Intermetallic Compounds – Principles and Practice – Volume 3: Progress.
Edited by J. H. Westbrook, R. L. Fleischer

ISBNs 0-471-49315-5 (Hardback); 0-470-84585-6 (Electronic)

Chapter 40

Computer Design of Materials with Artificial Intelligence

Nadezhda N. Kiselyova

A.A. Baikov Institute of Metallurgy and Materials Science
Russian Academy of Sciences, Moscow, Russia

1. Introduction

It is possible to divide the problem of the design of new materials (including metallic materials) into two parts: (1) searching for new substances with predefined properties and (2) development of optimum conditions for the production and treatment of new materials. This chapter is devoted, for the most part, to solution of the first problem and describes the use of computer program systems, called 'artificial intelligence' systems.

The necessity for such a materials design stems from several reasons. Principal among them are the following:

- most simple binary metal systems are well investigated, but substances based on them no longer supply the needs of industry;
- searching for and research on new ternary and higher-order substances, many of which have unique properties, require considerably more time and expenditure than do binary systems;
- while searching for new substances and materials through investigation of all 5– 6 – ... component alloys, the materials scientists would use, just for laboratory experiments, practically all the rare metals of our planet;
- the now common computer design of new machines, devices, buildings, etc. uses only databases (DBs)[1] on properties of existing materials. The further development of automated design must address the development of systems for computer design of new materials with predefined properties;
- science, as with any institution of a human community, cannot expand indefinitely. However, the increasing complication of materials requires either an increase of expenditures for research, or conversion from an extensive approach to an intensive one;
- the computer design of intermetallic substances is the most powerful way of speeding materials science investigations in this area.

At present the search for new inorganic materials is carried out, for the most part, on the basis of the experience and intuition of researchers. The problem of *a priori* prediction of compounds that have not yet been synthesized and evaluations of their properties is one of the most difficult problems of modern materials science. Here the term '*a priori* prediction' means predicting yet unknown substances with predefined properties from only the properties of constituent components – chemical elements or more simple compounds. The following methods offer possibilities for solution of the last problem:

- quantum-mechanical methods (Pettifor, 1983; Shah and Pettifor, 1993; Chelikowsky *et al*., 1993; Cohen, 1986, etc.);
- empirical criteria (including two-dimensional diagrams) (Darken *et al*., 1953; Girgis, 1983; Hume-Rothery and Raynor, 1962; Laves, 1956; Mathias,

[1]A 'Key Terms List' with definitions appears at the end of the chapter.

1955; Rabe *et al.*, 1992; Villars, 1995; Vozdvizhenskii, 1975, etc.);

– multi-dimensional classifying rules (Chen, 1988; Chen *et al.*, 1999; Gulyev and Pavlenko, 1973; Jackson *et al.*, 1998; Kiselyova, 1987, 1993a, 1993b, 1995, 1997; Kiselyova *et al.*, 1977, 1989, 1998a, 1998b; Kiselyova and Burkhanov, 1987, 1989; Kiselyova and Kravchenko, 1992; Kiselyova and Savitskii, 1977, 1979, 1981, 1982, 1983, 1984; Kutolin and Kotyukov, 1978, 1979a, 1979b; Kutolin *et al.*, 1978; Manzanov *et al.*, 1987; Pao *et al.*, 1999; Savitskii *et al.*, 1968, 1977, 1978, 1979, 1980, 1981, 1982a, b, 1990; Savitskii and Gribulya, 1985; Savitskii and Kiselyova, 1978, 1979, 1983, 1984, 1985; Talanov and Frolova, 1981; Villars *et al.*, 2001; Vozdvizhenskii and Falevich, 1973; Yan *et al.*, 1994; Zhou *et al.*, 1989).

These and other approaches are briefly reviewed in another chapter in this volume by Naka and Khan who also present their application to the design of some multiconstituent intermetallics.

The quantum-mechanical approach to calculation of intermetallic compounds has been the most attractive for most physical metallurgists using the methods of modern physics in their researches. However, over the course of the past half-century the achievements of quantum mechanics in *a priori* calculations of complicated intermetallic compounds evoke little enthusiasm, even among the most ardent followers of this approach. The low precision of results of calculations for known phases does not allow prediction of new substances and, at best, only makes it possible to explain known facts. This situation strengthens doubts as to the promise of the quantum-mechanical approach to the design of new metallic substances. Moreover, analysis of the results of quantum-mechanical calculations calls into question whether the differential and integral equations used are adequate to the complexity of condensed metal systems.

Empirical criteria for classification of known substances and for the subsequent *a priori* prediction of alloys, not yet investigated, are most commonly used in materials science. There is a common tendency in empirical sciences, because of the complexity of objects to be investigated, to substitute classification schemes for computational models. Some examples are: Laves' rule (Laves, 1956) for predicting the crystal structure types of some intermetallic compounds, the Hume-Rothery criterion (Hume-Rothery and Raynor, 1962) and Darken–Gurry diagrams (Darken *et al.*, 1953) for predicting mutual solubility of metals, the Mathias criterion (Mathias, 1955) for predicting new superconductors with the $A15$ crystal structure type, etc. (Girgis, 1983; Rabe *et al.*, 1992; Villars, 1995 and Vozdvizhenskii, 1975). Frequently these rules are named for their founders and are the result of labour-intensive analysis of experimental data but are not a consequence of any theoretical calculations. Moreover, in most cases theoretical physics cannot even explain the reason for the successful implementation of such rules. The principle of the development of empirical criteria of this type is a search for such properties of the chemical elements or analytical functions of these properties which would allow one to find a one- or two-dimensional space, in which it would be possible to divide known substances into distinct domains. The advantage of this approach consists doubtlessly in its simplicity and the ability to visualize the results with the help of one- or two-dimensional plots. The essential shortcomings, from our point of view are the following:

– these criteria quickly lose their reliability when new data do not easily fit within the framework outlined by the classification rule;

– the laboriousness of the development of the criteria;

– the criteria do not take into account the whole set of properties of the chemical elements (or simple compounds) which determine membership of a given substance in a certain class, a circumstance that frequently brings about intersection of classes.

Search for multidimensional classifying rules has become possible, using computers supplied with special programs for data analysis. This approach is a natural evolution of the above-mentioned empirical approach. The application of computers and programs to searching for multidimensional regularities in large volumes of data has allowed sharp reductions in the time of development of new criteria and revision of old criteria with the advent of new data. The solution of the problem of multidimensionality of the experimental data array to be analyzed is limited primarily by the progressively higher capabilities of computers and programs. The advantages of the simplicity in using one- and two-dimensional criteria became immaterial once compact computers were developed, which allow one to instantly predict new substances using multidimensional regularities. The excellent modern computer graphics allows visualization, in accordance with the user's desire, of any section or projection of the multidimensional property space. Thus the classification programs expand the investigator's possibilities

for solution of the problem of searching for multidimensional regularities in large volumes of information. The search for multidimensional rules, connecting possible formation of compounds and correlating their properties with the properties of the chemical elements, is based on the use of databases on inorganic substances and materials and programs for searching for complicated regularities.

2. Databases on Inorganic Substances and Materials Properties

At present thousands of DBs in materials science and chemistry are in operation in the world (Andersson *et al*., 1985; Ansara, 1991; Bale and Eriksson, 1990; Degtyaryov *et al*., 1999; Drago and Kaufman, 1993; Eriguchi and Shimura, 1990; Golikova *et al*., 1989; Ho and Li, 1993; Kiselyova *et al*., 1996; Kravchenko *et al*., 1991; Savitskii *et al*., 1984; STN, 1993; White, 1985; Yudina *et al*., 1996; Zemskov *et al*., 1995, 1998a,b, etc.). Only a small part of them are confined to purely bibliographic information, i.e. contain the abstract or the full texts of publications in a certain subject field as the basic document entries. However, the total information content of such bibliographic giants as the databases of CAS, VINITI, COMPENDEX, etc., containing hundreds and even thousands of gigabytes of data, is highly competitive with the total information content of many factual databases on materials, i.e. those computer-aided systems that contain information about the properties of substances and materials. The overwhelming majority of factual DBs contain information about properties of organic substances. However, a great number of large databases on metals and other inorganic substances and materials are maintained in the world. Shown in Figure 1 is the distribution of databases by subject scope of the information contained. The majority include information about thermal, engineering and physical-chemical properties of inorganic substances and materials. In recent years the tendency has been toward cooperation in the development of DBs and the integration of already developed DBs at national or international levels including cooperation within the frameworks of CODATA and UNESCO. This is because of the opportunities thereby to remove duplication and to cut down the considerable expenses for development and maintenance of such databases. Many DBs are accessible by remote access with the use of telecommunication networks (Drago and Kaufman, 1993; STN, 1993; Degtyaryov *et al*., 1999, etc.).

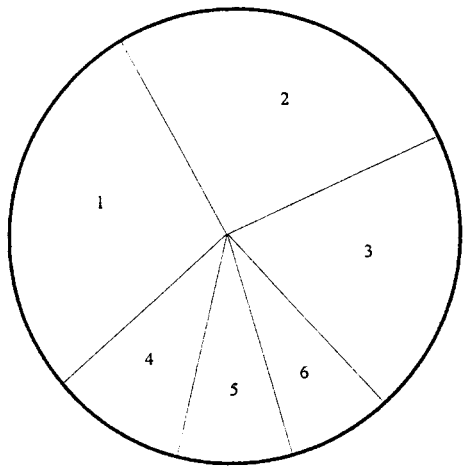

Figure 1 Distribution of databases on properties of inorganic substances and materials over subject scope: 1, thermodynamic properties; 2, engineering properties; 3, chemical and physical-chemical properties; 4, crystallographic and crystal chemical properties; 5, physical (electrical, magnetic, optical, etc.) properties; 6, other properties

The increase in the number and information content of DBs on substance and material properties is a natural tendency of the information age. However, the use of DBs only for information service does not reflect the requirements of society for the acceleration of scientific and technical progress and for the substitution of expensive experimental investigations by computer simulations. The problem of information processing for DBs on the properties of metals and other inorganic substances is particularly acute. The attempts to supply these information systems with programs of thermodynamical calculations, statistics, and so on (Andersson *et al*., 1985; Ansara, 1991, etc.), do not allow good predictions of the properties of inorganic compounds from 'first principles' using only the information of those databases.

We began to develop DBs on materials and material properties in the late seventies. By that time it had become clear that the software of the simplest information retrieval systems did not lend itself to subsequent computer simulation and also that this was an extremely archaic kind of information service. It has now been made obvious that it is necessary to develop DBs with complicated structures, directed towards both computer simulation and information service. Just such principles were assumed as the basis of the information systems of inorganic substances and material properties that we are developing.

The basic ideas, forming the foundation of our databases, are the following. First, databases are

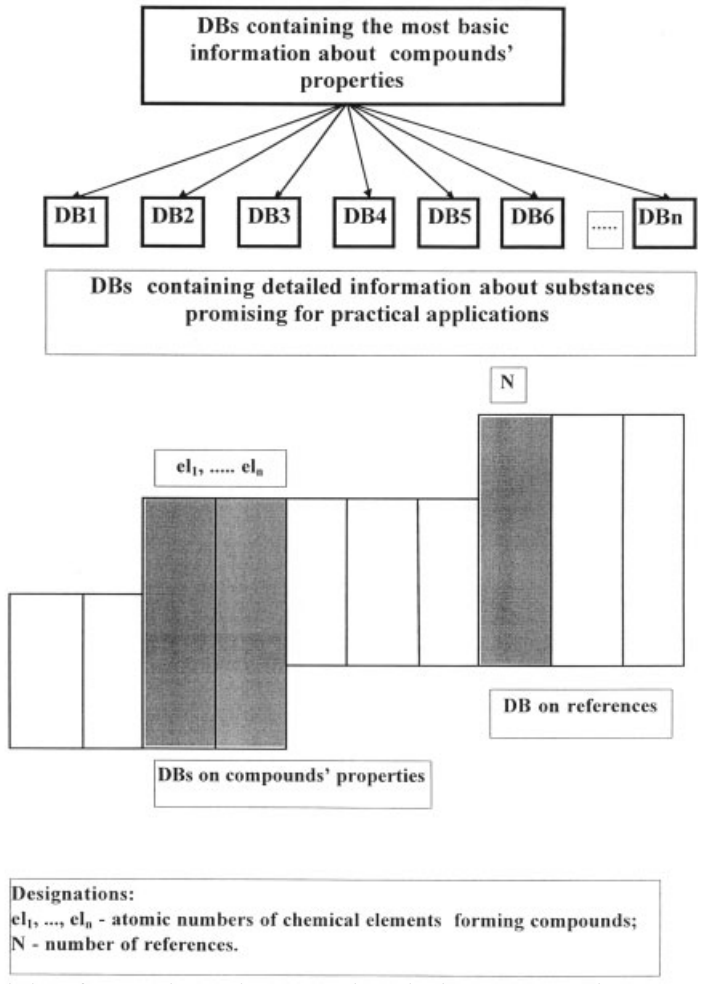

Figure 2 Principles of design of DBs on inorganic compounds' and substances' properties

divided into two classes: DBs containing the most common information about inorganic substances and physical-chemical systems, and DBs that include the most detailed information which was collected and assessed by experts about industrially vital substances (Figure 2). A database of the first type is a DB on the properties of inorganic compounds (Figure 3) containing information about thermal and crystal chemical parameters of compounds (Kiselyova, 1993b, 1997; Kiselyova *et al*., 1996; Kiselyova and Kravchenko, 1992; Savitskii *et al*., 1984). Databases of the second type are, for example, a DB on phase diagrams of systems with semiconducting phases (Zemskov *et al*., 1995, 1998a, b) (Figures 4–6) or a DB on properties of single crystals of acousto-optical, electro-optical and nonlinear-optical compounds (Degtyaryov *et al*., 1999; Golikova *et al*., 1989; Kravchenko *et al*., 1991; Yudina *et al*., 1996) (Figure 7). Secondly, we have completely departed from the philosophy of bibliographic DBs, whose factual elements for each document are data on the publication (paper, monograph or handbook) and in which any information about the substance concerned is added only as supplement. Such a philosophy increases the time for data retrieval for specific substances or systems, especially in the case of very large DBs, and does not provide full relevance (i.e. extraction of all documents which are appropriate to a certain request). The basic document of our DBs is

cumulative information about a particular system or substance, the key identifiers of which are the symbols (or atomic numbers) of the chemical elements forming the system (Figure 2). The references are collected in a separate DB (or relational table) with consecutive numbering. The databases (or tables) on the properties of the compounds (or systems) contain only the numbers of the references. Thus DBs (or the tables) on the properties of compounds (or systems) intersect each other at fields of constituent components (the quantitative composition of compounds and/or types of crystal structures) and with databases (or tables) of the references – at reference numbers. Such a DB structure optimizes fast data retrieval and preserves full relevance. These principles are used as the basis for the following DBs developed and maintained by us:

1. A DB on the properties of ternary inorganic compounds 'PHASES' (Kiselyova, 1993b, 1997; Kiselyova *et al.*, 1996; Kiselyova and Kravchenko, 1992; Savitskii *et al.*, 1984) contains the following information (Figure 3) on more than 39 000 ternary compounds in more than 16 000 systems including the elements from H(1) to No(102). Information is extracted from more than 12 000 publications. The retrospective data cover more than 70 years. The DB has been developed for IBM PCs operating under MS-DOS. The database management system (DBMS) is DATAREAL (Kiselyov, 1991). This DB is popular among

