Structured System of Concepts for Storing, Retrieving, and Manipulating Chemical Information

Paul E. van der Vet*,† and Nicolaas J. I. Mars

Knowledge-Based Systems Group, Department of Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Received February 25, 1993

Sharing and reuse of databases and knowledge bases with chemical information is seriously hindered by a lack of standards. The present paper stresses the need for standardization at the level of concepts. Standardization can be achieved by using limitative concept systems or ontologies (in the technical sense). The principle of the conceptual construction kit for building an ontology is introduced and utilized to construct an ontology for an important part of chemistry. This ontology is proposed as a first step toward a standardized ontology for chemistry.

1. INTRODUCTION

Chemical information, one of the resources essential for research and industrial activities, is increasingly stored and made available through electronic media. The field now witnesses a proliferation of databases and knowledge bases, each employing its own standards. This diversity seriously hinders use and reuse of information already stored in electronic form. The costs of building and maintaining a sizeable database or knowledge base can be very high. Sharing and reuse are therefore economically attractive and have become a major research issue.[1]

There are both technical and conceptual problems involved in enabling sharing and reuse. (Additionally, there are organizational and legal problems, which are not addressed in this paper.) Technical problems arise as a result of different storage formats and representation languages. These problems are typically solved through standardization efforts such as SQL, KQML, and KIF and, for chemistry, the important SMD format.[2] Conceptual problems arise because even when a database or knowledge base conforms to technical standards, it is hard to find out what precisely has been stored. The problem is particularly serious for knowledge bases because of their great expressive power. Conceptual problems, too, can be alleviated through standardization.

Standardization at the conceptual level can be achieved by employing a *limitative* system of concepts and by expressing all knowledge using those concepts only. The relations between concepts can be made explicit and certain concepts can be defined in terms of other concepts by imposing structure. In order to make a concept system sufficiently general to be of use as a standard, it has to abstract from representation issues. Hence, concept systems have to be designed at the *knowledge level*, *i.e.*, the level above that of representation languages and database schemes. The technical term for a structured concept system is *ontology*, a term introduced in this sense by Hayes.[3]

The advantage of using an ontology are as follows: conceptual discipline is enforced at the knowledge level; transparency is enhanced; the possibilities of reusing and sharing can be judged.[4,5] Various ontologies have been proposed. For chemistry, the work of Gordon[6-8] on organic chemistry and that of Mars[9] on measurement units can be cited as examples.

Recent standardization efforts for databases with chemical information[2,10,11] and the IUPAC recommendations are also aimed at sharing and reuse. However, we believe that these efforts can profit from a sharper distinction between terms and concepts. A *term* is a linguistic entity used as a name for a concept while a *concept* is language-independent (see also ref 12). There is a many-to-many relation between terms and concepts, so that terms are often ambiguous (hence the terminological problems identified in, *e.g.*, ref 13). For example, the terms "aluminium", "aluminum", "Al" and the Russian "alyuminii" all are used as names for two quite different concepts: (a) the metal and (b) the chemical element with atomic number 13. An ontology is required to be unambiguous and thus is a concept system rather than a term system.

A corollary of the distinction between terms and concepts is that between representation issues and conceptual issues. It is important to realize that the SMD, SMILES, WISWESSER, and MOLFILE representations of benzene[11] are, in a first approximation, just different names for the same concept. The differences between these representations are not such that they define different kinds of benzene; rather, they differ in the amount and the kind of additional information that can be stored.

In this paper, we present an ontology for parts of chemistry which was originally developed for a particular project, the Plinius project. In Plinius the ontology serves specific purposes leading to particular requirements. These requirements are quite similar to those imposed from a standardization point of view. Therefore, we believe that the Plinius ontology has wider relevance. In particular, we think it is a step toward standardization of concept systems for chemistry that allow reuse and sharing of chemical information stored in databases and knowledge bases.

2. THE PLINIUS PROJECT

The Plinius project[12] aims at semi-automatic knowledge extraction from natural-language texts. The input material consists of all abstracts from the 1990 volume of *Engineered Materials Abstracts* with subject identification code C1-C; the code refers to mechanical properties of ceramic materials.

