule is simply ignored if the "else" clauses are absent. The ⟨else⟩ part of the rule is just added for programming convenience.

Most of the premises consist of testing where certain parameters, i.e., ⟨parm⟩, have certain values, i.e., ⟨val⟩. The more important actions specified in the "then" or the "else" clauses include the following: sending a message to the user, requesting user the value of certain parameters, sending commands to the retrieval systems, and updating the global database\textsuperscript{14} or retrieving data from it. Some examples will be discussed in the next subsection.

\textsuperscript{14} The global database not only stores all the production rules which constitute the knowledge base of the system, but also stores information about the search process and values of various parameters.
3.4.2 Examples of Production Rules

The following is an example of production rule, with its English translation, used to encode knowledge in database selection.

**RULE d500**

**IF:** The field of the search topic is Cancer research and therapy  
**THEN:**  
1) It is definite (1.0) that CANCERLIT is the appropriate database, and  
2) There is strongly suggestive (0.8) evidence that CANCERPROJ is appropriate, and  
3) There is suggestive (0.7) evidence that MEDLINE is appropriate  
4) Activate RULE d1000 for next consideration.

(defrule d500  
  (same field Cancer_research_and_therapy)  
  =>  
  (advice_to CANCERLIT 1.0)  
  (advice_to CANCERPROJ 0.8)  
  (advice_to MEDLINE 0.7)  
  (activate d1000))

Notice that the above rule strongly resembles the MYCIN rules. The only difference is that it also encodes information to control invocation of next rules. The uniformity of encoding domain-specific knowledge (e.g. database selection, translation, search strategy planning etc.) and knowledge controlling invocation or refinement of rules simplifies implementation of the interpreter and cut down the amount of machinery required.

The next example is the goal rule used in handling the protocol for connecting retrieval systems. Notice that exactly the same format of production rule is used to encode the translation mechanism.
RULE 1100

IF: The current goal is to login in
THEN:
1) Send user the following message: "CONIT will try..."
2) Activate RULE 1200 for next consideration.

(defrule l100
  (equal goal "login")
  ==> (senduser (catenate "CONIT will try to connect to the retrieval system -- " *system*))
  (activate l200))

The next rule is a more complicated example that involves the use of the "else" clauses. This rule is the first to be activated by the "login" goal-rule above.

RULE 1200:
IF phone connection is OK
THEN
1) Send user the message: "Phone connection..."
2) Sleep for 1 second
3) Send the remote host a carriage_return
4) Sleep for 2 seconds
5) Send host the string, ";;"
6) Get next response line
7) Activate RULE 1300
ELSE
8) Send user the message: "Phone connection to Telenet failed!"
9) Send user the message: "We will try..."
10) Activate RULE 1200

(defrule l200
  phone
  ==> (senduser "Phone connection made to Telenet.")
  (sleep 1)
  (sendhost cr)
  (sleep 2)
  (sendhost ";;"
  (get_response)
  (activate l300)
ELSE (senduser "Phone connection to Telenet failed!")
  (senduser "We will try to connect again.")
(activate l200))

Appendix-II contains the listing of all the rules used in the current implementation.

3.5 Retrieval, Refinement & Execution of Rules

The control of use of knowledge in a rule-based representation becomes the problem of controlling the invocation of rules. Traditionally, rule invocation is viewed as occurring in three steps: retrieval, refinement and execution. In this section, I shall explain how each of this step is accomplished in the current implementation.

The program used a simple forward chaining mechanism in controlling the invocation of rules. Initially, a goal is set up, say, to plan the search strategy. The goal rule corresponding to this specific goal is activated, i.e., retrieved (by name) and put into the agenda for consideration by the interpreter. Since there is only one rule that matches this goal, and hence one rule in the agenda, no refinement (or reordering) of agenda is necessary. The premise of the goal rule is evaluated to true, and the "then" clauses of the rule are executed. The execution will result in further activation of rules for the interpreter to consider. The goal rule itself will be deactivated, i.e., removed from the agenda, after the execution.

In the current implementation, no refinement or reordering of rules in the agenda is necessary. Information about which rules to be activated next is distributed into the rules itself (as shown in the examples in the previous section). The number of rules that are potentially applicable at each cycle of retrieval-execution is very limited (at most five or six). Any activated rule whose premise(s) if evaluated to be true will be executed. Execution of multiple rules is a common situation during database selection.
3.6 Summary and Others

The expert system consists of three major components: the knowledge base made up of production rules of the form IF...THEN...ELSE..., an interpreter of rules, and a global database to store production rules as well as information about the search process. Currently, there are about 150 rules in the system, half of which are database selection rules. The design and implementation of the first version of the expert system have been completed. The system software is written in Maclisp running on the MIT Multics System which has access to various communication networks such as Telenet, Tymnet and Arpanet. The total size of the current system is about 100 pages\textsuperscript{15} of storage.

\textsuperscript{15} Each page consists of 1024 words.
4. Example of User/Computer-Based-Expert Interactions

In this chapter, I shall illustrate the main features of the computer-based Expert System by going through an example session. In this session, the user is looking for documents about the relevance of transformational grammar to language learning. The main point is to show how the various types of expertise discussed in Chapter 3 are built into the Expert System. I shall also discuss some of the limitations of the current version of the Expert System, which is called EXPERT-1.

4.1 An Experimental Session with EXPERT-1

What follows is a transcription of the actual example session. Each portion of the example is given in a separate section which begins with an italicized note explaining the nature of that session portion. User input is shown on lines beginning with "--->

4.1.1 Beginning of Search Session

Note 1: EXPERT-1 asks the user to state his search topic, and explains to him, by examples, how to decompose the search problems into meaningful concepts.

---

run

The first task in searching is to plan your search. There are three phases to planning a search:
1. stating and developing the search query;
2. setting up the search strategies needed; and
3. collecting data relevant to your search, e.g., time period to be covered, how many documents are desired, etc.
Let's proceed...
Please describe your search topic clearly in one sentence or two...
Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.
--->
Relevance of transformational grammar to language learning $

The immediate task is to set up an initial search strategy for your search problem. In order to do this, we have to identify the basic components (concepts) of your query statement. For example, in a search for documents about the toxic effects of polychlorinated biphenyls on liver, the major concepts could be:
(1) polychlorinated biphenyls;
(2) toxic effects; and
(3) liver.
EXPERT-1 will then attempt to find documents that contain ALL of these concepts. As a second example, in a search for documents about the design and implementation of Algol 60, the major concepts could be:
(1) design;
(2) implementation;
(3) Algol 60.
However, if what the searcher really wants are documents concerning either the design of Algol 60 or the implementation of Algol 60, then the above formulation is inappropriate since EXPERT-1 will attempt to find documents that contain ALL of the above concepts, namely, 'design', 'implementation' and 'Algol 60'. One way to formulate this problem is to use two concepts instead of three:
(1) design;
(2) Algol 60
and, later, add the term 'implementation' as a related term (i.e. keyword) to the concept 'design'. With this formulation, EXPERT-1 will find documents that contain the term 'Algol 60' and either 'design' or 'implementation'.
****
Note that a concept may be expressed as a single word or a set of words, but in either case, it should convey significant information relevant to your search problem. Usually, it is desirable to specify several concepts -- possibly two or three-- in order to get started on a meaningful preliminary search strategy.
4.1.2 Listing Concepts of the search problem

Note 2: The user characterizes his problem by giving two concepts -- linguistic theory and language learning.

Let's find out distinct major concepts of your query statement.
Please give a major concept of your search topic:
Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.
--->linguistic theory $
Let's denote <linguistic theory> by CONCEPT-1.
Give another major concept of your search topic (if there is none, answer nil): Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.
--->language learning $
Let's denote <language learning> by CONCEPT-2.
Give another major concept of your search topic (if there is none, answer nil) Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.
--->nil

The following is the list of distinct concepts of your query statement:
CONCEPT-1:linguistic theory
CONCEPT-2:language learning
4.1.3 Finding search terms for each concept

Note 3: EXPERT-1 begins to build the initial search strategy by asking the user to supply search terms for each concept.

After the major concepts are determined, the next step is to find out descriptive terms or phrases for each of these concepts which are to be used in actual searching. NOTE that: these descriptive terms or phrases are the so-called FREE (or uncontrolled) search terms and they will be searched in a TRUNCATED MODE, i.e., searching on their respective root-stems. For example, if 'intelligence' is a free term, EXPERT-1 will search on its root which is 'intellig:' where the colon (:) indicates truncation. This means that every term beginning with 'intellig' will be considered as a match. Truncated search for a phrase will be searched as an intersection of its component words. For example, 'artificial intelligence' will be searched as 'artifi: AND intellig:'. As a consequence, the TRUNCATED search is a kind of broad searching as opposed to the EXACT search which will be used later in the search process for searching INDEX (or controlled) terms. Let's now list such keywords (i.e. synonyms, abbreviations, jargon words or descriptive terms/phrases) for CONCEPT-1:linguistic theory Remember we are only building an initial search strategy, there is no need to give an exhaustive list of keywords; a few will be sufficient to start. As a start, the concept name itself --linguistic theory will be taken as a keyword.

Please give another keyword for this concept (if no more keyword, answer nil):
Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.
--- transformational grammar $

Please give another keyword for this concept (if no more keyword, answer nil):
Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.
--- generative grammar $

Please give another keyword for this concept (if no more keyword, answer nil):
Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.
--- nil

The following is a list of keywords for concept-1:
(1) linguistic theory
(2) transformational grammar
(3) generative grammar
4.1.4 Finding search terms (cont.)

*Note 4: User is asked to supply search terms for the second concept.*

----------

Let's now list such keywords (i.e. synonyms, abbreviations, jargon words or descriptive terms/phrases) for CONCEPT-2: language learning. Remember we are only building an initial search strategy, there is no need to give an exhaustive list of keywords; a few will be sufficient to start. As a start, the concept name itself -- language learning will be taken as a keyword.

