(12) Poole, S. C.; Bohman, V. R.; Young, J. A. Review of selenium in soils, plants, and animals in Nevada. *Great Basin Nat.* **1989**, *49* (2), 201-213.
(13) Callahan, M. A.; Slimak, M. W.; Gabel, N. W.; May, I. P.; Fowler, C. F.; Freed, J. R.; Jennings, P.; Durfee, R. L.; Whitmore, F. C.; Maestri, B.; Mabey, W. R.; Holt, B. R.; Gould, C. *Water-Related Environmental Fate of 129 Priority Pollutants: Volume I*; United States Environmental Protection Agency: Washington, DC, Dec 1979; EPA-44014-79-029a.
(14) Lutrick, M. C.; Robertson, W. K.; Cornell, J. A. Heavy applications of liquid-digested sludge on three ultisols: II Effects on mineral uptake and crop yield. *J. Environ. Qual.* **1982**, *11* (2), 283-287.

An Expert System for Analytical Data Management

J. R. LEE and T. L. ISENHOUR*

Department of Chemistry, Kansas State University, Manhattan, Kansas 66506

J. C. MARSHALL

Department of Chemistry, Saint Olaf College, Northfield, Minnesota 55507

Received October 15, 1991

Two universally important factors in research are domain expertise and data management ability. This paper describes a novel combination of a relational database management system with a rule-based expert system for research and quality control laboratories. An important advantage of this system is that it provides a user-friendly way to incorporate both domain expertise and data management abilities into a laboratory data management system. This paper presents a general strategy for archiving knowledge and data in a form immediately useful as a resource for an expert system. This strategy should facilitate the sharing of procedures among laboratories.

INTRODUCTION

Every year, there are a large number of papers concerning methods, apparatus, new techniques, and data analysis strategies from both quality control and research laboratories. Frequently, very similar work is reported by different laboratories and published in different journals. This paper will examine the use of analytical chemical techniques and expert database strategies to construct an expert analytical system. This system uses a relational data model to organize, store, and retrieve the data in such a way that it is readily and conveniently available to laboratory scientists. A novel feature of this system is the use of artificial intelligence (AI) techniques to convert the data and procedural information into a knowledge base that can be processed directly by an expert system. These techniques should help eliminate redundant work and lead to a greater standardization of methods among laboratories.

This paper will outline a knowledge base strategy that will be used to build an expert analytical system to integrate experimental data and published methods. This expert analytical system will generate and evaluate the data and knowledge base rules relating to a given procedure and then propose the best method for the analysis. With the laboratory environment standardized, the best conditions for an analysis from one laboratory can then be transferred to other laboratories.

This paper will demonstrate how this expert analytical system works by describing its application to the analysis of wastewater for soluble toxic metal ions using extraction followed by spectrophotometric analysis. Although the introduction of new instruments and new analytical methods in water analysis have had a significant impact, extraction is still a widely-used separation technique.

SYSTEM IMPLEMENTATION

System Domain. The purpose of this paper is to combine expert database strategies with traditional chemical techniques to create an expert system as an aid for the analytical research scientist. The analysis of toxic metal ions in water and wastewater is the general problem domain we have chosen.[1] This domain contains a great many standard methods frequently making it necessary for the system proposed to aid in the selection of an appropriate method from a list of alternatives.

Table I. Extraction Separation of Chelate Compounds

reagent	solvent	acidity	element extracted
DDTA	chloroform	5 N HCl	Pt(II), Hg, W, Cu
cupferron	ethyl acetate	2 N HCl	Th, V, Ti
8-H	chloroform	pH 5	Co, Ni, Pb, Cd
acetylacetone	chloroform	pH 5	Be
dithizone	chloroform	pH 9	Cu, Co, Mn, Pb, Ni

8-H is 8-hydroxyquinoline.

Analytical Techniques. The elimination of interferences for the quantitative analysis of specific metal ions is frequently necessary. Two separation techniques were used in this work to demonstrate the utility of expert database strategies in the system described.

