(11) A similar unique numbering also results from seeking the minimum binary string from the matrix (Randic, M. On Canonical Numbering of Atoms in a Molecule and Graph Isomorphism. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 171-180). Although this approach is of course equally unique as a canonical representation, it is substantially longer and requires more storage space, as discussed in ref 9.
(12) Times given are for the DEC Micro VAX 3500.
(13) Davies, R. E.; Freyd, P. J. $C_{167}H_{336}$ Is The Smallest Alkane With More Realizable Isomers than the Observed Universe Has 'Particles'. *J. Chem. Ed.* **1989**, *66*, 278-281.
(14) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969; pp 213-224.
(15) Parks, C. A., Brandeis University, unpublished results.
(16) The degree of each vertex in any K_n graph is $(n - 1)$, and the number of rings is equal to $(n^2 - 3n)/2 + 1$.
(17) Hendrickson, J. B. Fragmentations and Rearrangements in Organic Synthesis. *J. Am. Chem. Soc.* **1986**, *108*, 6748-6756.
(18) Smith, D. H.; Carhart, R. E. Structural Isomerism of Mono- and Sesquiterpenoid Skeletons. *Tetrahedron* **1976**, *32*, 2513-2519.

ELDAR, a Knowledge Base System on Microcomputer for Electrolyte Solutions. The Factual Knowledge of ELDAR

JOSEF BARTHEL* and HERIBERT POPP
Institute for Physical and Theoretical Chemistry of the University of Regensburg, 8400 Regensburg, Federal Republic of Germany

Received July 11, 1990

The knowledge base system ELDAR (**EL**ectrolyte **DA**ta **R**egensburg), consisting of data base, method base, rule base, and communication manager, classifies the knowledge of electrolyte solutions into factual, algorithmic, and rule knowledge. In this paper information is given on the factual knowledge of ELDAR and the mapping of facts in Codd's relational data model with an extension of its "1st Normal Form" to repeating attributes. ELDAR offers equal user interfaces for all factual knowledge services, such as literature, data, thesaurus, module, parameter, basic data, and rule retrieval.

INTRODUCTION

At all times the properties of aqueous electrolyte solutions have been of crucial interest in biology, medicine, geology, oceanography, limnology, and various chemical and electrochemical technologies. During the last decade the high flexibility of nonaqueous electrolyte solutions for tackling technical problems has increasingly focused the attention of applied research on suitable solvents and solvent mixtures for high-energy batteries, wet capacitors, electroplating, phase-transfer catalysis, electroorganic synthesis, solar cells, thin-film procedures, coating, etc.[1,2] An almost unlimited scale of solution properties offered by mixed solvent systems permits the realization of electrolyte solutions with properties planned on the drawing board. A most important obstacle to the search of suitable electrolyte solutions for special applications is the lack of data books or data bases on electrolyte solutions. The electrolyte data are widely spread in the literature, and many problems have already been solved in fields quite different from that of an actual research. However, a time-consuming literature research often does not yield the required information. The needed data must often be interpolated or estimated from similar systems, well-known procedures to every engineer.

ELDAR was developed in 1980 to assist scientists and engineers concerned with electrolyte solutions. A data, method, and rule base and their interaction, controlled by a communication manager make up this knowledge base system, see Figure 1.

ARCHITECTURE OF ELDAR

Past experience has shown that the knowledge produced and used by scientists and engineers cannot be represented by only one of the four elementary types of knowledge representation, i.e., logic programming, production rules, semantic networks, and frames.[4] In accordance with the quite recently developed object-orientated languages,[5] ELDAR manages the knowledge on electrolyte solutions in three categories.[6]

- Factual knowledge is represented in a relational data model and managed by a data base with the basic relations quoted in Figure 1 and explained by their attributes in Table I.
- Algorithmic knowledge is represented in normalized modules and managed by a method base as statistical, mathematical and physicochemical modules.
- Rule knowledge is represented in Horn clauses and managed by a rule base.

Table I. Definition of Relations Representing the Factual Knowledge[a]

LITERATURE	(*primary key*, classification, identification, CA number, language, document type, year, volume, issue, page, authors, title, publication, descriptors, remarks)
CHEMICAL SYSTEM	(*primary key*, *number*, system TAG, components, reaction products, reaction, variable characterization)
DATA	(*primary key*, *number*, *data number*, variable values)
THESAURUS	(*name key*, name, preference name, synonyms, superterms, subterms, associative terms, formula)
MODULE	(*name*, title, category, descriptors, source, date/version, externals, connections, programming language, accuracy, memory, files, standard, theory, remarks)
EFFECT	(*name*, *number*, entry name, program type, function domain, function unit, effect terms, effect, manual, instruction example, test data)
PARAMETER	(*name*, *number*, *parameter name*, data type, domain, usage type, terms, unit)
BASIC DATA	(*primary key*, system TAG, components, data source, degree of freedom, sign changes, error type, mean weight, estimated variance, used module, integer constants, real constants, state variables, state range, variable range, basic data)
RULE	(*primary key*, title, category, descriptors, source, date/version, externals, connections, language, probability, explanation, remarks)

[a] The italic attributes build the primary key.