The system designed in Plinius employs a number of information sources: a grammar for syntactic analysis; a lexicon storing the syntactic features and (a part of) the meanings of words; and a knowledge base with background

† E-mail address: vet@cs.utwente.nl.

0095-2338/93/1633-0564$04.00/0

knowledge. The intended output is a knowledge base which stores the knowledge conveyed by the abstracts, or at least those parts which are interesting from a chemical point of view. The Plinius ontology, fully described elsewhere,[14] is designed to fulfill two functions directly relevant to the process of converting natural-language texts into knowledge base entries. In the first place, the ontology enables co-operation between the semantic part of the lexicon and the knowledge base with background knowledge. In the second place, the ontology partially specifies the output of the process. The output knowledge base cannot be specified in advance since it is dynamic, so an indirect specification is needed. In both cases the ontology helps to determine the semantics (in ref 15, a similar point is made). The functions of the Plinius ontology determine its major features such as granularity and scope.

As a result of the primary functions of Plinius ontology, it is also instrumental in decisions on further usage of the output knowledge base. Because the conceptual vocabulary is specified, one can judge whether and, if so, how the output knowledge base can be utilized in systems for, *e.g.*, materials selection or combined with existing databases and knowledge bases for yet other purposes. For this reason, the Plinius ontology is a proposal for a part of a standard ontology for chemistry.

3. PRINCIPLES

3.1. The KL-ONE Approach. The KL-ONE family of knowledge representation languages has been, and still is, an important source of inspiration for ontology research. Therefore the ideas underlying the Plinius ontology are best explained by contrasting them with the KL-ONE approach. In KL-ONE-like systems,[16] the knowledge base is split into two parts: a terminological part, which is a representation of the ontology, and an assertional part, which stores representations of assertions using only concepts from the terminological part plus logical symbols.

In *knowledge level* terms, the KL-ONE approach structures the ontology as a taxonomic hierarchy much like the Linnean *genus–species* hierarchy in biology. The emphasis on the knowledge level is essential here. When KL-ONE is viewed merely as a language, there is no contrast with the Plinius approach. In fact, the Plinius ontology can be expressed in several existing KL-ONE languages.[17]

A KL-ONE structure is built top-down with a very general concept (often called "entity") as the starting point. Further concepts are added by recursive differentiation. Each concept except the top concept has at least one so-called IS–A-relation (subconcept–superconcept–relation) with a concept above. Differentiation, *i.e.*, a specification of the aspects in which a subconcept differs from its superconcept, is performed by links with other concepts. A concept corresponds to a set of objects in reality. The IS–A-relation corresponds to the subset relation in set theory and links correspond to relations (sets of tuples of objects). KL-ONE thus is a system for storing and manipulating intensional specifications of sets. This way of structuring is also manifest in other representation languages, for instance in order-sorted logic.

3.2. The Plinius Approach: The Conceptual Construction Kit. For Plinius, the KL-ONE strategy seems less suitable. Chemists do not approach the world in the top-down fashion of recursive differentiation, but rather in the bottom-up fashion of composite construction. The world is viewed as being composed out of basic building blocks at different levels of abstraction. This is not a matter of words. Consider the problem of a dynamic conceptual vocabulary that arises in Plinius: it cannot be predicted beforehand which concepts are needed to interpret an as yet unprocessed text. The solution is to employ an ontology which implicitly defines all concepts in a controlled way. In chemistry this is done bottom-up: pure substances are composed of elements, phases of pure substances, heterogeneous systems of phases.

The principle of composite construction is manifest in the grammar of chemical formulas specifying stoichiometric compounds, such as $Ca_5(PO_4)_3OH$. (Compare the extensive treatment of grammatical issues in the IUPAC recommendations on the nomenclature of inorganic compounds.[18]) The grammar is controlled in the sense that only chemical elements, numbers, and grouping symbols are allowed, and this at particular places only. Note that the grammar still allows the construction of a formula such as $Na_{17}O_{13}$ that does not correspond to a compound now known to exist. This is as it should be, for two reasons: (a) one does not want to burden a grammar with time-dependent knowledge about the stability of compounds; and (b) in order to express the fact that $Na_{17}O_{13}$ is not known to exist we need a concept for $Na_{17}O_{13}$.