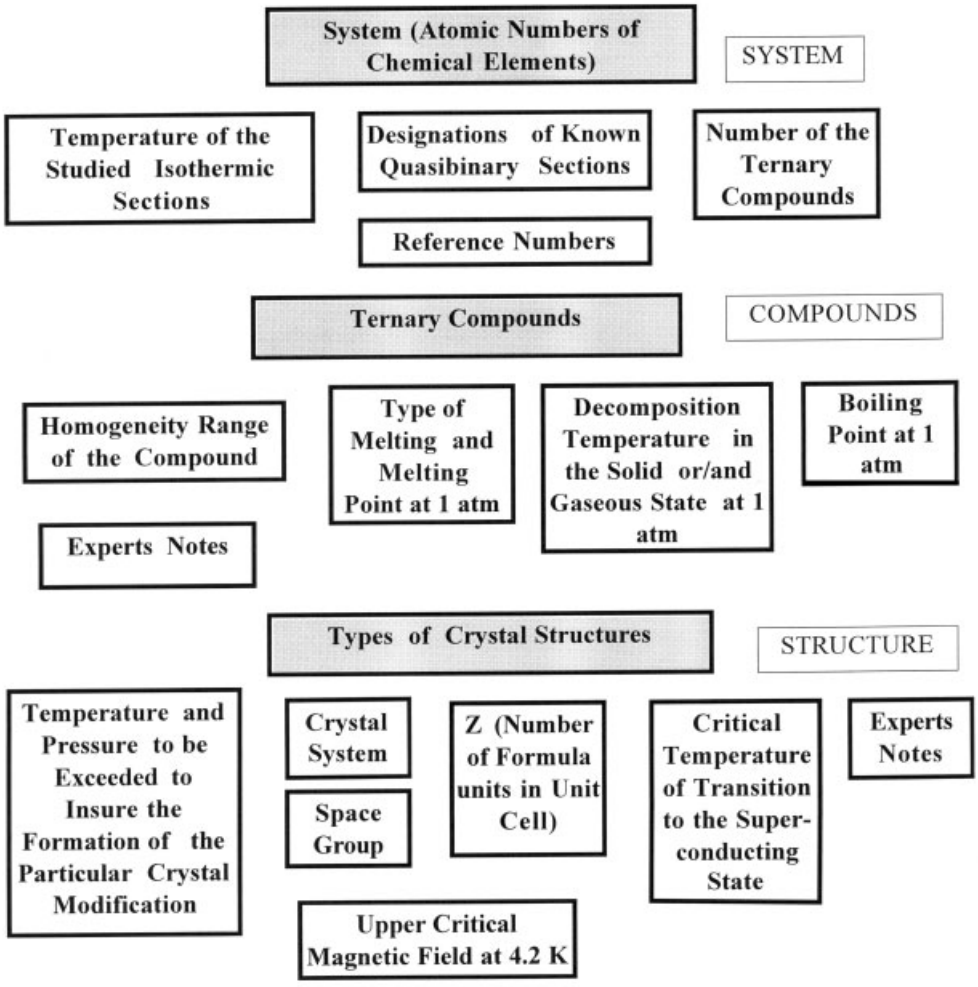

Figure 3 Structure of database 'PHASES' on ternary inorganic compounds

Russian users; therefore development of a new version of the DB with Internet-access has begun. Some of the data have been assessed by experts.

Total data assessment by experts is a feature of DBs on the properties of substances and materials for electronics, which we began to develop during the 1990s.

2. A database of phase diagrams of systems with intermediate semiconducting phases 'DIAGRAM' (Zemskov *et al.*, 1995, 1998a, b) contains information on the most important pressure–temperature–concentration phase diagrams of semiconducting systems evaluated by qualified experts and also on the physical-chemical properties of the intermediate phases. The figures of the diagrams are only presented after critical assessment, statistical optimization (using expert evaluation of the data of different researchers) or thermodynamical self-consistency. Figures show the information that is stored for every binary system (Figure 4) and every ternary system (Figure 5); Figure 6 shows the structure of the DB on references. The DB 'DIAGRAM' includes detailed analytical reviews for each system – mini-monographs of a sort – that reflect the extent of investigation of the system. Thermodynamic parameters and computational models, which were utilized for thermodynamic self-consistency or statistical optimization of data of the different investigators, are stored separately. Apart from information about semiconducting systems, this DB contains data about some binary metallic systems that are constituent parts of ternary systems with semiconducting phases, and also data about crystal structure of the chemical elements. The database of phase diagrams now contains information on several tens of semiconducting systems. Data retrieval is carried out by dialog-based menus and screening forms. Updating of the database is carried out every month. This DB has been developed for the IBM PCs operating under MS-DOS (DBMS = DATAREAL). We also developed a new Internet-version of this DB.
3. A database of crystals with significant acousto-, electro- and nonlinear-optical properties, 'CRYSTAL' (Degtyaryov *et al.*, 1999; Golikova *et al.*, 1989; Kravchenko *et al.*, 1991; Yudina *et al.*, 1996) contains information on crystals of the most important substances of this class as evaluated by experts. The information contained in the database about the properties of the crystals

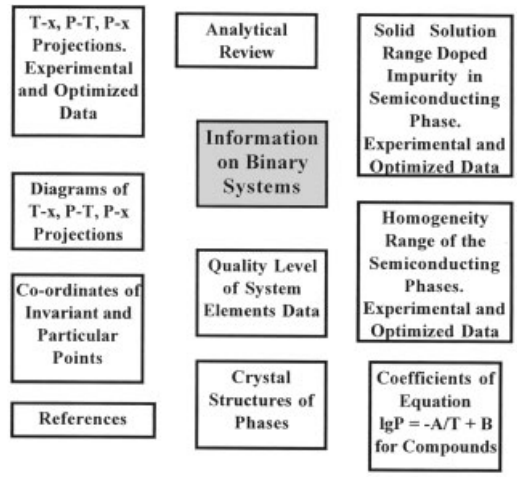

Figure 4 Structure of DB on binary semiconducting systems

is displayed in Figure 7. In addition, the database includes extensive graphical information about the properties of the substances. At present data on several tens of substances are stored in this database. A version of this DB for Internet-users is now in the making. In addition to a DB in Russian, a version in English has also been developed.

3. Application of Artificial Intelligence Methods for the Analysis of Information from Databases

Doubtlessly databases on the properties of substances and materials open new avenues for information

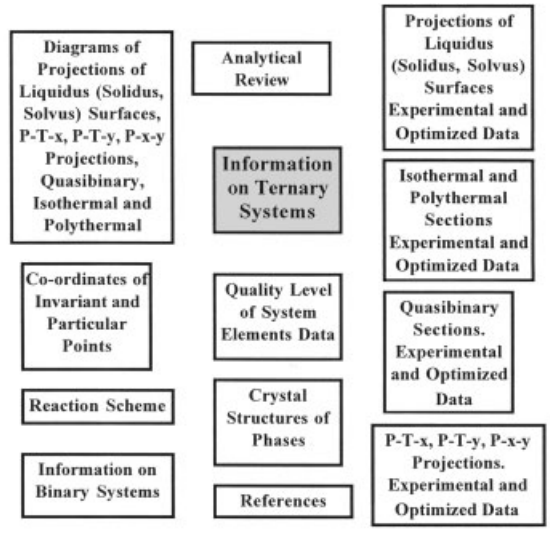

Figure 5 Structure of DB on ternary semiconducting systems

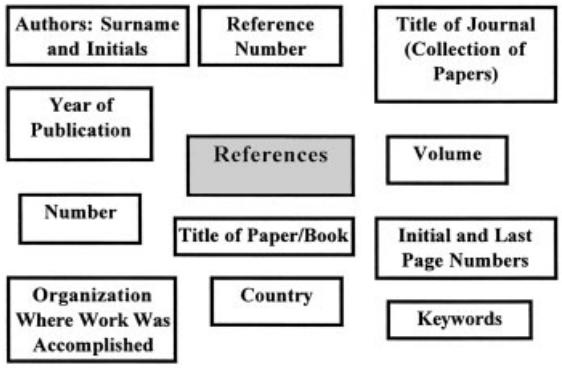

Figure 6 Structure of DB on references

service for specialists. However, it is but one of the aspects of the new information technology. Rational use of stored data implies their processing with the purpose of searching for regularities that could be applied: to the prediction of new substances with predefined properties; to the development of the technology of synthesis of new materials; and to the prediction of the behavior of materials under the effects of various factors, i.e. automation of the practice of materials science.

This problem can be decided easily in those rare cases where there is a good analytical description of the regularities to be sought, and the specialist needs only to insert the necessary information from the database into one of the chosen models. Among computer-aided systems of this kind are numerous databases on thermodynamic properties of substances provided with programs of thermodynamic calculations: IVTANTHERMO (Ansara, 1991), THERMOCALC (Andersson *et al.*, 1985), etc. However, the majority of materials science problems cannot be formalized with the application of only those simple algebraic structures that are used, for example, in thermodynamics. Prediction of new substances with predefined properties, interpretation of spectral information, selection of substances for certain purposes, development of optimal technological processes for the synthesis of materials, separation and identification of substances, etc. belong to this class of intractable problems. All

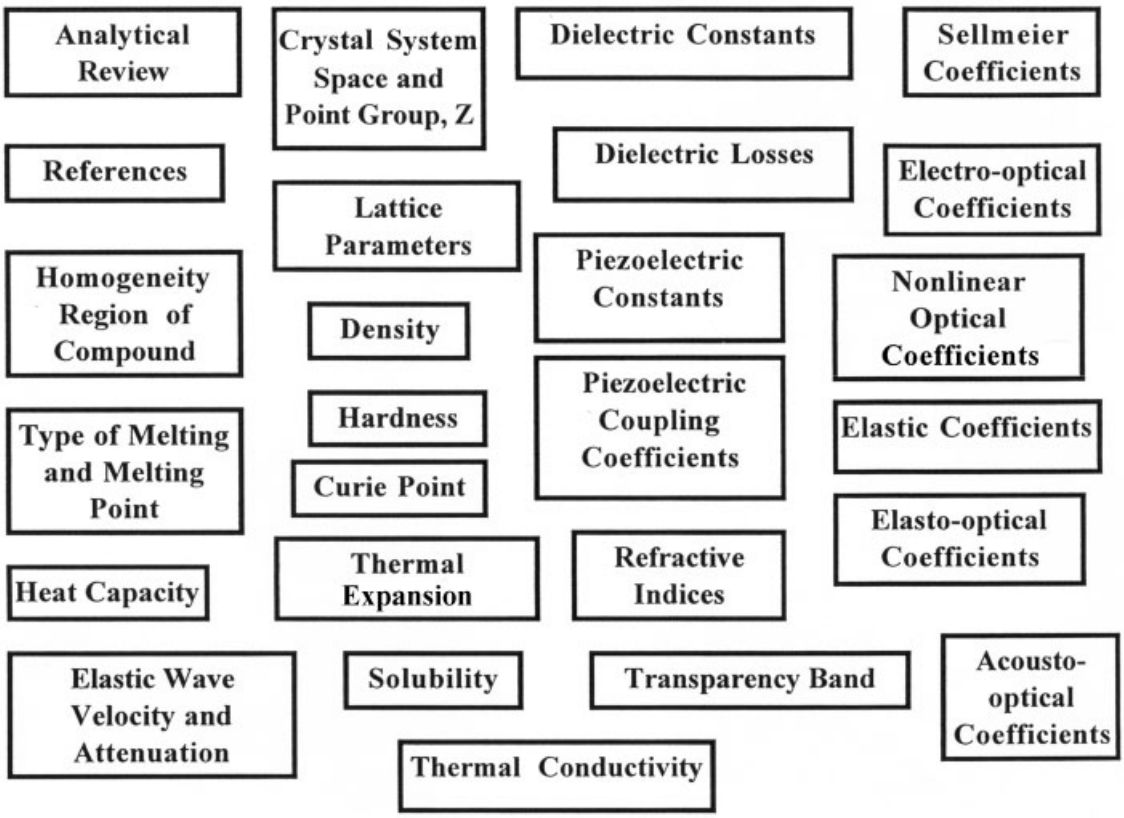

Figure 7 Structure of database 'CRYSTAL' on crystals with acousto-, electro- and nonlinear-optical properties

these problems are presently solved only from the experience and intuition of the investigators. One of the most effective ways to automate these fields is by the application of the ideas of artificial intelligence and knowledge bases.

What are the problems that appear with the intellectualization of materials databases (Gladun, 1995; Gladun and Vashchenko, 1995; Kiselyova, 1997; Popov, 1987; Pospelov, 1988; Pospelov and Pospelov, 1985; Zagoruiko, 1999)? First, it is necessary that the computer understand the professional language and the statement of the problem of the user, i.e. the intelligent system should have two sorts of knowledge: knowledge of language and knowledge of the field in which the user works. Knowledge of the first sort is realized at the level of a *conversational, or linguistic processor*. It can be a system of special programs or a complex including both hardware and software. As a rule, the knowledge of language is stored directly in the conversional processor. The knowledge about the subject field is stored in a special *knowledge base*.

The terms 'knowledge' and 'knowledge base' as applied to computer information appeared in the 1970s during the development of artificial intelligence systems. What are the distinctions between 'data' and 'knowledge'? Data are values used for solution of problems and metadata (Westbrook and Grattidge, 1991a) for descriptions of objects, situations, phenomena, and connections between objects. Knowledge is information about the processes of solution, the regularities which, applied to data, generate new information (Gladun, 1995). Particular features, distinguishing knowledge from data in connection with their representation in computers (Pospelov, 1988; Pospelov and Pospelov, 1985), are the following: *interior interpretability* which makes it possible for the computer to 'understand' the information to be input at a substantial level; *availability of structure* which provides a computer with the ability to form a hierarchy of concepts, to introduce new generalized concepts and to decompose concepts into constituent subconcepts and the relations between them; *availability of connections* that provides a computer with the possibility of including the connections not only between concepts but also between the facts, processes and phenomena; *activity* is the feature that relates the computer to *homo sapiens* and is connected with actions leading to a realization of procedures that can be useful for the solution of certain problems (for example, the detection of a contradiction between the prediction and an experimental fact becomes the stimulation for overcoming it and forming new knowledge). It is, however, impossible to demark an accurate division between knowledge and data. For example, interpretability is inherent to relational databases, and structuring is an integral part of all modern DBs. But availability of connections and activity have no parallel in a methodology of DBs.

Secondly, it is important to have a system that converts the description of a source problem into a running program. A complex of software for the solution of this problem is called a *program scheduler*, or simply, a *scheduler*. During its work the scheduler continually contacts the knowledge base, getting from the base the knowledge of the application domain, the methods for solution of tasks, and information about the possibility of an automated combination of programs from some of the basic programs which are stored in the knowledge base. A special system named the *monitor* realizes a control by interaction of all subsystems. The complex of conversational processor, knowledge base, scheduler, and monitor forms the *intelligent interface* of the computer (Pospelov, 1988; Pospelov and Pospelov, 1985).