0095-2338/91/1631-0107$02.50/0

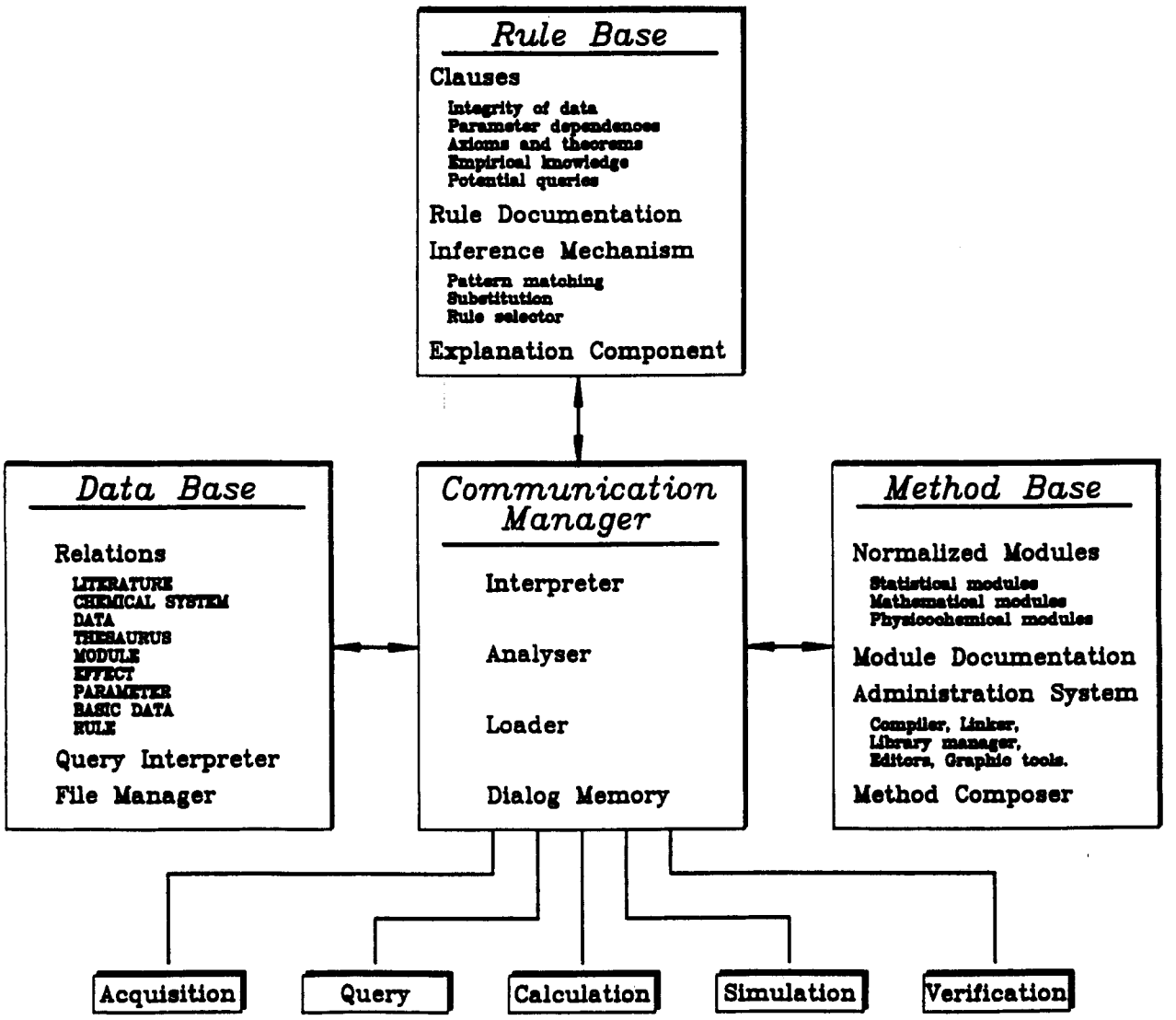

Figure 1. Architecture of the knowledge base system ELDAR.[3]

Since 1980 the *data base* of ELDAR (collection of facts = relations, query interpreter, and file manager) has been implemented on a microcomputer net. Actually about 12000 references with more than 400000 data tuples for more than 50000 chemical systems are stored in the data base and can be retrieved under various diversely formulated queries. A thesaurus with about 38000 terms in polyhierarchical order supports the user in his search for information. The data base also contains the documentation of the modules with their input/output parameters, the parameters (basic data) of those property equations which are contained in the method base, and the documentation of the rules.

The *method base*[7] consists of a collection of normalized modules (the documentation of the modules is mapped into the relations managed by the data base, see Table I), an administration system (compiler, linker, library manager, editors, graphic tools), and a method composer.

The *rule base* with a collection of Horn clauses (the documentation of the rules is mapped to the relation RULE, see Table I), inference machine, and explanation component is the object of the actual development.

A *communication manager* (interpreter, analyzer, loader, dialog memory) coordinates and regulates the wanted information of the three bases. It supplies the input parameters of the methods (method base), the predicates of the queries (data base), and the rules (rule base) with needed facts for retrieval, calculation of electrolyte solution properties, and verification of assertions. ELDAR offers equal user interfaces for all factual knowledge services, such as literature, data, thesaurus, module, parameter, basic data, and rule retrieval.

Data and method bases are running with the operating systems MS/DOS or OS/2 on microcomputer. A UNIX version is planned. The programs of the data base are coded in PASCAL, the modules of the method base in FORTRAN-77, and the rules of the rule base in PROLOG.

RELATIONAL DATA MODEL WITH REPEATING ATTRIBUTES

Different data base models have been developed in the past. Traditionally there are three concepts.[8] These are the relational, the hierarchical, and the network approach. Engineering data commonly have hierarchical structure. Therefore engineers often prefer a hierarchical data model. However, arguments can be raised against this choice, that a relational data model would lead to the independence of access paths and also to a high degree of data independence.[9] As a result of these features ELDAR is based on *Codd's relational data model*, which normalizes relations to normal forms (NF).

NF 1: every attribute is elementary
NF 2: functional dependencies
NF 3: transitive dependencies
a relation is in NF 2 and NF 3 if every attribute is either part of a key or provides a dependency from the whole key and from nothing else.
NF 4: multivalued dependencies[10]

In accordance with the results discussed on the "GI-Fachtagung 1985" on data base systems,[11] our experience shows that the theoretically well-defined relational model of Codd has weak points when applied to data bases. Since its beginning ELDAR has made some extensions to Codd's model, increasing its efficiency.

- *Expansion of Codd's NF 1 to repeating attributes.*[12] The domain of a repeating attribute is the power set of the domain of an elementary attribute. For instance, an attribute *authors*, containing all authors of a publication irrespective of their order, is a repeating attribute of the elementary attribute *author*. The algebra of repeating attributes[12] is a special case of the algebra of the non first normal form (NF^2)[13] which opens new possibilities for the relational data modeling. (NF^2) as defined by IBM Scientific Centre Heidelberg[14] permits the treatment of a complete relation as a single attribute. The use of repeating attributes makes the definition of NF 4 redundant.[12] It reduces the number of the basic relations of ELDAR from initially 29 to 9 which are quoted in Table I.
- *Definition of semantic connections by relations (THESAURUS).* The wish for an efficient representation of the semantics of data led to semantic data models.[15,16] Codd proposed to integrate the semantic demand in his relational model by the definition of special relations (E relation, P relation, PG relation, CG relation, AG relation, and generalization relation).[17] In ELDAR the THESAURUS relation, see Chapter 2, suffices to manage the necessary semantic connections.[18]
- *Predefinition of frequently used joins.* The join[8] is an operation of the relational data model which connects two relations partially containing equal attributes to yield another relation. Such joins requiring many sorting processes are computer-time consuming. ELDAR offers special predefined joins and controls their update processes, especially to avoid update anomalies. In particular the THESAURUS relation is joined by the attribute *synonym* with every relation containing the attribute *term* or *descriptor* or *component*. In this way no time is wasted during the retrieval process for the join of synonymous terms of the relations.

FACTUAL KNOWLEDGE AND RELATIONAL DATA MODEL

Everyone in charge of a search for information on electrolyte solutions needs a collection of literature references and measured data, calculation methods for the estimation of solution properties, and a collection of rules which control the application of these methods. In addition to modules and rules, the user-friendly consulting system ELDAR manages the following factual knowledge.

- literature and measured data
- thesaurus
- module and parameter documentation
- basic data
- rule documentation

The structure and coordination of factual knowledge will be presented and characterized in the following paragraphs.

Literature and Measured Data. There are good reasons to store only measured data in a data base. Smoothed data are the result of the submission of original data to actual models and hence are unsuited for estimations based on other models. ELDAR separates the measured data from the calculation methods, leaving the estimation of all solution properties open to the user who is free to combine data and calculation methods.

Literature and original data with their error limits are coded in the standardized form of the DECHEMA data base DETHERM,[19] extended for electrolyte solutions[18] to an universal scheme for the formated representation of different data types. This standardized form is mapped into the relational data model.

Literature and data documents, as shown by the graphical TOP DOWN[20] representation in Figure 2, consist of the elements "Publication, Bibliography, Description" and a repetition of "Data block". A data block consists of a chemical system and its data. The part referred to as "Chemical system" is subdivided into the attributes *system TAG* (classification), *components*, *reaction products*, *reaction*, and *variable characterization*. The attribute components (products), in turn, consist of the attributes *CA name*, *synonyms*, *superterms*, *subterms*, *CA registry number*, *ions* (if the component is an electrolyte), and *formula*. The part referred to as "Data" is a repetition of *data tupel* consisting of state and property variables with their values, units, errors, and error types; a data tupel may contain a maximum of 9 variables.