The idea underlying the organization of the subject-matter of chemistry is systematized in the principle of the conceptual construction kit. There are predefined *atomic concepts* or primitives and predefined *construction rules* for defining complex concepts in terms of atomic concepts. For convenience, atomic concepts might be distributed over several sets and construction rules might define complex concepts in terms of other complex concepts, but the principle remains unaffected. Expressions for complex concepts can always be expanded to yield a version in which only atomic concepts occur.

The Plinius ontology is built as a conceptual construction kit in a rather straightforward way. In a first approximation (elaborated below), there are two sets of atomic concepts: a set of chemical elements and a set of natural numbers; and the construction rule for concepts denoting pure substances is evident. From a technical point of view, the principle of the conceptual construction kit reconciles the apparently conflicting demands of conceptual parsimony and great expressivity. In another project of our group, the Sapiens project,[19] an ontology of diseases has been built as a conceptual construction kit. The principle thus seems to be amenable to disciplines other than chemistry.

3.3. The COMPOSES-Relation. In the KL-ONE approach, hierarchies are built top-down by means of the IS–A-relation. In chemistry, hierarchies are built bottom-up by means of a relation putatively called COMPOSES: chemical elements compose pure substances, pure substances compose phases, and so on. The converse of COMPOSES is PART-OF. "PART-OF" is avoided here for two reasons. First "COMPOSES" stresses the bottom-up approach. Second, the term "part-of" is confusing. There is a relation called part-of with a precise technical meaning in a formal theory called *mereology*. Mereology has recently become the subject of a vigorous debate in AI. The relation introduced here differs significantly from the mereological relation.

The identification of a relation as being COMPOSES in an as yet intuitive sense presupposes that there are parts and wholes made up out of those parts. For a start, a whole is just the collection of parts. But that is not enough, because by inspecting the collection of parts it cannot be decided whether the parts form a whole or are disconnected. (That is what distinguishes a car from a building kit to make that car.) Therefore, a whole is the collection of its parts plus what will be called *configuration information*, *i.e.*, information about

the way the parts are put together. The obvious chemical example identifies a molecule as a whole, the constituent atoms as its parts, and the molecular structure as the configuration information. The COMPOSES-relation of the Plinius ontology can now be made precise. A set of parts COMPOSES a whole if the whole is defined by specifying the set of parts and configuration information.

The hierarchy in the Plinius ontology is the result of propagating the strategy of identifying parts and configuration information upwards to the levels of phases and heterogeneous systems. This is done by choosing the appropriate construction rules.

3.4. Alternative Construction Replacing the IS–A-Relation. Thus the backbone of the ontology is formed by COMPOSES-relations (in the precise sense specified above). Hierarchical relations of the IS–A kind can also be defined in chemistry; for instance, organic compounds can be classified according to functional groups. The resulting hierarchy is not a taxonomy but rather a so-called *tangled hierarchy*. For example, every amino acid contains at least two functional groups and thus is a subconcept with respect to the two superconcepts corresponding to these functional groups.

The solution for this problem in the Plinius ontology is to build hierarchies bottom-up rather than in the KL-ONE top-down fashion. If a concept is constructed from building blocks (whether atomic or complex concepts), a superconcept can be constructed by leaving at least one of the building blocks unspecified. The interpretation is that for the unspecified building block every member of the relevant set can be substituted. If in the superconcept there still are specified parts, the procedure can be repeated. For instance, the concept for "amino acid" can be generated from the concept for serine (or any other amino acid) by leaving everything unspecified save for the parts specifying the NH_2- and COOH-groups. Next, the building block specifying the COOH-group can be replaced by an unspecified building block, resulting in a concept for "amino compound". If the procedure is carried to its logical conclusion, the result is a concept for "compound" in general.