Please give another keyword for this concept (if no more keyword, answer nil):
' Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.
  --> language acquisition $

Please give another keyword for this concept (if no more keyword, answer nil):
' Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.
  --> nil

The following is a list of keywords for concept-2:

1) language learning
2) language acquisition
4.1.5 Review Initial Search Strategy

Note 5: EXPERT-1 displays the initial strategy to the user who can modify it if desired.

----------

Having found out the major concepts and their respective keywords (i.e., descriptive terms/phrases), we are going to review the initial representation for them. This table (or representation) of the concepts of the query statement with their respective keywords is crucial to the success of searching. Depending on partial search results, good keywords (identified from relevant documents) may be added to the table whereas bad search terms (those that yield irrelevant documents) are deleted. Even the major concepts themselves may be added or deleted. Thus, as the search proceeds and partial results are obtained, this table of concepts and keywords is continuously refined until a sufficient number of good documents are retrieved.

****

As a reminder, the following table is only the initial representation of concepts and keywords. Addition or deletion of keywords and/or concepts are recommended only after initial searching is done.

****

<table>
<thead>
<tr>
<th>CONCEPT-1</th>
<th>AND</th>
<th>CONCEPT-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>/linguistic theory</td>
<td>/language learning</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>OR</td>
<td></td>
</tr>
<tr>
<td>transformational grammar</td>
<td>language acquisition</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>OR</td>
<td></td>
</tr>
<tr>
<td>generative grammar</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Does the above table correctly represent the major concepts and their respective search terms that you have supplied?

Please answer yes or no:

---y

How many relevant documents do you wish to find?
(If you want a comprehensive search, i.e., one that retrieves as many relevant documents as there are in a certain database, specify any number greater than 100.)

---30
4.1.6 Database Selection

Note 6: EXPERT-1 assists the user in selecting appropriate databases.

--------

The current task is to select appropriate databases for your search topic.

Select ONE of the subject area numbers in which of the following subject areas the information of your search topic is likely to be found:
(1) government activities, news and sports
(2) sciences
(3) applied sciences and industries
(4) engineering
(5) medical and health sciences
(6) energy and environment
(7) business, economics and law
(8) social sciences, education and humanities

--->

Select one or more of the following field numbers which are most closely related to your search topic. Multiple values are allowed, but terminate your answer with a space followed by the dollar sign symbol ($) and a carriage return.
(1) Education_for_exceptional_children
(2) Education_media_and_materials
(3) Education_issues_professions_and_research
(4) Modern_art_and_design
(5) Philosophy
(6) Psychology_and_behavior_sciences
(7) Public_affairs_administration
(8) US_and_World_Histories
(9) US_Political_sciences

---> 3 6 $

The following is a list of suggested databases together with their certainty factor. Please select ONE of these databases to start searching.
(1)(ERIC 1.0)
(2)(PSYCHOLOGY_ABSTRACTS 1.0)
(3)(AIM_ARM 0.4)
(4)(SOCIAL_SCISEARCH 0.4)

--->1
4.1.7 Logging into ERIC on SDC/ORBIT

Note 7: EXPERT-1 automatically connects to the retrieval system that contains the ERIC database selected by the user. All the network and login protocols are handled by the system.

----------

EXPERT-1 will try to connect to the retrieval system -- ORBIT
Telephone 1 will be used.
09/24/80  1835.2 cdt Wed
Phone connection made to Telenet.
Telenet Responding.
617 8C
213 33B CONNECTED
Logging onto ORBIT.
YOU ARE ON LINE L22

HELLO FROM SDC/ORBIT IV. (09/24/80 3:34 P.M. PACIFIC TIME)

PROG:
****
CHEMDEX2 NOW AVAILABLE! SEE NEWS.
YOU ARE NOW CONNECTED TO THE ORBIT DATABASE.

PROG:
ELAPSED TIME ON ORBIT: 0.01 HRS.
YOU ARE NOW CONNECTED TO THE ERIC DATABASE.
4.1.8 Actual Searching

SS 1 /C?

Note 8: EXPERT-1 translates the search strategy into a sequence of search commands. 1,650 documents are retrieved.

We have already connected to the desired database and are ready to start searching. EXPERT-1 will automatically translate the concepts and keywords in the above table into a sequence of search commands. The following will be the actual execution of these search commands and the response from the retrieval system. You can just wait and see the results.

FD ALL lingu: AND ALL theor:

Search 1 found 3583 documents.
SS 2 /C?
FD ALL transform: AND ALL grammar:

Search 2 found 1233 documents.
SS 3 /C?
FD ALL gener: AND ALL grammar:

Search 3 found 1839 documents.
SS 4 /C?
1 OR 2 OR 3

Search 4 found 4779 documents.
SS 5 /C?
FD ALL lang: AND ALL learn:

Search 5 found 17180 documents.
SS 6 /C?
FD ALL langu: AND ALL acquisit:

Search 6 found 1736 documents.
SS 7 /C?
5 OR 6

Search 7 found 17889 documents.
SS 8 /C?
4 AND 7

Search 8 found 1650 documents.
4.1.9 Suggestions to how to proceed

Note 9: EXPERT-1 suggests to the user that narrowing of the search is needed to reduce the number of documents retrieved.

The current search probably found too many documents than what you want. EXPERT-1 suggests you to NARROW your search in the next iteration of searching by either

(1) dropping out search terms that are either too general or do not convey
    the meaning of the concepts to which they belong from the Concept Table; or
(2) tightening the search requirement by adding a new concept to the Table.

But before you do that, it is a good idea to look at some of the documents found just then. This will give you an idea of which documents are relevant, and hence which search terms to keep or to delete.

SS 9 /C?

The current task is to examine some of the documents found in the previous iteration of searching. EXPERT-1 will first show the TITLES and AUTHORS of ten documents so that you can select from them some of the more promising documents, i.e., those documents that are likely to be relevant to your search statement. EXPERT-1 will then show more details on each of the selected documents for your further examination.
4.1.10 Sample Documents by Title & Authors

Note 10: EXPERT-I shows the titles and authors of the first ten documents for the user to consider.

---------

PROG:

-1-
TITLE: Mental Representation and Early Language Development: Directions for Exploring Relationships. Souvenir of Conversation Hour
AUTHORS: colich, Lorraine McCune; And Others

-2-
TITLE: Linguistics and Language Teaching
AUTHORS: Lehmann, Winfred P.

-3-
TITLE: TESL Reporter, Vol. 1, Nos. 1-4
AUTHORS: Pack, Alice C., Ed.

-4-
TITLE: TESL Reporter, Vol. 2, Nos. 1-4
AUTHORS: Pack, Alice C., Ed.

-5-
TITLE: Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett

-6-
AUTHORS: Harley, Birgit

-7-
TITLE: Outlines of Suggestopedia Applied to Teaching of Foreign Languages
AUTHORS: Racle, Gabriel L.

-8-
TITLE: Social and Psychological Factors in Second Language
Acquisition: A Study of an Individual. Proceedings of
the Los Angeles Second Language Research Forum
Jones, Rebecca A.

Allgemeine Sprachfähigkeit und Fremdsprachenerwerb.
Zur Struktur von Leistungsdimensionen und
linguistischer Kompetenz des Fremdsprachenlerners
(General Language Ability and Foreign Language
Acquisition. On the Structure of Performance
Dimensions and the Linguistic Competence of the
Foreign Language Learner). Diskussionsbeiträge aus
dem Institute für Bildungsforschung, No. 1
Sang, Fritz; Vollmer, Helmut J.

An Historical Overview of Second Language Acquisition
Research
Hatch, Evelyn
4.1.11 Relevance Judgement of Selected Documents

Note 11: The user selects documents which seem to be relevant and EXPERT-1 shows more information from these documents.

-------

SS 9 /C?

Select one or more of the above documents (by number) that seem likely to be relevant to your search topic. EXPERT-1 will show more detail on them for further examination. Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.

---$15$---

EXPERT-1 will now show more information on the documents that you have judged to be likely relevant. For each of these documents, EXPERT-1 will display its TITLE, AUTHORS, ABSTRACT, SOURCE OF PUBLICATION and a list of controlled terms used to index this document. Your task is to judge the relevance of the documents and to select additional good terms from them for next iteration of searching.

PROG:

-1-

TITLE
Mental Representation and Early Language Development: Directions for Exploring Relationships. Souvenir of Conversation Hour

AUTHORS
Nicolich, Lorraine McCune; And Others

ABSTRACT
This collection of conference abstracts focuses on new directions for research on mental representation and early language development. One page summaries are provided on the following topics: Mental Representation and Initial Language Learning, by Lorraine M. Nicolich; Critical Issues in Language and Cognitive Development, by Roberta Corrigan; Relationships Between Early Language and Other Symbolic Skills in Young Down’s Syndrome Children, by Patricia M. Hill; Communicative and Sensorimotor Development in Down’s Syndrome Children, by Carol Greenwald and Laurence B. Leonard; Sensorimotor Development and Pre-Linguistic Communication, by K. F. Steckol and L. B. Leonard; Contextual Support: A Perspective on the Development of Language and Action, by Lorraine Rocissano; and a statement on measures of representation by Kurt W. Fischer. (SS)

SOURCE
Mar 1979; 11pp
4.1.12 Feedback by Selecting Additional Search Terms

Note 12: EXPERT-1 shows a list of terms that have used to index the selected document. These terms are potentially useful for augmenting the search strategy.

---------

Select ONE or more of the following controlled terms (by number) which, you think, are good search terms that could be added to your list of keywords in your Concept Table:
Multiple values are allowed, but terminate your answer with a space followed by the dollar sign symbol ($) and a carriage return.
(1) Cognitive Development
(2) Cognitive Processes
(3) Communication Skills
(4) Developmental Stages
(5) Down's Syndrome
(6) Language Acquisition
(7) Linguistic Theory
(8) Perceptual Motor Coordination
(9) Symbolic Learning
(10) Young Children

--> 2 $

Please give additional search term, which you think is relevant to your search topic, from either the TITLE or the ABSTRACT of the document just shown (if there are none, answer nil):
Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.