This data structure is mapped into the relational data model of Codd extended to include repeating attributes. A hexagon in Figure 2 indicates a repetition of attributes specified at its right side. A repeating attribute is defined when the right side of a hexagon contains only one attribute, e.g., *authors*. The definition of a proper relation is necessary when the right side of a hexagon is a further repetition (hexagon) of an elementary attribute, e.g., *data tupel* in Figure 2 with the repetition of *variable*. In the same manner every hexagon with more than one attribute on the right side demands the definition of a particular relation for the attributes on the right side, e.g., the definition of the relation *chemical system* in Figure 2.

The definitions of the four relations LITERATURE, CHEMICAL SYSTEM, DATA, and THESAURUS for literature and data search are given in Table I. If we had not used repeating attributes, 12 relations would be needed instead of these four to represent the data structure, a very time-consuming and awkward problem for the formulation of queries.

Table II shows a typical example of a search for a literature document and its data. Here the concentration-dependent conductance and the corresponding literature are required for the chemical system lithium bromide in acetonitrile at 298.15 K.

The 400 000 data tupels actually stored in ELDAR belong to 50 000 solutions of about 2000 electrolyte compounds in more than 750 solvents.

The *solvents* are water, alcohols (91), aldehydes (10), amides (85), amines (50), cyclic compounds (40), esters (85), ketones (35), ethers (120), hydrocarbons (85), and others (150). For pure solvents and solvent mixtures, ELDAR contains characteristic properties such as density, viscosity, permittivity, isothermal compressibility, isobaric expansion coefficient, melting point, boiling point, dipole moment, polarizability, molecular dimensions, and empirical solvent parameters such as donor or acceptor number.

The 2000 pure *electrolyte compounds* Y composed of z_i valent ions $X_i^{z_i}$ with stoichiometric numbers ν_i, $Y = (X_1^{z_1})_{\nu_1}(X_2^{z_2})_{\nu_2}(X_3^{z_3})_{\nu_3}...$, are subdivided in 1-1 electrolytes (960), 2-2 electrolytes (110), 1-2 and 2-1 electrolytes (410), 3-1 electrolytes (200), other binary (270), and nonbinary electrolytes (50). Their characteristic quantities are the electrolyte structural formulas, stoichiometric numbers ν_i, ion valences z_i, ionic size parameters, polarizability, and electron distribution.

Table II. Example of Relations LITERATURE, CHEMICAL SYSTEM, and DATA

```
PRIMARY KEY: HA010
CLASSIFIC. : 4.4-ION-79
IDENTIFIC. : 017/52/811125
CA-NUMBER  : 90:211023F   LANGUAGE : ENGL    DOCUMENT TYPE : 02
YEAR       : 1979         VOLUME   : 8       PAGE          : 147-155
AUTHOR(s)  : HOPKINS,H.P.;JAHAGIRDAR,D.V.;NORMAN,A.B.
TITLE      : CONDUCTANCE STUDIES ON LITHIUM SALT- ACETONITRILE
             SOLUTIONS AT 25 C
PUBLICATION: J.SOLUTION CHEM.
DESCRIPTORS: CONDUCTANCE//FUOSS ONSAGER EQUATION//ASSOCIATION CONSTANT
             //LIMITING EQUIVALENT ELECTRICAL CONDUCTIVITY//LITHIUM
             SALTS
REMARKS    : CONCENTRATION/MR, 0.0004 TO 0.004

PRIMARY KEY: HA010
NUMBER     : 02
SYSTEM-TAG : E20
COMPONENTS : (1)   ACETONITRILE
             (2)   LITHIUM BROMIDE
VARIA.CHAR.: TSY00@000RUUUNA$TU
             XSY002000RLUUL A$TU
             ICD00@000RLU#UD$TU
```

DN	TEMPERATURE CELSIUS Error .05	CONCENTRATION MOL/L	EQUIV. CONDUCTANCE CM2/OHM.MOL
1	25.0	9.850E-4	142.46
2		12.902E-4	139.60
3		20.091E-4	130.60
4		21.542E-4	127.11
5		26.440E-4	124.34
6		26.800E-4	122.28
7		31.860E-4	118.19
8		36.605E-4	117.80
9		37.061E-4	116.13
10		41.030E-4	112.40
11		41.612E-4	113.18
12		45.240E-4	109.95

The *electrolyte solution* properties stored in ELDAR are the thermodynamic, dielectric, spectroscopic, and transport properties of the solutions as a function of temperature (T), pressure (p), and composition (molality, m; molonity, $\tilde{m}$; molarity, c; mole fraction, x; weight fraction, ξ); for spectroscopic properties also the frequency (ν) is reported.

In Table III a summary is given of the stored solution properties and the number of the corresponding data blocks. A data block may consist of 1–500 data tupels. The data are stored with the units of the original literature but are automatically transfered to SI units when used for calculations with the help of the method base.

The features of electrolyte solutions at given composition of the solution and external parameters are

- the limiting properties at infinite dilution
- the excess quantities resulting from ion–ion, ion–solvent, and solvent–solvent interactions, and their competition depending on the nature of solvents and solutes

These are the basic quantities for the chemical model equations[1] reproducing or simulating the property profiles of electrolyte solutions. Critically reviewed and updated basic data are of great interest for the user. Their determination from measured data is a task of the method base. They are stored as "best values" in the relation BASIC DATA; their updating takes place whenever new data are available.

Thesaurus. An important part of the consulting system of ELDAR is the thesaurus which contains the classification of solvents and electrolytes and the synonyms of terms. For the data administration the thesaurus supports the integration of independent data sources and the query translation (e.g., simplification of the translation by focusing to the preference name of attribute names). For the user it provides reconciliation of different nomenclatures, overlapping classification schemes (e.g., for solvents[1]), and different levels of data abstraction.

The thesaurus is implemented by the use of current data management technology. It is mapped to Codd's relational model with the extension to repeating attributes. The THESAURUS relation, see Table I, improves the semantic capacity of the relational data model. It contains the attributes *name key*, *name*, *preference name*, *superterm* (used for generalization), *subterm* (used for classification and formation of hierarchies yielding inheritances of properties), *associative term* (ions of electrolytes), and *formula* (only for components). The attributes *synonym*, *superterm*, *subterm*, and *associative term* are repeating attributes; more than one synonymous term may belong to one preference name. ELDAR manages 38 000 terms (key words or components) and their semantic references. Figure 3 sketches the polyhierarchical order of the thesaurus.

Table III. Statistics of Thermodynamic, Transport, and Dielectric Properties in ELDAR Including the Actual Number of Data Blocks of Solvents or Solvent Mixtures, and Electrolyte Solutions Stored in Data Base

	blocks	
	solvent	solution
thermodynamic properties		
compressibility (Pa^{-1})	24	11
density (kg m^{-3})	4071	1642
electromotive force (V)	0	350
enthalpies of dilution (J mol^{-1})	16	166
enthalpies of solution (J mol^{-1})	41	496
enthalpies of mixing (J mol^{-1})	13	1
expansion coefficient (K^{-1})	0	0
heat capacity (J mol^{-1} K^{-1})	97	274
molar volume (m^3 mol^{-1})	409	411
osmotic coefficient	0	319
solubility (mol m^{-3})		3045
surface tension (mol m^{-3})	86	14
vapor pressure (Pa)	207	224
transport properties		
diffusion coefficient (m^2 s^{-1})	0	38
molar conductance (m^2 Ω^{-1} mol^{-1})	202	5950
thermal conductivity (J K^{-1} m^{-1} s^{-1})	32	74
transference number	0	324
velocity of sound (m s^{-1})	159	82
viscosity (kg m^{-1} s^{-1})	2445	1094
dielectric properties		
absorption $\mathcal{J}(\hat{\varepsilon})^a$	172	81
dispersion $\mathcal{R}(\hat{\varepsilon})^a$	172	81
permittivity	2387	161
refractive index	854	73

[a] Dispersion and absorption are stored as real and imaginary parts of the complex permittivity.