Expert systems are the most widespread kind of artificial intelligence systems. They date back about 20 years and are intended to solve very complicated problems in particular application domains with the use of large volumes of special knowledge of a high quality. The latter are extracted from various sources, namely: books, papers, scientific and technical documents, domain experts, etc. Such knowledge also includes some procedures, strategies, empiric regularities, and so on. This knowledge is represented in a special manner and is stored in the knowledge base. It will be noted that expert systems use models based on special formalisms of artificial intelligence (Gladun, 1995; Gladun and Vashchenko, 1995; Popov, 1987; Pospelov, 1988; Pospelov and Pospelov, 1985; Zagoruiko, 1999). Unfortunately, many developers don't take this aspect into consideration. These developers use the fashionable term *expert system* for the definition of their program systems that use a conversational mode similar to the natural language. True expert systems are artificial intelligence systems that use knowledge represented as rules, frames, or semantic networks (Gladun, 1995; Gladun and Vashchenko, 1995; Popov, 1987; Pospelov, 1988; Pospelov and Pospelov, 1985; Zagoruiko, 1999). A particular feature of expert systems is a subsystem of explanations that is a constituent part of these systems. It controls the work of the scheduler and describes its functioning in a condensed form that is convenient for the specialist. It fosters trust in the work of the intelligent system and

answers for the user the questions 'How?' and 'Why?' one or another solution is accepted as an alternative choice.

A particular feature of artificial intelligence systems is an ability for automatic searching for regularities and use of them for prediction. Thus expert systems are systems with 'poor' intelligence because they use only those regularities (knowledge) which are extracted from the experts. As our experience shows, any attempt to extract rules, connecting the formation of a certain compound to the properties of its constituent elements, from chemists or physical metallurgists have been unsuccessful, because they prefer only to estimate the authenticity of data concerning the existence of the given compound or the values of the particular properties of the phase. While formulating the specific features of this application domain, we abandoned the idea of making the chemist or physical metallurgist outline the rules of the formation of metallic and other inorganic substances with desired properties and decided only to make use of their expert assessment of the data intended for computer learning. It should be noted that an expert system is a passive system that is not capable of obtaining new regularities or of searching for contradictions between knowledge and the facts. This circumstance, in combination with the difficulties of the extraction of knowledge from the experts, were the reasons for prior failures of the development of expert systems in physical metallurgy and inorganic materials science. Further development of intelligent systems requires a removal of these restrictions and realization of what are called, *partner systems* (Zagoruiko, 1999).

4. Information-Predicting System for the Design of New Inorganic Substances

The A. A. Baikov Institute of Metallurgy and Materials Science of the Russian Academy of Sciences is now making a version of a partner system – an information-predicting, computer-aided system. This system is intended for data retrieval on known compounds, the prediction of hypothetical inorganic compounds, not yet synthesized, and the forecasting of their properties. This system includes a prediction subsystem based on an artificial intelligence method – the method of concept formation by a growing network CONFOR (Gladun, 1995; Gladun and Vashchenko, 1995). Apart from the latter, the information-predicting system employs a database of properties of inorganic compounds as discussed above, a knowledge base, a conversational processor and monitor (see Figure 8). The system is developed for an IBM computer.

The principles of the working of a system of concept formation, which we use for the prediction of new intermetallic and other inorganic compounds, are detailed in Gladun (1995), Gladun and Vashchenko (1995), and Kiselyova *et al.* (1998). A distinctive feature of the system used is a special associative data structure in the computer memory in the form of a *growing pyramidal network*. This ensures a fast search for regularities in the information contained in the DB which completely classifies known physical-chemical systems, data for which are processed by the computer. During the prediction process the computer receives only the atomic numbers of the elements or designations of simple compounds. A brief review of the data analysis methods from the point of view of their applicability for searching for rules in the information in databases on the properties of inorganic substances and materials is given below (section 4.1).

The *knowledge base* of the information-predicting system stores the rules already obtained for various classes of intermetallic and inorganic compounds for use in the prediction of phases and to forecast the phase properties, unless the database has no such information about the particular physical-chemical system. Rules in the knowledge base are represented in the form of growing pyramidal networks or of their equivalent conjunctive-disjunctive logical expressions.

The *conversational processor* manages the conversation of the user, who may have little experience working with the computer or with the information-predicting system. It provides an expert in the given application domain also with a dialog with the information-predicting system. In the future, the employment of a linguistic processor in the software or some software-hardware support can be expected. It will allow the system to understand the problem-oriented language of the user.

The *monitor* controls the computation process and provides the interface between the functional subsystems as well as telecommunication access to the system. In addition, the monitor signals whenever new experimental data contradict existing classification rules. Such contradictions will be eliminated by including the new data in the computer learning and modifying the rule in the knowledge base.

The information-predicting system operates as follows (see Figure 8). The user requests information about an existing compound of a certain composition. If data about this phase are stored in the database,

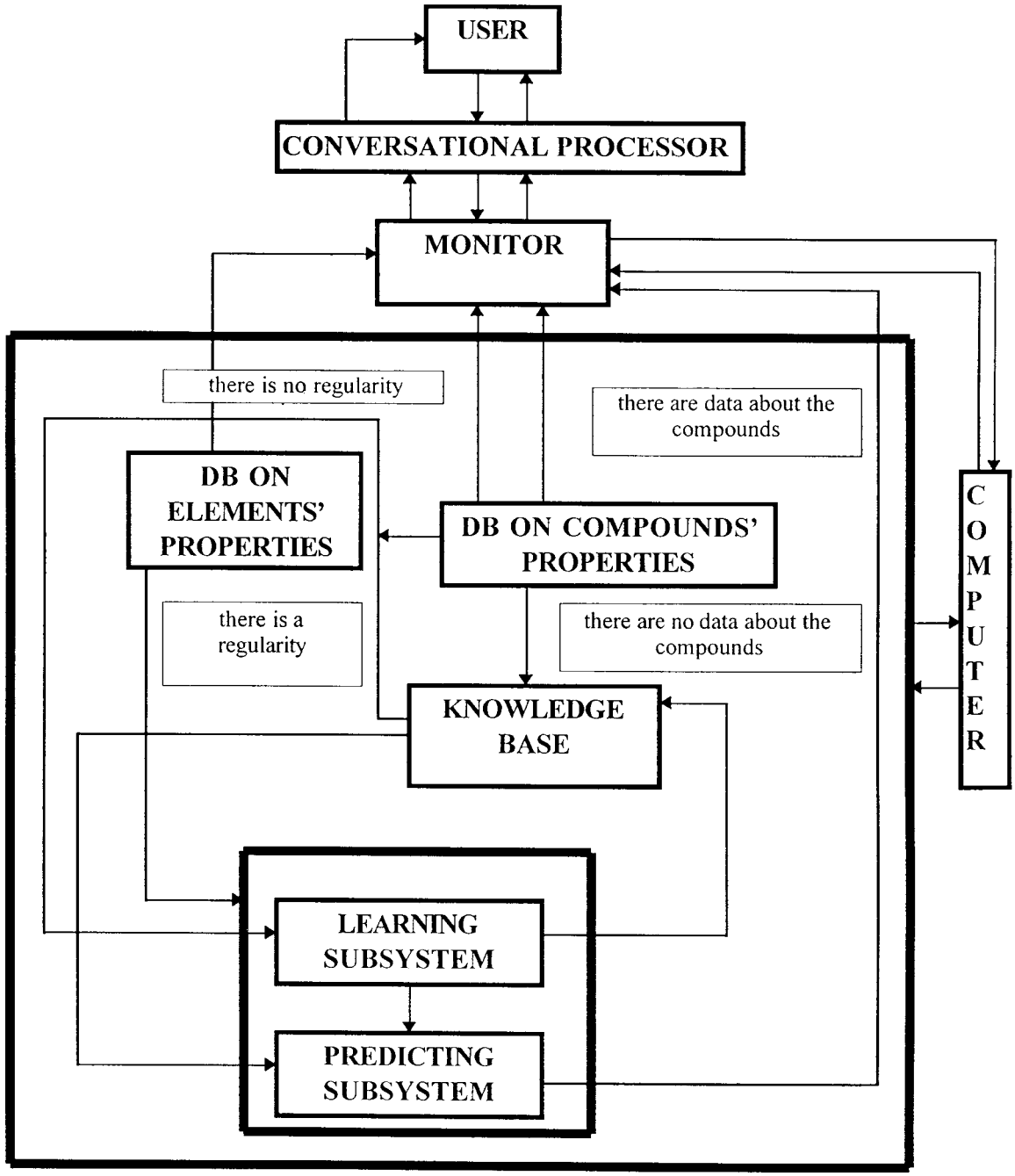

Figure 8 Schematic diagram of an information-predicting system

they can be extracted and used for further studies. If no information about the compound is stored in the database, or if the information available is incomplete, the computer, in response to the user's request, determines whether the rule (network) corresponding to the desired property for a compound of a certain type is present in the knowledge base. If the phase is present, the database supplies the appropriate set of component properties to predict the desired characteristic. If the knowledge base does not have the desired rule, then examples for the computer learning process are searched for in the database. The correctness and representativeness of these examples are estimated once more by the user; and, if the sample is found adequate for computer learning, the learning and prediction subsystems process them in turn. The resultant prediction is received by the user, while the classifying rule thus formed is stored in the knowledge base. The above example is the simplest of the problems that can be solved by an information-predicting system. A more complicated problem would be, for example, predicting all possible phases in ternary and multicomponent systems, combined with the estimation of their properties. Although the previous problem can be solved by real-time processing, the latter problem requires much more time.

The principles underlying the information-predicting system as developed at present have been tested successfully earlier on the prototype system supported by the BESM-6, second-generation computer (Kiselyova and Burkhanov, 1989). The employment of powerful personal computers will make it possible to build a version of the system that can be operated by users at large.

4.1 Application of Artificial Intelligence Methods to Data Analysis

Classical applications of mathematics to natural-science domains are associated with the calculation of one parameter of an object or phenomenon from knowledge of others. Examples are: Ohm's law, Fick's law, etc. Each model of the object is described by some analytical expression in the form of an equation or set of equations or inequalities that are approximations to reality. Models, expressed by formulas, and the use of accurate computing procedures frequently create an impression of high-quality models and hence the validity of the results obtained does not always represent the facts.

More 'inexact' from the point of view of the theoretical physicist, another approach is connected with a solution of problems with identified models (Zagoruiko, 1999). In this case the models of processes or objects to be investigated are known to parametric accuracy, i.e. sets of the most important features and a general view of the dependencies between them are known, but it is necessary to calculate coefficients, exponents and other parameters, using experimental data for the behavior of the objects or the course of the processes. This approach does not ensure comparable rigour of computations as in the conversion from one parameter set to another, because assumptions about the values of unknown parameters and their checking with available experimental data, connecting input and target features, are required. However, this approach gives good results in research on complicated chemical-technological and metallurgical processes, simulation of kinetics or diffusion, etc.

The emerging field of cybernetics with its concept of 'a black box' has allowed an approach to the solution of even more complicated problems in which the investigator has only a set of experimental data with a description of input and output features, and neither the form of the models nor the degree of influence of the input features on the subjects of interest are known. Such problems are decided most effectively with various methods of artificial intelligence. The challenges faced by physical metallurgists, chemists and technologists include: prediction of the formation of chemical compounds, evaluation of their properties, development of models of very complicated, multi-phase technological processes, etc. From the point of view, for example, of the theoretical physicist, the results are not sufficiently rigorously substantiated from physical theories, nor from the rigor of the mathematical methods used. Moreover, the models obtained do not give any possibility to get to the root of the investigated processes or phenomena. But in the absence of well-substantiated theoretical methods, an alternative could be to abandon the offered computational approach and use only experimental methods; that is absolutely incorrect under present conditions of the intensification of scientific research.

By convention, the problem of data analysis by artificial intelligence methods can be divided into three parts:

- automatic classification (or pattern recognition without computer learning);
- computer learning to classify objects (or computer learning in pattern recognition);
- searching for the most important features for classification (or conjunctions of sets of feature values).

The first problem is formulated as follows: it is necessary to divide a set of objects, specified by sets of features, into classes, such that points inside each are close to one another in feature space.

The algorithms to solve this problem are most sensitive to violation of the hypothesis of compactness, the realization of which is a general prerequisite for a correct solution of each of the three types of problems listed above. The objects of each class should form 'compact clusters' in feature space. The problem of automatic classification is the following: it is necessary to determine such clusters and to construct a dividing hypersurface, the use of which will allow determination of the membership of objects to the classes found. The most justified classification of algorithms of pattern recognition without computer learning is given in the review by Dorofeyuk (1971). A set of algorithms for an automatic classification are also reviewed in the monographs by Arkad'ev and Braverman (1971), Izerman *et al.* (1970), and Zagoruiko (1999). The applications of these methods in materials science and chemistry include, among the most interesting: automation of the interpretation of spectra (Jurs and Isenhour, 1975), refinding Mendeleev's law (Zagoruiko, 1999), a search for components of alloys (Gulyev and Pavlenko, 1973), etc.

The second problem – computer learning for pattern recognition – is formulated as follows: let a feature space be given. It is known that it contains a fixed number of domains, the boundaries of which are unknown, and there are no rules for the definition of membership of a particular point to one or another domain. In computer learning, the points, which are sampled randomly from these domains, are entered into the computer, and possibly relevant information about their belonging to one or another domain is given. The purpose of computer learning is the construction of hypersurfaces which divide not only the points shown but also all other points belonging to these domains. The purpose of the consequent prediction consists of an interpolation and extrapolation of the available data from a sample of small volume to the entire set. This problem differs from the previous one only in the rigid specification of classes. The most successful classification methods of this kind and examples of algorithms used are given in the monograph by Zagoruiko (1999). Algorithms of computer learning are widely applied to materials science including our investigations aimed at predicting the existence of new inorganic compounds and evaluating their properties. Here a specific algorithm is used: to teach the computer how to process data for concept formation.

We use the system of computer learning CONFOR (Gladun, 1995; Gladun and Vashchenko, 1995) that represents the initial information about known physical-chemical systems as growing pyramidal networks. A pyramidal network is an acyclic, oriented graph having no vertices with one entering arc. If the processes of concept formation are determined in the network, then the pyramidal network is designated as a growing one (Gladun, 1995). The network is built during the process of input of objects. Each object (a physical-chemical system or compound) is put in as a set of values of the component properties with an indication of the class to which the system belongs. The nearby values of properties of components are united into one interval using a special program or the experience of the researcher. The concept-formation process consists of the analysis of vertices in the network that is built and the choice of those vertices that are the most typical for each class. These vertices become the checking vertices. The resultant concepts (classifying regularities) can be stored in computer memory and printed out or read in the form of learned, growing pyramidal networks (if the resolution of the screen allows it) or an equivalent Boolean expression in which the intervals of values of the component properties constitute the variables. During the prediction process the computer receives only the atomic numbers of the elements or designations of simple compounds. The values of the properties of the appropriate elements or simple compounds are then automatically extracted from the DB, substituted into the growing pyramidal network, and the researcher can easily obtain the necessary predictions.

The third problem – the selection of the most important classifying features – has a dual purpose. First, it is necessary to minimize the initial feature set in order to reduce the time of data processing using pattern recognition algorithms and leave only the (most important) dividing features. Secondly, most practical problems are aimed at selecting those features which are the most typical for the given class and distinguish this class from others. Algorithms of this kind can be classified into two classes: algorithms of *a priori* weighting and algorithms of *a posteriori* weighting. In the first case the process of constructing a classifying regularity is carried out before the evaluation of the importance of the features, and on the contrary for the other class. The classification of algorithms to decide the importance of features is made possible by the type of criteria of importance for class division.