--> nil
4.1.13 Relevance Judgement (cont.)

Note 13: Repeat feedback procedure for the second document that the user has selected.

EXPERT-1 will show more information on the next document that seems to be likely relevant to your search topic for your further judgment of relevance and selection of additional good search terms.

PROG:

-5-

TITLE Sentence Processing: Psycholinguistic Studies
Presented to Merrill Garrett


ABSTRACT The subsections in this volume represent a type of current psycholinguistic research that focuses both on the nature of human information processing and the coding of linguistic structure. The subsections and authors are as follows: (1) "The Wherefores and Therefores of the Competence-Performance Distinction," by V. Valian; (2) "Levels of Processing and the Structure of the Language Processor," by K.I. Forster; (3) "Time-Compressed Speech and the Study of Lexical and Syntactic Processing," by M. Chodorow; (4) "Monitoring Sentence Comprehension," by a Cutler and D. Norris; (5) "Intonation and Ambiguity," by R. Wales and H. Toner; (6) "Perceptual Mechanisms and Formal Properties of Main and Subordinate Clauses," by T. Bever and D. Townsend; (7) "Some Hypotheses About Syntactic Processing in Sentence Comprehension," by V. M. Holmes; (8) "Superstrategy," by J. Fodor; (9) "Role of Efference Monitoring in the Detection of Self-Produced Speech Errors," by J. Lackner and B. Tuller; (10) "Speech Errors as Evidence for a Serial-Ordering Mechanism in Sentence Production," by S. Shattuck-Hufnagel; (11) "'Like' Syntax," by J. Ross and W. Cooper; and (12) "Three Cheers for Propositional Attitudes (Some Reflections on D.C. Dennett's 'Intentional Systems')," by J.A. Fodor. (AMH)

SOURCE (1979); 447pp
4.1.14 Selecting More Search Terms

Note 14: More search terms are to be selected from the index term list.

Select ONE or more of the following controlled terms (by number) which, you think, are good search terms that could be added to your list of keywords in your Concept Table:
Multiple values are allowed, but terminate your answer with a space followed by the dollar sign symbol ($) and a carriage return.
(1) Deep Structure
(2) Intonation
(3) Language Acquisition
(4) Language Processing
(5) Language Research
(6) Linguistic Competence
(7) Linguistic Performance
(8) Linguistic Theory
(9) Listening Comprehension
(10) Psycholinguistics
(11) Semantics
(12) Sentences
(13) Sentence Structure
(14) Speech Communication
(15) Surface Structure
(16) Syntax
(17) Vocabulary
---> 10 $

Please give an additional search term, which you think is relevant to your search topic, from either the TITLE or the ABSTRACT of the document just shown (if there are none, answer nil):
Terminate your input with a space followed by the dollar sign symbol ($) and a carriage return.
---> nil

Do you want to see more documents from this search set before beginning the next round of iteration of searching?

Please answer yes or no:
---> n

If you think the search set obtained in this iteration of searching is very relevant to your search topic, answer YES and EXPERT-1 will request offline printout of this search set.

Please answer yes or no:
---> no
4.1.15 Updating the Concept Table

Note 15: The search strategy is updated to include the new terms that have been selected.

For each of the following controlled (index) terms EXPERT-1 will ask you the concept(s) to which the term belongs so that they may be added to the Concept Table for next iteration of searching. Note that: ALL the concepts to which a term is related should be mentioned.

TERM-1: Cognitive Development
Select ONE or more of the following concepts (by number) to which the new term belongs:
Multiple values are allowed, but terminate your answer with a space followed by the dollar sign symbol ($) and a carriage return.
(1)CONCEPT-1
(2)CONCEPT-2
---$ 

TERM-2: Psycholinguistics
Select ONE or more of the following concepts (by number) to which the new term belongs:
Multiple values are allowed, but terminate your answer with a space followed by the dollar sign symbol ($) and a carriage return.
(1)CONCEPT-1
(2)CONCEPT-2
---1 $
### 4.1.16 Update Concept Table (cont.)

*Note 16: The user can further change the Concept Table.*

The current task is to examine the updated Concept Table which will include all the INDEX terms or FREE terms, if there is any, that you have supplied to EXPERT-1. The following are few suggestions to what you may want to do with the Concept Table:

1. **Add additional search terms to the Table.**
   - do it in case you have left out some search terms that you think are relevant; or
   - in case you want to BROADEN your search, i.e., to find more documents.

2. **Delete search terms.**
   - do it in case some of the previous search terms are irrelevant or too general for your search purpose; or
   - in case you want to NARROW your search, i.e., to find fewer documents.

3. **Replace search terms in the Table.**
   - do it in case you have mis-spelled some of the search terms; replace is also a quick way to do delete followed by addition of a search term within the same concept.

<table>
<thead>
<tr>
<th>CONCEPT-1</th>
<th>AND</th>
<th>CONCEPT-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>/linguistic theory</td>
<td>/language learning</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>OR</td>
<td></td>
</tr>
<tr>
<td>transformational grammar</td>
<td>language acquisition</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>OR</td>
<td></td>
</tr>
<tr>
<td>generative grammar</td>
<td>Cognitive Development</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>psycholinguistics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Does the above table correctly represent the r concepts and their respective search terms that you have supplied?

Please answer yes or no:

--->

y
4.1.17 Next Iteration of Searching

Note 17: EXPERT-1 performs the next iteration of searching based on the updated Concept Table.

------------------

EXPERT-1 will translate the updated Concept Table into sequence of search commands and execute them. The following will be the actual interactions with the retrieval system. You can just wait and see the results. Psycholinguistics

Search 9 found 2609 documents.
SS 10 /C?
1 OR 2 OR 3 OR 9

Search 10 found 6603 documents.
SS 11 /C?
Cognitive Development

Search 11 found 5131 documents.
SS 12 /C?
5 OR 6 OR 11

Search 12 found 22378 documents.
SS 13 /C?
10 AND 12

Search 13 found 2622 documents.
SS 14 /C?

The current search probably found too many documents than what you want. EXPERT-1 suggests you to NARROW your search in the next iteration of searching by either
(1) dropping out search terms that are either too general or do not convey
the meaning of the concepts to which they belong from the Concept Table; or
(2) tightening the search requirement by adding a new concept to the Table.
But before you do that, it is a good idea to look at some of the documents found just then. This will give you an idea of which documents are relevant, and hence which search terms to keep or to delete.
The current task is to examine some of the documents found in the previous iteration of searching.
EXPERT-1 will first show the TITLES and AUTHORS of ten documents so that you can select from them some of the more promising documents, i.e., those documents that are likely to be relevant to your search statement. EXPERT-1 will then show more details on each of the selected documents for your
4.2 Summary of Built-in Expertise of EXPERT-1

The previous section illustrates the main features of EXPERT-1 to assist an user in document retrieval. The following types of expertise (as discussed in Chapter 3) have been built into the system:

1. Characterize the user problem by composition of major concepts that convey significant information about the problem;

2. Formulate search strategy in terms of intersection of concepts and union of search terms of each concept;

3. Suggest a ranked list of databases that are may be relevant to the search topic;

4. Handle all network protocols and login procedures of the retrieval systems;

5. Translate the search strategy into a sequence of executable search commands that are appropriate for the connected database. The translation mechanism remembers previous search results, thus avoiding to repeat searches that have been done before.

6. Suggest to user how to reformulate search strategy, e.g., narrowing or broadening by addition or deletion of search terms etc.

7. Assist user to select potentially good search terms from the index terms of documents chosen by the user.

4.3 Limitations of EXPERT-1

Although EXPERT-1 has not been evaluated formally, we understand that it has certain limitations that could affect the success of such a system:

---

16. Notice that the user requested 30 relevant documents but we have found 1,650 documents in the first trial and 2,622 in the second. In order for this search to succeed, the user has to follow EXPERT-1's suggestions to narrow the search either by replacing more general search terms in the Concept Table with the more specific index terms selected or by adding a third concept to tighten the retrieval requirement.
1. EXPERT-1 handles only subject search with either truncated or exact mode; it does not handle author search, adjacency search, search with major descriptors, and search with selected fields (such as title, abstract, etc.).

2. EXPERT-1 cannot formulate database-dependent search strategy, e.g., not to use certain terms in certain databases.

3. Although EXPERT-1 suggests what kind of reformulation is desirable (e.g. narrowing) and how the reformulation can be done, it does not actively push the user into doing the reformulation.

4. EXPERT-1, unlike human search expert cannot suggest potential search terms and, more importantly, cannot evaluate how relevant a retrieved document is.

Further experimentation of EXPERT-1 is needed in order to determine to what extent these limitations will affect the retrieval effectiveness of the system.
5. Conclusions

5.1 Justifications of Design Goals of EXPERT-1

In this section, I shall defend the design decisions of EXPERT-1 with respect to the design goals stated in Chapter 3, namely, ease of use, efficiency and maintainability. As we have seen in Chapter 4, EXPERT-1 does include most of the functionality of a human expert searcher. Users of EXPERT-1 do not have to learn any command language. All information needed by EXPERT-1 to perform a search, such as search problem characterization, judgment on retrieved documents etc., is obtained by "menu-type" of question/answering; most of the time, the user is just required to input a number as his answer. EXPERT-1 formulates search strategy by intersection of major concepts, and union of search terms for each concept supplied by the user. It keeps track of the Concept Table which reflects its current "understanding" of the user's search problem. It embodies an algorithm to translate search strategy into a sequence of executable search commands. It can remember previous searches so that redundant searches are avoided. Finally, it can also make suggestions to the user to guide his selection of appropriate databases and reformulation of search strategy. With such an intermediary, even a naive user can get searches done within a relatively short period of time (in the order of 30 to 45 minutes for an entire search session).