In this way, the whole range of more general concepts at various levels of generality is available as soon as a specific concept is defined. Whether an IS–A-relation holds between two concepts can be determined automatically for all pairs, and explicit addition of IS–A-relations would be redundant. IS–A-relations, if used, lead to a tangled hierarchy orthogonal to the hierarchy built by means of COMPOSES. This approach advocated here can be compared with another approach[20] involving hierarchies of organic chemicals. The latter approach also views the hierarchies built by COMPOSES and IS–A as being orthogonal but is forced to build the IS–A hierarchy explicitly.

4. THE PLINIUS ONTOLOGY

4.1. Outline. Below, the core of the Plinius ontology is sketched. It is designed according to the principles discussed above and consists of

Atomic concepts distributed over several sets.

Definitions supplying construction rules for complex concepts.

Along with each set of atomic concepts there is an unspecified building block called *arbitrary*. These building blocks are typed, so that $arbitrary_X$ stands for any member of set X but for no other atomic concept. (To prevent confusion with chemical elements, elements of sets are called "members" here.) As an additional feature it is allowed to locally define proper subsets of sets of atomic concepts together with an unspecified building block standing for any member of the proper subset.

The construction rules lay down what is minimally needed for unambiguous identification of concepts, given the domain. More information could have been added in many cases. However, such extra information is not always necessary nor always available. Therefore incorporation of extra information in the construction rules would narrow the scope of the ontology and limit its practical use.

The Plinius ontology does not depend on a particular choice of representation language. It can and actually has been formalized and implemented in different ways. Currently, a formalization in first-order predicate logic implemented in Prolog and a formalization and implementation in the KL-ONE language CLASSIC[17] are available. Similar work with other KL-ONE languages and with conceptual graphs is underway.

4.2. Pure Substances. The first step in any ontology for a chemical domain evidently is a definition of concepts for pure substances. In the corpus studied for Plinius, organic compounds do not occur and inorganic compounds are rather simple. Therefore, a simple construction has been employed. A pure substance is built up out of elements and molecular sub-units, which in turn are built up out of elements. The introduction of the intermediate concept of *group* to embrace both elements and molecular sub-units results in a simple picture, with pure substances composed of groups.

For constructing concepts for pure substances, two sets of atomic concepts are needed: a set E of all 105 chemical elements, and a set N of natural numbers (*i.e.*, all integers > 0). Along with each of the two sets there is an unspecified building block, called $arbitrary_E$ and $arbitrary_N$, respectively. As every set of atomic concepts is accompanied by such an unspecified building block, this remark will not be repeated for other sets of atomic concepts that are introduced below.

Two definitions take care of pure substances:

Definition 1 (Group). *A* group *g is a set of tuples t such that for each t*

$$t = \langle e,n \rangle \quad \text{with} \quad e \in E \quad \text{and} \quad n \in N$$

such that no e occurs in more than one tuple. The set of all groups *is called GR.*

Definition 2 (Pure Substance). *A* pure substance *p is a set of tuples t such that for each t*

$$t = \langle g,n \rangle \quad \text{with} \quad g \in GR \quad \text{and} \quad n \in N$$

such that no g occurs in more than one tuple. The set of all pure substances *is called PS.*

Given the definition of a group, ions can be defined. We need extra atomic concepts to specify the charge. Ionic charges are always whole multiples of the elementary charge and can be positive or negative. Natural numbers have already been defined as atomic concepts, and we only need a set

$$C =_{Df} \{+,-\}$$

in order to fully specify charges.