Our experience in the application of various methods of artificial intelligence to data analysis

shows that it is impossible to assess the importance of an individual feature over the whole range of its values. As a rule, some feature is of importance for classification, only in a certain range and for certain kinds of substances. In most cases, chemical phenomena are determined by a set of features and thus, it makes sense to speak only about the importance of sets of features over specific intervals. We now explain this final proposition. Illustrated in Figure 9 is an example of a two-dimensional map for three classes of objects designated A, B and C. Separation of objects into classes is observed only within narrow ranges of the values of features X_1 and X_2. The features are not *generally* important for the separation of classes, if a mixture of classes is observed practically everywhere over the range of their variation. It can be seen in this example that the features are of importance only within certain intervals.

5. Prediction of New Intermetallics and Other Inorganic Compounds Using Artificial Intelligence Methods

The idea of the application of computer learning methods to searching for regularities of formation of binary intermetallic compounds was put forward for the first time by my colleagues from the Baikov Institute: E. M. Savitskii and V. B. Gribulya in 1966 (Savitskii *et al.*, 1968). They successfully resolved a great number of problems of predicting various types of binary systems (for example, those with full mutual solubility or with simple eutectics); they predicted the formation of thousands of binary compounds with compositions AB, A_2B, A_3B, etc. and evaluated some of their properties (type of crystal structure, melting point, critical temperature of transition to the superconducting state, etc.) (Savitskii and Gribulya, 1985; Savitskii *et al.*, 1968, 1979, 1980, 1981, 1982a, b, 1990) using simple methods of computer learning in pattern recognition (Devingtal, 1968, 1971). Data in the computer memory were represented only by information about the distribution of electrons in the energy levels of isolated atoms of the chemical elements. Even such a simple description of binary systems allowed the prediction of new compounds and forecasting of their properties with an average reliability of more than 90%. Thus the properties of binary systems depend strongly on the properties of the constituent chemical elements.

The further development of this approach has followed two interrelated directions: introduction of

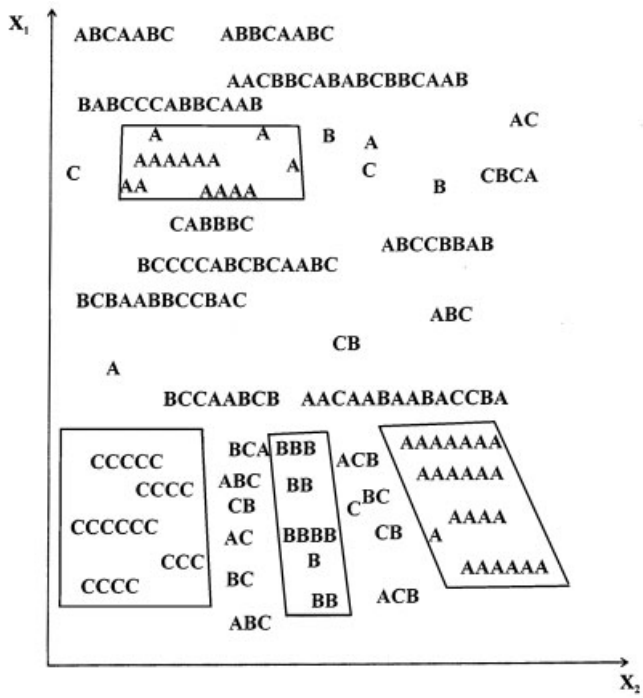

Figure 9 Map illustrating that features of importance are valid only over discrete intervals of parameters $\mathbf{X}_1$, and $\mathbf{X}_2$

the complication of the composition of the physical-chemical system, and the development of new predicting systems (Kiselyova, 1987, 1993a, 1993b, 1995, 1997; Kiselyova *et al.*, 1977, 1989, 1998a, 1998b; Kiselyova and Burkhanov, 1987, 1989; Kiselyova and Kravchenko, 1992; Kiselyova and Savitskii, 1977, 1979, 1981, 1982, 1983, 1984; Manzanov *et al.*, 1987; Savitskii *et al.*, 1977, 1978, 1979, 1980, 1981, 1982, 1990; Savitskii and Kiselyova, 1978, 1979, 1983, 1984, 1985; Talanov and Frolova, 1981).

Thousands of new compounds and their properties in ternary, quaternary and more complicated systems were predicted using this approach. The systems were described using diverse properties of the chemical elements (distribution of electrons in the energy levels of isolated atoms of the chemical elements, ionization potentials, thermal and thermodynamic properties, ionic or atomic radii, etc.) and properties of binary compounds, and also various functions of these properties (ratios of radii, electronic concentrations, etc.). The reliability of predictions depended on the algorithm used, on the correctness of classification of the analyzed data and how representative they were, and also on a good choice of properties for the description of certain classes of physical-chemical systems.

The search for, and development of, effective predicting systems were aimed at the creation of more powerful programs capable of analyzing, on the one hand, very large banks of experimental information, and, on the other hand, of allowing construction of multidimensional classification rules under the condition of small sets. Improvements in electronics allowed the development of systems with a user-friendly interface, working in real time (Chen *et al.*, 1999; Gladun, 1995; Gladun and Vashchenko, 1995; Pao *et al.*, 1999). The trend has been a transition from the simplest algorithms of pattern recognition (Gulyev and Pavlenko, 1973; Kutolin and Kotyukov, 1978, 1979a, 1979b; Kutolin *et al.*, 1978; Savitskii *et al.*, 1968; Savitskii and Gribulya, 1985; Talanov and Frolova, 1981; Vozdvizhenskii and Falevich, 1973) toward more powerful methods based on the use of neural and semantic networks (Kiselyova, 1987, 1993a, 1993b, 1995, 1997; Kiselyova *et al.*, 1977, 1989, 1998a, 1998b; Kiselyova and Burkhanov, 1987, 1989; Kiselyova and Kravchenko, 1992; Kiselyova and Savitskii, 1977, 1979, 1981, 1982, 1983, 1984; Manzanov *et al.*, 1987; Pao, 1999; Savitskii *et al.*, 1977, 1978, 1979, 1980, 1981, 1982a, b, 1990; Savitskii and Kiselyova, 1978, 1979, 1983, 1984, 1985; Villars *et al.*, 2001; Yan *et al.*, 1994).

However, the most important result of research in this field is that artificial intelligence methods have become an operating tool for searching for regularities in experimental data, and the use of these regularities for predicting new intermetallic and other inorganic substances has been achieved. The approach, which had been developed in the A. A. Baikov Institute of Metallurgy and Materials Science, now has followers in different countries. The most powerful groups work in the Baikov Institute in Russia, in the USAF Wright Laboratory in Ohio under the leadership of Dr. S. R. LeClair and in the Institute of Metallurgy of the Chinese Academy of Sciences and in Shanghai University under the leadership of Prof. N. Y. Chen.

Let us consider in greater detail the investigations of the Baikov Institute on predicting ternary intermetallic compounds. The problem of predicting new substances with desired properties can be divided into four consecutive problems:

- prediction of compound formation or non-formation for ternary systems;
- prediction of ternary compounds of desired composition;
- prediction of phases with a specific crystal structure type;
- estimation of quantitative properties of the phase (critical temperature of transition to the superconducting state, homogeneity region etc.).

These problems can be solved sequentially, or any one of these tasks can be solved separately, as examples of the complementary classes are added to the learning set. For example, in the case of predicting the crystal structure at room temperature and atmospheric pressure for compounds of a desired composition, it is necessary to include in the learning set examples of both the formation and non-formation of the compounds in such systems under these conditions.

In most cases prediction is carried out for systems at normal conditions, for example, the prediction of a phase at ambient pressure and temperature. In order to predict phases that exist under other conditions (e.g. high pressure), it is necessary to enter examples of known compounds that exist at high pressures into the learning set and add pressure as a parameter. Unfortunately, a majority of the experimental measurements of compound properties are carried out under conditions of incomplete equilibrium. In addition, the determination of the crystal structure of a given compound is not often undertaken in conjunction with its phase-diagram investigation. Therefore, it is not always clear under what conditions a specific crystal polymorph is stable. Standardization of the

presentation of data for compound properties is a task for the future; meanwhile, in extracting examples for computer learning, we have to run a risk – which is not always rewarding. Taking into account the dependence of compound properties on the conditions of synthesis, it is possible to enter processing parameters for the production of the substance into the learning set, and further to predict a method of synthesis which will enable the target properties to be achieved most efficiently. Various program versions of this concept formation method (Gladun 1995, Gladun and Vashchenko 1995) were used in the calculations to be mentioned below.

5.1 Prediction of the Formation of Compounds with Composition ABX_2 (X=S, Se, Te)

The chalcogenides with composition ABX_2 are a class of compounds that is promising for the development of new semiconducting, electro-optical, acousto-optical, and other materials for electronics.

Each system A-B-X (hereafter A and B indicate any chemical elements) was represented in the computer memory as a set of especially (Gladun and Vashchenko, 1995) coded values of properties of the components A and B, whose class (formation or non-formation of a compound with composition ABX_2 in various systems) was chosen as a target feature. Searches for regularities and predictions were carried out separately for the systems A-B-S, A-B-Se and A-B-Te. Based on physical and chemical understanding of the nature of substances of this kind, three sets of component properties were chosen for description of these chalcogenide systems:

1. The distribution of electrons in the energy levels of the isolated atoms of the chemical elements A and B and their formal valences in ABX_2 compounds.
2. The types of incomplete electronic shells, the Pauling electronegativities, the covalent radii according to Bokii and Belov, the formal valences of the elements A and B in these compounds, and the enthalpies of formation of appropriate simple chalcogenides.
3. The covalent radii by Bokii and Belov, the standard entropies and enthalpies of formation of appropriate simple chalcogenides.

Regularity classifications and the predictions of the formation of unknown compounds with composition ABX_2 were obtained separately for each of the three sets of properties of the constituent components. From such a dichotomy, the method of Gladun (1995) and Gladun and Vashchenko (1995) produces results of three types: (1) formation of a compound of composition ABX_2 in a system A-B-X is designated in Table 1 by the sign '+', (2) non-formation of a compound of composition ABX_2 in a system A-B-X under normal conditions is designated by the sign '−', and (3) in the event, that the unknown system does not bear similarities to any of the objects of the learning set, the prediction is indefinite. On the basis that the properties of the chemical elements should depend periodically on their atomic numbers, it is hoped that the results of the prediction with the use of the various sets of properties of the elements and of simple chalcogenides coincide. Failures (empty squares in tables of the predictions) arose from errors in the learning sets, unsuccessful coding of the initial properties of the components (Gladun, 1995; Gladun and Vashchenko, 1995), or unsuccessful classification in the corresponding space of component properties. Failures can be explained by the fuzzy nature of the concept 'chemical compound' or may also be due to the metastability of compounds under normal conditions. To improve the reliability of future predictions the conclusion about the formation of the compound should be definite only when all three of the feature sets agree.

Shown in Table 1 are some of the predictions of compounds with the composition ABX_2 (Savitskii and Kiselyova, 1979). In the last two decades 79 predictions were tested experimentally. Only three predictions, compounds with compositions $CsPrS_2$, $TlEuSe_2$ and $TlCeTe_2$, were in error. But, taking into account refined information about the objects for computer learning (in particular, the non-formation of compounds with the specified composition in the systems Cu_2Se-B_2Se_3, where B = Sc, Y, Tb-Lu, Tl, or Bi), the problem was solved by re-teaching the computer-aided system of artificial intelligence using new experimental data. More exact results are presented in a book (Savitskii *et al*., 1990). In this book the prediction of compounds with composition ABX_2 and crystal structure type α-$NaFeO_2$ is given. In our work (Kiselyova, 1995; Kiselyova *et al*., 1998) on a search for new semiconducting and electro-optical substances, new compounds were predicted for this composition with chalcopyrite, α- and β-$NaFeO_2$, α-$LiFeO_2$, and TlSe structures.

5.2 Prediction of the Formation of Compounds with Composition AB_2X_4 (X=S, Se, Te)

The chalcogenides of this composition belong to a class of compounds that is promising for semi-

Table 1 Part of a table illustrating the prediction of compounds with the composition ABX_2

X	S								Se								Te							
A / B	Li	Na	K	Cu	Rb	Ag	Cs	Tl	Li	Na	K	Cu	Rb	Ag	Cs	Tl	Li	Na	K	Cu	Rb	Ag	Cs	Tl
B			©	©	©		©	⊕			+	©	+	+	+	⊕							–	
Al			©	⊕	+	⊕	+	⊕	+	⊕	⊕	⊕	+	⊕	+	⊕		⊕	⊕	⊕		⊕		⊕
Sc	©	©	+	⊕	+	©	+		+	+	+	⊕	+	⊕	+	+		+				+		
Ti	⊕	©	©	©	⊕	+	⊕		©	©	+	+	+	+	+	+	+	+	+			+	+	
V	⊕	©	+	+	+	+	+	+	+	+	+	⊕	+	+	+	+		©				+		
Cr	⊕	⊕	©	⊕	⊕	⊕	+	⊕	+	⊕	+	⊕	⊕	⊕	+	⊕		+				©		©
Mn	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		©				+		
Fe	+	⊕	⊕	⊕	⊕	⊕	⊕	⊕	+	+	©	⊕	©	⊕	©	⊕	+	+	+	⊕		⊕	+	
Co	+	+	+	+	+		+		+	+	+	+	+	+	+	+	+	+	+			+	+	
Ni	+	©	+	⊕	+	+	+	+	+	©	+	+	+	©	+	+	+	+	+			©	+	
Ga	⊕	©	©	⊕	©	⊕	©	⊕	+	+	⊕	⊕	+	⊕	©	⊕		⊕	⊕	⊕		⊕		⊕
As	+	⊕	©	⊕	+	⊕	+	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕				↔		⊕		⊕
Y	©	⊕	©	⊕	+	©	+	⊕	©	©	+	⊕	+	⊕	+	©						⊕		⊕
Rh	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+						+		+
In	©	⊕	⊕	⊕	⊕	⊕	⊕	⊕	©	©	⊕	⊕	+	⊕	+	⊕			©	⊕	©	⊕	+	⊕
Sb	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	↔	↔	⊕	⊕	⊕	⊕
La	©	⊕	⊕	⊕	⊕	↔	⊕	↔		⊕		⊕		↔		↔		+		⊕		+		↔
Ce	©	⊕	⊕	⊕	⊕		⊕		©	⊕	+	⊕		↔	+	©		+		⊕		+		⊗
Pr	⊕	⊕	⊕	⊕	⊕		⊗	↔	+	⊕	+	⊕	+	↔	+	©		+	+	⊕	+	+		⊕
Nd	⊕	⊕	⊕	⊕	⊕		+	↔	+	⊕	+	⊕	+	↔	+	©		+		⊕		+		⊕
Pm	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+		+		+		+		+
Sm	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	+	⊕	+	⊕	+	↔	+	⊕		+		⊕		+		⊕
Eu	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	+	⊕	+	⊕	+	↔	+	⊗		+		+		+		⊕
Gd	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	©	⊕	+	⊕	+	©	+	⊕		+		⊕		⊕		⊕
Tb	⊕	⊕	⊕	⊕	©	⊕	+	⊕	©	⊕	+	⊕	+	⊕	+	©		+		⊕		+		⊕
Dy	⊕	⊕	⊕	⊕	©	⊕	+	⊕	©	⊕	+	⊕	+	⊕	+	©		+		©		⊕		⊕
Ho	⊕	⊕	⊕	⊕	©	⊕	+	⊕	©	⊕	+	⊕	+	⊕	+	©		+		©		⊕		⊕
Er	⊕	⊕	⊕	⊕	©	⊕	+	⊕	©	⊕	+	⊕	+	⊕	+	©		+		⊕		⊕		⊕
Tm	⊕	⊕	⊕	⊕	©	⊕	+	⊕	+	+	+	⊕	+	⊕	+	©		+		⊕		⊕		⊕
Yb	⊕	⊕	⊕	⊕	©	⊕	+	⊕	+	+	+	⊕	+	⊕	+	⊕		+		©		+		⊕
Lu	⊕	⊕	⊕	⊕	©	©	+	⊕	+	+	+	⊕	+	⊕	+	©		+		©		⊕		⊕
Tl	+	+	©	⊕	⊕	©	⊕		+		+	⊕	⊕	⊕	+			+	+	⊕		⊕	+	
Bi	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Ac	+	+	+	+	+	+	+	–	+	+	+	+	+	+	+	–		+		+		+		–
Th	+	+	+	+	+		+		+	+	+	+	+	+	+	+	+		+	+		+		
Pa	+	+	+	+	+	+	+	–	+	+	+	+	+	+	+	–		+		+		+		–
U	+	+	©	+	+	+	+		+	+	+	+	+	+	+	+						+		

Designations: +, predicted formation of a compound with composition ABX_2; –, prediction of no formation of a compound with composition ABX_2; ⊕, compound ABX_2 is known to be formed and this fact is used in the computer learning process; ↔, compound ABX_2 is not known to be formed and this fact is used in the computer learning process; ©, predicted formation of a compound with composition ABX_2 which is confirmed by experiment; ⊗, predicted formation of a compound with composition ABX_2 which is not confirmed by experiment; empty square, indeterminate result.

conducting, electro-optical, and other electronic materials.

