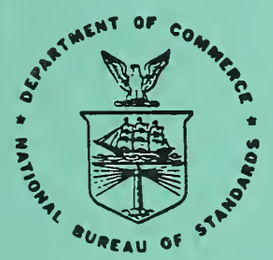
NBSIR 88-3689

A Prototype Expert System: An Automated Advisor to Select Data Sources From Chemical Information Databases

Elizabeth N. Fong
Christoper E. Dabrowski

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Institute for Computer Sciences and Technology
Information Systems Engineering Division
Gaithersburg, MD 20899

January 1988



U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

NBSIR 88-3689

A PROTOTYPE EXPERT SYSTEM: AN AUTOMATED ADVISOR TO SELECT DATA SOURCES FROM CHEMICAL INFORMATION DATABASES

Elizabeth N. Fong
Christoper E. Dabrowski

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Institute for Computer Sciences and Technology
Information Systems Engineering Division
Gaithersburg, MD 20899

January 1988

U.S. DEPARTMENT OF COMMERCE, C. William Verity, *Secretary*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

A PROTOTYPE EXPERT SYSTEM: AN AUTOMATED ADVISOR TO SELECT DATA SOURCES FROM CHEMICAL INFORMATION DATABASES

Elizabeth N. Fong
Christopher E. Dabrowski

Information Systems Engineering Division
Institute for Computer Sciences and Technology
National Bureau of Standards

ABSTRACT

A prototype expert system, called "Automated Advisor," was built as a part of a competency project within the Institute for Computer Sciences and Technology. The system conducts dialogue with the end-users and recommends a list of data sources from chemical information databases.

This report describes the problem domain and documents the knowledge engineering process.

Key words: database management system, expert system, knowledge acquisition, knowledge engineering, knowledge-based system, vapor pressure.

1. INTRODUCTION

This report describes a competency project within the Institute for Computer Sciences and Technology for research and development in the area of knowledge-based applications.

Knowledge-based systems (KBS) represent a new software methodology which can broaden the scope of computer applications [CUGI87]. Typically, such applications are those for which the number of decisions to be made is rather large, and the order in which decisions are made is unpredictable.

An important class of KBS applications is that of Expert Systems (ES) [HAYE83]. An expert system is a computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solution. The knowledge necessary to perform at such a level, plus the inference procedures used, can be thought of as a model of the expertise of practitioners in the field.

1.1 Purpose

The main purpose of this project is to gain competency in knowledge engineering, i.e., the practice of building an expert system. Our goal is to build a prototype in which it is possible to model a scenario having the following characteristics:

* there are incoming requests for information which require discussions, strategy development and refinement,

* there are many possible solutions but it is necessary to provide a "best" solution based upon the judgement of a domain expert.

1.2 Project Definition

In defining an application for the development of an expert system which will incorporate the scenarios mentioned above, Dr. David Jefferson, Chief of the Information Systems Engineering Division, conceived the idea of building an intelligent front-end -- that is, an expert system application which behaves as an expert which can assist and intelligently select data sources from a collection of databases. This goal proved to be of interest to the NBS, Office of Standard Reference Data (OSRD), to assist in coordinating computerized reference information and providing rapid public access to scientific information.

The domain of application for this prototype expert system was suggested by Dr. John Rumble of OSRD who coordinates many data

centers and disseminates up-to-date evaluated scientific information to the technical community.

The expert system to be prototyped, called the "Automated Advisor," is an intelligent assistant for selecting data sources from a collection of multiple databases.

1.3 Scope

This project has four main parts:

* Reviewing state-of-the-art commercial expert systems software and selecting a tool to be used to build the prototype Automated Advisor expert system.

* Acquiring knowledge from data center experts who respond to inquiries for scientific information from the technical community.

* Prototyping the Automated Advisor expert system using the expert system development software tool.

* Demonstrating and documenting the results.

The scope of the problem domain is limited to identifying and recommending information sources for a small subset of chemical thermodynamic properties of pure chemical substances.

1.4 Disclaimers

The project is a research venture in the area of knowledge engineering. Hence, it is driven by the designers' interests rather than by the objective of producing a deliverable expert system.

Certain commercial products are identified in this report in order to adequately specify the procedures being described. In no case does such identification imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the product identified is necessarily the best for the purpose.

1.5 Acknowledgement

We would like to thank the domain expert, Dr. Malcolm Chase, Group Leader of the Chemical Thermodynamics Data Center, who spent many hours talking into our tape recorder. We would like to thank the secondary domain expert, Dr. Howard White of OSRD, who furnished us with ancillary information on how to handle chemical information queries.

We also would like to thank Joan Sauerwein, Thomas Jobe, Neil Olien, Dr. Steve Stein, and Ted Selover who provided supplemental domain knowledge.

Scientists at NBS were invited to review the prototype system for the design of the user interfaces. Among those are Dr. David Garvin and Dr. David Neumann, both of the Chemical Thermodynamics Data Center.

The roles of Dr. David Jefferson, Dr. David Lide, and Dr. John Rumble, who helped us plan the project and provided us with actual problem domain experts, were critical for the realization of this project.

Thanks are due to Jon Orwant, a summer student, who assisted us in the implementation and testing of this expert system.

2. DESCRIPTION OF THE APPLICATION

In this information age, scientists, engineers and technicians need rapid access to reliable reference data. The Office of Standard Reference Data (OSRD) of the National Bureau of Standards provides up-to-date scientific information to the technical community. The Office, mandated by the Standard Reference Data Act (Public Law 90-396) coordinates the activities of about 23 data centers and approximately 40 other data evaluation projects [SAUE85]. Each data center monitors an important scientific area and maintains one or more databases. These databases are usually available to the technical community in two forms: published literature or computer tapes.

The Chemical Thermodynamics Data Center (CTDC) is one of the NBS data centers. The responsibilities of the CTDC are (1) to collect, maintain and analyze data on thermodynamic properties of chemical substances, and (2) to answer public inquiries relating to properties for specific chemicals. The CTDC collection includes data on the thermodynamic properties of more than 15,000 substances.

The purpose of this competency project is to assess the feasibility of using knowledge-based systems technology in a distributed data environment to provide computer assistance in understanding the user's data requirements and to provide relevant data sources.

2.1 The Problem Domain

The problem domain of interest is the identification of sources of chemical thermodynamics information. The problem centers around selection and recommendation of appropriate information sources for individual scientists or engineers who require data on specific thermodynamic properties for research or industrial use. Normally, this function is performed by scientists within the CTDC who interact with individual end-users to fulfill requests for selection of sources of data. The intent of this expert system is to simulate a scientist within the data center environment. The data center itself maintains a large store of chemical information located in various publications, files, and computer databases. However, a significant amount of relevant information may not be available at the CTDC. Instead, it may be found in collections located at other institutions, or through various electronic subscription services.

Recommending data sources to end-users means not merely providing citations to publications and electronic services, but also means using the scientist's knowledge about the information sources to recommend one or possibly a few sources which best match the end-user's requirements.

In addition, the scientist at the data center must deal with various issues and problems in locating and recommending data sources for the end-user.

* Search in large databases or literary collections is often time-consuming. To find correct sources quickly and reliably requires an in-depth knowledge of the research literature.

* In situations involving collected data on chemical properties, information may actually be uncertain and incomplete.

* Inquirers who need information usually require assistance in articulating the request so that problems can be stated accurately and the end-users requirements made explicit.

* There is also a need to provide advice and guidance to the end-users on how to use and interpret the selected data sources.

The functionalities of the Automated Advisor Expert System are to understand end-user's requests, to select appropriate data sources, to provide access to databases where references to data sources exist, and to give advice on the use of recommended data sources. The current prototype does not retrieve data from the recommended data sources; such a capability would be very useful but would require considerably more work and probably much more powerful and expensive hardware and software. In addition, some of the sources do not exist in machine-readable form.

2.2 The Scope of the Prototype

To illustrate the problem, the focus of the prototype is directed to a very small area within the domain. The scope is limited to vapor pressure properties of pure chemical substances.

Vapor pressure is one of many thermodynamic properties of a chemical substance. It can be described as the pressure exerted when a solid or a liquid is in equilibrium with its own vapor. The vapor pressure is a function of the substance and the temperature.

The Chemical Thermodynamics Data Center often gets inquiries such as "what is the vapor pressure of chemical compound X in temperature range T1 to T2?" Vapor pressure data is important under the following circumstances:

* For complying with federal regulations regarding storage and transportation of chemical substances.