Definition 3 (Ion). *An* ion *i is a tuple consisting of a group and a charge, so that*

$$i = \langle g,m,s \rangle \quad \text{with} \quad g \in GR, \quad m \in N, \quad \text{and} \quad s \in C$$

where the charge is expressed by m (magnitude) and s (sign). The set of all ions *is called IO.*

A simple example involving barium titanate ($BaTiO_3$) might illustrate the ideas. By using definition 1, two groups called g_{Ba} and g_{TiO_3} are defined:

$$g_{Ba} =_{Df} \{\langle Ba, 1\rangle\}_{GR}$$

$$g_{TiO_3} =_{Df} \{\langle Ti,1\rangle, \langle O,3\rangle\}_{GR}$$

The subscript GR is added for convenience. The concept for barium titanate becomes by definition 2

$$\{\langle g_{Ba},1\rangle, \langle g_{TiO_3},1\rangle\}_{PS}$$

and the concept for Ba^{2+}, by definition 3

$$\langle g_{BA},2,+\rangle_{IO}$$

4.3. Discussion. The principle of composite construction is clearly visible in definitions 1 and 2. For a group, the parts are atoms. Configuration information is lacking except for the fact that through the specification of the group the atoms involved are identified as forming a whole. Analogous remarks apply to pure substances.

For purposes of presentation, an automatic process could convert concepts into formulas or names according to IUPAC recommendations.[18] Recently, an automatic process has been proposed to convert formulas into names;[21] the process would be easier for us because parsing of formulas is not needed.

The present ontology covers every stoichiometric compound encountered in the set of input texts for Plinius. In this sense, it is complete. When the scope is widened, extensions will have to be considered. First, more complicated compounds involving nested grouping levels can be accommodated by making definition 1 recursive. Second, if isotopes have to be distinguished, chemical elements can no longer be atomic concepts. The way out here is to define atom types as tuples $\langle p,n\rangle$ with p the number of protons and n the number of neutrons, both of course members of N. An element now appears as a defined concept, e.g., the element Al as $\langle 13,arbitrary_N\rangle$. Finally, the construction involving groups fails to distinguish between isomers such as the pair cyanate/isocyanate. Only by including some configuration information can cyanate and isocyanate be told apart.

These observations make clear that standardization in general is an ill-defined task. Expressivity and completeness of an ontology can only be judged when an operational specification of the domain is supplied. For Plinius, the domain is that of the mechanical properties of ceramic materials. This specification is operationalized by means of the set of input texts.

4.4. Nonstoichiometric Compounds. In order to cover all examples of nonstoichiometric compounds, a new set of atomic concepts is needed: R, the set of expressions of type $a + \sum_i b_i x_i$, where a and all b_i are real-valued constants and all x_i are real-valued variables. Since all b_i can be zero, the set of real numbers is a subset of R. By a similar argument, the introduction of R formally obviates the need for a separate set N, as $N \subset R$. However, there is a need to distinguish between specified numbers of type integer and those of type real. In a formula such as "$UO_{2.04}$" the (omitted) integer 1 for uranium indicates that there is no excess or deficiency of uranium while the real 2.04 for oxygen indicates excess oxygen (see also ref 10). Therefore, N is kept. In the notation below the distinction will be made by writing the number as "1" for the integer-typed and as "1.0" for the real-typed version.

With the help of these atomic concepts, nonstoichiometric groups and compounds can be defined. It is immaterial whether the substance in question is properly regarded as a compound or as a mixture, mirroring common practice in chemistry. The definitions below cover both stoichiometric and nonstoichiometric groups and pure substances. The distinction between stoichiometric and nonstoichiometric can be made by inspecting whether the numbers are members of N or R.

Definition 4 (Generalized Group). *A* generalized group n *is a set of tuples* t *such that for each* t

$$t = \langle e,r\rangle \quad \text{with} \quad e \in E \quad \text{and} \quad r \in R$$

such that no e *occurs in more than one tuple. The set of all* generalized groups *is called GG.*

Definition 5 (Generalized Chemical Substance). *A* generalized chemical substance c *is a set of tuples* t *such that for each* t

$$t = \langle g,r\rangle \quad \text{with} \quad g \in GG \quad \text{and} \quad r \in R$$

such that no g *occurs in more than one tuple. Where there is a choice to put in the real-valued expression for proportion, it is always put in at the level of the* generalized chemical substance. *The set of all* generalized chemical substances *is called GS.*

A problem solved in definition 5 is that in some cases the real-valued proportion can be specified in two ways. Consider wüstite with the formula $Fe_{0.95}O$:

1. "$Fe_{0.95}$" is regarded as a nonstoichiometric group, so that at the level of the compound only integer-valued numbers appear.