(1) Extraction. (*a*) *Use of Chelates.*[2] The distribution of a solute between two immiscible solvents is an equilibrium process and can be written as

$$A_{aq} \xrightarrow{K} A_{org}$$

where A_{aq} and A_{org} refer to the solute concentration in the aqueous and organic solvents, respectively. $K = [A_{org}]/[A_{aq}]$ is the distribution constant. Because the distribution constant varies widely among solutes, it is possible to separate many solutes by extraction.

The extraction of a metal ion using a chelating reagent may be described as follows:

$$M^{+n}{}_{aq} + nHL_{org} \rightarrow ML_{n,org} + nH^{+}{}_{aq}$$

The extraction constant K_e is written as

$$K_e = (ML_n)[H^+]^n/[M^{+n}](HL)^n$$

0095-2338/92/1632-0148$03.00/0

Table II. U.S. Public Health Drinking Water Standards (1962)

metals	ppm	metals	ppm
Ba	1.0	Pb	0.05
Cd	0.01	Ag	0.05
Cr	0.05	Zn	5.0
Cu	1.0		

Table III. EPA Interim Primary Drinking Water Standards

metals	ppm	metals	ppm
Ba	1.0	Pb	0.05
Cd	0.01	Hg	0.002
Cr	0.05	Ag	0.05

where parentheses indicate the concentrations in the organic phase and brackets represent the concentrations in the aqueous phase. Generally, the uncharged metal chelate will be much more soluble in organic solvents than in water, so it will transfer into the organic solvent. The distribution coefficient of the metal ion is given by:

$$D = (ML_n)/[M^{+n}] = K_e\{(HL)/[H^+]\}^n$$

$$\log D = \log K_e + n \log (HL) + n\text{pH} \quad (1)$$

The value of D can be calculated from known values of K_e, (HL), and pH.

Some versatile chelating extractants are listed in Table I.[3] Most of the chelating extractants form a colored complex with the metal, and some extractants can be used for up to 60 different metal ions. Therefore, such extraction processes not only can separate the specific metal ions but also can be used to quantitatively analyze these metal ions. Dithizone will be used in this work because all metal dithizonates (and, unfortunately, dithizone itself) absorb strongly in the visible region. Dithizone is known to react with 20 metals. Nine out of 15 toxic metal ions in natural water can be extracted with dithizone. These ions are shown in Tables II and III. Dithizone in carbon tetrachloride will be the major extractant used in this system.

(*b*) *Degree of Extraction.* The incompleteness of a single extraction, E1, can be written as

$$E1 = 1/\{1 + ((V_{org}/V_{aq})D)\}$$

where V_{org} and V_{aq} are the volumes of the organic phase and the aqueous phase, respectively.

After n extractions, the incompleteness, E_n, will become

$$E_n = 1/\{1 + ((V_{org}/V_{aq})D)\}^n \quad (2)$$

Three principle factors determine the extraction efficiency of the metal ions with an organic chelating agent. They are the distribution coefficient (D), the ratio between volumes of aqueous and organic solvents (V_{org}/V_{aq}), and the number of successive extractions (n).

(2) Masking. The purpose of a masking reagent is to react with the interfering metal ions present and block their extraction. The conditions of both metal ions and interfering metal ions may be written as follows:

$$K_e' = K_e/R_m(x) \quad (3)$$

$$R_m(x) = [M']/[M] = [M] + [MR] + [MR2] + [MR3] + .../[M]$$

$R_m(x)$ is called the side-reaction coefficient. [M'] is the concentrations of all forms of the M ion, and R refers to the masking ligand. Note that the log of the side-reaction coefficient will subsequently be referred to as $\propto$. The difference between the log of K_e and $\propto$ is the log of K_e', effective extraction constant for the metal ion in the presence of the masking agent.

By using eqs 1 and 3, a new equation concerning the effect of the masking reagent can be written:

$$\log D = \log K_e' + n \log (HL) + n\text{pH} \quad (4)$$

The efficiency of the extraction may be determined from eqs 2 and 4.