Retrieval effectiveness of a search can be estimated by the number of good documents found per unit time.\(^{17}\) Efficiency of an intermediary is then measured by the amount of time (hence cost) it requires to assist the user to retrieve an adequate number of good documents. Since the time required by EXPERT-1

\(^{17}\) In most existing retrieval systems, the cost for a search is directly proportional to the time required for the search.
to translate search strategy into search commands and to execute the resulting commands is much less than that of a human counterpart, the effectiveness of the search strategy becomes the crucial factor in determining the efficiency of EXPERT-1. As we have noted, EXPERT-1 formulates search strategy by intersection of concepts, and union of search terms. Initially, "free" search terms are supplied by the user. Later, after some good documents are retrieved, EXPERT-1 assists the user in finding additional search terms from the list of index terms (i.e. controlled vocabularies) which are used to describe the contents of these documents. EXPERT-1, like many human expert searchers, tends to prefer searching in index terms in order to achieve higher precision. Alternatively users of EXPERT-1, if they choose, can always select search terms from the titles and abstracts (i.e. free terms) with which truncated searches are performed by EXPERT-1. In its present state of development, EXPERT-1 does not exhaust all possible search techniques (e.g. author search, string search, major descriptor search, citation search etc.), but, it does embody the two most commonly used search techniques, namely, exact search with index terms and truncated search with free terms. Whether such techniques are too simple to achieve high retrieval effectiveness is still an open question, which requires more experiments and evaluation of the expert system to resolve.

The last major design goal of EXPERT-1 is maintainability. It is well-known in software engineering literature that modularity is the most effective means to control complexity of software system as well as to enhance understandability of programs. The decision to embody knowledge of a human expert searcher in terms of production rules is largely a consequence of the modularity principle. Each of the production rules represents a modular chunk of knowledge, and is at a sufficiently high-level for programmers to understand and program in. The "IF...THEN...ELSE" structure of a production rule and its Lisp-like notation (i.e. predicates followed by arguments) are extremely convenient to Lisp programmers.
Moreover, the modularity of production rules simplifies the task of extension of the knowledge base to incorporate new knowledge.

5.2 Remaining Issues

There are a number of issues that have not been addressed adequately in this thesis. They are crucial to the success of an operational computer-based intermediary for document retrieval.

First, the issue of characterization of user's search problem. Currently, EXPERT-1 asks the user to formulate his/her search problem in terms of a boolean intersection of concepts. Two examples and their characterizations are shown to the user to assist him in his formulation. Informal experiments, however, show that the examples are not sufficient to guide a user into doing the appropriate characterization. For example, in one experiment, a user was looking for documents about "treatment of pigmentary glaucoma and elevated intraocular pressure". He identified three major concepts from his search statement, namely, (1) treatment, (2) pigmentary glaucoma, and (3) intraocular pressure. EXPERT-1 takes this to mean that the user wants documents that contain ALL of these concepts. Nevertheless, the user is asking for documents that are about either "treatment of pigmentary glaucoma" or "treatment of elevated intraocular pressure". Hence, the appropriate characterization should be an intersection of two concepts, i.e., "treatment" and "pigmentary glaucoma", and putting "elevated intraocular pressure" or "intraocular pressure" as a related term or synonym for "pigmentary glaucoma".

Another issue related to search problem characterization in terms of concepts is the granularity of concepts. For instance, in the above "glaucoma" search, should "pigmentary glaucoma" be one concept or two, i.e. "pigmentary" and "glaucoma"? Should one always decompose a multiple-word concept into intersection of single-word concepts? Or, is it better to decompose multiple-word concepts for certain
classes of search problems only? And, is it sufficient to explain the meaning of "concept" to a user just by showing him/her a few examples?

A third issue is about search terms and their search modes. EXPERT-1 asks the user to give only a few search terms initially to start searching. Could it be a better idea to ask the user to supply as many search terms as possible from the beginning, perhaps, with the aid of a thesaurus? And if a thesaurus is used, should it be stored in EXPERT-1's database and be referred automatically? Should the user be given the choice to specify search modes (e.g. truncated search, exact search, adjacency search etc.) for each search terms? What are the tradeoffs involved in terms of simplicity of the system and retrieval effectiveness?

The last important issue is the procedure of relevance feedback and reformulation of search strategy. EXPERT-1 shows the titles and authors of the first ten documents (if the retrieved set has more than ten documents) to the user and asks him to select the more promising ones for further examination. Is this sampling procedure desirable? Is the sample size appropriate? Is the information shown adequate for the user to perform his selection? As regards search reformulation, EXPERT-1 only asks the user to select "good" search terms from relevant documents. What about the idea of using negative information, e.g., examine why irrelevant documents are retrieved?

Some of the unresolved issues cited above are just a matter of "fine-tuning" the design of EXPERT-1. However, a deeper issue is the merit of the whole approach (i.e. simulating human expert searchers) vis-a-vis alternate approaches discussed in section 2.4. This latter issue cannot be adequately addressed, I believe, unless further experimentation of EXPERT-1 is performed.
5.3 Future Directions

There are two directions in which more work can be done. The first direction is improvement of EXPERT-1 in terms of functionality and human aspects. For example, addition of author search, more appropriate suggestions to reformulate search strategy, better wording of questions and messages etc. It is also necessary to perform experiments with EXPERT-1 in order to measure its effectiveness and find out its strength and weaknesses.

The second line of work is towards the construction of an "ideal" intermediary, i.e., one that can adapt its behavior to different classes of users. EXPERT-1 is designed specially for inexperienced users or people who prefer not to learn any command language in order to do searching. For experienced users, it might be more appropriate to offer a different functionality. For instance, the system could allow user to characterize his search problem in a more flexible language (instead of just using "AND" and "OR"). Or, the user could have the option of specifying search modes for individual search terms. In other words, an "ideal" system is one that can offer different functionalities depending on the type of particular user currently using it. In the simplest case, the intermediary can be just an interface for transmitting user commands directly to remote retrieval host without any intermediate processing. At the other extreme, the intermediary resembles the current EXPERT-1 and offers maximal assistance to an user. One approach to this problem is to modularize the functionalities provided by the intermediary. For example, a user can select an intermediary which suggests appropriate databases and performs search strategy translation while the task of problem characterization or strategy reformulation is left entirely to the user. Whether such an approach is fruitful is, of course, an open question which needs further investigation.
6. References

[Buchanan 1971]

[Dattola 1979]

[Davis 1980]

[Davis & King 1977]

[Deane 1980]

[Doszkocs 1979]

[Feigenbaum 1977]

[Little 1972]

[Marcus & Reintjes 1977]

[------------------ 1979]

[Martin & Fateman 1971]

[McGill et al. 1977]

[Meadow 1977]

[Oddy 1977]

[Pople 1977]

[Salton 1968]
[1971]


[1978]

SIDC Search Service Orbit News, 6(8):1, October 1978.

[Shortliffe 1976]


[Shortliffe & Buchanan 1977]


[Smith 1976]


[Williams 1969]


[Williams & Preece 1975]

Appendix I - DataBase Selection Rules

;;;---* drule *---
;;;This file contains all the inference RULES for the database selection.

;;;Rules for asking user information to select databases...
;;;Rule d001 is the goal-rule for selecting databases.
;;;Rule d1000 is the "conclusion" of the task of database selection.
;;;Rules starting from d100 correspond to rules in SUBJECT AREA 1
;;;Rules starting from d200 correspond to rules in SUBJECT AREA 2
;;;and so on until SUBJECT AREA 8

(defrule d001
  (equal goal "database selection")
  => (senduser "The current task is to select appropriate databases for your search topic.")
  (askuser subject_area)
  (set-options-in-field)
  (askuser field)
  (activate-field-rule-list))

(defrule d1000
  true
  => (set-suggestions-in-dblist)
  (askuser dblist)
  (set-dbinfo)
  (setq goal "login")
  (activate l100))

;;;Rules about GOVERNMENT_ACTIVITIES_NEWS_AND_SPORT
(defrule d100
  (same field Canadian_news)
  =>
  (advice_to CANADIAN_NEWS_INDEX 1.0)
  (activate d1000))

(defrule d101
  (same field Current_research_projects)
  =>
  (advice_to SSIE 0.9)
  (advice_to NTIS 0.8)
  (activate d1000))

(defrule d102

(same field Congressional_information)
   =>
(advice_to CIS_INDEX 1.0)
(advice_to FEDERAL_INDEX 0.8)
(activate d1000))

(defrule d103
  (same field Doctoral_dissertations)
  =>
(advice_to DISSERTATION_ABSTRACTS 1.0)
(activate d1000))

(defrule d104
  (same field Federal_government_activities)
  =>
(advice_to FEDERAL_INDEX 1.0)
(advice_to CIS_INDEX 0.9)
(advice_to CRECORD 0.8)
(advice_to PAIS 0.7)
(activate d1000))

(defrule d105
  (same field Foundations_and_grants)
  =>
(advice_to FOUNDATION_GRANTS_INDEX 0.9)
(advice_to FOUNDATION_DIRECTORY 0.8)
(advice_to GRANTS 0.8)
(advice_to NATIONAL_FOUNDATION 0.7)
(activate d1000))

(defrule d106
  (same field Conference_papers)
  =>
(advice_to CONFERENCE_PAPERS_INDEX 1.0)
(activate d1000))

(defrule d107
  (same field Sports_and_recreation)
  =>
(advice_to SPORTS 1.0)
(advice_to NEWSPAPER_INDEX 0.8)
(activate d1000))

(defrule d108
  (same field US_government_publications)
(defrule d109
  (same field US_news_and_magazines)
  =>
  (advice_to NEWSPAPER_INDEX 1.0)
  (advice_to INFORMATION_BANK 0.8)
  (activate d1000))

;;;Rules concerning SCIENCES

(defrule d200
  (same field Agriculture_and_Food_sciences)
  =>
  (advice_to AGRICOLA 1.0)
  (advice_to CAB_ABSTRACTS 0.9)
  (advice_to CRIS 0.9)
  (advice_to BIOSIS_PREVIEW 0.6)
  (activate d1000))