2. "Fe" is regarded as a (normal) group according to definition 1, so that the number 0.95 enters at the level of the substance.

The latter possibility is weakly preferred because it is closer to what chemists would regard as the proper way of putting the matter. This is reflected by the demand in definition 5 that the real-valued expression be put in at the level of the substance if there is a choice.

Definitions 4 and 5 allow fine-grained distinctions between, for instance, uranium dioxides such as UO_2 (the stoichiometric variant), $UO_{2.14}$, $U_{0.99}O_2$, and $U_{0.99}O_{2.14}$ (compare the CAS registry records described in ref 10).

4.5. Phases. The definition of a homogeneous system or phase is now straightforward. The term *phase* is preferred here because it emphasizes the role played in heterogeneous systems. A phase is characterized by its chemical composition (the parts) and by its aggregation state (the configuration information). We need a new set of atomic concepts, A, comprising aggregation states. The members of A correspond to the different aggregation states (solid, liquid crystal, liquid, vitreous, vaporous). A finer distinction of "solid" into atomic concepts for crystal symmetries might be made. The introduction of both a vitreous and a liquid state obviates the need for a theoretical discussion on the nature of these states (distinct or not).

The following definition can now be given for a phase:

Definition 6 (Phase). *A* phase p *is a tuple*

$$p = \langle s,a\rangle \quad \text{with} \quad a \in A$$

Here, s *is a set of tuples* s_i *such that for each* s_i

$$s_i = \langle c_i,w_i\rangle$$

Here, the c_i *stand for components, each being a* generalized chemical substance, *and the* w_i *stand for proportions such that their sum is 1. The set of all phases is called PH.*

4.6. Heterogeneous Systems. The next step consists of a definition of a heterogeneous system. This definition takes the form of a set of phases and lists their proportions. By convention in this ontology, a system can be heterogeneous even when it consists of just one phase. The demarcation

between homogeneous and heterogeneous systems is made by inspecting the system for the presence of boundary planes: if they are absent, the system is homogeneous, else heterogeneous. In this way, a monocrystalline system (homogeneous) can be distinguished from a polycrystalline system (heterogeneous), even if the system consists of just one pure substance or nonstoichiometric compound.

Definition 7 (Heterogeneous System). *A* heterogeneous system *h is a set of tuples* h_i *such that for each* h_i

$$h_i = \langle p_i, w_i \rangle$$

Here, each p_i *is a phase and the* w_i *stand for proportions such that their sum is 1. The set of all heterogeneous systems is called HS.*

5. CONCLUSION

It has been demonstrated that a relatively rich ontology for the chemical domain can be built with modest means by making use of the principle of a conceptual construction kit. The construction rules for concepts denoting pure substances, phases, and heterogeneous systems are sufficiently general to be of use as a first step toward a standardized ontology of chemistry. The construction rules mirror the way chemists organize their subject matter and can in fact be regarded as a systematized version of this organization. Future work will have to extend the present framework to embrace all concepts of interest. One extension deals with quantities and measurement units[9] and will be supplemented by construction rules for properties. Another extension has to deal with concepts denoting materials and samples of materials. These extensions result in a complete ontology for Plinius. This work is almost completed. Finally, operational specification of other domains will have to result in further extensions to the ontology. This work is presently underway.

ACKNOWLEDGMENT

Engineered Materials Abstracts is published by Materials Information, a joint service of the Institute of Materials, London, and ASM International, Metals Park, OH. We are indebted to Materials Information for providing us with the source material.

REFERENCES AND NOTES

(1) (a) *Proceedings ECAI-Workshop W4. Knowledge Sharing and Reuse: Ways and Means*, Vienna, 1992; Mars, N. J. I., Ed.