* In manufacturing or production applications, vapor pressure data is needed to effect correct chemical separation procedures.

* In chemically reactive systems involving mixtures of several different compounds, vapor pressures of the reactants and the products must be known so that the vapor pressure of the mixture may be calculated as the reaction proceeds.

* In research applications, scientists may need to know the vapor pressure of a known compound, or may need to estimate the vapor pressure of a new compound on the basis of properties of similar known compounds so that safe research procedures can be performed.

2.3 Types of Queries

The queries that come into the data centers are usually telephone calls or written inquiries. From the data gathered at the OSRD, the bulk of the inquiries (about 60%) are specific. These inquiries are requests for information about the vapor pressure of a specific substance, requests for references to a specific publication, or requests for a computer tape. Initially the requests are handled by a clerk.

Based upon the statistics gathered at OSRD, on the average, there are approximately 5 to 10 telephone calls a day consisting of simple inquiries of the types listed above. The duration of the dialogue is typically less than 3 to 5 minutes.

On the average, there are 1 to 2 requests a day consisting of inquiries that are complex and cannot be handled by a clerk. The complex inquiries are generally referred to a scientist who specializes in the domain of the subject query. Complex queries require more discussion between the expert and the inquirer in order to clarify the request. A final answer may not be determined during the discourse. The expert may have to do further research to answer the question. Typically, the scientist will deliver several references to the user. The user may then contact the expert for more data. This type of iteration can last for a considerable length of time.

About 10% to 15% of the inquiries cannot be answered. This inability to provide an answer may mean that no data source

exists to fulfill the query, or it could mean that the data is not available in the current bibliographic collections.

2.4 Types of Data Sources

The information sources used by the expert to answer an inquiry exist in many different forms. Those typically used by the CTDC are:

* computerized bibliographic data collected and maintained by the Data Center,
* OSRD publications and tapes which are for sale,
* handbooks or articles containing chemical data that are generally found in technical libraries,
* the scientist's personal collection of data sources and bibliographies including his/her own research,
* access to on-line database subscription services, i.e., Numerica, TDS, etc.

The handbook or article may list the desired chemical data in a table or graph, or provide one or more equations which can be used to calculate the data. Also, a reference to a bibliographic work may be provided which the inquirer can use to find other data sources.

3. EXPERT SYSTEMS TECHNOLOGY

This section provides a brief description of the elements of expert systems technology and a discussion of the application of expert knowledge in rule form.

3.1 Representation of Knowledge in Expert Systems

In many cases, knowledge about how to solve a problem can be obtained from human experts and conveniently translated into rule form. Rules can be thought of as chunks of expert knowledge. The rules associated with the problem solving task performed by an expert system are known collectively as a knowledge base. The problem to be solved is itself represented internally as a group of facts about which the rules can reason. During the execution of an expert system, application of the rules results in examination of part or all of the facts associated with a problem leading to the systematic conclusion of new facts ultimately including the problem solution.

3.2 Rule-based Systems

For the purposes of this report, rule-based systems may be considered a specialization of expert systems which rely primarily on rules for representing and applying knowledge. Rule-based systems have several important aspects which are described in this section.

3.2.1 Rules

Rules consist of IF --> THEN condition action pairs. Rules are internal data structures used to represent small pieces of knowledge about what action to take or what to conclude under a particular set of conditions. Rules have two parts: the IF part, or antecedent, lists one or more conditions which must hold true; the THEN part, or consequent, contains conclusions which are reached if the conditions in the IF part are satisfied. Individual conditions and conclusions are represented internally as clauses or expressions which are patterns to be matched against actual data. For example, the following rule has an antecedent portion consisting of two conditions, and one consequent conclusion.

```
        IF the animal has wings
        AND the animal has feathers
    THEN
        CONCLUDE the animal is a bird.
```

3.2.2 Rule Sets

Rules may be organized into rule sets by topic or category. They may be segregated into groups on the basis of subject matter,

types of conclusions reached, and problems addressed, among other criteria. These groups may be applied to the problem individually at specific times. Determination must be made when to apply a particular rule set and the conditions under which it may operate. This determination may be made by a controlling module responsible for overall problem processing. Typically, this module is itself a set of rules.

3.2.3 Inference Engine

Rules are applied by an inference engine. Inference engines are computer programs which match the patterns in rules against existing information to make conclusions. They are responsible for applying the rules of a knowledge base or a subdivision of the knowledge base, and for controlling the execution or "reasoning" of a knowledge-based system. Two strategies are generally recognized:

* Backward chaining begins with a top level goal: to prove that a premise is implied by existing facts. Backward chaining does this by working "backwards" through a series of (hopefully) simpler subgoals which will establish the premise. The procedure is simple: if a required fact is not already known, a rule is sought which includes that fact in the THEN part. The conditions in the IF part of the rule must then be satisfied; each condition becomes a new subgoal. The procedure continues until either the top level goal is established, or no new subgoals can be generated.

* Forward chaining is generally used to determine the consequences of facts. That is, the IF portions of rules are examined to see whether or not they are true. If they are, the facts in the THEN part are concluded and added to the knowledge base. The IF portions are then examined again to see if new facts can be concluded. The process continues until no more new facts can be concluded.

3.2.4 Confidence Factors

Confidence factors are a numeric measure of the degree to which a fact is believed to be true (or false) by the knowledge-based system. For instance, absolute confidence may be 1.0; absolute denial may be -1.0. If a consequent is established by a given rule, then the confidence of that consequent may be provided by a confidence factor associated with that rule, or may be derived

from the confidence factors of the facts which satisfied the antecedent portion of that rule. The inference engine is responsible for the derivation of confidence factors. Experts determine confidence factors associated with the rules.

Confidence factors are useful in making judgmental conclusions, taking into account different and possibly conflicting evidence. Use of confidence factors allows the expert system to make "best guess" approximations of the best choice. For example, the rule shown above is modified below to reflect a confidence level of 0.90 on the scale from -1.0 to 1.0.

```
IF the animal has wings
   AND the animal has feathers
THEN
   CONCLUDE the animal is a bird.
   CONFIDENCE FACTOR = 0.90
```

According to this rule, if both the conditions are met we can be very confident the animal is a bird.

3.3 Expert System Shells

It is important to distinguish between the tool used to build the expert system and the expert system itself. The expert system building tool includes both the language used to represent and access the knowledge contained in the system, and the support environment. These tools, used by the knowledge engineer, differ from conventional programming languages in that they provide convenient ways to represent knowledge.

Expert system shells are software tools which can be used to build expert systems. Expert system shells provide a software development environment in which knowledge engineers can develop individual expert systems. Typically, shells have facilities for representing knowledge internally (most commonly as rules), inference engines to apply the rules, a user interface to allow knowledge engineers to develop the expert system, and software facilities to develop an external user interface to the finished expert system for end-users. The shell provides much of the software required for implementation of an expert system, sparing the knowledge engineer much time and effort in writing the code.

Currently there is a wide variety of commercially available expert system shells, ranging from inexpensive micro-computer based software to sophisticated development environments

available only on Lisp machines [WATE85]. The expert system shell used for this research project is INSIGHT 2+[1], produced by Level Five Research [OXKO86].

INSIGHT 2+ is a rule-based shell using backward chaining inference technique. It operates on any compatible IBM PC/XT/AT under MS-DOS version 2.0 or later. The development language of INSIGHT 2+ consists of IF-THEN statements. Among the important features of INSIGHT 2+ for this application are its support in interfacing to database management systems dBASE II or III[2]. The interface is a PASCAL program under the control of INSIGHT 2+ inference engine. The PASCAL program contains "fetch" and "receive" statements in order to access the dBASE II or III databases.

---

[1] INSIGHT 2+ is copyrighted. Use of this product does not imply recommendation or endorsement by the National Bureau of Standards.

[2] dBASE III is copyrighted. Use of this product does not imply recommendation or endorsement by the National Bureau of Standards.

4. DESCRIPTION OF THE SYSTEM

The purpose of this system is to model a data center environment. The data center serves the public in answering queries about a specific disciplinary area, e.g., chemical thermodynamics information. The data center uses many different databases as sources of information. Some databases are kept manually in file cabinets, some data are automated on a database management system (DBMS) but privately collected and maintained, and some data can be in the form of a handbook. The data center also has access to subscription services, e.g., Chemical Abstracts Service.