Module Documentation. A high barrier to the acceptance of calculation programs by a user is the lack of understanding of the background theory of the algorithms. Which equation should be chosen for data analysis? Is this equation valid for the entire range of the data in question? What is the confidence interval for the method and the resulting data?

ELDAR offers the complete documentation and a calculation example for every module in the data base. This module documentation[7] consists of an explanation of the administration and the programming, a documentation of the theory, and a description of the effects. A module may have more than one effect depending on the controlling parameters, and one effect may have more than one input/output parameter. This structure is shown in Figure 4. Module documentation is mapped into the relational data model producing the basic relations MODULE (example in Table IV), EFFECT (example in Table V), and PARAMETER (example in ref 7), with the attributes shown in Table I. The repeating attributes of Figure 4 are *category*, *descriptor*, *term of effect*, and *term*.

A large variety of differently structured equations must find entrance into the method base to meet the requirements of the potential users. There is a need to represent the temperature, pressure, or concentration dependence of electrolyte solution properties with the help of polynomials in x, $x^{1/2}$, ln x, etc. of one or two variables; or to adopt the statistical mechanical

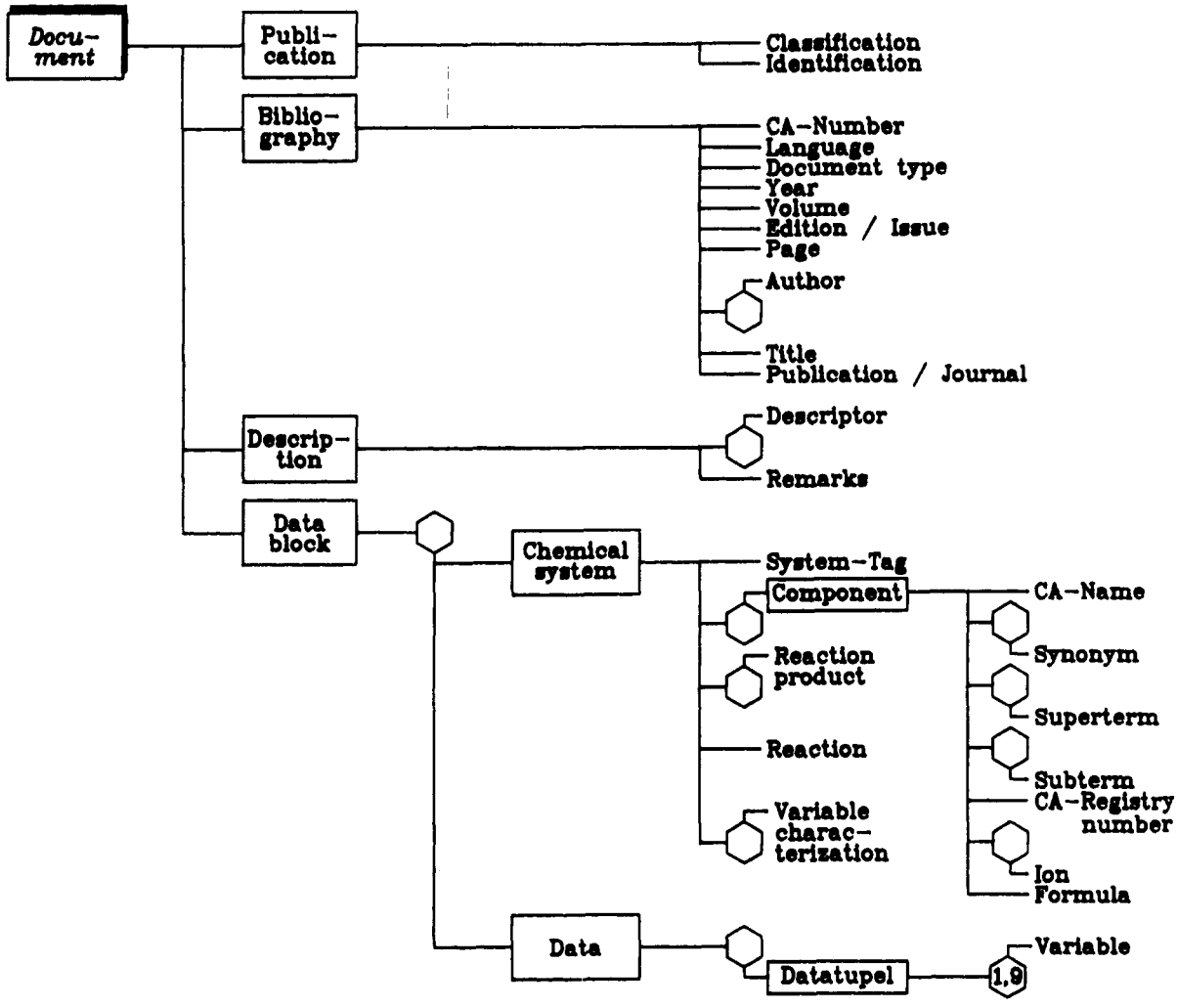

Figure 2. Data structure of a literature and data document in ELDAR.

equations of Hamiltonian models at the McMillan–Mayer level such as the limiting laws at vanishing concentrations or the *chemical models*[1,21] at low to high concentrations; or to reproduce data for practical purposes with the help of empirical and semiempirical equations such as the Pitzer equations for thermodynamic properties, Vogel–Fulcher–Tammann equations for transport properties, the Casteel–Amis equation for the concentration dependence of specific conductivity, or permittivity equations for the absorption and dispersion curves. The method base contains modules for all these equations. Actually the documentation of the following modules is stored:[3]

- Modules of numerical mathematics[22] and statistics[23]
- Modules of thermodynamic properties (including ion association for symmetrical electrolytes) such as mean activity coefficient, osmotic coefficient, molar volume, enthalpy of dilution, enthalpy of solution, heat capacity, electromotive force
- Modules of transport properties (including ion association for symmetrical electrolytes) such as equivalent conductivity, transference number, and single-ion conductivity
- Integral equation modules for the estimation of osmotic coefficients
- Modules based on empirical equations[22]
- Modules for different purposes, such as concentration scale conversion, conversion of units to the SI system.

Basic Data. The result of the interaction of the modules and the relation DATA is the relation BASIC DATA. The process for the generation of basic data under the control of the communication manager has already been described.[23] Basic data are the coefficients of reliable property equations calculated from measured data with the help of the modules of the method base, in particular the properties of the solutions at infinite dilution, association constants, and pair correlation functions. The relation BASIC DATA provides information on the data selected for the particular calculation, the modules used with their input parameters, and the statistics of the calculation process. Table VI shows an example of this relation, the calculation of the basic data of the conductivity equation by the chemical model.

As an important application of the knowledge base system, the user is provided with data and property profiles which are not stored as measured data. ELDAR solves this problem with the help of the BASIC DATA relation which may be used to yield property equations and their coefficients suitable for interpolation, thus acting as the source of actually best data. Data such as electrolyte properties at infinite dilution, association constants, or pair correlation functions are valid not only for the actually investigated property, but can be stored and used for the generation of data in the chemical model calculation of other properties.[21] Ionic properties at infinite dilution are additive and may be used to calculate unknown data of electrolyte solutions which have not been investigated so far. Figure 5 shows simulations of electric conductivity using the chemical model.