(defrule d201
  (same field Biology_and_Life_sciences)
  =>
  (advice_to BIOSIS_PREVIEW 1.0)
  (advice_to SCISEARCH 0.8)
  (advice_to AQUATIC_SCIENCES_AND_FISHERIES_ABSTRACTS 0.6)
  (activate d1000))

(defrule d202
  (same field Chemistry)
  =>
  (advice_to CHEMICAL_ABSTRACTS 1.0)
  (advice_to CHEMNAME 0.8)
  (activate d1000))

(defrule d203
  (same field Geosciences)
  =>
  (advice_to GEOARCHIVE 1.0)
  (advice_to GEOREF 0.9)
  (advice_to METEOROLOGICAL_ABSTRACTS 0.7)
(advice_to SCISEARCH 0.6)
(activate d1000))

(defrule d204
  (same field Mathematics)
  =>
  (advice_to SCISEARCH 0.9)
  (advice_to NTIS 0.8)
  (advice_to INSPEC-A 0.7)
  (activate d1000))

(defrule d205
  (same field Meteorology)
  =>
  (advice_to METEOROLOGICAL_ABSTRACTS 1.0)
  (advice_to OCEANIC 0.6)
  (activate d1000))

(defrule d206
  (same field Nuclear_sciences)
  =>
  (advice_to ENERGY 0.9)
  (advice_to COMPENDEX 0.9)
  (advice_to NTIS 0.8)
  (advice_to NUCLEAR SCIENCE_ABSTRACTS 0.6)
  (activate d1000))

(defrule d207
  (same field Physics)
  =>
  (advice_to SEARCHABLE_PHYSICS_INFORMATION_NOTICES 1.0)
  (advice_to INSPEC-A 0.8)
  (advice_to NTIS 0.7)
  (activate d1000))

;;; Rules concerning APPLIED SCIENCES AND INDUSTRIES

(defrule d300
  (same field Aluminum_production_and_uses)
  =>
  (advice_to WORLD_ALUMINUM_ABSTRACTS 1.0)
  (advice_to METADFX 0.9)
  (advice_to COMPENDEX 0.7)
  (activate d1000))
(defrule d301
  (same field Food_technology_and_packaging)
  =>
  (advice_to FOODS_ABSTRACTS 1.0)
  (advice_to FOOD_SCIENCE_AND_TECHNOLOGY_ABSTRACTS 0.9)
  (advice_to AGRICOLA 0.6)
  (activate d1000))

(defrule d302
  (same field Metallurgy)
  =>
  (advice_to METADEX 1.0)
  (advice_to COMPENDEX 0.9)
  (advice_to WORLD_ALUMINUM_ABSTRACTS 0.7)
  (activate d1000))

(defrule d303
  (same field Paper_printing_and_packaging)
  =>
  (advice_to PIRA 1.0)
  (advice_to PAPERCHEM 0.8)
  (activate d1000))

(defrule d304
  (same field Rubber_and_Plastics)
  =>
  (advice_to RAPRA_ABSTRACTS 1.0)
  (activate d1000))

(defrule d305
  (same field Textile)
  =>
  (advice_to TITUS 1.0)
  (advice_to WORLD_TEXTILE_ABSTRACTS 0.9)
  (advice_to CHEMICAL_ABSTRACTS 0.8)
  (activate d1000))

(defrule d306
  (same field Transportation_and_automobile)
  =>
  (advice_to SAFETY_SCIENCE_ABSTRACTS 0.9)
  (advice_to AUTOMOTIVE_INDUSTRIES 0.9)
  (advice_to COMPENDEX 0.6)
  (activate d1000))
(defrule d307
  (same field Water_resources)
  ==> 
  (advice_to WATER_RESOURCES_ABSTRACTS 1.0)
  (advice_to WATER_RESOURCES_RESEARCH 0.8)
  (advice_to ENVIRONMENTAL_IMPACT_STATEMENTS 0.6)
  (activate d1000))

(defrule d308
  (same field Welding)
  ==> 
  (advice_to WELDSEARCH 1.0)
  (advice_to METADEX 0.8)
  (activate d1000))

;;; Rules concerning ENGINEERING

(defrule d400
  (same field Aerodynamics_and_fluid_engineering)
  ==> 
  (advice_to BHRA_F1_UDP_ENGINEERING 0.9)
  (advice_to COMPENDEX 0.8)
  (advice_to SAFETY_SCIENCE_ABSTRACTS 0.6)
  (advice_to NTIS 0.5)
  (activate d1000))

(defrule d401
  (same field Chemical_engineering)
  ==> 
  (advice_to COMPENDEX 0.9)
  (advice_to CHEMICAL_ABSTRACTS 0.8)
  (advice_to NTIS 0.7)
  (activate d1000))

(defrule d402
  (same field Civil_engineering)
  ==> 
  (advice_to COMPENDEX 1.0)
  (advice_to SAFETY_SCIENCE_ABSTRACTS 0.8)
  (advice_to NTIS 0.8)
  (activate d1000))

(defrule d403
  (same field Electronics_computers_and_systems)
  ==> 

(advice_to COMPENDEX 0.9)
(advice_to INSPEC-B&C 0.9)
(advice_to NTIS 0.7)
(activate d1000))

(defrule d404
  (same field Material_sciences)
  ==> 
  (advice_to CHEMICAL_ABSTRACTS 0.9)
  (advice_to METADEX 0.8)
  (advice_to COMPENDEX 0.6)
  (advice_to INSPEC-A 0.7)
  (activate d1000))

(defrule d405
  (same field Mechanical_engineering)
  ==> 
  (advice_to ISMEC 1.0)
  (advice_to COMPENDEX 0.9)
  (advice_to NTIS 0.8)
  (activate d1000))

(defrule d406
  (same field Nuclear_engineering)
  ==> 
  (advice_to COMPENDEX 0.9)
  (advice_to ENERGY 0.8)
  (advice_to NTIS 0.8)
  (advice_to INSPEC-A 0.7)
  (activate d1000))

(defrule d407
  (same field Space_technology)
  ==> 
  (advice_to COMPENDEX 0.9)
  (advice_to NTIS 0.8)
  (advice_to SAFETY_SCIENCE_ABSTRACTS 0.6)
  (activate d1000))

;;;Rules concerning MEDICAL_AND_HEALTH_SciENCES

(defrule d500
  (same field Cancer_research_and_therapy)
  ==> 
  (advice_to CANCERLIT 1.0)
(defrule d501
  (same field Clinical_medicine_and_medical_research)
  =>
  (advice_to MEDLINE 1.0)
  (advice_to EXCERPTA_MEDICA 0.8)
  (advice_to BIOSIS_PREVIEWS 0.7)
  (advice_to TOXLINE 0.6)
  (activate d1000))

(defrule d502
  (same field Drug_industry)
  =>
  (advice_to PHARMACEUTICAL_NEWS_INDEX 1.0)
  (advice_to INTERNATIONAL_PHARMACEUTICAL_ABSTRACTS 0.8)
  (activate d1000))

(defrule d503
  (same field Drug Uses_and_pharmacology)
  =>
  (advice_to INTERNATIONAL_PHARMACEUTICAL_ABSTRACTS 1.0)
  (advice_to MEDLINE 0.9)
  (advice_to TOXLINE 0.8)
  (advice_to HISTLINE 0.6)
  (activate d1000))

(defrule d504
  (same field Ethics_and_policies_in_medicine_and_health_care)
  =>
  (advice_to BIOETHICSLINE 1.0)
  (advice_to MEDLINE 0.7)
  (activate d1000))

(defrule d505
  (same field Health_planning_and_administration)
  =>
  (advice_to HEALTH_PLANNING_AND_ADMINISTRATION 1.0)
  (advice_to BIOETHICSLINE 0.8)
  (advice_to MEDLINE 0.6)
  (activate d1000))

(defrule d506
  (activate d1000))
(same field Toxicology)
   =>
(advice_to TOXLINE 1.0)
(advice_to CHEMICAL_ABSTRACTS 0.8)
(advice_to BIOSIS_PREVIEWS 0.7)
(advice_to MEDLINE 0.6)
(activate d1000))

;;; Rules concerning ENERGY_AND_ENVIRONMENT

(defrule d600
   (same field Economic_aspects_of_oil_and_petroleum)
   =>
(advice_to P/F_NEWS 1.0)
(advice_to INFORM 0.8)
(advice_to ENERGYLINE 0.7)
(activate d1000))

(defrule d601
   (same field Energy_alternatives_and_impact)
   =>
(advice_to ENERGYLINE 1.0)
(advice_to ENERGY_GENERAL_AND_PRACTICAL 0.9)
(advice_to ENERGY_DATABASE 0.8)
(advice_to ENERGY_RESEARCH_IN_PORGRESS 0.6)
(activate d1000))

(defrule d602
   (same field Energy_related_research)
   =>
(advice_to ENERGY_RESEARCH_IN_PROGRESS 0.9)
(advice_to ENERGYLINE 0.8)
(advice_to ENERGY_DATABASE 0.7)
(advice_to NTIS 0.5)
(activate d1000))

(defrule d603
   (same field Environmental_issues)
   =>
(advice_to ENVIRONMENTAL_IMPACT_STATEMENTS 1.0)
(advice_to ENVIROLINE 0.9)
(advice_to ENERGYLINE 0.8)
(advice_to ENVIRONMENTAL_PERIODICALS_BIBLIOGRAPHY 0.7)
(activate d1000))
(defrule d604
  (same field Human_ecology_land_and_water_resources)
  =>
  (advice_to ENVIRONMENTAL_PERIODICALS_BIBLIOGRAPHY 1.0)
  (advice_to ENVIROLINE 0.9)
  (advice_to ENERGYLE 0.8)
  (advice_to POLLUTION_ABSTRACTS 0.7)
  (activate d1000))