4.1 System Overview

The overall architecture of the prototype "Automated Advisor" is shown in Figure 1.

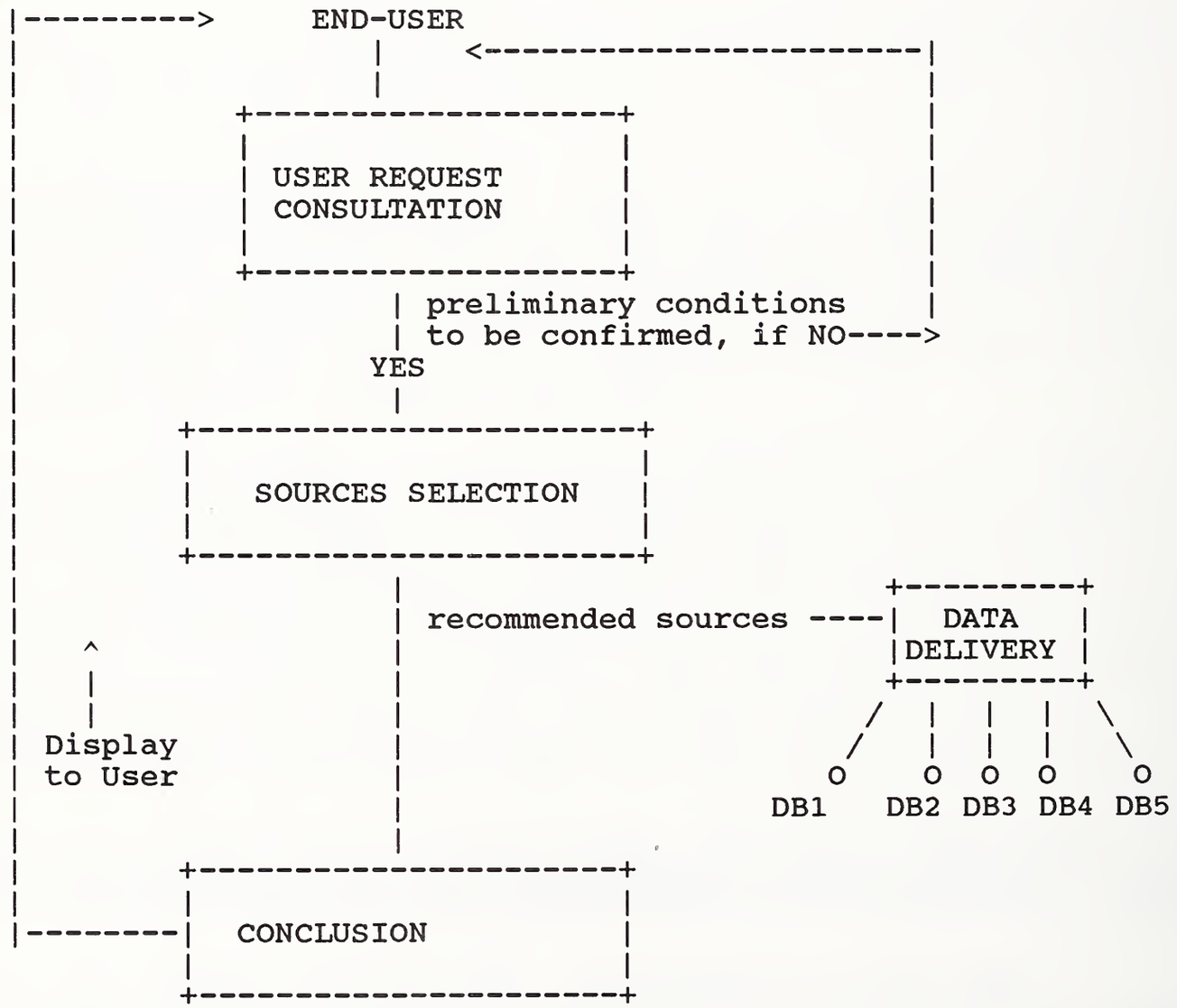


Figure 1 - Overall Architecture of Automated Advisor

The architecture assumes that there are several databases with DBMSs installed in a loosely-coupled manner. These databases are physically separate but logically integrated via a Global Data Source (GDS). Each database that participates in the Automated Advisor must have a portion of the data source information defined within the GDS.

We assume a human end-user interacting with the Automated Advisor in consultation mode. The end-user is assumed to be a subject-matter specialist who requires specific information about vapor pressure properties but does not understand the structure and content of the databases. Typically, we see the end-user as an engineer or scientist engaged in industrial research, or even a reference librarian from a technical organization.

A sample script of the dialogue between the end-user and the Automated Advisor is presented in Appendix A.

4.2 User Request Consultation Module

The User Request Consultation Module consists of a set of rules organized in a goal oriented fashion and implemented in INSIGHT 2+. The goal of this module is to determine the end-user's basic requirements for vapor pressure data so that appropriate data sources may be selected. To achieve this goal, a simplified version of the criteria and methods used by data center scientists to fulfill an end-user's inquiry is captured in rule form. The application of these rules by the inference engine results in a dialogue with the user. For examples of specific rules, see Appendix B.

The knowledge base also includes parts of the Global Data Source. By conducting a dialogue with the end-user, the User Request Consultation Module obtains a precise description of the request and generates a set of parameters (user criteria). These parameters include the name of the substance, the chemical class the substance belongs to, whether the user requires a limited number of (perhaps approximated) data points or vapor pressure equations, including equations of derivatives, and the type of source desired (written publication, user subscription service, etc). The User Request Consultation Module then presents the preliminary criteria to the user for verification. If the end-user is not satisfied, then the system repeats the dialogue. The list of criteria is then transferred to another software module called Sources Selection Module.

4.3 Sources Selection Module

The Sources Selection Module consists of several submodules which may be invoked from the User Request Consultation Module. As the name implies, the purpose of these submodules is to select data

sources which satisfy the end-user's requirements. Each submodule is a rule set implemented in INSIGHT 2+. Individual submodules are associated with particular data source types and chemical classes. For instance, there are separate submodules for inorganic written sources, for on-line subscription service data sources, and for data sources available in tape form. Each module asks the user further questions about the end-user's problem and about the intended use for the vapor pressure data in order to determine additional criteria for data source selection. Rules for selection of data sources are intended to:

* discriminate between potential data sources allowing for a finer determination of the applicability of a data source.

* rank the recommended sources based on the applicability to the user's problem.

* generate pieces of "advice" and "cautions" on the use of each source being recommended.

The list of selected data sources and confidence factors is passed to the Data Delivery Module.

4.4 Data Delivery Module

The Data Delivery Module is a Pascal program which accepts the list of selected sources and their confidence factors from the Sources Selection Module and locates the corresponding citations in the distributed DBMSs. The retrieved data sources are sorted in descending order of confidence factor values. The end-user has the option to either view the best sources, i.e. a limited number of sources with high confidence values, or to view all of the selected sources. Control is then returned to the expert system to display the conclusion screen.

4.5 The Conclusion Module

The conclusion module is the final display by the INSIGHT 2+ program. The list of recommended sources and all of the related advice is displayed to the end-user. This ends the session dialogue for this particular chemical substance. The user may choose to exit or start another session.

4.6 Databases

The data that are accessed by the prototype consist of the following dBASE III databases:

* Chemical Thermodynamics Data Center's bibliographic data on organic and inorganic chemical compounds.

* Office of Standard Reference Data Center publication lists and database tapes list.

* Interactive database services on chemical thermodynamic properties.

* Chemical information services which are available through subscription services.

* Other bibliographies or handbooks that are typically available in technical libraries.

The structures of these databases are slightly different using different data element names. The data definitions of the five databases are presented in Appendix C.

4.7 Global Data Source

The Global Data Source (GDS) consists of the union of all the data sources of the five separate databases plus some global attributes. The global attributes are additional criteria which are needed to discriminate between sources. The data definition of the GDS is presented in Appendix D. This GDS is implemented as another database on dBASE III. In the current implementation, the GDS is used only for representing descriptive information about each data source, and is not used to select data sources. Instead, the selection criteria of the GDS are incorporated in rules within the expert system.

5. KNOWLEDGE ENGINEERING PROCESS

The process of building an expert system is called "knowledge engineering." Knowledge engineering addresses the problem of building a computer system, aiming first at extracting the expert's knowledge and then at organizing it in an effective implementation. The procedure of extracting knowledge from an expert and encoding it in program form is called "knowledge acquisition" [FREI85, KAHN85]. This transfer and transformation of problem-solving expertise from a knowledge source to a program is the heart of the expert system development process.