Each A-B-X system was input to computer memory as a set of coded values (Gladun and Vashchenko, 1995) of the properties of components A and B, whose class (a compound with composition AB_2X_4, formation or non-formation in the system) is indicated as the target feature. As in the previous case (section 5.1) the search for regularities and the prediction were carried out separately for systems A-B-S, A-B-Se and A-B-Te. Based on physical and chemical information on the nature of compounds of this class, three sets of component properties were chosen to describe the chalcogenide systems as listed in section 5.1.

The classifying regularities and predictions of formation of unknown compounds with composition AB_2X_4 were obtained separately for each of the three sets of component properties. In Table 2 are listed the predictions of compounds with composition AB_2S_4 (Kiselyova and Savitskii, 1979) and in Table 3 are

Table 2 Part of a table illustrating the prediction of compounds with the composition AB_2S_4

A / B	Mg	Ca	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ge	Sr	Cd	Sn	Ba	Sm	Eu	Yb	Hg	Pb	Ra
Al	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	©	⊕	⊕	⊕	⊕	
Sc	⊕	⊕	+	+	©	⊕	⊕	+	+	+	⊕	+	⊕	⊕	+	⊕	+	⊕	+	+	©	+
Ti	⊕	⊕					⊕	⊕	⊕	⊕	⊕	+	⊕	⊕		⊕		©		⊕	+	
V	⊕	⊕		⊕	©		⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕		⊕		⊕		⊕	+	
Cr	+	+	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	⊕	©	⊕	+	⊕	+	⊕	⊕	+
Mn							+			+	©	+						+		+	+	
Fe	⊕	⊕			+		⊕			⊕	⊕	⊕	⊕	⊕		⊕		+		⊕	+	
Co	⊕	⊕	+	+	⊕	+	+	⊕	⊕	⊕	⊕	+	⊕	⊕	+	⊕	+	+	+	⊕	+	+
Ni	⊕	⊕	+	+	+	+	⊕	©	⊕	+	⊕	+	⊕	⊕		⊕	+	+	+	⊕	+	+
Ga	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	⊕	⊕	⊕	⊕	⊕	©	⊕	⊕	⊕	⊕	
As	+	+	+	+	+	+	+	+	+	+	+	+	©	⊗	©	⊕	©	+	©	↔	⊕	
Y	⊕	⊕	+	+	⊕	⊕	⊕	+	+	+	↔	⊕	⊕	⊕		⊕	⊕	⊕	⊕	+	⊕	+
Rh						↔	⊕	⊕		⊕	+	+		+				+		+	+	
In	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	+	⊕	⊕	+
Sb	⊕	⊕	©	+	+	⊕	⊕	+	+	⊕	⊕		⊕	⊕	⊕	⊕	©	⊕	©	⊕	⊕	+
La	↔	⊕								©	↔	⊕	⊕	⊕		⊕	⊕	⊕	⊕	©	⊕	+
Ce	↔	⊕		+	+	↔		+		+	↔	⊕	⊕	⊕		⊕	⊕	⊕	⊕	©	⊕	+
Pr	↔	⊕	+	+	©	↔	+	+	+	+	↔	⊕	⊕	⊕	+	⊕	⊕	⊕	⊕	©	⊕	+
Nd	↔	⊕	+	+	+	↔	⊗	+	+	+	↔	⊕	⊕	⊕	+	⊕	⊕	⊕	⊕	©	⊕	+
Pm	–	+	+	+	+	–	+	+	+	+	–	+	+	+	+	+	+	+	+	+	+	+
Sm	↔	⊕	+	+	+	↔	+	+	+	+	⊕	⊕	⊕	⊕	+	⊕	⊕	⊕	⊕	©	⊕	
Eu	↔	+	+	+	+	+	+	+	+	+	↔	+	+	⊕	+	+	+	⊕	+	+	+	+
Gd	↔	⊕	+	+	+	↔	⊗	+	+	+	↔	⊕	⊕	⊕	+	⊕	⊕	⊕	⊕	©	⊕	
Tb	⊕	⊕	+	+	+	⊕	+	+	+	+	↔	⊕	⊕	⊕	+	⊕	⊕	⊕	⊕	©	⊕	+
Dy	⊕	⊕	+	+	⊕	⊕	⊕	+	+	+	↔	⊕	⊕	⊕	+	⊕	⊕	⊕	⊕	©	⊕	+
Ho	⊕	⊕	+	+	⊕	⊕	⊕	+	+	+	↔	⊕	⊕	⊕	+	⊕	⊕	⊕	⊕	©	⊕	+
Er	⊕	⊕	+	+	⊕	⊕	⊕	+	+	+	↔	©	⊕	⊕	+	⊕	⊕	⊕	⊕	©	⊕	
Tm	⊕	⊕	+	+	⊕	⊕	⊕	+	+	+	⊕	+	⊕	⊕	+	⊕	⊕	⊕	⊕	©	⊕	+
Yb	⊕	⊕	©	©	⊕	⊕	⊕	+	+	+	⊕	+	⊕	⊕	+	⊕	+	⊕	⊕	©	⊕	+
Lu	⊕	⊕	+	+	⊕	⊕	⊕	+	+	+	⊕		⊕	⊕		⊕	⊕	⊕	⊕	©	⊕	
Tl	+	+					+	+		⊗	+	+	+	+	©	+	+	+	+	+	⊗	
Bi	⊕	⊕					↔			⊕	⊕	↔	⊕	⊕	⊕	⊕	©	⊕	©	⊕	⊕	
Ac							+			+	+	+		+				+		+	+	
Th							+			+	+	+		+				+		+	+	
Pa	+	+					+			+	+	+	+	+	+	+	+	+	+	+	+	
U	+	+					+		⊕	+	+	+	+	+	+	+	+	+	+	+	+	

Designations: +, predicted formation of a compound with composition AB_2S_4; −, prediction of no formation of a compound with composition AB_2S_4; ⊕, compound AB_2S_4 is known to be formed and this fact is used in the computer learning process; ↔, compound AB_2S_4 is known not to be formed and this fact is used in the computer learning process; ©, predicted formation of a compound with composition AB_2S_4 which is confirmed by experiment; ⊗, predicted formation of a compound with composition AB_2S_4 which is not confirmed by experiment; empty square, indeterminate result.

some of the predictions of compounds with composition AB_2Te_4 (Savitskii *et al*., 1978). In the last two decades 43 predictions of sulfide compounds and 39 predictions of telluride compounds were tested experimentally. Only five predictions of complicated sulfides (predictions of compounds with composition $CdAs_2S_4$, $FeNd_2S_4$, $FeGd_2S_4$, $CuTl_2S_4$, and $PbTl_2S_4$) and six predictions of complicated tellurides (prediction of compounds with compositions CaR_2Te_4 (R = La-Nd, Sm) and Yb_3Te_4) were in error. More recently it was decided to reteach the computer system using new experimental data. More exact results are presented in the book by Savitskii et al. (1990). Predictions under normal conditions of compounds with composition AB_2X_4 and the crystal structure types Th_3P_4, $CaFe_2O_4$, $NiCr_2S_4$, or spinel are given also. Kiselyova (1995) reports on the search for new semiconducting and electro-optical substances; we predicted new compounds of this composition with the structures of chalcopyrite, spinel, olivine, $PbGa_2Se_4$, Yb_3S_4, Th_3P_4, Yb_3Se_4, $CaFe_2O_4$, or $NiCr_2S_4$ (at room temperature and atmospheric pressure).

Table 3 Part of a table illustrating the prediction of compounds with the composition AB_2Te_4

A / B	Mg	Ca	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ge	Sr	Cd	Ba	La	Sm	Eu	Yb
Al	⊕		+	⊕					⊕		⊕	⊕	⊕				
Sc							−		−		+		+	+	+	+	+
Ti		−	©				−		−	−							
Cr			©		⊕	⊕		⊕	+		+	+	+	+	+	⊕	⊕
Mn					+						+	+	+	+	+	+	+
Fe			+	+					+								
Co									+								
Ga	⊕	+	+	⊕					⊕	↔	⊕	⊕	⊕				
As							−		↔	−	+	−	+	+	+	©	©
Y	+	©	+	+				+		+	+		+	+	+	+	+
Rh	+	⊕	©	+	+	⊕	⊕	+			+		+	+	+	+	+
In	⊕	©	+	⊕					⊕	↔	⊕	⊕	⊕				
Sb		+	+		+	+		+		⊕	+		+	+	©	⊕	©
La	+	⊗	+	+				+		+	+		+	©	+	©	+
Ce	+	⊗	+	+				+		+	+		+	+	+	©	+
Pr	+	⊗	+	+				+		+	+		+	+	+	©	+
Nd	+	⊗	+	+				+		+	+		+	+	+	©	+
Pm	+	+	+	+				+		+	+		+	+	+	+	+
Sm	+	⊗	+	+				+		©	+		+	+	©	©	+
Eu	+	+	+	+				+		+	+		+	+	+	⊕	+
Gd	+	+	+	+				+		+	+		+	+	+	©	+
Tb	+	+	+	+				+		+	+		+	+	+	©	+
Dy	+	©	+	+				+		+	+		+	+	+	©	+
Ho	+	©	+	+				+		+	+		+	+	+	©	+
Er	+	©	+	+				+		+	+		+	+	+	©	+
Tm	+	©	+	+				+		+	+		+	+	+	+	+
Yb	+	©	+	+				+		+	+		+	+	+	©	⊗
Lu	+	©	+	+				+		+	+		+	+	+	+	+
Bi					−	−			O	⊕		O				⊕	

Designations: +, predicted formation of a compound with composition AB_2Te_4; −, prediction of no formation of a compound with composition AB_2Te_4; ⊕, compound AB_2Te_4 is known to be formed and this fact is used in the computer learning process; ↔, compound AB_2Te_4 is known not to be formed and this fact is used in the computer learning process; ©, predicted formation of a compound with composition AB_2Te_4 which is confirmed by experiment; ⊗, predicted formation of a compound with composition AB_2Te_4 which is not confirmed by experiment; O, predicted absence of a compound with composition AB_2Te_4 which is confirmed by experiment; empty square, indeterminate result.

5.3 Prediction of New Compounds with Composition AB_2X_2 (X=Si or Ge) and Crystal Structure Type $ThCr_2Si_2$

Compounds with crystal structure type $ThCr_2Si_2$ are promising for new magnetic and superconducting materials.

Each system A-B-X was represented in the computer memory as a set of especially coded values (Gladun and Vashchenko, 1995) of the properties of elements A and B, whose class (a compound of composition AB_2X_2 with crystal structure type $ThCr_2Si_2$ and formation or non-formation in the system) is indicated as the target feature. The searches for classifying regularities and predictions were carried out separately for systems A-B-Si and A-B-Ge. Two sets of properties of elements were chosen for the description of the systems:

1. The distribution of electrons in the energy levels of isolated atoms of the chemical elements A and B.
2. The first three ionization potentials, the metal radii by Bokii and Belov, the standard entropies of individual substances, the melting points, the number of complete electronic shells, the number of electrons in incomplete s-, p-, d- or f-electronic shells for the atoms of elements A and B.

The classifying regularities and predictions of formation of unknown compounds of composition AB_2X_2 with the $ThCr_2Si_2$ crystal structure were obtained separately for each of the two sets of component properties.

Table 4 Part of a table illustrating the prediction of the crystal structure type $ThCr_2Si_2$ for compounds with the composition AB_2Si_2

B / A	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ru	Rh	Pd	Ag	Os	Ir	Pt	Au
Ca	+	⊕	↔	⊕	⊕	⊕	⊕	+	+	©	⊕	+	+	+	⊕
Sr		⊕		⊕	⊕	⊕	⊕	−	−		⊕			©	⊕
Y	⊕	⊕	⊕	⊕	⊕		−				−	+	©	+	Ø
Zr	+	+	⊕	⊕	⊕						−	+	+		
Ba	+	⊕	+	⊕	⊕	⊕	⊕	+	Ø	Ø	⊕	+	Ø	+	⊕
La	+	⊕	⊕	⊕	⊕	⊕	⊕	⊕	©	©	⊕	+	©	⊕	⊕
Ce	↔	⊕	⊕	⊕	⊕	⊕	⊕	⊕	©	⊕	⊕	©	+	⊕	⊕
Pr	+	⊕	⊕	⊕	⊕	©	⊕	©	+	©	⊕	+	+	+	⊕
Nd	©	⊕	⊕	⊕	⊕	©	⊕	⊕	+	©	⊕	+	+	⊕	⊕
Pm	+	+	+	+	⊕	+	+	+	+	+	+	+	+	+	+
Sm	©	⊕	⊕	⊕	⊕	©	+	⊕	+	©	⊕	+	+	Ø	⊕
Eu	+	+	⊕	⊕	⊕	⊕	+	+	©	©	⊕	+	+	⊕	⊕
Gd	©	⊕	⊕	⊕	⊕	©	+	©	©	⊕	⊕	©	⊕	⊕	⊕
Tb	©	⊕	⊕	⊕	⊕	©	+	⊕	©	©	+	©	©	+	⊕
Dy	©	⊕	⊕	⊕	⊕	©	+	⊕	©	©	+	+	+	⊕	⊕
Ho	©	⊕	⊕	⊕	⊕	©	+	©	+	©	+	©	+	Ø	⊕
Er	©	⊕	⊕	⊕	⊕	©	+	⊕	©	©	+	©	©	⊕	⊕
Tm	©	©	⊕	⊕	⊕	©	+	©	+	©	+	+	+	⊕	+
Yb	©	⊕	⊕	⊕	⊕	©	+	⊕	⊕	⊕	⊕	+	+	⊕	⊕
Lu	©	+	⊕	⊕	⊕	©	+	©	+	©	+	+	+	⊕	+
Hf	+	+	↔	⊕	⊕										
Ac	+	+	+	+	+	+		+	+	+	+	+	+	+	+
Th	⊕	⊕	⊕	⊕	⊕	⊕		©	©	©	+	©	©	©	©
Pa	+	+		+	+	+		+	+	+	+	+	+	+	+
U	©	©		©	⊕	©		©	©	©	+	©	⊕	©	©
Np	⊕	⊕	⊕	⊕	⊕			+	+	+	+	+	+	+	+
Pu	+	+		+	+	+		+	+	+	+	+	+	+	+
Am	+	+		+	+	+	+	+	+	+	+	+	+	+	+
Cm	+	+		+	+	+		+	+	+	+	+	+	+	+

Designations: +, formation of a compound with the crystal structure type $ThCr_2Si_2$ is predicted; −, formation of a compound with the crystal structure type $ThCr_2Si_2$ is not predicted; ⊕, a compound with the crystal structure type $ThCr_2Si_2$ was synthesized and appropriate information was used in the computer learning process; ↔, a compound with the crystal structure type $ThCr_2Si_2$ does not exist under normal conditions and this fact was used in the computer learning process; ©, predicted formation of a compound with the crystal structure type $ThCr_2Si_2$ is confirmed by experiment; ⊗, predicted formation of a compound with the crystal structure type $ThCr_2Si_2$ is not confirmed by experiment; Ø, predicted absence of a compound with the crystal structure type $ThCr_2Si_2$ which is not confirmed by experiment; empty square, indeterminate result.