Expertise. Based on eqs 2 and 4, nine key factors determine the efficiency of the extraction. These factors are listed below:

1. type of the extractant
2. acidity of aqueous phase
3. chelating reagent concentration
4. masking agent
5. K_e value
6. number of extractions
7. type of organic solvent
8. solubility of the metal chelate
9. stability of the metal chelate

Using these nine factors as the basis for the extraction database, the database tables for the system can be constructed. For example, silver has the following properties:

a. Silver can be extracted easily from water by dithizone.

b. Dithizone can be dissolved into carbon tetrachloride or chloroform.

c. D value for dithizone/CCl_4 depends on pH value, K_e value, and the masking agents present.

d. Cu^{2+} can be masked in dilute acid with EDTA.

Based on this information, the database for Ag will include [chelate type], [solvent type], [D value], [pH, K_e, masking agent], and [number of extraction]. Square brackets indicate the attributes or tables for the database.

DATABASE IMPLEMENTATION

Even if the nine extraction factors noted above are all specified for the system, there could still be problems with the quantitative analysis for metal ions. For example, if Ag^+ and Hg^{2+} both exist at the trace level, Hg^{2+} will interfere with the extraction process for Ag^+. The interference could be worse with more than two ions present. In general, a separation designed using only those nine extraction factors will not work in most cases. However, there is much data in the literature concerning separations prior to colorimetric determination. To construct this expert analytical system, these data should be available to the system as the procedure is designed.

A database maintenance problem with redundant data might occur if all the data on the same subject are collected from different sources without being properly organized. For example, consider how many combinations of solvents, chelates, and masking reagents can be used for the extraction separation of Ag^+, Hg^{2+}, and Cu^{2+}. Even if the extraction problem is limited to two ions, the size of the data file for this particular problem will be tremendously large. Therefore, an intelligent and dynamic database has to be set up for the system.[4]

Relational Database. A relational database management system (RDBMS) is a database structured in accordance with the relational data model. A relational data model has two main functions: (1) data are stored in tables and (2) associations between tables are represented within the table data, not in overhead data structures. To apply a RDBMS in this system we have used Paradox (Borland International, Scotts Valley, CA) as the tool for database manipulation.[5,6] Using Paradox, a series of tables of the metal ions and their relationships according to the nine extraction/separation factors can be created. The user can create, access, and update the data in these tables using the facilities of the Paradox RDBMS facilities.

In RDBMS two interface subsystems, Structure Query Language (SQL) and Query by Example (QBE), are commonly used to enable the user to access the data. These two

interface systems have similar functions, with the QBE being the more graphic of the two. In this system, we have used the QBE system from Paradox. Users can make direct queries to the database, extract data, and display results with no programming required. Using the QBE, the user indicates what data are to be extracted from the database by completing entries in a template displayed on the screen. To utilize more fully the capabilities of Paradox, we also used the manipulation language, Paradox Application Language (PAL), to customize the QBE process.

Using the Paradox software package, and based on the nine factors noted above as the attributes, several data tables can be created. In this relational database the tables will contain chelates, extractants, masking reagents, and so on. Using this system, if we want to obtain the chelates for Ag^+ in the database, the following simple query can be used.

Query Process from SQL.

SELECT chelates, solvents, interferences
FROM chelates_table
WHERE metal_ions = Ag^+

From this query solvents, chelates, and information concerning interferences for Ag^+ can be obtained. Appendix 1 shows the same query process using the standard paradox query language (QBE). With the commands available to QBE or SQL, the database can be accessed and edited quite easily. The data can be assembled from the books, journals, and experimental data. Much of the data required would likely already be available in an analytical laboratory.

KNOWLEDGE IMPLEMENTATION

Much of the data required is only available in unprocessed form and will likely have little intrinsic meaning as originally obtained. For example, the peak heights, peak areas, and retention times that come from a chromatogram or the inflection points in an acid–base titration are typical raw data. To make these data useful, additional experimental information must be linked to the original data. For example, retention times are valuable for compound identification only when there exist tables of retention time data relevant to the conditions of the experiment. Similarly, peak area information is valuable only when relevant detector response data for the experiment are available. In sum, most raw data must be converted to domain-related information. This frequently requires that expertise be implemented in the form of additional information linked to the original data. For example, in quantitative analysis of Ag^+ and Hg^{2+}, the interference between these two ions create problems for the extraction. Back-extraction is frequently necessary for the process, and a rule for the process could be as follows:

Rule 4.
IF Metal_Ions = Ag+ **AND** Metal_Ions = Hg2+
AND Extractant = Dithizone
THEN Acid_Thiocyanide_Add
AND Ag_Back_Extraction

The rule listed above along with other related rule sets can be used to construct a knowledge base. A knowledge base is a database combined with a set of rules. Usually, the knowledge base is used to construct an expert system in some specific problem domain. In this work, we use the QBE/PAL to manipulate a knowledge base. Therefore, the knowledge base can be updated and edited like a standard database. Another part of the rules for extraction and separation of Cu^{2+}, Hg^{2+}, and Ag^+ metal ions is listed in Appendix 2.

RELATIONAL KNOWLEDGE BASE/DATABASE INTERFACE

To make access to the data/knowledge base operations more efficient, it is necessary to implement both the database and the relational knowledge into the system at the same level. To ensure that chemists can use the knowledge from the expert system, both the domain expertise and research related data must be stored and available. A computer program written in C is used as the interface to link the database and knowledge base. This program acts as the bridge between the expert system and relational database management system. Figure 1 shows this interface.

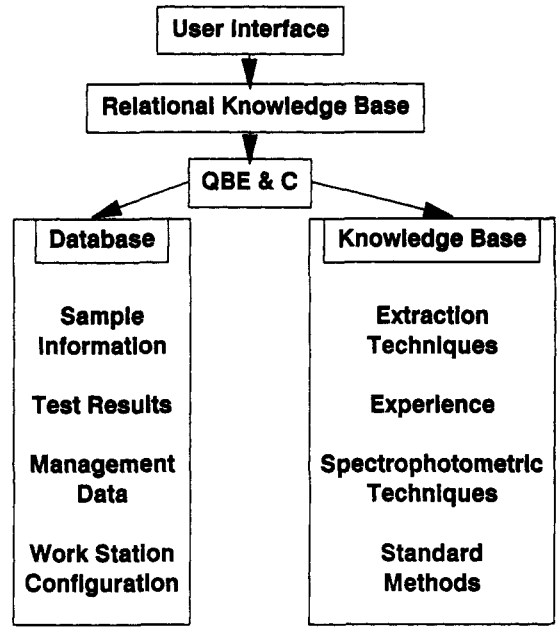

Figure 1. Hierarchy of a relational knowledge base.

SYSTEM STRATEGIES

It is generally possible to create special purpose expert systems using one of the many currently available expert system shells. An expert system shell is a software package that allows the user to create expert systems with a minimum involvement in the programming details. However, the expert system that results is frequently difficult to modify and therefore not useful for this work. In this paper we describe our own strategy for building a flexible expert analytical system. Four steps are taken to construct this system.

1. Apply ID3 algorithm.
2. Embed QBE with PAL/C.
3. Use forward chain reasoning as inference engine.
4. Provide conflict resolution to organize knowledge base.

ID3 Algorithm. The use of chemical data to build an expert system directly using the ID3 algorithm has been described by Schlieper et al.[7] The ID3 algorithm has been applied by several companies to create expert system shells.[8] If the information available can be easily divided into several yes/no categories, then this algorithm can be used to create the expert system. To simply apply the ID3 to create the expert system directly from the data will likely result in at least two problems: (1) It is difficult to check for equivalence. (2) It is not suitable as a calculation environment. The first problem is a limitation of ID3 algorithm; the second is a general limitation of expert systems. Furthermore, the lack of the ability to access the data directly and the elimination of some of the properties of the original data are also disadvantages of using ID3 algorithm.

To address these disadvantages, the ID3 algorithm for our expert analytical system is used in conjunction with other methods that can access and manipulate the original database.

Embedded QBE. A useful database must be easy to access and manipulate. To address the problems that might result from the direct conversion of the data to knowledge using the

ID3 algorithm, in this paper we embed the QBE with the PAL/C. Not only is the QBE easy to use, but it has the additional advantage of supplying RDBMS support. With the QBE, the original database can be accessed easily. Also, the use of the relational algebra expression functions of the QBE can solve calculation problems by directly manipulating the requested data from database. These specific features complement the capabilities of the ID3 algorithm. But the QBE is a user graphic subsystem which can only be controlled directly by the user. To link the expert system with the QBE we use both PAL and a C program.