The knowledge engineering process for this project consists of:

* Initial problem definition.
* Project team organization.
* Preliminary knowledge acquisition.
* Selection of the expert system shell.
* Creation of segments of the system to be verified by the domain expert.
* Acquisition of further knowledge and expansion of the knowledge base.
* Intermediate verification.
* Design and implementation of the user interface with users' input.
* Evaluation of the prototype.

5.1 Initial Problem Definition

The first step in formulating an application for an expert system is to characterize the problem and determine that the problem is appropriate to the use of expert system technology. In our case, we believe that building an automated advisor to select data sources from chemical information databases is a proper expert system application because the knowledge involved is vast and expert opinion needs to be applied in giving a "best" solution.

5.2 Project Team Organization

Before the knowledge acquisition process can begin, the participants must be selected and their roles defined. We used a single domain expert, the Group Leader of the Chemical Thermodynamics Data Center. While multiple domain experts could be used in a production system, this was not done in our research

prototype. Use of multiple domain experts might have resulted in different "expert's opinions" which would have introduced a complicating factor that was felt unnecessary in our research prototype.

In the initial knowledge acquisition stage, we also talked to people who might potentially be the end-users of this prototype. Interviews with these people provided statistics about the frequency of incoming queries, the percentage of queries that are simple and quickly fulfilled, and the percentage of queries that are too complex or ill-defined to be answered without further discussions.

The knowledge engineers for this project consisted of two computer scientists, one having a background in knowledge-based systems and the other having a background in database issues. We discovered during the knowledge acquisition process that there were two types of knowledge: 1) the types of data available and 2) the nature and extent of rules that underlie the human solutions.

5.3 Preliminary Knowledge Acquisition

The objective of this initial phase was to identify and understand the basic problem of selecting chemical data sources for external users. This was accomplished through discussions with people from the NBS Chemical Thermodynamics Data Center. Second, it was necessary to become familiar, in general, with chemical terminology, and, in particular, with the terminology associated with the taxonomy of chemical substances. This familiarity was necessary to understand the domain and to conduct knowledge acquisition sessions with domain experts. Third, by gaining an understanding of the basic problem to be solved and by obtaining familiarity with chemical terminology, we hoped to gain some understanding of important issues and potential difficulties which might be encountered in developing expert system software to solve the problem.

Among the problems and issues immediately obvious to us were: the sheer size of the domain (over 6,000,000 compounds), the necessity of having a good user interface to conduct a dialogue with the user, and the requirement to provide fast response time so that users would not have to wait long periods for questions during the dialogue. But perhaps most important was the need to be able to construct a prototype quickly so that a system could be demonstrated, thereby facilitating knowledge acquisition, further understanding of the domain, and system development.

The preliminary knowledge acquisition phase was also necessary to provide a basis for making decisions on selection of an expert system shell.

5.4 Selection of the Expert System Shell

Picking the right tool for building the expert system is an important but difficult decision. While none of the tools reviewed may be perfect for a given task, there may be a number of tools that will perform equally well.

Our goals in selecting a shell included the following:

* to gain experience with a member of the genre of inexpensive microcomputer-based expert system shells that are commercially available and cost less than $500.00.

* to gauge the shell's capabilities, and to select a shell which provides a rapid prototyping capability so that a small version of a complete expert system could be quickly constructed for demonstration.

* to permit automatic coupling into a database management system which would store actual citations.

We chose INSIGHT 2+ because it is inexpensive, it has an interface to the dBASE III DBMS, and it has "easy to use" characteristics which make rapid prototyping possible.

5.5 Small Example and Verification

In this phase, the objective was to focus on a small segment of the chemical thermodynamics domain as a target area for the prototype. After talking with the domain expert, data on vapor pressure properties was selected as an appropriate topic, because:

* The vapor pressure "sub-area" is of moderate size, i.e., not too large for prototyping purposes but big enough and complex enough to demonstrate the advantages of using the expert system approach.

* Preferably, the "sub-area" should be of importance to both experts and users so that it will generate interest. Vapor pressure data has this characteristic.

* The area chosen should be typical of the rest of the domain. Atypical areas should be avoided.

* If a rule-based system is contemplated, the "sub-area" knowledge should be expressible in no more than 200 rules for purposes of the prototype. We felt a useful prototype of this size could be created.

Next we conducted preliminary knowledge acquisition sessions and created a small sample knowledge base. The purpose of this was to determine the nature of dialogue between the domain expert and the end-user and to provide sample rules containing domain knowledge for the expert to review. These sessions consisted of extensive interviews in which the domain expert was asked detailed questions to determine the problem solving techniques and to obtain examples of problem solutions. The sessions were tape recorded.

We concentrated on one very small area within the vapor pressure domain. By concentrating on the dialogue which occurs between the inquirer and the expert, we were able to identify several important data sources as well as criteria used to differentiate and select sources.

Using this information, a small rule set (about 20 rules) was created. The information in the rule set was presented to the domain expert for verification with emphasis on the type of information contained in the rules, and types of conclusions reached by the rules. The purpose was to ascertain if the rules we constructed captured the expertise.

The rules were then revised to take into account the expert's comments. We then proceeded to create a small knowledge base consisting entirely of rules using INSIGHT 2+. This system conducted a short dialogue with a user and selected data sources within the limited domain area chosen.

The initial knowledge acquisition sessions gave us an understanding of the organization of vapor pressure data domain and identified the major classes of data sources. These major classes included written sources (books and journals) available in chemical libraries, compilations of data stored on tape and available for a fee, data sources available exclusively within the Office of Standard Reference Data (OSRD), dial-up subscription services, and micro-computer DBMS available for a cost. These classes provided the basis for the architecture of the prototype and for the databases of citations. At this point, five separate dBASE III databases were defined. Those data sources deemed relevant were entered into the databases.

We also acquired a basic understanding of the process by which domain experts select sources for users. As indicated above, this process is reflected in the dialogue conducted with the user. The expert uses a number of criteria, including knowledge

about the user's application for the vapor pressure data, to gain an understanding of the user's requirements and help select the sources which best satisfy the user request. The expert may immediately know an appropriate data source or may consult one or more sources to determine if they have the required information on the substance in question. If sources are not directly available to the expert, he/she may recommend sources in other collections known to have the desired information for the substance or chemical class to which the substance belongs.

In many cases, the problem is not as simple as the mere identification of one or more sources for a user. The user must also be provided with advice on how to use the source. Often, the chemical data is not explicitly provided in a source but must be derived through algebraic manipulations. In addition, the quality of the data contained in a source determines how the source should be used. Quality is based on a number of factors including the method by which the data is obtained, method of evaluation of data, prior sources used which the work is based on, if any, and the purity of the sample. Knowledge about the quality and characteristics of each data source is vital to making reliable recommendations to end users.

5.6 Acquisition of Further Knowledge and Expansion of the KB

Once a basic understanding of the dialogue was obtained and the format and content of the rules were clarified, we proceeded to divide the vapor pressure domain into several areas and to conduct knowledge acquisition for each area.

Initially, the domain was partitioned on the basis of the major classes of data sources. However, as more knowledge was acquired, it became clear that this partition would result in a disproportionately high number of rules concerning written sources. It was therefore necessary to further subdivide knowledge about written sources on the basis of chemical taxonomy. Sources for organic and inorganic substances were grouped separately with further subdivisions within each class. Furthermore, it became clear that for many substances only a single or very small group of data sources were available. In these cases it was expedient to identify these sources immediately without proceeding though the entire dialogue. The number of such exceptions quickly grew and ultimately became regarded as a separate subdivision within the domain.

During this phase the prototype grew from the initial size of about 15 rules to about 150. We concentrated on incorporating into the databases information about a large number of written sources concerning inorganic materials as well as sources from among the other major classes of data sources. The total database grew from a handful to about 60 records.

5.7 Intermediate Verification

As the prototype grew, the primary domain expert as well as other persons from OSRD were invited to test and review the system. During this period, the method of knowledge acquisition changed from conducting interviews using a tape recorder to having the expert operate the system and suggest additions and modifications to the knowledge base directly.