The determination of reliable basic data is a time-consuming task. The selection process yielding the "best data source" requires expert knowledge and demands many calculation cycles for the iterative adjustment of parameters. Helpful proposals for the selection processes are stored in the rule base.

Rule Documentation. The expert rules such as helpful rules for the choice of initial values for iteration processes, the choice

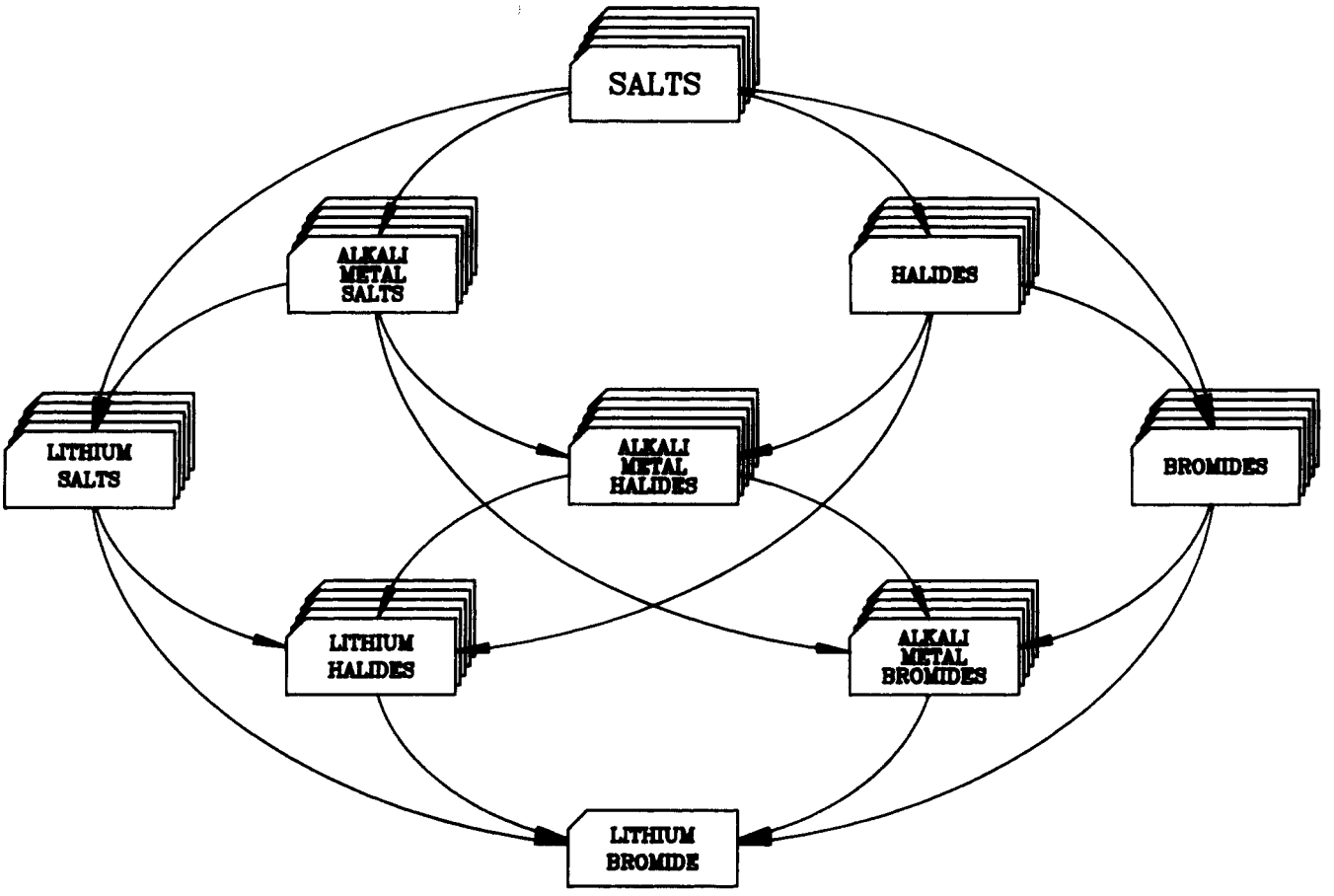

Figure 3. Extract of the ELDAR thesaurus net showing different levels of hierarchical connections.

Table IV. Example of Relation MODULE: Conductivity Module

NAME	: CONEQU
TITLE	: IONIC CONDUCTANCE OF DILUTE SOLUTIONS CALCULATED WITH THE HELP OF THE CHEMICAL MODEL
CATEGORY	: TRANSPORT PROPERTY//ELECTROLYTE SOLUTIONS// CHEMICAL MODEL//LEAST SQUARES FIT
DESCRIPTOR	: IONIC CONDUCTANCE // LIMITING CONDUCTIVITY// ASSOCIATION CONSTANT//EQUIVALENT CONDUCTIVITY FIT// CHEN EFFECT
SOURCE	: Barthel,J.; Buchner,R.;Wittmann,H.J.; Z. Phys. Chem. NF, VOL.139, 1984, p23-37.
DATE-VERS.	: 873007(1.00)
EXTERNALS	: QUAFRO(QUAFRO)//ACOEFF(ACOEFF)
CONNECTION	: It is mostly used from module LSQFIT
PROG.LANG.	: MS FORTRAN
MEMORY	: 27KB SOURCE; 21KB OBJECT
ACCURACY	: REAL*8
FILES	:
STANDARD	: FORTRAN 77
THEORY	: Transport equations of equivalent electrolyte conductance L = L^∞ - S * SQRT(c) + E*c*log(c) + J1*c* + J2*c**(3/2) with L^∞ as the limiting electrolyte conductance at infinite dilution and coefficients S, E, J1, J2 contain contribution due both to the relaxation and electrophoretic effect. Only the free ions in the solution are supposed to transport charges in the applied external field.For associating electrolytes the equation above is transformed into the set of equation L = alpha*[L^∞ - S*SQRT(alpha*c) + E * (alpha * c) * log(alpha*c) + J1*(alpha*c) + J2*(alpha*c)**(3/2). KA = (1-alpha)/(alpha**2*c*(y'+-)**2) and y('+-) = EXP[-kappa'* Q/(1 + kappa'*R)], where KA is the association constant; y('+-) is the activity coefficient of the dissociated part of the electrolyte; R = a+S is the distance parameter, alpha = degree of dissociation, Q = (e0**2*ABS(z(+)*z(-)))/(8*PI*eps*eps0*k*T); kappa'**2 = e0*SUM(i)(nue(i)* z(i)**2) / eps * eps0 * k * T). S = S1 *L^∞ + S2; and S1 = 2*Q*kappa'/(3*(2 + SQRT(2)); S2 = NL*e0**2*Z *SQRT(kappa')/(3* PI * eta). E = E1*L^∞ -chen* E2; E1=(Q*kappa')**2/6; E2= S2 * kappa' * Q / 8; chen=1 or 2 J1= SIGM1*L^∞+SIGM2; J2= SIGM3*L^∞ +SIGM4, with SIGM1 = E1 * (2 * R/Q + (R/Q)**2 - 0.25 * (R/Q)**3 + 1.814 + 2*log(kappa'*R), SIGM2 = E2 * (35/ (6*R/Q)+0.5*(R/Q)**2 - 2.0684 - 4*log(kappa' *R)
REMARKS	: Application only for binary symmetrical electrolytes; range of concentration: kappa*R ≪ 1 → c(max)

Table V. Example of Relation EFFECT for Conductivity Module[a]