(defrule d605
  (same field Nuclear_safety_and_regulation)
  =>
  (advice_to NUCLEAR_SAFETY_INFORMATION_CENTER 1.0)
  (advice_to ENERGY 0.7)
  (activate d1000))

(defrule d606
  (same field Petroleum_technology)
  (activate d1000))
  =>
  (advice_to APILIT 1.0
    (activate d1000))
  (advice_to TULSA 0.9)
  (advice_to APIPAT 0.8
    (activate d1000))
  (advice_to COMPENDEX 0.6))

(defrule d607
  (same field Pollution_sources_and_control)
  =>
  (advice_to POLLUTION_ABSTRACTS 1.0)
  (advice_to APTIC 0.9)
  (advice_to ENVIROLINE 0.8)
  (advice_to ENVIRONMENTAL_PERIODICALS_BIBLIOGRAPHY 0.7)
  (activate d1000))

;;;Rules concerning BUSINESS_ECONOMICS_AND_LAW

(defrule d700
  (same field Accounting_and_finance)
  =>
  (advice_to ACCOUNTANTS_INDEX 1.0)
  (advice_to INFORM 0.9)
  (advice_to MANAGEMENT CONTENTS 0.8)
(activate d1000))

(defrule d701
  (same field Business_statistics_and_economic_forecasts)
  =>
  (advice_to US_STATISTICAL_ABSTRACTS 0.9)
  (advice_to INTERNATIONAL_STATISTICAL_ABSTRACTS 0.8)
  (advice_to US_ANNUAL_TIME_SERIES 0.7)
  (activate d1000))

(defrule d702
  (same field Current_business_news)
  =>
  (advice_to DOW_JONIES_NEWS_RETRIEVAL 1.0)
  (advice_to MANAGEMENT CONTENTS 0.9)
  (advice_to INFORM 0.8)
  (activate d1000))

(defrule d703
  (same field General_business_and_economics)
  =>
  (advice_to MANAGEMENT CONTENTS 1.0)
  (advice_to INFORM 0.9)
  (advice_to ECONOMIC_ABSTRACTS INTERNATIONAL 0.9)
  (activate d1000))

(defrule d704
  (same field Industrial_relations_labor_and_human_rights)
  =>
  (advice_to LABORDOC 1.0)
  (advice_to MANAGEMENT CONTENTS 0.8)
  (activate d1000))

(defrule d705
  (same field International_trade)
  =>
  (advice_to PTS_MARKET_ABSTRACTS 1.0)
  (advice_to PROMPT 0.9)
  (advice_to FOREIGN_TRADERS_INDEX 0.8)
  (advice_to INTERNATIONAL_STATISTICAL_ABSTRACTS 0.7)
  (activate d1000))

(defrule d706
  (same field Manufacturing_plants_and_industries)
  =>
  )
(advice_to ECONOMIC_INFORMATION_SYSTEMS_INDUSTRIAL_PLANTS 1.0)
(advice_to PROMPT 0.8)
(advice_to INTERNATIONAL_STATISTICAL_ABSTRACTS 0.7)
(activate d1000))

(defrule d707
(same field Marketing_management_and_administration)
  =>
(advice_to MANAGEMENT_CONTENTS 1.0)
(advice_to INFORM 0.9)
(advice_to ECONOMIC_ABSTRACTS INTERNATIONAL 0.8)
(advice_to PTS_MARKET_ABSTRACTS 0.7)
(activate d1000))

(defrule d708
(same field Non-manufacturing_industries)
  =>
(advice_to EIS_NON-MANUFACTURING_ESTABLISHMENTS 1.0)
(activate d1000))

(defrule d709
(same field law)
  =>
(advice_to PUBLIC_AFFAIRS_INFORMATION_SERVICE_NATIONAL 0.8)
(activate d1000))

;;;Rules concerning SOCIAL_SCIENCES_EDUCATION_AND_HUMANITIES

(defrule d800
(same field Education_for_exceptional_children)
  =>
(advice_to EXCEPTIONAL_CHILD_EDUCATION RESOURCES 1.0)
(advice_to ERIC 0.8)
(advice_to NIMIS 0.8)
(activate d1000))

(defrule d801
(same field Education_media_and_materials)
  =>
(advice_to AIM_ARM 1.0)
(advice_to NICEM 0.9)
(advice_to NIMIS 0.8)
(advice_to ERIC 0.7)
(activate d1000))
(defrule d802
   (same field Education_issues_professions_and_research)
   \implies
   (advice_to ERIC 1.0)
   (advice_to AIM_ARM 0.8)
   (advice_to PSYCHOLOGY_ABSTRACTS 0.7)
   (advice_to NIMIS 0.6)
   (activate d1000))

(defrule d803
   (same field Modern_art_and_design)
   \implies
   (advice_to ART_BIBLIOGRAPHIES_MODERN 1.0)
   (activate d1000))

(defrule d804
   (same field Modern литература_language_and_linguistics)
   \implies
   (advice_to MODERN_LANGUAGE_ASSOCIATION_BIBLIOGRAPHY 0.8)
   (advice_to LANGUAGE_AND_LANGUAGE_BEHAVIOR_ABSTRACTS 0.7)
   (activate d1000))

(defrule d805
   (same field Philosophy)
   \implies
   (advice_to PHILOSOPHER_INDEX 1.0)
   (advice_to SOCIAL_SCISEARCH 0.8)
   (advice_to ERIC 0.7)
   (activate d1000))

(defrule d806
   (same field Psychology_and_behavior_sciences)
   \implies
   (advice_to PSYCHOLOGICAL_ABSTRACTS 1.0)
   (advice_to SOCIAL_SCISEARCH 0.8)
   (advice_to SOCIOLOGICAL_ABSTRACTS 0.7)
   (activate d1000))

(defrule d807
   (same field Public_affairs_administration)
   \implies
   (advice_to PUBLIC AFFAIRS_INFORMATION_SERVICE_INTERNATIONAL 1.0)
   (advice_to CRECORD 0.8)
   (advice_to FEDERAL_INDEX 0.7)
   (activate d1000))
(defrule d808
  (same field US_and_World_Histories)
  ==> 
  (advice_to AMERICA:_HISTORY_AND_LIFE 0.9)
  (advice_to HISTORICAL_ABSTRACTS 0.8)
  (activate d1000))

(defrule d809
  (same field US_Political_sciences)
  ==> 
  (advice_to US_POLITICAL_SCIENCE_DOCS 1.0)
  (advice_to PUBLIC_AFFAIRS_INFORMATION_SERVICEINTERNATIONAL 0.8)
  (activate d1000))
Appendix II - Listing of Production Rules

;;; --* Production rules of the Knowledge Base *--

;;; Rules for developing search query and constructing initial Conceptual Representation...

(defrule p100
    (equal goal "planning the search")
    => (senduser
      "The first task in searching is to plan your search. There are three phases to planning a search:
1. stating and developing the search query;
2. setting up the search strategies needed; and
3. collecting data relevant to your search, e.g., time period to be covered, how many documents are needed, etc.
Let's proceed...")
      (askuser search_topic)
    (senduser
      "The immediate task is to set up the initial search strategy for your search problem. In order to do this, we have to identify the basic components or concepts of your query statement. For example, in a search for documents about the toxic effects of polychlorinated biphenyls on liver, the major concepts could be:
(1) polychlorinated biphenyls;
(2) toxic effects; and
(3) liver.
CONIT will then attempt to find documents that contain ALL of these concepts.
As a second example, in a search for documents about the design and implementation of Algol 60, the major concepts could be:
(1) design;
(2) implementation;
(3) Algol 60.
However, if what the searcher really wants are documents concerning either the design of Algol 60 or the implementation of Algol 60, then the above formulation is inappropriate since CONIT will attempt to find documents that contain ALL of the above concepts, namely, 'design', 'implementation' and 'Algol 60'. One way to formulate this problem is to use two concepts instead of three:
(1) design;
(2) Algol 60
and, later, add the term 'implementation' as a related term (i.e. keyword) to the concept 'design'. With this formulation, CONIT will find documents that
contain the term 'Algol 60' and either 'design' or 'implementation'.

Note that a concept may be expressed as a single word or a set of words, but in either case, it should convey significant information relevant to your search problem.

Usually, it is desirable to specify several concepts -- possibly two or three -- in order to get started on a meaningful preliminary search strategy. Let's find out distinct major concepts of your query statement..."

(senduser

"After the major concepts are determined, the next step is to find out descriptive terms or phrases for each of these concepts which are to be used in actual searching. NOTE: that: these descriptive terms or phrases are the so-called FREE (or uncontrolled) search terms and they will be searched in a TRUNCATED MODE, i.e., searching on their respective root-stems. For example, if 'intelligence' is a free term, CONIT will search on its root which is 'intellig' where the colon (:) indicates truncation. This means that every terms beginning with 'intellig' will be considered as a match. Truncated search for a phrase will be searched as an intersection of its component words. For example, 'artificial intelligence' will be searched as 'artifi: AND intellig:'. As a consequence, the 'TRUNCATED search is a kind of broad searching as opposed to the EXACT search which will be used later in the search process for searching INDEX (or controlled) terms."

(senduser

"Having found out the major concepts and their respective keywords (i.e. descriptive terms/phrases), we are going to review the initial representation for them. This table (or representation) of the concepts of the query statement with their respective keywords is crucial to the success of searching. Depending on partial search results, good keywords (identified from relevant documents) may be added to the table whereas bad search terms (those that are either too general or does not convey the meaning of a particular concept) are deleted. Even the major concepts themselves may be added or deleted. Thus, as the search proceeds and partial results are obtained, this table of concepts and keywords is continuously refined until a sufficient number of good documents are retrieved.

As a reminder, the following table is only the initial representation of concepts and keywords. Addition or deletion of keywords and/or concepts are recommended only after initial searching is done.

(show-conceptual-representation)

(askuser-conceptual-representation-OK?)