Shown in Table 4 are some of the predictions of compounds with composition AB_2Si_2 and structure type $ThCr_2Si_2$ (Kiselyova and Savitskii, 1983), and in Table 5 are predictions of compounds with composition AB_2Ge_2 and structure type $ThCr_2Si_2$ (Savitskii and Kiselyova, 1984). Because of the great promise of this class of crystal phases, these compounds have recently been studied intensively. An experimental check showed that out of 79 predictions of silicides checked, only six were wrong and of 37 predictions of germanides only five results did not fit our predictions.

5.4 Prediction of New Crystal Phases with Al, Ga and In

More than 10 years ago we predicted hundreds of new compounds of aluminum, gallium and indium with compositions: AB_2X, ABX and AB_2X_2 (X = Al, Ga or In) and with crystal structures that resemble the Heusler alloys, TiNiSi (E phase), ZrNiAl, $CaAl_2Si_2$ and $ThCr_2Si_2$ (Kiselyova and Burkhanov, 1989). We used the prototype of our information-predicting system for the first time.

Table 5 Part of a table illustrating the prediction of the crystal structure type $ThCr_2Si_2$ for compounds with the composition AB_2Ge_2

A B	Cr	Mn	Fe	Co	Ni	Cu	Ru	Rh	Pd	Ag	Os	Ir	Pt	Au
Li									+					+
Na									+					+
K							+	+	+		+	+	+	+
Ca						⊕			+	⊕				⊕
V	–	+	–	–	–	–	–	–	–	–	–	–	–	+
Rb										–	+	+	+	+
Sr	+					⊕			+	⊕				⊕
Y		⊕		⊕	⊕	⊕				–				
Nb						–				–	+	+		
Cs	+					+				–				
Ba	+					+	+	+	+	⊕			+	
La	+	⊕	⊕	⊕	⊕	⊕	©	©	⊕	+	+	Ø	⊕	+
Ce	+	⊕	⊕	⊕	⊕	⊕	©	©	⊕	©			⊕	
Pr	+	⊕	⊕	⊕	⊕	⊕	©	©	+	+			Ø	
Nd	+									+				
Pm	+	+				+	+	+	+	+			+	
Sm	+	⊕		⊕	⊕	⊕	©	©	+	+			Ø	
Eu	+	©			⊕	©	©	©	©	©			Ø	
Gd	+	⊕	⊕	⊕	⊕	⊕	©	⊕	+	+	+	+	⊕	+
Tb	+	⊕	⊕	+	⊕	⊕	©	©	+	+	+	+	Ø	+
Dy	+	⊕	⊕		⊕	⊕	©	©	⊕	+			⊕	
Ho	+	⊕		⊕	⊕	⊕	©	©	©	+			+	
Er	+	⊕	⊕	⊕	⊕	⊕	©	©	+	+			+	
Tm	+	+		⊕	⊕	⊕	©	©	+	+			+	
Yb	+	⊕	⊕	⊕	⊕	⊕	©	©	⊕	+			+	
Lu	+	+	+	⊕	⊕	+	+	+	+	+	+	+	+	+
Hf				+		+				+				+
Ta				+		+				+				+
Tl				+		+								+
Pb				+		+								+
Bi				+		+								+
Fr	+		+	+	+	+	+	+	+		+	+	+	+
Ra				+		+								+
Ac	+		+	+	+	+	+	+		+	+	+	+	+
Th	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	⊕	+	+	+	⊕	⊕
Pa	+		+	+	+	+	+	+	+		+	+	+	+
U	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	⊕		+	⊕	⊕	+
Np	+		©	+	+	+	+	+	+		+	+	+	+
Pu	+		+	+	+	+	+	+	+		+	+	+	+
Am	+		+	+	+	+	+	+	+		+	+	+	+
Cm	+		+	+	+	+	+	+	+		+	+	+	+

Designations: see Table 4.

Two sets of element properties that had allowed us to obtain good results in the solution of similar problems (see section 5.3) were chosen for description of the systems. Prediction of the regularities of formation of phases with a definite composition and crystal structure type were obtained separately for each of the two sets of component properties. The use of these regularities has allowed us to obtain two tables of predictions of new compounds for each composition and each crystal structure type. Analysis of these predictions was published by Kiselyova and Burkhanov (1989).

Table 6 shows some of the predictions of compounds with composition ABAl and crystal structure type TiNiSi. Of 16 predictions that were checked, all agreed with the new experimental data. Table 7 contains past predictions of compounds with composition AB_2In and crystal structure type resembling the Heusler alloys. Out of 22 predictions that were checked only three were wrong.

Table 6 Part of a table illustrating the prediction of the crystal structure type TiNiSi for compounds with the composition ABAl

A / B	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Ru	–	–	–	–	–	+	–	+	+	+	+	+	+	–	–
Rh	–	–	–	–	–	+	–	+	+	+	+	+	+	–	–
Os	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ir	©	©	©	©	+	©	+	©	©	©	©	©	©	+	©
Pt	+	©	©	©	+	⊕	+	⊕	⊕	⊕	⊕	⊕	⊕	©	⊕

Designations: +, formation of a compound with the crystal structure type TiNiSi is predicted; –, formation of a compound with the crystal structure type TiNiSi is not predicted; ⊕, a compound with the crystal structure type TiNiSi was synthesized and appropriate information was used in the computer learning process; ©, the predicted formation of a compound with the crystal structure type TiNiSi is confirmed by experiment.

5.5 Prediction of New Heusler Alloys with Composition ABD_2 (D=Co, Ni, Cu or Pd)

One further successful result of the suggested approach is a prediction of new cubic Heusler alloys (crystal structure of the type $AlCu_2Mn$) (Kiselyova, 1987).

Two sets of properties of elements, that yielded good results for the solution of similar problems of predicting crystal structure types of intermetallic compounds (see sections 5.3 and 5.4), were chosen for description of these systems. The regularities of formation of Heusler alloys with a definite composition were obtained separately for each of the two sets of properties of the components. Use of these regularities for prediction has allowed us to obtain two tables of predictions of new compounds with crystal structure type resembling the Heusler alloys. The results of comparison of these predictions for each pair of regularities were published by Kiselyova (1987).

Table 8 shows some of these results for predicting $ABCo_2$ compounds and Table 9 shows other results for $ABCu_2$ compounds, which have a crystal structure type resembling the Heusler alloys. Of the four checked predictions for compounds with cobalt, all agreed with the new experimental data. Three predictions of Heusler compounds with copper, and three predictions of the non-formation of Heusler alloys in the systems A-B-Cu, coincided with the new experimental data.

The results that have been shown thus far, in sections 5.1–5.5 do not exhaust the possibilities for prediction of new intermetallic compounds with our approach. The results of prediction of the crystal structure type of new equiatomic ternary compounds with composition ABAl (the crystal structure type ZrNiAl was predicted), ABSi (the crystal structure types ZrNiAl, PbFCl, or TiNiSi were predicted), ABGe (the crystal structure types ZrNiAl or TiNiSi

Table 7 Part of a table illustrating the prediction of a crystal structure type resembling the Heusler alloys for compounds with the composition AB_2In

A / B	Ru	Rh	Pd	Ag	Os	Ir	Pt	Au
Ti			⊕	+			⊕	⊕
Sr				+				
Y			⊕	©				⊕
Zr	+	+	⊕	+	+	+	⊕	⊕
Nb	+	+	+	+	+	+	+	+
Tc	+	+	+	+	+	+	+	+
La	+	+	⊗	©	+	+	+	⊕
Ce				©				
Pr	+	+	+	©	+	+	+	⊕
Nd	+	+	+	©	+	+	+	⊕
Pm	+	+	+	+	+	+	+	+
Sm	+	+	+	©	+	+	+	⊕
Eu	+	+	+	+	+	+	+	+
Gd	+	+	©	©	+	+	+	+
Tb	+	+	+	©	+	+	⊗	⊕
Dy	+	+	©	©	+	+	+	⊕
Ho	+	+	+	©	+	+	⊗	⊕
Er	+	+	©	©	+	+	+	⊕
Tm	+	+	©	©	+	+	+	©
Yb	+	+	+	+	+	+	+	⊕
Lu	+	+	©	+	+	+	+	©
Hf	+	+	⊕	+	+	+	⊕	⊕
Ta				+				
Re	–			+				

Designations: +, formation of a compound with a crystal structure type resembling the Heusler alloys is predicted; –, formation of a compound with a crystal structure type resembling the Heusler alloys is not predicted; ⊕, a compound with a crystal structure type resembling the Heusler alloys was synthesized and appropriate information was used in the computer learning process; ©, predicted formation of a compound with a crystal structure type resembling the Heusler alloys which is confirmed by experiment; ⊗, predicted formation of a compound with a crystal structure type resembling the Heusler alloys which is not confirmed by experiment; empty square, indeterminate result.

Table 8 Part of a table illustrating the prediction of a crystal structure type resembling the Heusler alloys for compounds with the composition $ABCo_2$

B \ A	Al	Si	Ga	Ge	In	Sn	Sb	Tl	Pb
Li			+			+			
Be	+	⊕	+	⊕	+	+	+	+	+
Mg			+			+			
K		+	−	−	−	+	−	−	−
Sc		+	−	−	−	©	−	−	−
Ti	⊕	©	⊕	⊕	+	⊕	+	+	+
V	©	⊕	⊕	+	+	⊕	+	+	+
Cr	+	+	↔			+			
Fe	⊕	↔	−	−		+			
Ni			−	−	−	+	↔	−	−
Y			+		−	+	−	−	−
Zr	⊕	+				⊕			
Nb	⊕	+	⊕			⊕			
Mo	+	+	+			+			
Ru	+	+	+			+			
Rh			+		−	+	−	−	−
Pd			+		−	+	−	−	−
Ag			+		−	+	−	−	−
Lu	+	+	+		+	+	+		
Hf	⊕	+	©	+	+	⊕	+		
Ta	⊕	+	⊕	+	+	+	+		
Au	+	+	+		+	+	+		

Designations: ↔, a compound with a crystal structure type resembling the Heusler alloys does not exist under normal conditions and this fact was used in the computer learning process; see Table 7 for other symbols.

were predicted), ABP (the crystal structure types ZrNiAl or TiNiSi were predicted), ABPd (the crystal structure ZrNiAl was predicted) are presented in a paper by Savitskii and Kiselyova (1985). The predictions of the crystal structure type at normal conditions for complicated pnictides with compositions: ABP and ABAs (the crystal structures types ZrNiAl, PbFCl, or TiNiSi were predicted), ABSb (the crystal structure types MgAgAs (half Heusler phase) or PbFCl were predicted), ABBi (the crystal structure type MgAgAs was predicted), AB_2P_2, AB_2As_2 and AB_2Sb_2 (the crystal structure types $CaAl_2Si_2$ or $ThCr_2Si_2$ were predicted) were published in a paper by Kiselyova and Burkhanov (1987). The predictions of hundreds of new compounds with various compositions and crystal structures in chalcogenide systems are presented in the book by Savitskii *et al.* (1990). Apart from intermetallic compounds, we have also predicted the formation of thousands of new compounds in the oxide and halogenide systems A-B-O, A-B-D-O, A-B-Hal, and A-B-D-Hal. The results of experimental checking of the predictions are presented in Table 10. Comparison of the predictions with experimental data, obtained recently, shows (Table 10) that the average reliability of the prediction of metal and other inorganic compounds exceeds 80%. Such a high accuracy for *a priori* predictions of new inorganic compounds has not been attained by any other known theoretical method.

6. Cybernetical-Statistical Approach to the Design of New Materials

The search for an optimal technology for the production of a material having extreme values of target parameters is an integral part of the design of new substances. As already noted (section 5), it is possible to predict, not only the formation of compounds with a certain composition and to estimate their properties, but also to predict the best method of their production.

Just such a sequence: the information system → the predicting system → the design of multifactorial

Table 9 Part of a table illustrating the prediction of a crystal structure type resembling the Heusler alloys for compounds with the composition $ABCu_2$

B \ A	Li	Be	Al	K	Sc	V	Cr	Fe	Co	Ni	Ga	Ge	Y	Nb	Mo	Ru	Rh	Pd	Ag	In
Zn			©									−	−	−	−	−	−	−	−	
Ga	+	+		+	⊕	+	+	+	+	+		−							−	+
In	+	+		+	©	+	+	+	+	+	+	+	©	+	+	+	+	+	+	
Sn	O	−	O	−	−	−		⊕	⊕	⊕	−	−	−	↔	−	−	−	−	↔	⊕
Lu	−	−		−	−	−	−	−	−	−		−	−	−	−	−	−	−	−	+
Ta		−			−	−	−	−	−	↔		−	−	↔	−	−	−	−		+
Au	−	−		−	−	−	−	−	−	−		−	−	−	−	−	−	↔	↔	+
Tl												−							−	+
Pb	−	−	O	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+

Designations: O, predicted absence of a compound with a crystal structure type resembling the Heusler alloys which is confirmed by experiment; see Tables 7 and 8 for other symbols.