We noticed that having the domain expert operate the system proved to be an effective means to accurately build and refine the knowledge base. It became apparent that ultimately it would be desirable to be able to allow the domain expert to develop the knowledge base without the aid of knowledge engineers. However, this would require a dynamic data management coupling capability with the expert system shell. This capability is unavailable in the expert system shell we selected for the prototype.

5.8 Development of the User Interface

Reviewing the system with domain specialists as well as potential users also resulted in improvements to the user interface. The review determined if the questions were correctly and clearly stated. We also focused on whether the flow of questions was logically sequenced, and whether the messages and instructions from the system were understandable to the end-users without any verbal help. The explanation facilities contained in INSIGHT 2+ proved particularly useful in providing explanatory messages to assist the user in understanding the dialogue.

5.9 Evaluation of the Prototype

Currently, evaluations of an expert system are most often done by comparison with human performance. However, this raises the issue of whether a "correct solution" to a user question is one that a human expert would give, or one that represents the ideal solution based on the formal rules established by the domain experts. Among the reasons for a difference in solutions could be the possibility that the human did not apply all known rules in order to reach his or her solution, or the possibility that the expert system had an incomplete set of rules from which to derive its solution. The types of reasons that cause a difference in solutions make it difficult to arrive at a quantified methodology for evaluation that would determine whether or not an expert system is providing "correct solutions." At this time no one knows how to fully evaluate human expertise adequately, let alone how to evaluate an expert system that is attempting to recreate that expertise, and may, in fact, actually be an improvement on that expertise.

Although the demonstrated prototype does what it is supposed to do, it is not likely to be fielded as a production system. This

is because the expert system shell used does not have a sophisticated pattern matching capability. This severely limits the representation of complex chemical compounds. However, we found that when the substances selected were restricted to elements alone, the expert system performed much better.

The prototype system demonstrates that an expert system can be used to effectively identify and interpret requests for chemical thermodynamic information and to recommend appropriate data sources. The expertise of a human specialist can be incorporated in rules and applied during user consultation sessions to provide useful answers. According to one member of the Chemical Thermodynamics Division, the prototype, despite the limited amount of knowledge it contained, is capable of screening 50% of the simple user requests received. We expect that an expanded full scale version would increase the number of cases handled, thus providing a reliable automated capability for identification and selection of chemical data sources.

The prototype provides the ranking of data sources based upon the expert knowledge of the quality of sources and provides advice about each source. Therefore, it goes beyond the capability of an ordinary bibliographic retrieval system.

6. CONCLUDING REMARKS

The conclusions drawn from the prototype implementation of the expert system can be evaluated from different perspectives: the experience gained from the knowledge engineering process, the characteristics of the problem domain, the choice of the shell, and the assessment of the finished system.

6.1 Knowledge Engineering Process

Experience gained in knowledge engineering proved to be very valuable. The project started in January 1987 with two persons working about 50% of the time. By June 1987, we completed the initial knowledge acquisition phase. A summer student was added to the implementation team to develop the expert system to DBMS interface program. By mid July 1987, we gave our first demonstration to our domain expert. Improvements in the sequence of dialogue and other refinements were incorporated, and a second demonstration to the domain expert occurred in mid August 1987. By that time, we were ready to invite users at NBS (CTDC Division) to review the system with emphasis on the design of the user interface.

The most difficult and time consuming task of the whole knowledge engineering process was knowledge acquisition. It took almost four meetings before a simple case could be articulated with both sides understanding each other's terminology.

6.2 The Characteristics of the Problem Domain

The domain which was selected concerns the solution of a problem for which knowledge is subjective, ill-codified and judgmental. Providing expert opinion and recommendations on data sources is the main function of this prototype. However, within the domain of thermodynamic chemical properties, we encountered many difficulties including the following:

* The vast amount of chemical substances (over six million) are organized in a hierarchy of chemical classes. A full understanding of this hierarchy and its terminology involved many time-consuming knowledge acquisition sessions.

* The large, diverse set of data sources is sometimes unclear, uncertain and requires interpretation. The problem is compounded not only by the constantly growing list of sources but also because the correspondence of sources to chemical classes is not exact.

* Expert opinion on data sources changes with time as better data sources appear and research establishes more reliable data.

* Within the population of potential end-users of this expert system, the backgrounds of individual end-users can be quite diverse. Scientists and engineers who require information on vapor pressure data in research and industry may find that certain questions asked by the prototype system have obvious answers. Yet, the same questions may be necessary to conduct an "expert" dialogue with a reference librarian.

6.3 The Choice of the Shell

The choice of an expert system shell was based on the criteria that it run on micros, be inexpensive, and have an automatic interface to a DBMS. Our underlying goal was to have an easy to learn, easy to use, flexible software system which could be used to construct and modify a prototype expert system quickly, without exhausting financial resources on more expensive software and hardware.

Our choice, INSIGHT 2+, fulfilled these criteria but did not provide all the functionality required to construct a full scale expert system. Two features which were needed for this application but were lacking including:

* A pattern matching capability with variable substitution is required for representation of complex chemical compound names. This capability is necessary for creation of generalized rules which can be used for making inferences about entire classes of substances instead of having individual rules for each substance.

* The interface to the DBMS (dBASE III) in INSIGHT 2+ is through a PASCAL program. The PASCAL program calls the dBASE databases using the database name and the record number. Therefore, the retrieval condition must be coded in PASCAL rather than formulated as a conditional query in the dBASE query language. This type of coupling of an expert system with a DBMS is referred by Jarke and Vassiliou in [JARK83] as "Loose Coupling" which means that the communication channel between the two systems occur

statically as opposed to "Tight Coupling" where access to the database could occur on-the-fly dynamically during the same session.

6.4 The Finished System

As the prototype grew, our conception of what the knowledge representation system should be also changed. Initially, we incorporated decisions on selection of specific sources directly into the rules. This meant each rule was based on specific criteria for selection of a particular source, together with advice on how the source was to be used. The databases installed within this prototype were based on a limited schema holding only minimal bibliographic information about each source together with some limited explanatory comments. Since there were many possible criteria, the result was the creation of a large number of rules necessary to cover different combinations of criteria alternatives.

We discussed an alternative approach in which we would use the global data source (GDS) for direct selection of data sources. Currently, this function is performed by a large number of rules. The GDS is a database containing information on the data sources in the five individual databases plus additional attributes about each data source. These attributes can be used in selection conditions in queries for data sources fulfilling various criteria. A smaller set of rules would be used to formulate the queries and to determine the selection conditions for queries to the GDS. The resulting retrievals would yield the selected sources. Separate rules would exist to make more discriminating judgments on the selected data sources, assign confidence factors, and provide the user with advice on how the selected sources should be used. We could not implement this approach in INSIGHT 2+ but its advantages were obvious. By using the GDS for selection of data sources instead of rules, part of the task of the expert system is transferred to the database management system. This is important because much of the knowledge for data source selection can be more easily represented and updated in a database than in a collection of rules. The resulting smaller number of rules would also be more manageable and allow concentration on development of specific rules for providing valuable advice about sources.

7. REFERENCES

[CUGI87] Cugini, John V. "Programming Languages for Knowledge-Based Systems," NBS Special Publication 500-145, Feb. 1987.

[FREI85] Freiling, M. et al., "Starting a Knowledge Engineering Project: A Step-by-Step Approach," The AI Magazine, Fall 1985, pp 150-164.

[HAYE83] Hayes-Roth, F., Waterman, D.A., and Lenat, D.B. (Eds.) Building Expert Systems, Addison-Wesley Publishing Company, Inc., Reading, MA, 1983.

[KAHN85] Kahn, G., Nowlan, S., and McDermott, J., "Strategies for Knowledge Acquisition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-7, No. 5, Sept 1985.

[JARK83] Jarke, Matthias and Vassiliou, Yannis, "Coupling Expert Systems with Database Management Systems," in Proceedings of the NYU Symposium, May 1983.

[OXKO86] Oxko Corporation, "Level Five Research Insight 2+ Expert System Shell Course," 1986.

[SAUE85] Sauerwein, Joan C. and Dalton, Geraldine R., "Standard Reference Data Publications 1964-1984, NBS Special Publication 708, Dec. 1985.

[WATE85] Waterman, Donald A., A Guide to Expert Systems, Addison-Wesley Publishing Company, Inc., Reading, MA, 1985.