(activate p200))
(deffacts p200
(same_conceptual_representation-OK? yes)
(askuser number_of_documents_desired)
(setq goal "database selection")
(activate d001)
(else
(askuser modifications-to-CR)
(make-modifications-to-CR)
(senduser "The following table represents the updated representation.
(show-conceptual_representation)
(activate p200))

;;;Rules for handling LOGIN

(setq terminalid "")

(deffacts i100
(equal goal "login")
(=> (senduser (catenate "CONIT will try to connect to the retrieval system -- " *system*))
(activate i200))

(deffacts i200
(phone
(=> (senduser "Phone connection made to Telenet.")
(sleep 1)
(sendhost cr)
(sleep 2)
(sendhost ";")
(get_response)
(activate i300)
(else
(senduser "Phone connection to Telenet failed ")
(senduser "We will try to connect again.")
(activate i200))

(deffacts i300

(appear resline "TELENET")
(=> (senduser "Telenet Responding.")
(get_response)
(activate i400)
(else
(senduser resline)
(get_response)
(activate i300)))
(defrule i400
  (appear resline "617")
  => (senduser resline)
  (get_response)
  (activate i500)

(else (error "Unknown response line: " resline))

(defrule i500
  (blank resline)
  => (senduser resline)
  (sleep 3)
  (get_partial_response 9)
  (activate i600))

(defrule i600
  (appear resline "TERMINAL =")
  => (sendhost terminalid)
  (sleep 2)
  (get_response)
  (senduser resline)
  (sleep 3)
  (activate i700o)
  (activate i700s)
  (activate i700n))

(defrule i700o
  (equal *system* "ORBIT")
  => (sendhost "C 213 33")
  (get_response)
  (activate i800o))

(defrule i700s
  (equal *system* "SUNY")
  => (sendhost "C 518 20")
  (get_response)
  (activate i800s))

(defrule i700n
  (equal *system* "NLM")
  => (sendhost "C 301 20")
  (get_response)
  (activate i800n))

(defrule i800o
(blank resline)
==> (senduser resline)
   (get_response)
   (activate 1800n)
ELSE  (activate 1900n))

(defrule 1800n
 (blank resline)
 ==> (senduser resline)
   (get_response)
   (activate 1800n)
ELSE  (activate 1900n))

(defrule 1800s
 (blank resline)
 ==> (senduser resline)
   (get_response)
   (activate 1800s)
ELSE  (activate 1900s))

(defrule 1900o
 (appear resline "CONNECTED")
 ==> (senduser resline)
   (senduser "Logging onto ORBIT.")
   (sendhost "/LOGIN SDCNET3")
   (get_response)
   (activate 11000n))

(defrule 1900n
 (appear resline "CONNECTED")
 ==> (senduser resline)
   (senduser "logging onto NLM.")
   (sleep 5)
   (sendhost "/LOGIN")
   (get_response)
   (activate 11000n))

(defrule 1900s
 (appear resline "CONNECTED")
 ==> (senduser resline)
   (senduser "logging onto SUNY.")
   (sendhost "/LOGIN")
   (get_response)
   (activate 11000n))
(activate s100))

;;; Rules for searching...

(defrule s100
  (appear resline "USER:"))
  => (construct-search)
  (senduser
  "We have already connected to the desired database and are ready to start searching. CONIT will automatically translate the concepts and keywords in the above table into a sequence of search commands. The following will be the actual execution of these search commands and the response from the retrieval system. You can just wait and see the results."
  )
  (nullify *index_term_list*)
  (nullify *free_term_list*)
  (activate s200)
ELSE (senduser resline)
  (get_response)
  (activate s100))

(defrule s200
  (null *query-agenda*)
  => (setq goal "feedback")
  (activate f100)
ELSE (setq outline (eval-query (pop *query-agenda*)))
  (activate s210))

(defrule s210
  (equal outline "")
  => (activate s200)
ELSE (sendhost outline)
  (senduser outline)
  (sleep 5)
  (get_response)
  (activate s300))

(defrule s300
  (appear resline "PROG:"))
  => (get_response)
  (activate s310)
  (activate s320)
  (activate s330)
(activate s340)

(activate s350)

ELSE (senduser resline)
(activate s300))

(defrule s310
  (or (appear resline "GE\NTRM OVFL")
  (or (appear resline "PROCPSTG OVFL")
  (appear resline "PSTG OVFL")))
  ==> (senduser resline)
  (senduser "CONIT will redo the search in exact mode.")
  (push (retranslate-query (car *search*) *query-agenda*)
  (get_response)
  (activate s200))

(defrule s320
  (or (appear resline "SS ")
  (or (appear resline "NP")
  (or (appear resline "NONE")
  (or (appear resline "NO MATCH")
  (appear resline "NO POSTINGS"))))
  ==> (fill-search-result resline)
  (activate s400))

(defrule s330
  (or (appear resline "STORPSTG OVFL")
  (or (appear resline "PERM OVFL")
  (appear resline "KE\BTRM OVFL")))
  ==> (see 2)
  (senduser
  (catenate " We have overflowed the storage space that the " *system*)
  (senduser
  "system allows. In order to continue, CONIT will clear all the previous searches that have been done. And the Concept Table will be retranslated into a sequence of searches to avoid overflow again by replacing TRUNCATED searches which found more than 50,000 documents to EXACT searches."
  )
  (nullify *search*)
  (nullify *query-agenda*)
  (nullify-queries)
  (construct-search)
  (nullify *index_term_list*)
(nullify *free_term_list*)
(sendhost "ERASEALL")
(sleep 5)
(get_response)
(activate s400))

(defrule s340
  (appear resline "TIME OVFLW")
  =>
  (get_response)
  (activate s360))

(defrule s350
  (blank resline)
  =>
  (get_response)
  (activate s310)
  (activate s320)
  (activate s330)
  (activate s340)
  (activate s350))

(defrule s360
  (appear resline "USER:")
  =>
  (sendhost "Y")
  (sleep 4)
  (get_response)
  (activate s300)
 ELSE
  (get_response)
  (activate s360))

(defrule s400
  (appear resline "USER:")
  =>
  (activate s200)
 ELSE
  (senduser resline)
  (get_response)
  (activate s400))

;;; Rules for feedback...

(defrule f100
  (equal goal "feedback")
  =>
  (setq last_document_shown 0)
  (setq init-feedback-info
"CONIT will now show more information on the documents that you have judged to be likely relevant. For each of these documents, CONIT will display its
"The above iteration of searching does not find any document that contains all of the concepts in your search statement. This has two explanations: (1) Either you have not included sufficient search terms for your concepts; or (2) The requirement that only documents containing 10% of the concepts of your search statement be retrieved is too tight.

CONIT, therefore, suggests you to revise your Concept Table by (a) adding additional keywords under each concept; or (b) loosening the concept requirement by dropping a less important concept from the table.

The following will be the Concept Table for your previous search; please modify it."

(show-conceptual-representation)
(askuser conceptual-representation-OK?)
(activate f2300))

"The current search probably found too many documents than what you want. CONIT suggests you to NARROW your search in the next iteration of searching by either (1) dropping out search terms that are either too general or do not convey the meaning of the concepts to which they belong from the Concept Table; or (2) tightening the search requirement by adding a new concept to the Table. But before you do that, it is a good idea to look at some of the documents found just then. This will give you an idea of which documents are relevant, and hence which search terms to keep or to delete."

(activate f250))

(defrule f220)
(lessp total_document (get-slot-value number_of_documents_desired value))

=> (senduser

"The current search probably found too few documents than what you want. CONIT suggests you to BROADEN your search in the next iteration of searching by either
(1) adding additional search terms to the Concept Table; or
(2) loosening the search requirement by deleting a concept from the Table.
But before you do that, it is a good idea to look at some of the documents found just then. It will give you an idea of which documents are relevant, and hence what search terms to choose from them.
"

(activate f240)
(activate f250))

(defrule f230

(lessp total_document (times
 (get-slot-value number_of_documents_desired value) 10))
 (greatep total_document
 (get-slot-value number_of_documents_desired value))

=> (activate f240)
(activate f250))

(defrule f240

 (greatep total_document 0)
 (lessp total_document 10)

=> (senduser

"The current task is to examine all of the documents found in the previous iteration of searching. CONIT will first show the TITLES and AUTHORS of these documents so that you can select from them some of the more promising documents, i.e., those documents that are likely to be relevant to your search statement. CONIT will then show more details on each of the selected documents for your further examination."

(activate f310o)
(activate f310n))

(defrule f250

(greatep total_document 9)

=> (senduser

"The current task is to examine some of the documents found in the previous iteration of searching. CONIT will first show the TITLES and AUTHORS of ten documents so that you can select from them some of the more promising documents, i.e., those documents that are likely to be relevant to your search statement. CONIT will then show more details on each of the selected documents for your further examination."
)
(activate f300n)

(defrule f300n
    (greaterp total_document (plus last_documentShown 9))
    (or (equal *system* "NLM")
        (equal *system* "SUNY"))
    =>
    (setq outline (catenate """"PRINT INDENTED 10 SKIP"
        (g pname last_documentShown)))
    (sendhost (catenate outline "; Tl, AU"))
    (setq last_documentShown (plus last_documentShown 10))
    (activate f400))

(defrule f300o
    (greaterp total_document (plus last_documentShown 9))
    (equal *system* "ORBIT")
    =>
    (sendhost (catenate """"PRINT INDENTED Tl AU 10 SKIP"
        (g pname last_documentShown)))
    (setq last_documentShown (plus last_documentShown 10))
    (activate f400))

(defrule f310n
    (greaterp total_document last_documentShown)
    (lessp total_document (plus last_documentShown 10))
    (or (equal *system* "NLM")
        (equal *system* "SUNY"))
    =>
    (setq outline (catenate """"PRINT INDENTED"
        (g pname (difference
            total_document last_documentShown))))
    (setq outline (catenate outline " SKIP "))
    (setq outline (catenate outline (g pname last_documentShown)))
    (sendhost (catenate outline " Tl, AU"))
    (setq last_documentShown total_document)
    (activate f400))