Because PAL is a high-level language which can be used only under Paradox's environment, another program must be used to link both database and knowledge base. A C program which can be used to link the PAL and expert system was written to link both the database and knowledge base. The use of the C program can also take advantage of additional features of the Paradox Engine (Borland International), which provides a set of C functions to directly access Paradox utilities. The actual linkage is as follows. Depending on the initial expert system query process, conditions will be requested and then the action which responds to those conditions will be followed. Usually, the knowledge base uses IF–THEN rules to organize the knowledge. In this system, the IF–THEN formats can apply to both database and knowledge base. The condition-driven function will allow the system to access the resources it wants. The query process of the expert system will trigger a search for the right knowledge in the knowledge base or exit to PAL, which uses the QBE to access the original database.

Forward-Chain Reasoning. The knowledge base and database are the two fundamental components of the expert system. The use of the ID3 algorithm makes it possible to convert the original data to rule-formatted information. The use of the embedded QBE system allows the user to access and manipulate these data from the database. To interpret the knowledge base/database, we also need to have a control strategy to respond to the user requests. There are two different ways of reasoning. One is forward-chaining reasoning; the other is backward-chaining.[9]

Forward-chaining is also called event-driven or data-driven reasoning. Backward-chaining reasoning is called goal-driven reasoning. The choice of the type of chaining will depend on the number and sequence of the rules in the knowledge base. In this research, the situations encountered are most likely to fit the data-driven model, so forward-chaining was used.

Conflict Resolution. Usually, in the use of an expert system, several rules apply simultaneously because these rules all have conditional parts that satisfy the request of the system. The redundant activation of procedures should be prevented. To solve this rules conflict problem, several conflict resolution methods can be applied.[10] These methods are (1) rule order, (2) data order, (3) generality order, and (4) recency order. In this work, recency order was the most appropriate.

EXPERT ANALYTICAL SYSTEM IMPLEMENTATION

Based on the data retrieved from the relational database, the system will organize both knowledge and data to build an expert system. The outline of how this is done is listed below.

1. Extract parameters from chemical techniques.
2. Use nine determinants (see above) as attributes for database.
3. Retrieve the relational database.
4. Convert the data into rules.
5. Combine other expertise.
6. Classify the rules, store separately.
7. Apply the forward-chain reasoning as inference engine.
8. Test/save the expert system for later use.

In addition to laboratory automation, an analytical research laboratory should standardize data representations strategies. Presently, procedural information is difficult to transfer between laboratories because there is no recognized standard form for such information. However, once data are converted to the form of IF–THEN rules, as advocated by this work, then such data can be transferred readily and be immediately useful in the environment of another laboratory.

To implement this system, we also need to define a standard form for analytical procedures. A way to do this is to follow the sequence used by most collections of standard methods. These collections use a hierarchical structure to catalogue methods and procedures. The expert analytical system hierarchy proposed here is similar to the one used in most standard methods collections. We suggest that this system can be used to share data and procedures and to standardize method interpretation.

DEMONSTRATION CASE

To illustrate this system, two examples will be detailed below. The first example will show the traditional method of expert system query. Depending on the analytical problem, the expert system will recommend the best action based on information available to it in its knowledge base and database. To find a method to separate two metal ions, the expert system will need to ask the user to input the metal ions that are to be separated and (optionally) the extractant to be used. After the user inputs the requested information, the system will examine the IF–THEN rule base and suggest appropriate actions.