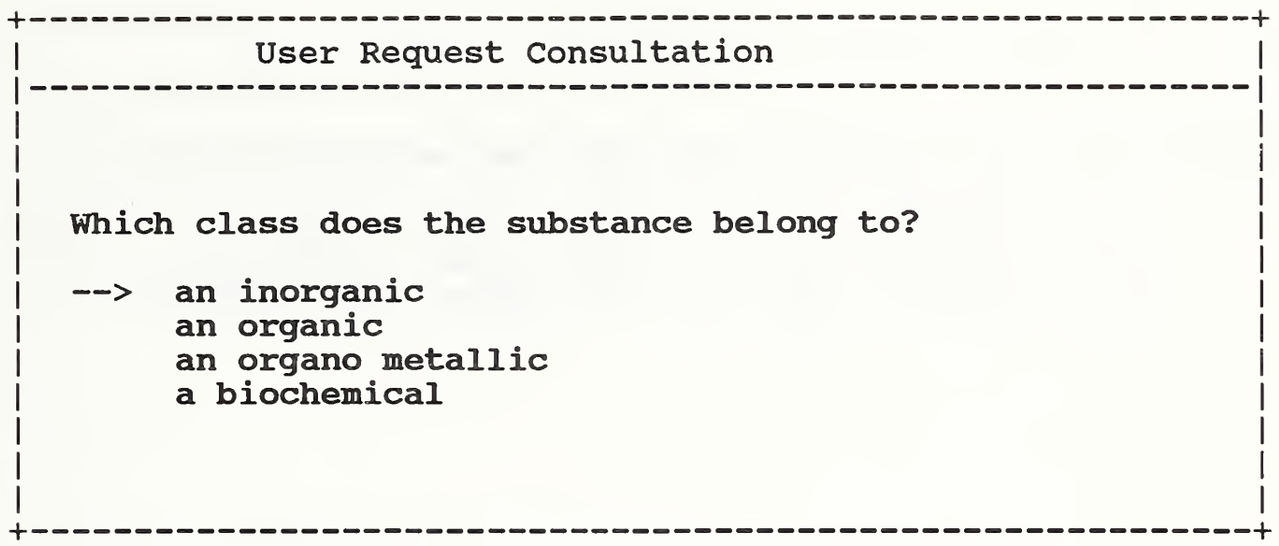
APPENDIX A - A SAMPLE SCRIPT

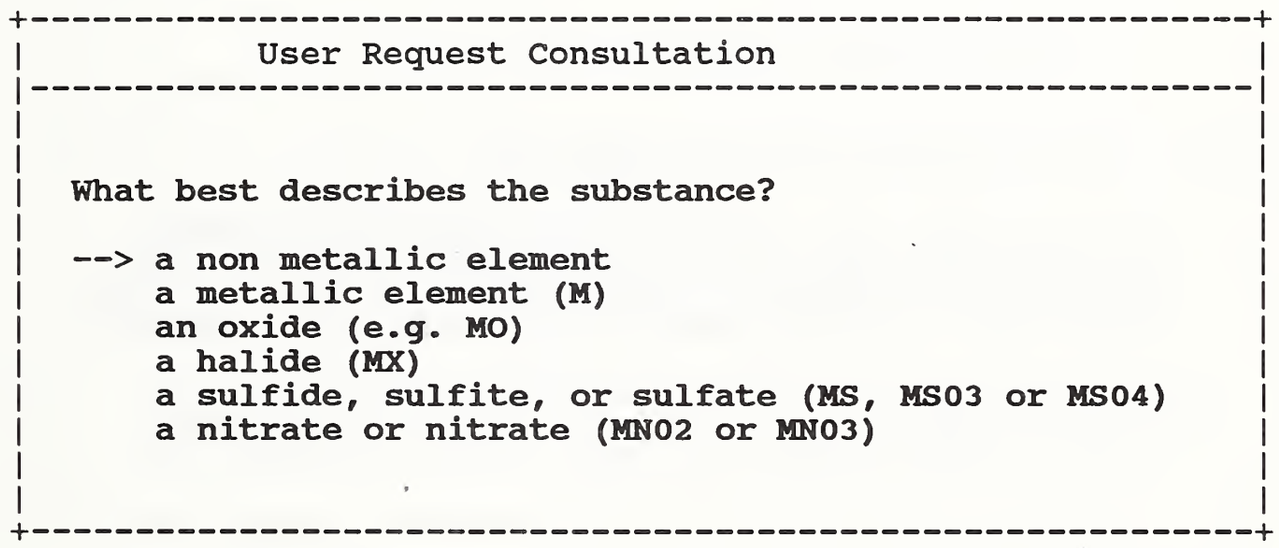
This appendix presents a sample dialogue conducted by the prototype expert system "Automated Advisor." The screen will vary depending upon the input compound name.

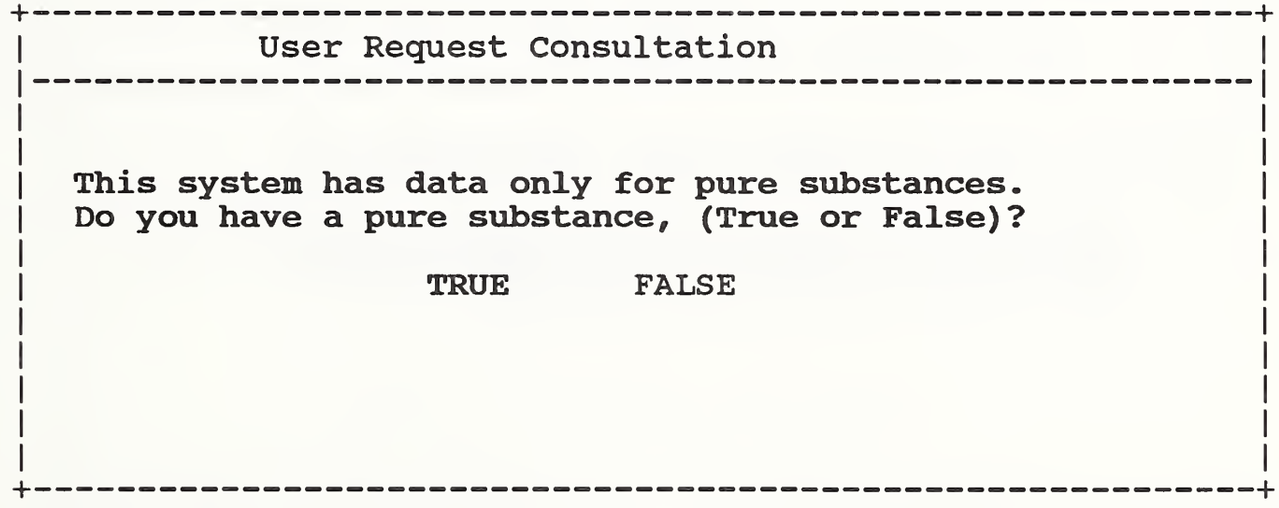
```
+-------------------------------------------------------+
|        User Request Consultation                      |
|-------------------------------------------------------|
| You are entering the AUTOMATED ADVISOR                |
|                                                       |
| This ADVISOR uses a number of distributed databases on|
| the Chemical Thermodynamics Properties:               |
|                                                       |
|                        VAPOR PRESSURE                 |
|                                                       |
| This is Dr. Mal Chase of NBS, CTDC talking to you.    |
| Please note that I am not as smart as I should be.    |
|                                                       |
| Ask me any question on Vapor Pressure and I will try to|
| recommend the "best" sources for you.                 |
| I cannot look up actual numbers for you.              |
|                                                       |
|                                                       |
+-------------------------------------------------------+
```

```
+-------------------------------------------------------+
|           User Request Consultation                   |
|-------------------------------------------------------|
|                                                       |
| (Make your selection by hitting UP or DOWN arrows,    |
|  and hit ENTER to continue.)                          |
|                                                       |
|   Are you seeking vapor pressure data for?            |
|                                                       |
|                                                       |
|  -->  a single substance                              |
|       a class of chemical compounds                   |
|       a mixture                                       |
|                                                       |
|                                                       |
+-------------------------------------------------------+
```

```
+-------------------------------------------------------+
|           User Request Consultation                   |
|-------------------------------------------------------|
|                                                       |
|                                                       |
|  What is the name of the substance?                   |
|       (enter the name in lower case.)                 |
|                                                       |
|    calcium                                            |
|                                                       |
|                                                       |
|                                                       |
|                                                       |
+-------------------------------------------------------+
```