Table 10 Comparison of predictions with new experimental data

Compounds/Systems	Characteristics to be predicted	Experimental tests as of January 2000	Error of prediction (%)
ABX (X = Se,Te)	Compound formation	100	44
ABX_2 (X = O,S,Se,Te)	Compound formation	337	10
ABX_3 (X = O,F,S, Cl,Se,Br,Te,I)	Compound formation	420	11
ABX_4 (X = O,F,Cl,Br,I)	Compound formation	393	5
A_2BX_2 (X = S,Se)	Compound formation	24	9
AB_2X_4 (X = O,F,S, Cl,Se,Br,Te,I)	Compound formation	761	16
$A_2B_2X_7$ (X = O,S,Se)	Compound formation	97	26
$A(Hal)_2 - B(Hal)$	Systems w/ compounds	108	10
AB_2X_4 (X = O,S,Se,Te)	Structure type	381	7
ABX (X = Al,Si,P,Ga,Ge,As,Pd,In,Sb,Bi)	Structure type	78	35
ABO_3	Perovskite structure	186	13
$A_2B_2O_7$	Pyrochlore structure	74	15
AB_2X_2 (X = Al,Si,P,Ge,As,Sb)	Structure type	200	8
ABX_2 (X = Co,Ni,Cu,Pd)	$MnCu_2Al$ structure	28	14
AB_2X (X = Al,Ga,In)	$MnCu_2Al$ structure	24	13
$A_x(SO_4)_y$–$B_z(SO_4)_w$ and $A(NO_3)_x$–$B(NO_3)_y$	Compound formation	130	4
$ABDO_4$	Compound formation	28	4
			Average = 14%

experiments – was used by Savitskii *et al.* (1982) in a simplified variant for predicting superconducting Chevrel phases of composition $A_xMo_6S_8$ and for optimization of the technology of their synthesis. Any mathematical optimizing method can be used for the solution of the latter problem. As a rule we use statistical methods of the design of multifactorial experiments in our work.

First we predicted new compounds of this kind with critical temperatures for the transition to the superconducting state (T_c) greater than the boiling point of helium, using for computer learning the information from the bibliographic database on Chevrel phases. We predicted a new phase with composition $Ag_xMo_6S_8$ with a T_c above 4.2 K. Also, optimal conditions of synthesis of this phase have been sought using statistical methods of experimental design. At that time there was but one production procedure: a single sintering of the powder elements in an evacuated quartz ampoule. T_c was the parameter of optimization. We supposed that the phase $Ag_xMo_6S_8$ had a sufficiently large homogeneity range for cation A, (silver) as do the majority of Chevrel phases. The silver content, the annealing temperature and the annealing time were chosen as independent variables. A very simple plan using those three factors at two levels each was chosen from the catalogue of plans for the design of experiments. The gradient method of Box and Wilson was selected for reaching the highest T_c. A maximum T_c, equal to 7.8 K, was reached after two steps consisting of five experiments.

Why are statistical methods for design of multifactor experiments attractive for chemists and for materials scientists? In the first place, the number of experiments for a search for the extremum of a desired property is reduced sharply, owing to the simultaneous variation of all the independent variables. Second, the researcher derives an analytical model that can be used for the automation of the process. Third, a quantitative knowledge of the influence of the technological conditions on the parameters of optimization can be acquired. Furthermore, a physical-chemical model of the process can be developed on the base of this knowledge. And, finally, the series of planned experiments can be carried out by unsophisticated staff.

We have considered a sequential procedure for the design of new compounds with selected properties, namely: a database → a predicting system based on artificial intelligence methods → an optimization of the technology of synthesis of the predicted compounds. The predicting of compounds, the forecasting of the desired property, using information from the database, and the prediction of the best type for technology of synthesis can be considered collectively as a strategy for the search for substances with specified properties. The intrinsic properties of the chemical elements and of simple compounds are used to describe multicomponent physical-chemical systems. The experimental data for substances (which are similar to the predicted ones) and for the technology of their synthesis are analyzed using the computer. Design of

experiments to search for optimal conditions of synthesis of substances, predicted in stage one, can be considered as the tactics of the search for new materials. The researcher uses the technological parameters during this stage. The first stage is a theoretical procedure, but the second one is an active experiment using a formal plan. In the future these experiments may be carried out by a robot programmed for the implementation of the procedures of the design of multifactorial experiments. We considered an automation of the search for new substances on the base of new information technologies. Besides the intellectualization of the scientific work, this approach allows promotion of the search for new substances with specified properties.

7. Problems and Perspectives of Computational Materials Design Using Methods of Artificial Intelligence

What problems confront the computer design of metallic and other inorganic substances by artificial intelligence methods? The most important problem is the quality of experimental data for computer learning. The trouble is that the proposed approach eventually assumes a search of physical-chemical systems for learning sets and sets of predictions, which have similar features. If any physical-chemical system from the learning set has an erroneous character and if the set is small, then it is quite possible that it will yield erroneous predictions. Our experience is that the number of erroneous predictions varies proportionally with the number of errors in the experimental data processed, and the reliability of the prediction grows with an increase of the initial volume of data. (However, reliability approaches a limit with an increase in size and improvement in the representativeness of the learning set.) We use databases containing extensive volumes of qualitative information for overcoming these difficulties. With this aim in mind, we have developed DBs containing data assessed by qualified experts. This allows both an increase in quality and in the volume of the learning sets. However, it should be noted that an infinity of knowledge never leads to 100% reliability of prediction. The use of our information-predicting system will allow the enlistment of a user-expert for an assessment of data for computer learning. Usually he can solve the problems of the computer design of substances via analogs which are well known to the expert.

One of the problems of any computer classification in inorganic materials science is the search for those properties of the elements and simple compounds, that are the most important for separation of physical-chemical systems into certain classes. This procedure can hardly be completely formalized, but the system we use, CONFOR (Gladun, 1995; Gladun and Vashchenko, 1995), automatically rejects those properties that have no importance for the classification process. The initial set of properties for computer-aided analysis is prepared by the material scientists, and it is desirable that the artificial intelligence system extrapolates information from this representative set of initial features.

We have achieved good predictions of the qualitative properties of physical-chemical systems: formation of compounds, their crystal structure type, etc. However, the problem becomes even more complicated if it is necessary to predict some *quantitative* property (e.g., the melting point, homogeneity range, etc.). The hypothesis of class compactness, based on methods of computer learning, presupposes that the different classes are located compactly in the multidimensional feature space and that there are no intersections between these classes. But we found some sets of properties whose space occupancy contradicts this hypothesis. The application of cluster analysis to the exemplar learning set, in combination with the grouping of features according to a statistical correlation, allows us to decrease the intersections of classes, but only slightly, owing to the selection of the natural threshold values (for a certain learning set) of the predicted quantitative properties. Note that these natural threshold values are less a consequence of the nature of the phases and more a consequence of the set of examples used for the computer learning method. These observations are based upon the examples of learning sets that we have thus far obtained and investigated.

Therefore, as a consequence of the above interaction problem, the attempt to predict certain threshold values that are important for technological applications, e.g. boiling-point temperatures of helium and nitrogen for superconducting compounds, is justified only from a practical standpoint. The error of this prediction will be high, but it will be possible to predict (with high reliability) those objects which are widely spaced in the features space. *A priori* identification of these objects by a researcher seems to be a great problem. One possibility to solve this problem is to visualize a two-dimensional projection of points, which correspond to the objects of the learning set, in

combination with the cluster analysis of objects and grouping of features according to their statistical correlations. The algorithms for this system involve cluster analysis based on the method of potential functions (Arkad'ev and Braverman, 1971; Izerman *et al*., 1970) and the extreme grouping of parameters (Arkad'ev and Braverman, 1971).

As stated above, the prediction accuracy of quantitative properties depends strongly on the volume and representativeness of the learning set. Our experience shows that the number of examples in the learning sets must be in the hundreds or even in the thousands in order to have an acceptable estimation of the quantitative property. The future of this approach, using DBs and AI methods, is connected with the development of information-predicting systems. That is a very expensive and time-consuming procedure. However, such systems allow us to cut down the time and expense of a search for and the development of new materials with specified properties. This kind of simulation requires DBs containing only 'good' information.

Let's imagine the laboratory of the future. A materials scientist, who must solve the problem of searching for new materials with desired properties, makes a request to the computer to find the necessary substances. If the set of substances that the researcher receives from a database does not satisfy the request, he asks for a prediction of new substances having the property sought. He chooses the best prediction, from his point of view, and asks the computer to develop an optimal plan for the synthesis of the substance to meet an extreme target property. Such a 'virtual' laboratory is a tool of automation for searching for new substances on the basis of the use of new information technologies. The proposed approach will allow us to speed up considerably the search for new substances with desired properties.

8. Conclusions

In the process of automating scientific research ranging from the development of databases to the building of systems of artificial intelligence, the historical process of cognition is repeated: from collection and processing of the empirical source data to the generalization of the experimental facts. The latter is used as a basis for constructing scientific theories that reflect the basic relations and the correlations between the processes and phenomena studied. Development of an artificial intelligence system, such as an information-predicting system, is indicative of the advent of a new type of modeling of cognitive activity, namely knowledge engineering. Such modeling will play an important role in those fields of science and technology where mathematical simulation and computer-aided experimentation have proven to be inadequate (for example, in physical metallurgy, chemistry, science of materials, and the like).

9. Acknowledgments

Partial financial support from RFBR (Grant N.99-07-90040) is gratefully acknowledged. I should like to thank my colleagues Prof. Victor P. Gladun and Dr. Neonila D. Vashchenko of the Institute of Cybernetics of the National Academy of Sciences of Ukraine for their help and support.

10. Key Term List

artificial intelligence – an artificial system, usually constructed on the basis of computer technology, which simulates a human solution of complicated tasks. It is intended for perception, processing and storage of information, and also for forming solutions of problems in an expedient manner.

class – a set of objects chosen according to some property (properties).

classification – a separation of objects according to some essential property (properties).

classification scheme – a set of rules determining a certain classification.

computer learning (a method of artificial intelligence) – a process of the modification of the parameters of a classifying system on the basis of the use of experimental data with the purpose of improving the quality of the classification.

concept – a generalized model of some class of objects that provides for recognizing and generating models of specific elements of this class.

CONFOR (*CON*cept *FOR*mation) – a set of software tools intended for the logical analysis of large volumes of experimental data (Gladun, 1995; Gladun and Vashchenko, 1995) with the purpose of searching for regularities.

***data** – scientific or technical measurements, values calculated therefrom, observations, or facts that can be represented by numbers, tables, graphs, models, text, or symbols and which are used as a basis for reasoning or further calculation. Note: 'data' is a plural form; 'datum' is the singular.

feature – a property of a constituent component of the physical-chemical system.

***information** – a collection of data and facts, so selected, arranged and interrelated that they give relevance, coherence and utility within a defined sphere of interest and are communicable.

information-predicting system – a system intended for data retrieval on known compounds, prediction of inorganic compounds not yet synthesized, and the forecasting of their properties. This system employs a database of properties of inorganic compounds, a database of element properties, the system CONFOR, a knowledge base, a conversational processor and a monitor (Figure 8).

***knowledge** – acquaintance or awareness of factual information or data together with understanding of their relationships and implications for utilization.

***knowledge base** – (1) a collection of interrelated information, facts, or statements (IEEE 610.12); (2) in artifical intelligence, a representation of information about human experience in a particular field of knowledge and data resulting from solution of problems that have been previously encountered (ISO per ANSI X3 (modified)).

learning set – a multidimensional array of feature values and a column vector of the desired property. Each row corresponds to some physical-chemical system already known, whose class is indicated by the row position of the column vector.

***metadata** – data about data. Consists of descriptors of data in a database to provide systematic information for users, application programs, and database management software. Metadata may also be manipulated and searched.

object – a physical-chemical system which is described as a set of property (feature) values of the constituent elements.

physical-chemical system – a system (e.g. compound or solid solution) which is formed from chemical elements.

prediction – an identification (classification) of a new object belonging to a certain class in compliance with a fixed classification scheme.

qualitative property – an object or element property which can be described as a qualitative concept (e.g. a multi-element system with compound formation or non-formation of a crystal structure type, possibility of forming compounds of desired composition, and so on).

quantitative property – an object or element property which has a numeric value taken from some continuum (or quasi-continuum) set of numbers (e.g. melting point, birefringence, index of refraction, and so on).

set for prediction – a multidimensional array of feature values. Each row corresponds to some unknown physical-chemical system, whose class it is necessary to predict.

Note: Terms with an * are from Westbrook and Grattidge (1991b).

11. References

Andersson, J.-O., Jansson, B., and Sundman, B. (1985). THERMO-CALC: a data bank for equilibria and phase diagram calculations, *CODATA Bull.*, **58**, 31.

Ansara, I. (1991). Génération et application des bases de données thermochimiques, *Entropie*, **27**, 74.

Arkad'ev, A. G. and Braverman, E. M. (1971). *Computer Learning and Object Classification*. Nauka. Moscow (Russ.).

Bale, C. W. and Eriksson, G. (1990). Metallurgical thermochemical databases, a review, *Can. Met. Quart.*, **29**, 105.

Chelikowsky, J. R., Binggeli, N., and Keskar, N. R. (1993) First principles methods for structural trends in oxides: applications to crystalline silica, *J. Alloys & Compounds*, **197**, 137.

Chen, N. Y. (1988). Chemical pattern recognition research in China, *Anal. Chim. Acta*, **210**, 175.

Chen, N. Y, Lu, W. C., Chen, R., and Qin, P. (1999). Software package 'Materials Designer' and its application in materials research, *Proc. Second Int. Conf. Intelligent Processing & Manufacturing of Materials*, **v.2**, July 10–15, 1999. Honolulu, Hawaii. p. 1417.

Cohen, M. L. (1986). Predicting new solids and superconductors, *Science*, **234**, 549.

Darken, L. S., Gurry, R. W., and Bever, M. B. (1953). *Physical Chemistry of Metals*, McGraw-Hill, New York.

Degtyaryov, Yu. I., Podbel'skii, V. V., Kiselyova, N. N., *et al.* (1999). Database on the properties of acousto-optical, electro-optical, and nonlinear-optical properties of crystals with Internet access, *Izvestiya VUZov. Materialy elektron. tekhniki*, **3**, 35 (Russ.).

Devingtal, Yu. V. (1968). About optimal coding of objects and their classification by pattern recognition methods, Izvestiya Akad. Nauk SSSR, Tekhn. kibernetika, **1**, 162 (Russ.).

Devingtal, Yu.V. (1971). Coding of objects with the application of a dividing hyperplane in their classification, *Izvestiya Akad. Nauk SSSR, Tekhn. kibernetika*, **3**, 1 (Russ.).

Dorofeyuk, A. A. (1971). Algorithims of pattern recognition without a teacher, based on the potential function method, *Avtomatika i Telemekhanika*, **12**, 78 (Russ.).

Drago, V. J. and Kaufman, J. G. (1993). Technical features of the chemical and materials property data network services on STN International, *J. Chem. Inf. and Comput. Sci.*, **33**, 46.

Eriguchi, K. and Shimura, K. (1990). Factual databases for materials design and manufacturing, *ISIJ Int.*, **30**, 409.

Girgis, K. (1983). Structure of intermetallic compounds in *Physical Metallurgy*,Part I, 3rd Edition (eds R. W. Cahn, and P. Haasen). North-Holland, Amsterdam, p. 219.

Gladun, V. P. (1995). *Processes of Formation of New Knowledge*. SD 'Pedagog 6'. Sofia. (Russ.).

Gladun, V. P, and Vashchenko, N. D. (1995). Local statistical methods of knowledge formation, *Cybernetics and Systems Analysis*, **31**, 207.

Golikova, M. S., Burkhanov, G. S., Kiselyova, N. N., *et al.* (1989). Databank on acousto-optical properties of inorganic compounds, *Izvestiya Akad. Nauk SSSR, Neorganicheskie Materialy*, **25**, 700 (Russ.).