Example 1:

⟨**system query**⟩INPUT THE IONS: ⟨**user response**⟩-Hg^{2+}, Cu^{2+}

⟨**system query**⟩INPUT THE EXTRACTANT: ⟨**user response**⟩Dithizone

⟨**system response**⟩RECOMMEND:

1. Add KI to mask Hg^{2+}
2. Add EDTA and thiocyanide to mask Cu^{2+}

An expert system for this example is very easy to construct. But this simple expert system cannot solve a problem like "Can bromide be used to help separate small amount of Cu^{2+} from Hg^{2+}?" To solve this type of problem, illustrated with example 2 below, other features of this system were needed. Several procedures were taken from literature sources.[2,3,11] Appendix 3 shows these procedures in text format and their equivalent rule format. These are the generic rules for extraction and separation of the metal ions.

Example 2: To find the solution for example 2, the first step is to use QBE/PAL to obtain the necessary information from the system's database and knowledge base. This initial query process is shown below:

⟨**user query**⟩SELECT Ke − α
FROM BROMIDES,CHELATES
WHERE CHELATES.METAL = BROMIDES.METAL
AND CHELATES.METAL = 'Cu2+'

⟨**system response**⟩

METAL	Ke	CHELATE	α	Ke − α
Cu^{2+}	10.5	DITHIZONE	0.1	10.4
Hg^{2+}	26.8	DITHIZONE	21	5.8

This query consulted the database and returned the effective formation constants ($K_e' = K_e - \alpha$) for Cu^{2+} and potentially interfering metal ions in a bromide environment as shown

Table IV. Database for Expert System

metal	chelate	solvent	conc, M	K_e	interference	abs, mu
Ag^+	dithizone	CCl_4	2×10^{-4}	7.18	Cu, Pd, Hg	426
Cu^{2+}	dithizone	CCl_4	2×10^{-4}	10.5	Pd, Hg, Ag, Bi	550
Zn^{2+}	dithizone	CCl_4	2×10^{-4}	2.3	all	535
Pb^{2+}	dithizone	CCl_4	2×10^{-4}	0.04	all	520
Ni^{2+}	dithizone	CCl_4	2×10^{-4}	1.18	all	670
Hg^{2+}	dithizone	CCl_4	2×10^{-4}	26.8	Ag, Cu, Pd	492

Table V. Data Tables for Ag^+

chelate	interference	masking
dithizone	Cu^{2+}	EDTA
dithizone	Pd^{2+}	dimethyloxime
dithizone	Hg^{2+}	acid/thiocyanide

above. According to rule 1 in Appendix 3, the K_e' values of the two metal ions are compared.

⟨**rule 1**⟩ IF Dis_Coe_Diff ≥ 4.5
THEN Separable

Because the K_e' value of the target metal ion differs from other metal ions by more than 4.5, the system tentatively assumes a separation method has been found, and rules 2 and 3 of appendix 3 will be invoked.

⟨**rule 2**⟩ IF Separable
THEN Check_D AND Not_Extract_D ≤ 0.01

⟨**rule 3**⟩ IF Not_Extract_E_Value ≤ 0.004
THEN Masking_Reagent_Ok

In this process, the concentration of the chelating reagent is required. From the query process illustrated in Appendix 5, followed by an easy calculation by the system, the extraction conditions are established (pH ≤ 0.7, bromide as the masking reagent to separate Cu^{2+} and Hg^{2+}). Once the system has found a useful set of conditions, the system will trigger rule 4 and create a new rule. Appendix 5 shows the rule created for the system.

DISCUSSION

In this work, expertise concerning the extraction and spectrophotometric determination of metal ions is used to create a set of rules which function both as a knowledge base and a database for an expert system. Both the knowledge base and the database can be accessed using QBE/PAL. This system allows the user to input, modify, and retrieve all the necessary data for a specific problem domain. These retrieved data can then be used to build an expert system according to the user's need. In our problem domain, information for six metal ions are required. The tables of the required data are shown in Tables IV and V.

These tables represent a useful way to organize and store the required data. The information in these tables is derived from the traditional extraction techniques which can be used for designing separation procedures for the toxic metal ions. A problem-specific expert system can be implemented by using the keywords from the analysis problem as search attributes or search keys for the databases. The retrieved "IF.. AND..OR..THEN" rules can be used by an expert system to design an appropriate extraction procedure.

The important features of this system are as follows: (1) It provides a simple way to organize raw data along with domain expertise to construct an expert system. (2) It has the ability to handle numeric calculations including the calculation of the amount of solvent, concentration of the chelate, amount of masking reagent, and number of extractions. (3) It can generate new rules based on its functions.