User Request Consultation
Which class does the substance belong to?
--> an inorganic
an organic
an organo metallic
a biochemical

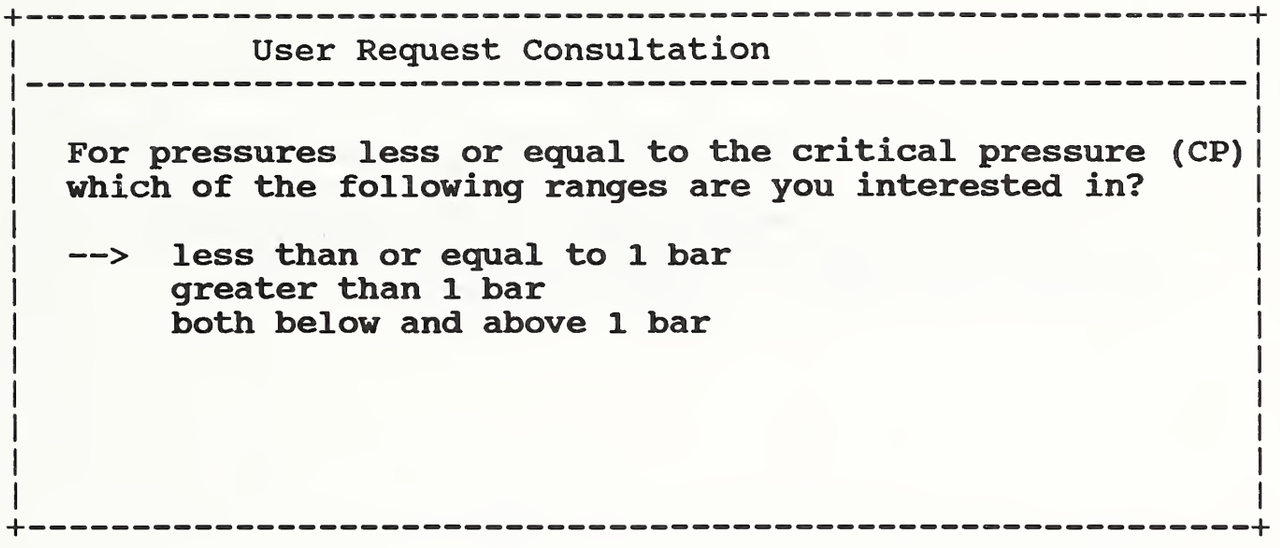

User Request Consultation
What best describes the substance?
--> a non metallic element
a metallic element (M)
an oxide (e.g. MO)
a halide (MX)
a sulfide, sulfite, or sulfate (MS, MS03 or MS04)
a nitrate or nitrate (MN02 or MN03)

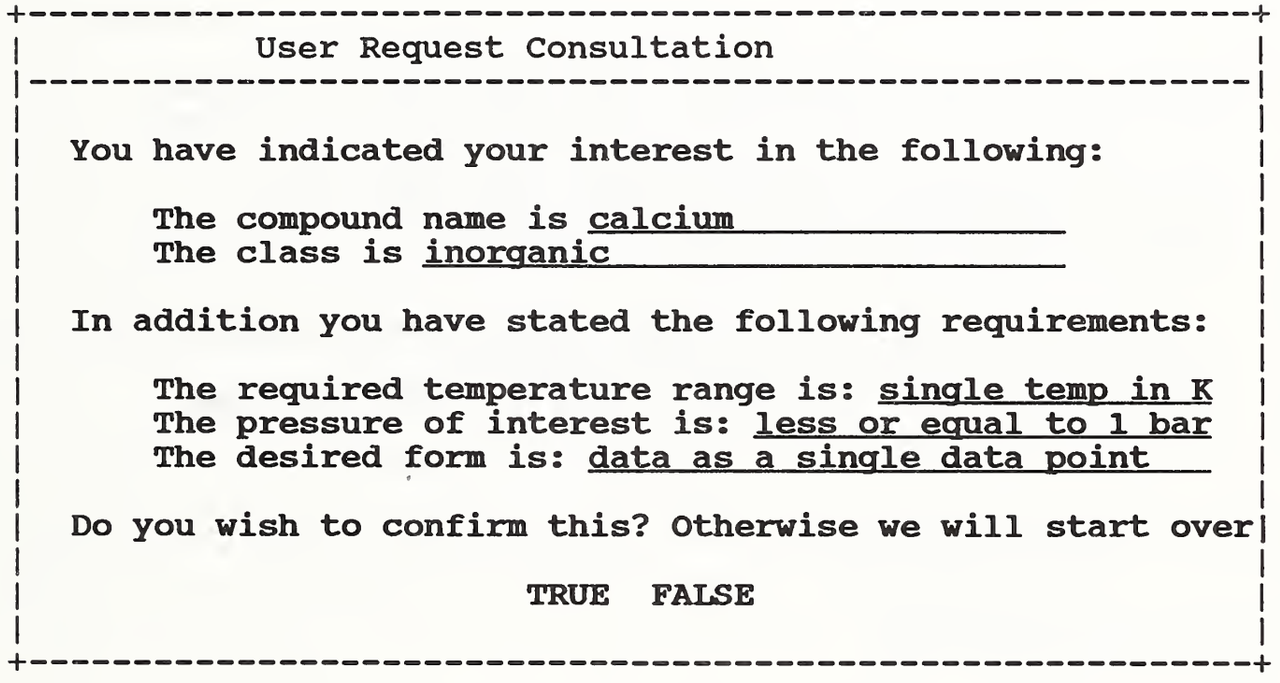

User Request Consultation
This system has data only for pure substances.
Do you have a pure substance, (True or False)?
TRUE
FALSE

```
+-----------------------------------------------------------+
|           User Request Consultation                       |
|-----------------------------------------------------------|
|                                                           |
|  How much vapor pressure data will you require?           |
|                                                           |
|  -->  Vapor pressure data as a single data point          |
|       Vapor pressure data in a small number of isolated pts|
|       Vapor pressure data as a continuous interval of pts |
|       Vapor pressure equation                             |
|                                                           |
|                                                           |
|                                                           |
|                                                           |
|                                                           |
+-----------------------------------------------------------+
```

```
+-----------------------------------------------------------+
|           User Request Consultation                       |
|-----------------------------------------------------------|
|                                                           |
|  What form would you like your data in?                   |
|                                                           |
|  --> Do you prefer the data reference in printed form?    |
|      in tape form (e.g. OSRD sells FORTRAN program tapes) |
|      subscription service of databases (e.g. Gmelin, etc) |
|      online dial up retrieval to a service (e.g. DCAPII)  |
|                                                           |
|                                                           |
|                                                           |
|                                                           |
|                                                           |
+-----------------------------------------------------------+
```

```
+-----------------------------------------------------------+
|           User Request Consultation                       |
|-----------------------------------------------------------|
|                                                           |
|   Which of the following are you interested in?           |
|                                                           |
|   -->  Vapor (sublimation) data in crystal region         |
|        Vapor pressure data in the liquid region           |
|        Both crystal and liquid regions                    |
|                                                           |
|                                                           |
|                                                           |
|                                                           |
|                                                           |
+-----------------------------------------------------------+
```


User Request Consultation
For pressures less or equal to the critical pressure (CP)
which of the following ranges are you interested in?
--> less than or equal to 1 bar
greater than 1 bar
both below and above 1 bar


User Request Consultation
You have indicated your interest in the following:
The compound name is calcium
The class is inorganic
In addition you have stated the following requirements:
The required temperature range is: single temp in K
The pressure of interest is: less or equal to 1 bar
The desired form is: data as a single data point
Do you wish to confirm this? Otherwise we will start over
TRUE FALSE

```
+-----------------------------------------------------------+
|                  Sources Selection                        |
|-----------------------------------------------------------|
|                                                           |
| Will the VP data be used in any of the following ways?    |
|                                                           |
| -->  In separation procedure using distillation columns   |
|      In reactive system in industrial application         |
|      None of the above                                    |
|                                                           |
|                                                           |
|                                                           |
|                                                           |
+-----------------------------------------------------------+

+-----------------------------------------------------------+
|       DATA DELIVERY                                       |
|-----------------------------------------------------------|
|                                                           |
|   I will show you the sources that have been selected.    |
|   The sources will be ranged by confidence factor (CF)    |
|                                                           |
|   After the sources are displayed, you will see a         |
|   conclusion screen containing a summary of all sources   |
|   together with advice on the use.                        |
|                                                           |
|   If you would only like to see sources recommended with  |
|   a confidence of 80 or greater, type G.                  |
|   Otherwise, press any key to see all of the sources      |
|                                                           |
+-----------------------------------------------------------+

+-----------------------------------------------------------+
|          DATA DELIVERY                                    |
|-----------------------------------------------------------|
| Source information:                                       |
|                                                           |
| This is source 1 of 1.                                    |
| This source is recommended with confidence 95             |
|                                                           |
| Authors: Hultgren, Ralph et al.                           |
| Title: Selected Values of Thermo. Properties of Elements  |
| Citation: American Society for Metals, Ohio 44073         |
| Year:                                                     |
| Remark:                                                   |
|                                                           |
| Class:                                                    |
|                                                           |
| Press any key to continued, R to restart or B to move back|
|                                                           |
+-----------------------------------------------------------+
```