MODULE NAME	: LSQFIT
ENTRY NAME	: LSQFIT
TYPE	: SUBR
DOMAIN	:
UNIT	:
TERM	: LEAST SQUARE FIT//CHEMICAL MODEL//GAUSS-NEWTON METHOD// MARQUARDT METHOD//IONIC CONDUCTANCE//ASSOCIATION CONSTANT//LIMITING IONIC CONDUCTIVITY//CHEN-EFFECT
EFFECT	: CHEMICAL MODEL EQUIVALENCE CONDUCTIVITY BY LEAST SQUARE FIT WITH GAUSS-NEWTON METHOD OR MARQUARDT METHOD
MANUAL	: <<How to use this module yielding the effect described in EFFECT ? Explanation of the parameters which are used >>
INSTR.EXA	: <<Definition of problem and description of a FORTRAN-program CALL (list of the module parameters) that solves the problem>>
TEST DATA	: <<The complete input data of a problem and the corresponding output data resulting from this module>>

[a] For the sake of simplicity the text of the attributes *manual*, *instruction example*, and *test data* which would need several pages is suppressed and replaced by a short ≪ information on content ≫.

of appropriate approximations, or the production of simulated values, etc. are coded in Horn clauses of PROLOG. The understanding of the Horn clauses is made easier by the documentation accompanying every rule. These documentations are facts and therefore are managed in the data base by the relation RULE, see Table I. The attribute *name* acts as the primary key of the relation RULE. The attribute *title* describes the effect of the rule; *category* shows the superterms and *descriptor* offers the key words; *source* indicates where the rule knowledge comes from; *date-version* shows the date of installation; *connection* points to the commonly used application of this rule; *language* is the logic language PROLOG. The attribute *probability* shows the degree of uncertainty of the assertion, *explanation* provides the full text of the rule, and *remark* is devoted to additional information.[3] Table VII shows an example.

Rules also result from the data model, since every searchable attribute must be defined in the rule base by the relation name to which it belongs. The integrity constraints of the attributes

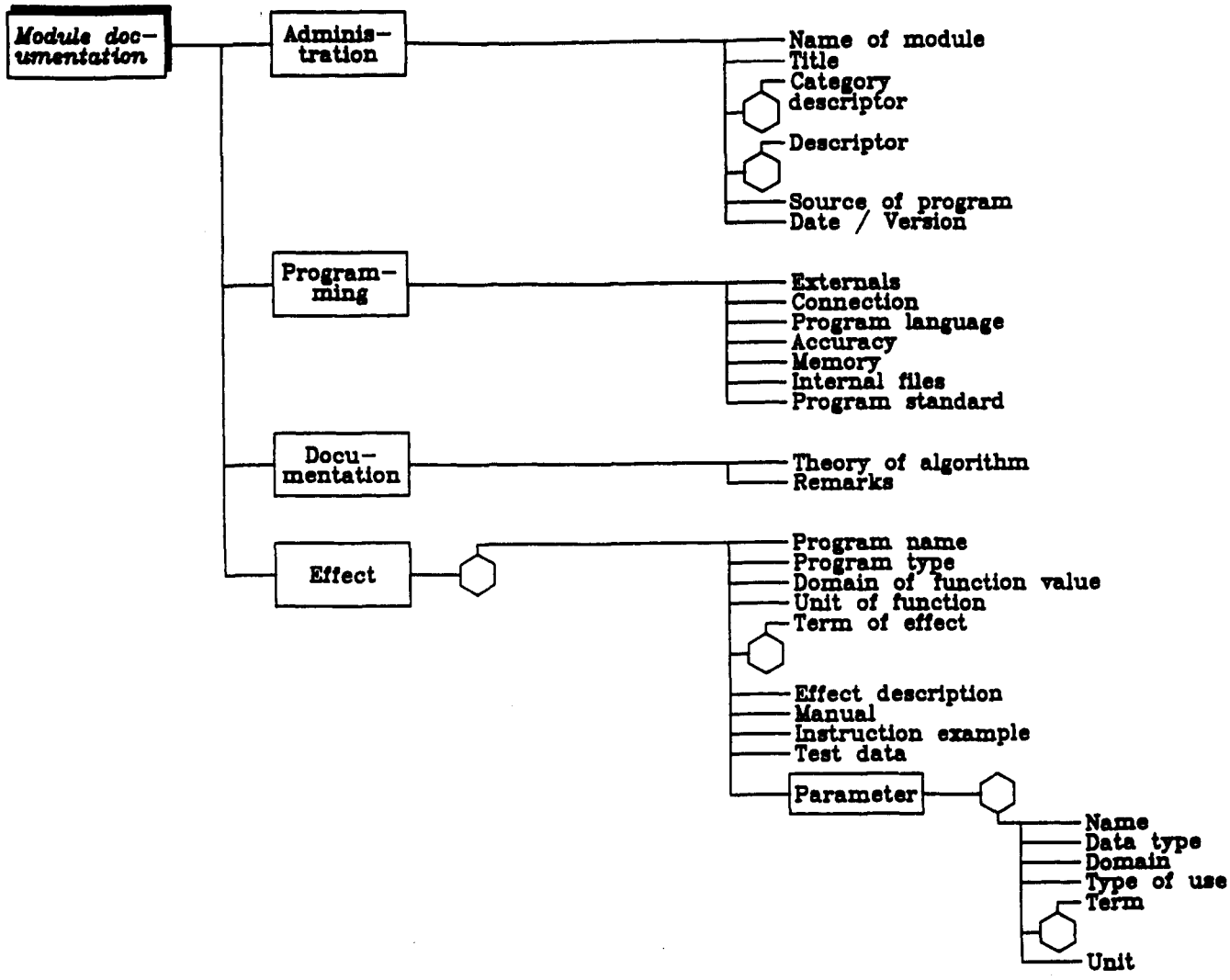

Figure 4. Data structure of the module documentation.

Table VI. Example of Relation BASIC DATA[a]

PRIMARY KEY:	ICDAL
SYSTEM TAG :	E20
COMPONENTS :	ACETONITRILE//LITHIUM BROMIDE
DATA SOURCE:	02HA010(12)04HD661(6/1-5)05HH130(4)
DEG.FREEDOM:	18
SIGN CHANGE:	15
ERROR TYPE :	1
MEAN WEIGHT:	1.0
VARIANCE :	0.2201E-3
USED MODULE:	CONEQU 2 4 1 1 1 1 TFTT
INT. CONST.:	1 1 1 0
REAL CONST.:	1. 5.8E-10 8.9E-10 298.15 35.891 3.4303E-4
STATE VAR. :	PRESSURE 101325 PA
	TEMPERATURE 298.15 K
STATE RANGE:	CONCENTRATION,LOWER 0.66E-1 ML/M3
	CONCENTRATION,UPPER 0.45E1 ML/M3
VAR. DOMAIN:	CONDUCTIVITY,LOWER .111E-1 M2/OML
	CONDUCTIVITY,UPPER .176E-1 M2/OML
BASIC DATA :	LIMITING CONDUCTIVITY 0.0176 M2/OML .18E-3
	J1-TERM 3299.4 .0
	J2-TERM -71661 4.1
	ASSOCIATION CONSTANT 0.265 M3/ML .35E-2

[a] Coefficients of the conductance equation obtained by the application of the conductance equation (the module is given in Table IV) on the literature data for solutions of lithium bromide in acetonitrile is given here.