In example 1, we illustrated the traditional way to query and obtain the advice from an expert system. This process provided an example showing how an expert system can be constructed and used. This process would most likely be carried out inside an expert's brain. It is the purpose of this paper to provide several strategies to allow this data/expertise conversion process to be carried out with a computer.

In example 2, we demonstrated how the expert system can be used in a numerical environment. The original data from the database as well as rules from literature resources and domain expertise were stored at the same level. All of this information can be accessed using the utilities of a relational database. In this environment, numerical calculations can be performed as required. We have used this ability to make the calculations required in the spectrophotometric determination of metal ions using diphenylthiocarbazone. The unique linkage between knowledge base and database provides a means to generate new rules for the system automatically. Once confirmed, the rules generated can be saved for later use.

The system developed is of general applicability and could be used in a wide variety of analytical environments. While examples given are for a system that gives advice on traditional separation techniques, the design of the system is independent of any particular technique. The general strategies should apply to a wide variety of techniques. It is only necessary for the user to supply the appropriate knowledge base and database.

CONCLUSION[12]

This paper presents an expert analytical system strategy that can help a user extract knowledge from data and organize that knowledge in a rule-formated knowledge base. Knowledge stored in such a rule-formated form is not only immediately useful as a target for expert system interrogation but it provides a convenient and immediately useful format for data exchange between laboratories.

APPENDIX 1. QUERY PROCESS FROM QBE FOR AG^+

Chelates_table

metal_ions	chelates	solvents	interferences
Ag^+	√	√	√

APPENDIX 2. KNOWLEDGE BASE FOR INFORMATION SYSTEM

```
RULE 1.
  IF Metal_Ions = Ag+
    AND Extractant = Dithizone
  THEN EDTA_Add AND Ag_Back_Extraction
RULE 2.
  IF Metal_Ions = Cu2+ AND Extractant = Dithizone
  THEN KI_Add
RULE 3.
  IF Metal_Ions = Hg2+ AND Extractant = Dithizone
  THEN EDTA_Add AND Thiocyanide_Add
RULE 5.
  IF Metal_Ions = Ag+ AND Metal_Ions = Cu2+
    AND Extractant = Dithizone
  THEN Two_Samples_Separated AND KI_Add
    AND EDTA_Add
    AND Ag_Back_Extraction
RULE 6.
  IF Metal_Ions = Hg2+ AND Metal_Ions = Cu2+
    AND Extractant = Dithizone
  THEN KI_Add AND EDTA_Add
```

AND Thiocyanide_Add

APPENDIX 3. DIFFERENT FORMATS FOR EXPERTISE

Expertise in Text Format.

1. After adding the masking reagents, if the difference between two metal ions distribution coefficient is larger than 4.5, then these two metal ions can be separated.
2. If these two metal ions can be separated, then check their distribution coefficients. The one needed to mask its distribution coefficient must be smaller than 0.01.
3. If the masked metal ion's extraction efficiency is smaller than 0.004, then this masking reagent can be used.
4. If this masking reagent can be used, then create the rules for the masking reagent and relevant metal ions.

Expertise in Rules Format.

1. IF Dis_Coe_Diff ≥ 4.5
   THEN Separable
2. IF Separable
   THEN Check_D AND Not_Extract_D ≤ 0.01
3. IF Not_Extract_E_Value ≤ 0.004
   THEN Masking_Reagent_Ok
4. IF Masking_Reagent_Ok
   THEN Create_Rule

APPENDIX 4. QUERY PROCESS TO FIND CONDITIONS FOR HG^{2+} AND CU^{2+} SEPARATION

⟨**user query**⟩SELECT (Ke- ∝ + 2 Log_Chelate_Conc[a] + 2[b])/2
FROM BROMIDES,CHELATES
WHERE CHELATES.METAL = BROMIDES.METAL
AND CHELATES.METAL = 'Hg2+'
⟨**system response**⟩pH = 0.7

Footnotes for the query above. (*a*) Log_Chelate_Conc stands for the log value of the concentration of the dithizone in carbon tetrachloride. (*b*) 2 is the minus log value of 0.01; the value of 0.01 is from rule 2 from Appendix 3.