Conclusion
The following conclusions have been reached:
--> RECOMMENDED SOURCE: Hultgren CF = 95
Advice: Hultgren data critically evaluated from multiple source

APPENDIX B - EXAMPLES OF RULES IN INSIGHT 2+

This appendix illustrates several rules from the Sources Selection Module. The first rule is used for preliminary selection of a scholarly article: Miller, R. W., "Vapor Pressure of Some Liquid and Solid Metals", IND. ENG. CHEM., 17, 34-5, CA 19 758; VP Review, 1925. The source is simply referred to as Miller in the rules below. The rule states that if the end-user wants vapor pressure equations for any of the nine elements in published form, then Miller is a preliminary (potential) selection.

The rule is followed by data quality rules. These rules make determinations on the quality of vapor pressure data required by the end-user in his/her application. For example, the first data quality rule states that if the end-user application requires small amounts of the chemical substances for reactive applications, then the end-user data quality requirement is not critical.

Finally, there are rules which make final recommendations about Miller including a confidence factor to reflect the degree of belief that this is a good recommendation and advice about how to use the source. These rules rely on conclusions about data quality requirements and usage of vapor pressure equations to make recommendations.

```
Rule preliminary select Miller1
IF class = inorganic
AND compoundname = cadmium
OR compoundname = manganese
OR compoundname = magnesium
OR compoundname = aluminum
OR compoundname = gold
OR compoundname = copper
OR compoundname = iron
OR compoundname = cobalt
OR compoundname - nickel
OR form of data = equations
AND preferred way of acquiring data IS a publication
THEN preliminary IS Miller

RULE data quality 1
IF vapor pressure usage IS in a reactive system
AND substance amount IS For storage and handling of small amounts
THEN required data quality IS not critical

RULE data quality 2
IF substance amount IS For storage and handling of large amounts
OR substance amount IS An unknown quantity of the substance
AND NOT vapor pressure usage IS in a separation procedure
THEN required data quality IS important

RULE data quality 3
IF vapor pressure usage IS in a separation procedure
THEN required data quality IS critical

RULE final select Miller 1
IF preliminary IS Miller
AND required data quality IS important
OR required data quality IS critical
THEN Source IS Miller CF 100
AND Advice IS Miller is an excellent source for substance in
question

RULE final select Miller 2
IF preliminary IS Miller
AND required data quality IS not critical
THEN Source IS Miller CF 50
```

APPENDIX C - DATA DEFINITIONS OF THE FIVE DATABASES

This Appendix presents the data definitions of the five databases in dBASE III. The names of these databases and brief descriptions are also presented.

CTDC - A bibliographic database of organic and inorganic chemical reference materials privately collected by the Chemical Thermodynamics Data Center.

```
. use ctdc
. display structure
Structure for database:  C:ctdc.dbf
Number of data records:     33
Date of last update    :  01/09/87
Field   Field name       Type          Width         Dec
    1   RECNUM           Numeric           4
    2   AUTHORS          Character        50
    3   TITLE            Character        70
    4   CITATION         Character        70
    5   YEAR             Numeric           4
    6   REMARK           Character       254
    7   CLASS            Character        40
```

OSRD - A bibliographic database of publications of pure chemical substances that are distributed by OSRD.

```
. use osrd
. display structure
Structure for database:  C:osrd.dbf
Number of data records:     13
Date of last update    :  01/09/87
Field   Field name       Type          Width         Dec
    1   RECNUM           Numeric           4
    2   AUTHORS          Character        50
    3   TITLE            Character        70
    4   CITATION         Character        70
    5   YEAR             Numeric           4
    6   REMARK           Character       254
    7   CLASS            Character        40
```

ONLINE - A database of interactive on-line retrieval services on thermodynamic properties of pure chemical substances.

```
. use online
. display structure
Structure for database:  C:online.dbf
Number of data records:        4
Date of last update    :  01/09/87
Field   Field name       Type            Width         Dec
    1   RECNUM           Numeric             4
    2   SUPPLIER         Character          50
    3   TITLE            Character          70
    4   CITATION         Character          70
    5   YEAR             Numeric             4
    6   REMARK           Character         254
    7   CLASS            Character          40
    8   COST             Character          40
```

SUBS - A database of all the subscription services on chemical literature.

```
. use subs
. display structure
Structure for database:  C:subs.dbf
Number of data records:        3
Date of last update    :  01/09/87
Field   Field name       Type            Width         Dec
    1   RECNUM           Numeric             4
    2   SUPPLIER         Character          50
    3   TITLE            Character          70
    4   CITATION         Character          70
    5   YEAR             Numeric             4
    6   REMARK           Character         254
    7   CLASS            Character          40
```

LIBS - A bibliographic database of chemical literature that is available in big institutional libraries.

```
. use lib
. display structure
Structure for database:  C:lib.dbf
Number of data records:      0
Date of last update    :  01/09/87
Field   Field name       Type            Width          Dec
    1   RECNUM           Numeric             4
    2   AUTHORS          Character          50
    3   TITLE            Character          70
    4   CITATION         Character          70
    5   YEAR             Numeric             4
    6   REMARK           Character         254
    7   LOCATION         Character          40
```

APPENDIX D - DATA DEFINITION OF GLOBAL DATA SOURCES

This appendix presents the data definition of the global data sources which are represented in dBASE III.

GDS - A database of total data sources used for this prototype Automated Advisor system. This database is used for collecting data sources during the knowledge acquisition phase and are incorporated as rules within the INSIGHT 2+ program.

```
. use gds
. display structure
Structure for database:  C:gds.dbf
Number of data records:      54
Date of last update    :  01/09/87
Field   Field name       Type            Width          Dec
    1   RECNUM           Numeric             4
    2   NODE             Character          10
    3   AUTHORS          Character          50
    4   TITLE            Character          70
    5   CITATION         Character          70
    6   YEAR             Numeric             4
    7   REMARK           Character         254
    8   FORM             Character          20
    9   TEMPRANGE        Character          20
   10   EVALUATED        Character          20
   11   CLASS            Character          40
   12   OTHERINFO        Character          60
   13   AVAILABLE        Character          60
   14   ADVICE           Character          60
   15   PRANGE           Character          20
   16   LOCATION         Character          20
   17   COST             Character          10
```

NBS-114A (REV. 2-8C)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET (See instructions)	1. PUBLICATION OR REPORT NO. NBSIR 88-3689	2. Performing Organ. Report No.	3. Publication Date January 1988

4. TITLE AND SUBTITLE

A Prototype Expert System: An Automated Advisor to Select Data Sources from Chemical Information Databases

5. AUTHOR(S)

Elizabeth N. Fong and Christopher E. Dabrowski

6. PERFORMING ORGANIZATION (*If joint or other than NBS, see instructions*)

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract/Grant No.

8. Type of Report & Period Covered

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (*Street, City, State, ZIP*)

10. SUPPLEMENTARY NOTES

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT (*A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here*)

A prototype expert system, called "Automated Advisor", was built as a part of a competency project within the Institute for Computer Sciences and Technology. The system conducts dialogue with the end-users and recommends a list of data sources from chemical information databases.

This report describes the problem domain and documents the knowledge engineering process.

12. KEY WORDS (*Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons*)

database management system, expert systems, knowledge acquisition, knowledge engineering, knowledge-based system, vapor pressure

13. AVAILABILITY

☒ Unlimited

☐ For Official Distribution. Do Not Release to NTIS

☐ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

☒ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES

47

15. Price

$11.95

USCOMM-DC 6043-P80