Table VII. Example of Relation RULE

NAME	: R0001
TITLE	: DISTANCE PARAMETER OF AQUEOUS AND ALCOHOL SOLUTIONS OF LITHIUM SALTS
CATEGORY	: CHEMICAL MODEL//ELECTROLYTE SOLUTION
DESCRIPTOR	: DISTANCE PARAMETER//UPPER LIMIT//ALCOHOLS//WATER// LITHIUM SALTS
SOURCE	: A1330 (means ref. Barthel,J.Gores,H.;Schmeer,G.; Wachter,R.; 1983)
DATE-VERS.	: 893008(1.00)
CONNECTION	: Modules of the chemical model need this distance parameter
LANGUAGE	: PROLOG
PROBABILITY:	.95
EXPLANATION:	In aqueous and alcohol solutions the lithium cation contributes to the distance parameter of closest ionic approach by the sum of the lithium radius and the length of an orientated solvent molecule.
REMARKS	:

are rules which are managed in the rule base.

USER INTERFACE

Ergonomics for user interfaces applies the knowledge of human perception, learning, and cognition, the results of behavioristic sciences, and industrial experience. It plays an important role in the design of user-friendly knowledge base systems. The ELDAR user interfaces are realized with the help of different dialog types. For a given task the most appropriate technique is always used, such as form technique for the input of relations, command technique for corrections, menu technique for the selection of services, and a combination of graphic and mapping-based languages for information retrieval.

During the *fact input*, especially data input, the typist is led through large and sometimes complex data tables of the publications by the offer of the actual value of the state variables and an isomorphic form of the actual table.

The *fact retrieval* process permits two possibilities, the "occasional user mode" which allows the input with the help of a quasi-natural language and the "experienced user mode"

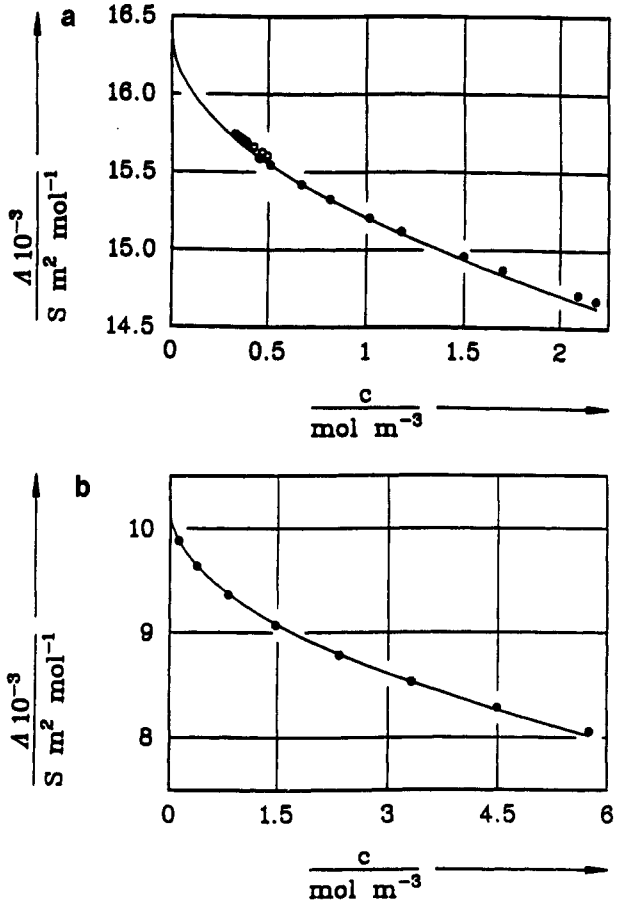

Figure 5. Simulation of electric conductivity by chemical model calculation for the systems Bu_4NI in acetonitrile (a) and Bu_4NI in methanol (b) at 298.15 K. Limiting conductivity is calculated from single-ion conductivities (acetonitrile)[24] and (methanol),[25] and the association constants are taken from vapor pressure measurements.[26] For comparison the data from conductance measurements from ref 25 (methanol) and refs 27,28 (acetonitrile) are added (full and open circles).

permitting the input of short associative abbreviations for the commands. The retrieval screen in Figure 6 is subdivided in the following areas:

- the top line shows the dialog state, e.g., INFORMATION RETRIEVAL
- the bottom lines show the command menu
- the lines above the command menu show an input example of the actual command
- the central area is the input and output area
- submenus are offered on the right sight of the screen

The commands of the command menu are as follows:

FIND	opens a query specifying the wanted out of the offered relations and connecting the predicates (attributes and their values). The complete query appears on the screen in a natural language (actual situation of Figure 6).
SHOW	shows on screen all or parts of the last hits and makes projections on the specified attributes.
PRINT	prints on a printer all or parts of the last hits and makes projections on the specified attributes.
BUFFER	buffers on file all or parts of the last hits.
DISPLAY	shows the query table or explains abbreviations.
MS/DOS	allows the execution of MS/DOS commands of the operating system, e.g., starting of a compiler or an executable operator.
WORK	prepares the hits of a query (data or modules) for the use in the method base and writes the information on a communication file.
EXECUTE	executes the operations of a module of the method base; the input demands of this module are supplied by a communication manager.[23]
COMPOSE	composes a complex module from elementary modules and produces a main program

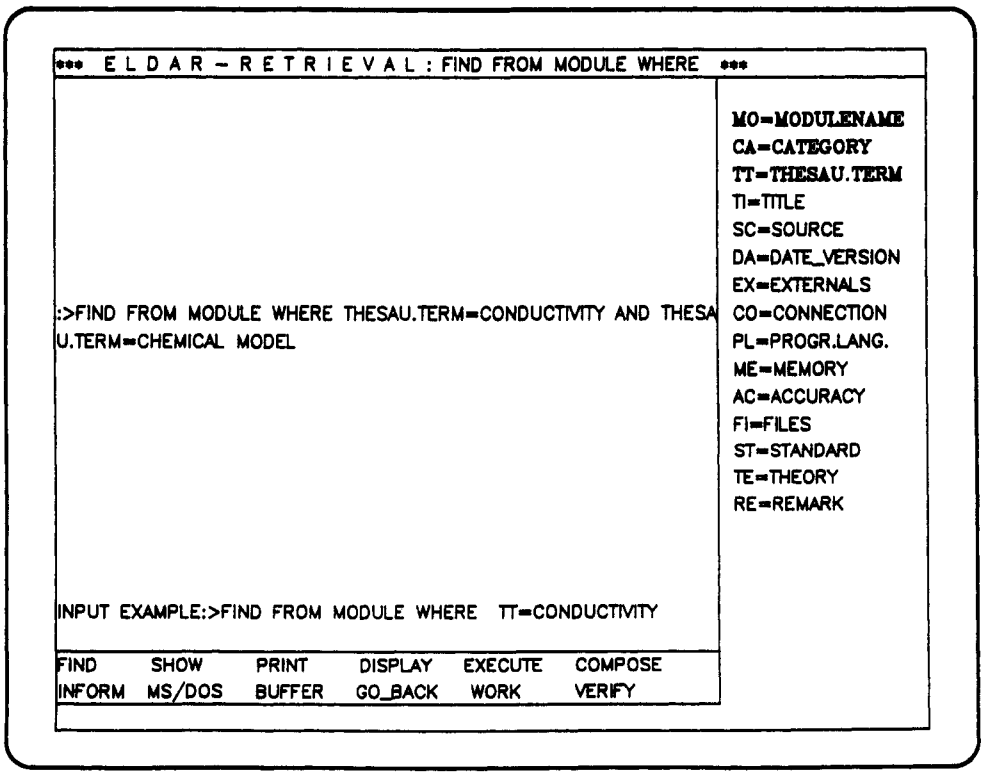

Figure 6. Fact retrieval screen for a module retrieval.

frame for this module (planned).