APPENDIX 5. RULE CREATED FOR HG^{2+} - CU^{2+} SEPARATION BY USING BROMIDE AS MASKING REAGENT

RULE CREATED:
IF Metal_Ions = Cu2+ AND Condition = In_Hg2+
AND Extractant = Dithizone
THEN Bromide_Add AND pH ≤ 0.7

REFERENCES AND NOTES

(1) Sawyer, C. N.; McCarty, P. L. *Chemistry for Environmental Engineering*, 3rd ed.; McGraw Hill: New York, 1980; p 514.
(2) Irving, H. *The Solvent Extraction of Metal Chelates*; The Macmillan Company: New York, 1964.
(3) Chalmers, R. A. *Solvent Extraction of Metals*; Van Nostrand Reinhold: London, 1970.
(4) Keyes, J. *AI Expert* **1989**, *4* (2), 60.
(5) Kitayama, S. *DEC Professional* **1988**, *July*, 52.
(6) Naecker, P. *DEC Professional* **1989**, *May*, 58.
(7) Schlieper, W. A.; Marshall, J. C.; Isenhour, T. L. Using Analytical Data to Build Expert Systems. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 159.
(8) Hu, D. *C/C++ for Expert Systems*; MIS: Portland, 1989; p 497.
(9) Barr, A.; Feigenbaum, E. A.; Cohen, P. R. *The Handbook of Artificial Intelligence*; William Kaufmann: Los Altos, CA, 1981; Vol. 1, p 23.
(10) Barr, A.; Feigenbaum, E. A.; Cohen, P. R. *The Handbook of Artificial Intelligence*; William Kaufmann: Los Altos, CA, 1981; Vol. 1, pp 192, 197.
(11) Inczédy, J. Tyson, J., Translation Ed.; *Analytical Applications of Complex Equilibria*; John Wiley & Sons Inc.: New York, 1976.
(12) Dehne, T. *Anal. Chem.* **1990**, *62*, 565A.

Automorphism and Equivalence Classes

JOHN FIGUERAS
65 Steele Road, Victor, New York 14564

Received November 19, 1991

Previously described determinations of equivalence classes of atoms in molecules have been based upon the use of graph invariants. Because invariants do not adequately convey all of the information present in a connection table, more or less complicated methods of refinement must be applied to compensate. A method for finding equivalence classes is presented that is theoretically complete because it is based on all possible mappings of a graph onto itself. This method, usually considered to be computationally impractical, is in fact a useful approach to the determination of topological symmetry. The procedure uses preliminary processing based on a few graph ivariants and the Morgan algorithm to induce an initial partition, which markedly accelerates subsequent mappings of a graph onto itself. As new symmetries are discovered, they are used to induce new partitions, which accelerates the process even more. In recalcitrant cases, where the usual graph invariants of atom identity and bond distribution are not sufficient to induce partitioning, the determination of ring membership may be useful.

INTRODUCTION

The determination of equivalence classes for atoms in a structure is equivalent to finding mappings of a structure onto itself (automorphisms), a special case of determining graph isomorphism. This can be seen from the following argument. Two atoms x and y are defined as equivalent if and only if their interchange leaves the structure unaltered. Operationally, this may be implemented by showing that the original structure is superimposable on an exact copy when atom x in the original is overlaid on atom y in the copy (automorphism). Methods for determining isomorphism (automorphism) are discussed in a paper by Read and Corneil.[1] Their work suggests that the only theoretically sound determination of isomorphism is that based upon atom-by-atom matching using a backtracking search algorithm, generally considered a brute-force solution. To avoid this approach, previous workers have tried to define different calculable atom parameters (graph invariants) that rigorously distinguish between nonequivalent atoms. If such a graph invariant could be found, then atoms having common values of such an invariant could rigorously be assigned to the same equivalence class. Read and Corneil showed that definition of such an invariant is equivalent to determining iso-

0095-2338/92/1632-0153$03.00/0