Gulyev, B. B., and Pavlenko, L. F. (1973). Simulation of the search for components of alloys, *Avtomatika i Telemekhanika*, **1**, 131 (Russ.).

Ho, C. Y., and Li, H. H. (1993). Numerical databases on materials property data at CINDAS/Purdue University, *J. Chem. Inf. and Comput. Sci.*, **33**, 36.

Hume-Rothery, W., and Raynor, G. V. (1962). *The Structure of Metals and Alloys*. The Institute of Metals, London, p. 198, 230.

Izerman, S. A., Braverman, E. M., and Rozonoer, L. I. (1970). *Method of Potential Functions in Computer Learning Theory*. Nauka, Moscow (Russ.).

Jackson, A. G., Pawlak, Z., and LeClair, S. R. (1998). Rough sets applied to the discovery of material knowledge. *J. Alloys & Compounds*, **279**, 14.

Jurs, P. C., and Isenhour, T. L. (1975). *Chemical Applications of Pattern Recognition*, John Wiley & Sons, New York.

Kiselyov, N. N. (1991). System for relational database management DATAREAL-1810, *Upravlyayushchie Sistemy i Mashiny*, **6**, 105 (Russ.).

Kiselyova, N. N. (1987). Prediction of Heusler-phases with composition ABD_2 (D = Co, Ni, Cu, Pd), *Izvestiya Akad. Nauk SSSR, Metalli*, **2**, 213 (Russ.).

Kiselyova, N. N. (1993a). Information-predicting system for the design of new materials, *J. Alloys & Compounds*, **197**, 159.

Kiselyova, N. N. (1993b). Prediction of inorganic compounds: experiences and perspectives, *MRS Bull.*, **28**, 40.

Kiselyova, N. N. (1995). *The Design of Inorganic Compounds: Searching for New Electro-optical, Ferro-electric, Superconducting, and Semiconducting Materials (final report)*. EOARD SPC-94-4097.

Kiselyova, N. N. (1997). Application of artificial intelligence methods to inorganic compounds design, *Perspektivnye Materialy*, **4**, 5 (Russ.).

Kiselyova, N. N., and Burkhanov, G. S. (1987). Prediction of crystal phases in ternary systems with elements of 5th group using computer learning methods, *Izvestiya Akad. Nauk SSSR, Neorganicheskie Materialy*, **32**, 2006 (Russ.).

Kiselyova, N. N., and Burkhanov, G. S. (1989). The search for new ternary phases with Al, Ga and In using an information-predicting system, *Izvestiya Akad. Nauk SSSR, Metalli*, **1**, 218 (Russ.).

Kiselyova, N. N., Gladun, V. P., and Vashchenko, N. D. (1998a). Computational materials design using artificial intelligence methods, *J. Alloys and Compounds*, **279**, 8.

Kiselyova, N. N., and Kravchenko, N. V. (1992) Database on ternary inorganic compounds' properties as a basis for computer design of new substances, *Zh. Neorgan. Khimii*, **37**, 698 (Russ.).

Kiselyova, N. N., Kravchenko, N. V., and Petukhov, V. V. (1996). Database system on the properties of ternary inorganic compounds (IBM-PC version), *Inorganic Materials*, **32**, 567.

Kiselyova, N. N., Lutsyk, V. I., Vorob'yova, V. P., *et al.* (1989). Cybernetic prediction of new compounds in the systems $AO\text{-}B_2O_3$, *Dokladi Akad. Nauk SSSR*, **304**, 657 (Russ.).

Kiselyova, N. N., Pokrovskii, B. I., Komissarova, L. N., and Vashchenko, N. D. (1977). Application of a cybernetic concept formation system to predicting the formation of multi-component chemical compounds, *Zh. Neorgan. Khimii*, **22**, 883 (Russ.).

Kiselyova, N. N., Vashchenko, N. D., Gladun, V. P., *et al.* (1998b). Prediction of inorganic compounds promising for the search for new electro-optical materials, *Perspektivnye Materialy*, **3**, 28 (Russ.).

Kiselyova, N. N., and Savitskii, E. M. (1977). Prediction of inorganic compounds with the composition ABO_2 using computer learning, *Dokladi Akad. Nauk SSSR*, **235**, 1367 (Russ.).

Kiselyova, N. N., and Savitskii, E. M. (1979). Prediction of chalco-spinels with the composition AB_2X_4 by using a computer learning method, *Zh. Neorgan. Khimii*, **24**, 1427 (Russ.).

Kiselyova, N. N., and Savitskii, E. M. (1981). Cybernetic prediction of Chevrel phases with selenium, *Zh. Neorgan. Khimii*, **26**, 1364 (Russ.).

Kiselyova, N. N., and Savitskii, E. M. (1982). Prediction of compounds with the composition $ABHal_4$ in systems of halides, *Zh. Neorgan. Khimii*, **27**, 2880 (Russ.).

Kiselyova, N. N. and Savitskii, E. M. (1983). Prediction of the formation of new silicides with the $ThCr_2Si_2$-structure, *Izvestiya Akad. Nauk SSSR, Neorganicheskie Materialy*, **19**, 489 (Russ.).

Kiselyova, N. N. and Savitskii, E. M. (1984). Prediction of new phases in the systems $AHal_2$-BHal, *Zh. Neorgan. Khimii*, **29**, 3104 (Russ.).

Kravchenko, N. V., Burkhanov, G. S., Kiselyova, N. N., *et al.* (1991). Databank on the properties of crystals for the control of lasers, *Izvestiya Akad. Nauk SSSR, Neorganicheskie Materialy*, **27**, 164 (Russ.).

Kutolin, S. A., and Kotyukov, V. I. (1978). Chemical affinity function and computer prediction of binary compositions and properties of rare earth compounds, *Zh. Phys. Chem.*, **52**, 918 (Russ.).

Kutolin, S. A., and Kotyukov, V. I. (1979a). Prediction of the properties of binary compounds of rare earth elements on the base of their simplified electron configurations, *Izvestiya Akad. Nauk SSSR, Neorganicheskie Materialy*, **15**, 96 (Russ.).

Kutolin, S. A., and Kotyukov, V. I. (1979b). Computer prediction of the compositions of compounds in ternary systems and their properties as a function of electron configurations of their components, *Izvestiya Akad. Nauk SSSR, Neorganicheskie Materialy*, **15**, 1389 (Russ.).

Kutolin, S. A., Vashukov, I. A., and Kotyukov, V. I. (1978). Prediction of binary compounds of rare earth elements and their properties using a computer, *Izvestiya Akad. Nauk SSSR, Neorganicheskie Materialy*, **14**, 215 (Russ.).

Laves, F. (1956). Crystal structure and atomic size; in *Theory of Alloy Phases*. ASM. Cleveland. p. 124.

Manzanov, Ye. E., Lutsyk, V. I., and Mokhosoev, M. V. (1987). Influence of the features of system selection on the prediction of compound formation in the systems A_2MoO_4-$B_2(MoO_4)_3$ and A_2MoO_4-$CMoO_4$, *Dokladi Akad. Nauk SSSR*, **297**, 646 (Russ.).

Mathias, B. T. (1955). Empirical relation between superconductivity and the number of valence electrons per atom, *Phys. Rev.*, **97,** 74.

Pao, Y. H., Duan, B. F., Zhao, Y. L., and LeClair, S. R. (1999). Analysis and visualization of category membership distribution from multivariant data, *Proc. Second Int. Conf. Intelligent Processing & Manufacturing of Materials*, vol. **2**, July 10–15, 1999. Honolulu, Hawaii. p. 1361.

Pettifor, D. G. (1983). Electron theory of metals in *Physical Metallurgy*, Part I, 3rd Edition (eds R. W. Cahn, and P. Haasen). North-Holland, Amsterdam, p. 73 .

Popov, E. V. (1987). *Expert Systems*. Nauka. Moscow (Russ.).

Pospelov, G. S. (1988). *Artificial Intelligence as a Base of New Information Technology*. Nauka. Moscow (Russ.).

Pospelov, G. S., and Pospelov, D. A. (1985). *Artificial Intelligence – Applied Systems*. Znanie. Moscow (Russ.).

Rabe, K. M., Phillips, J. C., Villars, P., and Brown, I. D. (1992). Global multinary structural chemistry of stable quasicrystals, high-T_c ferroelectrics, and high-T_c superconductors, *Phys. Rev.*, **B45**, 7650.

Savitskii, E. M., Devingtal, Yu. V., and Gribulya, V. B. (1968). About recognition of binary phase diagrams of metallic systems using the computer, *Dokladi Akad. Nauk SSSR*, **178**, 79 (Russ.).

Savitskii, E. M., Gladun, V. P., and Kiselyova, N. N. (1977). Prediction of compounds with composition $A_2B_2O_7$ and the pyrochlore-structure, *Dokladi Akad. Nauk SSSR*, **233**, 657 (Russ.).

Savitskii, E. M., and Gribulya, V. B. (1985). *Application of Computer Techniques in the Prediction of Inorganic Compounds*. Oxonian Press, New Delhi.

Savitskii, E. M., Gribulya, V. B., and Kiselyova, N. N. (1979). Cybernetic prediction of superconducting compounds, *CALPHAD*, **3**, 171.

Savitskii, E. M., Gribulya, V. B., and Kiselyova, N. N. (1980). On the application of cybernetic prediction systems in the search for new magnetic materials, *J. Less-Common Met.*, **72**, 307.

Savitskii, E. M., Gribulya, V. B., and Kiselyova, N. N. (1981). Forecasting of superconducting compounds, *Phys. Stat. Sol. (a).*, **63**, K67.

Savitskii, E. M., Gribulya, V. B., and Kiselyova, N. N. (1982a). Cybernetic prediction of inorganic compounds and its correlation with experiment, *Crystal Res. & Technol.*, **17**, 3.

Savitskii, E. M., Gribulya, V. B., Kiselyova, N. N., *et al.* (1990). *Prediction in Materials Science Using the Computer*. Nauka, Moscow, (Russ.).

Savitskii, E. M., and Kiselyova, N. N. (1978). Prediction of superconducting Chevrel phases, *Dokladi Akad. Nauk SSSR*, **239**, 405 (Russ.).

Savitskii, E. M., and Kiselyova, N. N. (1979). Cybernetic prediction of the formation of phases with the composition ABX_2, *Izvestiya Akad. Nauk SSSR, Neorganicheskie Materialy*, **15**, 1101 (Russ.).

Savitskii, E. M., and Kiselyova, N. N. (1983). Search criteria for chemical compound formation in salt systems using cybernetic methods, *Dokladi Akad. Nauk SSSR*, **272**, 652 (Russ.).

Savitskii, E. M., and Kiselyova, N. N. (1984). Cybernetic prediction of new AB_2Ge_2 germanides with crystal structure type $ThCr_2Si_2$, *Izvestiya Akad. Nauk SSSR, Metalli*, **1**, 191 (Russ.).

Savitskii, E. M., and Kiselyova, N. N. (1985). Prediction of the crystal structure type of some equiatomic ternary compounds, *Izvestiya Akad. Nauk SSSR, Metalli*, **5**, 182 (Russ.).

Savitskii, E. M., Kiselyova, N. N., Pishik, B. N., *et al.* (1984). Development of an automatized data bank on the properties of ternary inorganic compounds, *Dokladi Akad. Nauk SSSR*, **279**, 627 (Russ.).

Savitskii, E. M., Kiselyova, N. N., and Shkatova, T. M. (1982b). Cybernetic-statistical approach to the search for substances with predefined properties, *Izvestiya Akad. Nauk SSSR, Neorganicheskie Materialy*, **18**, 235 (Russ.).

Savitskii, E. M., Kiselyova, N. N., and Vashchenko, N. D. (1978). Application of computer learning to the search for new compounds with composition AB_2Te_4, *Dokladi Akad. Nauk SSSR*, **239**, 1154 (Russ.).

Shah, M., and Pettifor, D. G. (1993). Structural stability, local topology and electron count in small s-valent clusters, *J. Alloys & Compounds*, **197**, 145.

STN (1993). *10 years of STN International. Databases in Science & Technology*, FIZ Karlsruhe. STN Service Center Europe.

Talanov, V. M., and Frolova, L. A. (1981). Investigation of chalco-spinel formation using the method of potential functions, *Izvestiya VUZ. Khimiya i Khimicheskaya Technologiya*, **24**, 274 (Russ.).

Villars, P. (1995). Factors converning crystal structures. In *Intermetallic Compounds. Principles and Practice*, Vol. 1. (eds J. H. Westbrook, and R. L. Fleischer). John Wiley and Sons, Chichester, p. 227.

Villars, P., Brandenburg, K., Berndt, M. *et al.* (2001). Binary, ternary, and quaternary compound former/non-former prediction via Mendeleev number. *J. Alloys & Compounds*, **317/318**, 26.

Vozdvizhenskii, V. M. (1975). *Prediction of Binary Phase Diagrams*. Metallurgiya. Moscow (Russ.)

Vozdvizhenskii, V. M., and Falevich, V. Ya. (1973). Application of a computer pattern-recognition method to the identification of phase diagram type of binary metal systems, in *Basic Regularities in the Constitution of Binary Metallic Systems, Phase Diagrams*. Moscow. Nauka. p. 119 (Russ.).

Westbrook, J. H., and Grattidge, W. (1991a), The role of metadata in the design and operation of a materials database, in *Computerization and Networking of Materials*

Databases, Vol. 2, ASTM STP 1106 (eds J. G. Kaufman, and J. S. Glazman). American Society for Testing and Materials, Philadelphia, PA (1991) 84.

Westbrook, J. H., and Grattidge, W. (1991b). A glossary of terms relating to data, data capture, data manipulation, and databases, *CODATA Bull.*, **23**, Nos. 1 and 2, p. 196.

White, H. J. (1985). Thermodynamic databases in the US national standard reference data system, *CODATA Bull.*, **58**, 45.

Yan, L. M., Zhan, Q. B., Qin, P., and Chen, N. Y. (1994). Study of the properties of intermetallic compounds of the rare earth metals by artificial neural networks, *J. Rare Earths*, **12**, 102.

Yudina, N. V., Petukhov, V. V., Cheremushkin, E. A., *et al.* (1996). Data bank on the acousto-optical, electro-optical, and nonlinear optical properties of materials, *Crystallography Reports*, **41**, 464.

Zagoruiko, N. G. (1999). *Applied Methods of Data and Knowledge Analysis*. Institute of Mathematics Publishing (Russ.).

Zemskov, V. S., Kiselyova, N. N., Kiselyov, N. N., *et al.* (1995). DIAGRAMMA a database on the phase diagrams of semiconducting systems, *Inorganic Materials*, **31**, 1096.

Zemskov, V. S., Kiselyova, N. N., Petukhov, V. V., *et al.* (1998a). Database on phase diagrams of semiconducting systems, *Izvestiya VUZ. Materialy elektron. tekhniki*, **3**, 17 (Russ.).

Zemskov, V. S., Kuznetsov, F. A., and Ufintsev, V. B. (1998b). Databanks on semiconducting and other materials for electronics and the technology of their production, *Izvestiya VUZ. Materialy elektron. tekhniki*, **3**, 13 (Russ.).

Zhou, B., Jin S. M., Shao J., and Chen, N. Y. (1989). IMEC – an expert system for retrieval and prediction of binary intermetallic compounds, *Acta Metallurg. Sin.*, **B2**, 428.