(2) Bebak, H.; Buse, C.; Donner, W. T.; Höver, P.; Jacob, H.; Klaus, H.; Pesch, J.; Römelt, J.; Schilling, P.; Woost, B.; Zirz, C. The Standard Molecular Data Format (SMD Format) as an Integration Tool in Computer Chemistry. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 1–5.

(3) Hayes, P. J. In *Formal Theories of the Commonsense World*; Hobbs, J. R., Moore, R. C., Eds.; Ablex: Norwood, NJ, 1985; pp 71–107.

(4) Neches, R.; Fikes, R.; Finin, T.; Gruber, T.; Patil, R.; Senator, T.; Swartout, W. R. Enabling Technology for Knowledge Sharing. *AI Mag.* **1991**, *12* (3), 36–56.

(5) Gruber, T. R. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, Cambridge, MA, April 22–25, 1991; Allen, J., Fikes, R., Sandewall, E., Eds.; 1991; pp 601–602.

(6) Gordon, J. E.; Brockwell, J. C. Chemical Inference. 1. Formalization of the Language of Organic Chemistry: Generic Structural Formulas. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 117–134.

(7) Gordon, J. E. Chemical Inference. 2. Formalization of the Language of Organic Chemistry: Generic Systematic Nomenclature. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 81–92.

(8) Gordon, J. E. Chemical Inference. 3. Formalization of the Language of Relational Chemistry: Ontology and Algebra. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 100–115.

(9) Mars, N. J. I. In *Proceedings of the International Workshop on Formal Ontology in Conceptual Analysis and Knowledge Representation*; March 17–19 1993, Padova, Italy; Guarino, N., Poli, R., Eds.; 1993; p 297–303.

(10) Moulton, C. W. Composition: a Critical Property for Chemical and Material Databases. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 27–30.

(11) Gragg, C. E. Recent ASTM Standardization Developments for Chemical Information. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 18–21.

(12) Mars, N. J. I.; Vet, P. E. van der. In *TKE'90: Terminology and Knowledge Engineering*; Czap, H., Nedobity, W., Eds.; Indeks Verlag: Frankfurt a.M., 1990; pp 352–362.

(13) Westbrook, J. H. Problems in the Computerization of Chemical Information: Capture of Tabular and Graphical Data. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 6–17.

(14) Vet, P. E. van der; Mars, N. J. I. *An Ontology of Ceramics*; Version 2; Memorandum UT-KBS-99-13; University of Twente: Enschede, The Netherlands, 1993.

(15) Raskin, V. In *Advances in Artificial Intelligence: Natural Language and Knowledge Based Systems*; Golumbic, M. C., Ed.; Springer: New York, 1990; pp 114–128.

(16) Brachman, R. J.; Schmolze, J. G. An Overview of the KL-ONE Knowledge Representation System. *Cognit. Sci.* **1985**, *9*, 171–216.

(17) Speel, P. H.; Vet, P. E. van der; Stal, W. G. ter; Mars, N. J. I. In *Proceedings of the Poster Session of ISMIS'93*; Ras, Z. W., Ed.; Oak Ridge National Laboratory: Oak Ridge, TN, 1993, to be published.

(18) *Nomenclature of Inorganic Chemistry. IUPAC Recommendations 1990*; Leigh, G. J., Ed.; Blackwell: Oxford, U.K., 1990.

(19) Speel, P. H.; Mars, N. J. I.; Vet, P. E. van der. In *Proceedings of the Workshop on Language & Information Processing*, Oct 27, 1991, Washington, DC, held at the 54th ASIS Annual Meeting; McCray, A. T., Ed.; 1991; pp 49–58.

(20) Napoli, A. In *Proceedings Tenth European Conference on Artificial Intelligence*, Vienna, Aug 3–7 1992; Neumann, B., Ed.; Wiley: Chichester, 1992; pp 425–429.

(21) Eggert, A. A.; Jacob, A. T.; Middlecamp, C. H. Converting Chemical Formulas to Names: An Expert Strategy. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 227–233.