VERIFY verifies predicates of the user by the rule base (in development).

INFORM explains the result of a calculation or verification and the way to get results (in development).

To build up a query the user selects the command FIND and specifies the wanted relation. Then ELDAR shows Figure 6 on screen; the attributes of the relation are listed on the right side. Typing behind FIND FROM MODULE WHERE the abbreviation of the selected attribute, for the example given in Figure 6 this is TT = Thesaurus Term, the computer replaces the abbreviation by the complete name.

The thesaurus is a helpful tool for the query. It allows queries for classes of terms or components and within reasonable limits also the usage of different spelling.

If a query for data yields no hits, the communication manager asks the user to calculate the property using data-, method-, and rule-base interaction.

SUMMARY

The concerns of the data base of ELDAR presented here are the acquisition of factual knowledge on aqueous and nonaqueous electrolyte solutions, the generation of reliable basic data for science and technology, and the presentation of property profiles. User friendliness is achieved by an efficient fact management based on a data model reducing the number of basic relation from 29 to 9. The modular method base, managing the algorithmic knowledge, the rule base, managing heuristic knowledge, and the communication manager controlling the interaction of the different "bases" of ELDAR will be presented in following contributions.

The information on literature and data of ELDAR is available online in the DECHEMA data base DETHERM; FIZ (Fachinformationszentrum) Chemie, Berlin, and TDS (Numerica Technical Database Services), New York, manage a PC version of the ELDAR data and method base.

ACKNOWLEDGMENT

We are grateful to the Bundesministerium für Forschung und Technologie for financial support and to the co-workers of the Institute of Physical and Theoretical Chemistry of the University of Regensburg, especially Professor G. Schmeer, for valuable help.

REFERENCES AND NOTES

(1) Barthel, J.; Gores, H. J.; Schmeer, G.; Wachter, R. Non-Aqueous Electrolyte Solutions in Chemistry and Modern Technology. *Top. Curr. Chem.* **1983**, *111*, 33–144.

(2) Gores, H. J.; Barthel, J. Non-Aqueous Electrolyte Solutions. *Naturwissenschaften.* **1983**, *70*, 495–503.

(3) Barthel, J.; Popp, H. In *Data for Discovery, CODATA*; Glaeser, P. S., Ed.; Elsevier: North-Holland, 1991; (in press).

(4) Mylopoulos, J.; Levesque, H. In *GWAI-83*; Neuman, B., Ed.; Springer: Berlin, 1983; p 143.

(5) Schlageter, G.; Unland, R.; Wilkes, W.; Zieschang, R.; Maul, G.; Nagl, M.; Meyer, R. In *Datenbanksysteme in Büro, Technik und Wissenschaft, GI Fachtagung Darmstadt*; Scheck, H. J., Schlageter, G., Eds.; Springer: Berlin, 1987; p 132.

(6) Barthel, J.; Popp, H. In *Abstracts of Lectures and Posters, 8th International Symposium on Solute–Solute–Solvent Interactions*; Barthel, J., Schmeer, G., Eds.; Regensburg, 1987; p 255.

(7) Barthel, J.; Popp, H.; Schmeer, G. In *Softwareentwicklung in der Chemie 2*; Gasteiger, J., Ed.; Springer: Berlin, 1988; p 127.

(8) Date, C. J. *An Introduction to Database Systems*, 3rd ed.; Addison-Wesley: Reading, MA, 1981.

(9) Lochovsky, F. H. Data Base Management System and Performance. Ph.D. Dissertation, University of Toronto, Canada, 1978.

(10) Kent, W. A Simple Guide to Five Normal Forms in Relational Database Theory. *Commun. ACM* **1983**, *26*, 120–125.

(11) Blaser, A.; Pistor, P. *Datenbanksysteme in Büro, Technik und Wissenschaft, GI Fachtagung Karlsruhe.* Springer: Berlin, 1985.

(12) Popp, H. Mensch—Mikrocomputer Kommunikationssystem. Management Expertensystem in der Chemischen Industrie auf der Basis eines universellen Daten und Prozeduralmodells auf einem Mikrocomputerverbundsystem. Dissertation, Regensburg, 1984.

(13) Scheck, H. J.; Schlageter, G. *Datenbanksysteme in Büro, Technik und Wissenschaft, GI Fachtagung Darmstadt.* Springer: Berlin, 1987.

(14) Jaeschke, G.; Scheck, H. J. In *Principles of Database Systems. ACM SIGACT-SIGMOD Proceedings*; ACM Order No. 475820, 1982, p 128.

(15) Chen, P. P. The Entity-Relationship Model—Toward a Unified View of Data. *ACM TODS* **1976**, *1*, 9–36.

(16) Hammer, M.; Mc Leod, D. Database Description with SDM: A Semantic Database Model. *ACM TODS* **1981**, *6*, 351–386.

(17) Codd, E. F. Extending the Database Relational Model to Capture more Meaning. *ACM TODS* **1979**, *4*, 397–434.

(18) Barthel, J.; Popp, H.; Schmeer, G. Stoffdaten binärer nichtwässriger Elektrolytlösungen. Ein Mikrocomputerverbundsystem zur Bereitstellung von Literatur und Stoffdaten. Report BMFT-NTS 2936, 1983.

(19) Eckermann, R. Ein Stoffdaten-Informationssystem für die Chemische Technik. *Chem.-Ing.-Tech.* **1981**, *53*, 31–38.

(20) Ferstl, O. K. Flowcharting by Stepwise Refinement. *SIGPLAN Notices* **1978**, *13*, 34–42.

(21) Barthel, J. In *The Role of Data in Scientific Progress, CODATA*; Glaeser, P. S., Ed.; Elsevier: North-Holland, 1985; p 337.

(22) Barthel, J.; Popp, H. Die Methodenbank des wissensbasierten Systems ELDAR zur (statistischen) Analyse von Elektrolytlösungen auf einem Mikrocomputer. Report BMFT-FB 106 3211 0, 1989.

(23) Barthel, J.; Popp, H.; Schmeer, G. In *Software Development in Chemistry 4*; Gasteiger, J., Ed.; Springer: Berlin, 1990; p 101.

(24) Barthel, J.; Iberl, L.; Rossmaier, J.; Gores, H. J.; Kaukal, B. Conductance of 1,1-Electrolytes in Acetonitrile Solutions from −40° to 35 °C. *J. Solution Chem.* **1990**, *19*, 321–338.

(25) Barthel, J.; Krell, M.; Iberl, L.; Feuerlein, F. Conductance of 1-1 Electrolytes in Methanol Solutions from −45° to 25 °C. *J. Electroanal. Chem. Interfacial Electrochem.* **1986**, *214*, 485–505.

(26) Barthel, J.; Kunz, W.; Lauermann, G.; Neueder, R. Calculation of Osmotic Coefficients of Nonaqueous Electrolyte Solutions with the Help of Chemical Models. *Ber. Bunsen-Ges. Phys. Chem.* **1988**, *92*, 1372–1380.

(27) Singh, D.; Aggarwal, I. P. Electrolyte–Solvent Interaction: Conductance of Tetraethylammonium Iodide and Tetrabutylammonium Iodide in Acetonitrile-Dioxan. *Acta Chim. Budapest* **1972**, *73*, 325–333.

(28) Treiner, C.; Fuoss, R. M. Conductance and Viscosity of some Quaternary Salts in Acetonitrile at 25 °C. *Z. Phys. Chem. (Leipzig)* **1965**, *228*, 343–353.