(defrule f310o
    (greaterp total_document last_documentShown)
    (lessp total_document (plus last_documentShown 10))
    (equal *system* "ORBIT")
    =>
    (setq outline (catenate """"PRINT INDENTED Tl AU"
        (g pname (difference
            total_document last_documentShown))))
    (setq outline (catenate outline " SKIP "))
    (sendhost (catenate outline " Tl, AU")))
(setq last_document_shown total_document)
(activate f400))

(defrule f400
  true
  =>  (sleep 5)
       (get_response)
       (activate f500))

(defrule f500
  (appear resline "USER:"))
  =>  (askuser apparent_good_documents)
      (setq doclist (get-slot-value apparent_good_documents value))
      (activate f530)
    ELSE (activate f510))

(defrule f510
  (appear resline "CONTINUE PRINTING?")
  =>  (senduser resline)
      (get_response)
      (activate f520)
    ELSE (senduser resline)
      (get_response)
      (activate f500))

(defrule f520
  (appear resline "USER:"))
  =>  (sendhost "Y")
      (senduser "Y")
      (sleep 5)
      (get_response)
      (activate f500)
    ELSE (senduser resline)
      (get_response)
      (activate f520))

(defrule f530
  (equal doclist 'unknown)
  =>  (senduser
       "There is probably too few good documents in this search set. CONIT suggests
you to modify your Concept Table by adding more search terms or deleting
search terms that are too general or do not convey the meaning of the concepts
to which they belong. You can either look at muf documents before you do the
modifications, or stop browsing and proceed directly to change the Table."
       )
      (activate f1800)
ELSE (activate f600))

(deerule f600
   (or (equal doclist 'unknown)
        (null doclist))
   => (activate f1700)
ELSE (senduser init-feedback-info)
   (setq init-feedback-info
   "CONIT will show more information on the next document that seems to be
likely relevant to your search topic for your further judgement of relevance
and selection of additional good search terms."
   (setq outline (g_pname (difference (pop doclist)
                    1))
   (setq doc$title " ")
   (setq doc$authors " ")
   (setq doc$abstract " ")
   (setq doc$source " ")
   (setq doc$terms " ")
   (activate f800o)
   (activate f800n))

(deerule f800o
   (equal *system* "ORBII")
   => (sendhost (catenate "PRINT INDENTED TI AU AB IT SO 1 SKIP "
                  outline))
   (sleep 5)
   (get_response)
   (activate f900))

(deerule f800n
   (or (equal *system* "NLM")
        (equal *system* "SUNY")
   => (setq outline (catenate """"PRINT INDENTED 1 SKIP " outline))
   (sendhost (catenate outline ", TI, AU, AB, MH, SO"))
   (sleep 5)
   (get_response)
   (activate f900))

(deerule f900
   (appear resline "TITLE")
   => (setq doc$title (catenate doc$title resline))
   (senduser resline)
   (get_response)
   (activate f1000)
ELSE (senduser resline)
(get_response)
(activate f900))

(defrule f1000
  (appear resline "AUTHOR")
  = => (setq doc$authors (catenate doc$authors resline))
  (senduser resline)
  (get_response)
  (activate f1100)
ELSE (setq doc$title (catenate doc$title resline))
  (senduser resline)
  (get_response)
  (activate f1000))

(defrule f1100
  (appear resline "ABSTRACT")
  = => (setq doc$abstract (catenate doc$abstract resline))
  (senduser resline)
  (get_response)
  (activate f1300)
ELSE (activate f1200))

(defrule f1200
  (or (appear resline "INDEX TERM")
      (appear resline "MESH HEADING"))
  = => (activate f1400o)
  (activate f1400n)
ELSE (setq doc$authors (catenate doc$authors resline))
  (senduser resline)
  (get_response)
  (activate f1100))

(defrule f1300
  (or (appear resline "INDEX TERM")
      (appear resline "MESH HEADING"))
  = => (activate f1400o)
  (activate f1400n)
ELSE (setq doc$abstract (catenate doc$abstract resline))
  (senduser resline)
  (get_response)
  (activate f1310))

(defrule f1310
  (appear resline "CONTINUE PRINTING?")
  = => (see 2)
(sendhost "Y")
(sleep 5)
(get_response)
(activate fl300)
ELSE (activate fl300))

(defrule fl400o
    (equal *system* "ORBIT")
    => (setq doc$terms (catenate doc$terms resline))
    (get_response)
    (activate fl500o))

(defrule fl400n
    (or (equal *system* "NLM")
        (equal *system* "SUNY"))
    => (nullify MHlist)
    (activate fl500n))

(defrule fl500o
    (appear resline "SOURCE")
    => (senduser resline)
    (setq doc$source (catenate doc$source resline))
    (get_response)
    (fill-expect-in-index-terms (parse-IT doc$terms))
    (activate fl600))
ELSE (setq doc$terms (catenate doc$terms resline))
    (get_response)
    (activate fl500o))

(defrule fl500n
    (appear resline "MESH HEADING")
    => (setq inline (catenate " " resline))
    (setq doc$terms (catenate doc$terms resline))
    (get_response)
    (activate fl510n))

(defrule fl510n
    (appear resline "SOURCE")
    => (setq MHlist (append MHlist (parse-MH inline))
    (senduser resline)
    (setq doc$source (catenate doc$source resline))
    (get_response)
    (fill-expect-in-index-terms MHlist)
    (activate fl600)
ELSE  (activate f1520n))

(derule f1520n
  (appear resline "MESH HEADING")
  = ==> (setq doc$terms (catenate doc$terms resline))
          (setq MHlist (append MHlist (parse-MH inline)))
          (setq inline (catenate " " resline))
          (get_response)
          (activate f1510n)
)

FI.SE (setq inline (catenate inline resline))
      (setq doc$terms (catenate doc$terms resline))
      (get_response)
      (activate f1510n))

(derule f1600
  (appear resline "USER:"))
  = ==> (create-good-document)
          (askuser-good-terms)
          (activate f600)
ELSE  (activate f1610))

(derule f1610
  (or (blank resline)
      (appear resline "SS "))
  = ==> (get_response)
          (activate f1600)
ELSE  (setq doc$source (catenate doc$source resline))
          (get_response)
          (activate f1600))

(derule f1700
  (equal total_document last_document_shown)
  = ==> (senduser
        "All of the documents found in the previous iteration of searching have been
        shown already. Please modify your Concept Table to prepare for the next
        iteration of searching."
        (activate f1810))
ELSE  (askuser sample_more?)
          (activate f1800))

(derule f1800
  (same sample_more? yes)
  = ==> (activate f300o)
          (activate f300n)
          (activate f310o)
(activate f310n)
ELSE (activate f1810))

(defrule f1810
  true
  ==> (askuser good_set?)
       (activate f1900))

(defrule f1900
  (same good_set? yes)
  ==> (activate f1920)
ELSE (activate f2200))

(defrule f1920
  (greaterp total_document 300)
  ==> (senduser
       "The search set has more than 300 documents. CONIT can only request offline
print for a maximum of 300 documents.")
       (askuser proceed?)
       (activate f1930)
ELSE (activate f1940))

(defrule f1930
  (same proceed? yes)
  ==> (setq total_document 300)
       (activate f1940)
ELSE (activate f2200))

(defrule f1940
  true
  ==> (setq outline (catenate "PRINT " (g pname total_document)))
       (sendhost (catenate outline " OFFLINE FULL INDENTED"))
       (sleep 5)
       (get_response)
       (activate f2000))

(defrule f2000
  (or (appear resline "NAME")
      (appear resline "REQUESTER"))
  ==> (senduser "CONIT is now sending the mailing address to the retrieval system.")
(sendlhost
"R Marcus, ADDRESS = MIT 35-406, CITSTATZIP = Cambridge MA 02139, REQUESTER = Same, TITLE = NONE")
(sleep 5)
(get_response)
(activate f2100)
ELSE (senduser resline) (get_response) (activate f2000))

(defrule f2100
 (appear resline "OK")
 => (see 2)
 (sendhost "Y")
 (sleep 3)
 (see 6)
 (askuser STOP?)
 (activate f2110)
 ELSE (get_response) (activate f2100))

(defrule f2110
 (same STOP? yes)
 => (sendhost "STOP Y")
 (see 5)
 (senduser "Thank for using CONIT.")
 (quit)
 ELSE (activate f2200))

(defrule f2200
 true
 => (where-to-add *index_term_list*)
 (where-to-add *free_term_list*)
 (senduser
 "The current task is to examine the updated Concept Table which will include all the INDEX terms or FREE terms, if there is any, that you have supplied to CONIT. The following are few suggestions to what you may want to do with the Concept Table:

1. Add additional search terms to the Table.
   - do it in case you have left out some search terms that you think are relevant; or
   - in case you want to BROADEN your search, i.e., to find more documents.

2. Delete search terms.
   - do it in case some of the previous search terms are irrelevant or too general for your search purpose; or
   - in case you want to NARROW your search, i.e., to find fewer documents.

3. Replace search terms in the Table.
   - do it in case you have mis-spelled some of the search terms; replace
     is also a quick way to do delete followed by addition of a search term within the same concept.

4. Add a new Concept

"
- do it in case your previous search found too many documents.

(5) Delete a Concept
- do it in case your previous search found too few documents.
"

(show-conceptual-representation)
(askuser conceptual-representation-OK?)
(activate f2300))

(defrule f2300
  (same conceptual-representation-OK? yes)
  => (construct-search)
  (nullify *index_item_list*)
  (nullify *free_term_list*)
  (senduser
"CONIT will translate the updated Concept Table into sequence of search commands
and execute them. The following will be the actual interactions with the
retrieval system. You can just wait and see the results.
"
)
  (activate s200)
ELSE (askuser modifications-to-CR)
  (make-modifications-to-CR)
  (senduser "The following Concept Table represents the updated changes."
)(show-conceptual-representation)
(askuser conceptual-representation-OK?)
(activate f2300))