INTEGRATION OF MATERIALS DATA SYSTEMS FOR MATERIALS DESIGN

H. CHEN and S. IWATA
Department of Nuclear Engineering, Faculty of Engineering, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo, 113 Japan

ABSTRACT

To integrate multiple data base components for the purpose of materials design, a prototype metadata system containing well defined materials computerized dictionary to connect data from different data bases and levels of abstraction is developed. A relational data base management system(RDBMS) INGRES6.3 is used to create both the metadata system and a sample data base for integrated data systems. One of the integration examples of materials data systems is described to show the effectiveness of the computerized dictionary.

INTRODUCTION

Using computers to deliver materials information is now becomes a reality. Many materials data systems have been developed and many new materials data bases of different features are now being developed. Thus, materials information and knowledge are supplied in a variety of forms, e.g. tables, diagrams, natural language and pictures, and recently in so-called hyper-media forms. Each form has been historically developed because of its usefulness to describe each content. In spite of this fact, there is still no standardized data model for describing materials data in general due to the complexity of materials information as well as trade-offs of costs and accuracies in description, and to the efficiency of DBMS to describe them. The situation becomes more complex if piecewise information as in case of advanced materials are included.

As it is very difficult to provide information by one data system for wide variety of applications, ad hoc integration of relevant systems is usually needed as a practical compromise. Even for one simple application of materials, different levels and forms of information and knowledge are usually involved, we would integrate different data sources to increase reliabilities. Such an integrated system has to include the capability to deal with a wide variety of information so that all of the relevant information can be used effectively.

Many approaches to integrate data systems have been made but most of them have stayed at the prototype stage[1][2][3][4] with a few exceptions for simple retrievals[5]. In MD-GEN(Materials Data system with Graphical user interface, Extended data types and Networking) developed by T.Ashino, the system can produces a "map" table on all data models to integrate distributed data bases of different locations and data models managed by different RDBMSs[3]. Generic files concerning fundamental properties and industrial standards are also built to create a set of maps on all materials with respect to designated properties in the material data base. From such a generic map, users can go into the depth of experimental data. For the purpose of more intelligent integration of materials data systems, effective utilization of computerized dictionary and their suitable management are needed to be solved.

As one of complementary system to MD-GEN, a prototype design called Computerized Dictionary System of Materials Metadata (CDSMM) is implemented. A tool INGRES/Windows4GL is used to frame the menus of CDSMM. The main feature of this dictionary system is that the contents of the dictionary are extended for materials design more than conventional thesauruses. The dictionary itself is managed using the same RDBMS used to create the sample data base for integration.

COMPUTERIZED DICTIONARY SYSTEM

The information about application data, here, materials characterizations, properties, etc., are referred to as metadata. Application data are usually defined, managed, and controlled by the DBMS via their metadata mainly to keep the data independence. For distributed data systems, the diversity of data sources and the heterogeneity of data representations are to be included in the data independence. In a similar way, the metadata themselves need to be managed and controlled at the level of data base administrators. In addition to storing, managing, and controlling of the data, management of networks becomes essential.

An efficient approach in designing a dictionary system is to make it as an integral part of the whole DBMS. This section describes the schema of a computerized dictionary system of material metadata implemented by INGRES6.3.

Dictionary Schema

Metadata about different data base components and levels of abstraction of materials technical terms, e.g. materials classifications and designations, processing methods, product forms, test methods, properties and independent and dependent variable names, units and even data value domains, are summarized and then classified into several groups, e.g. "Material Description", "Property &

Test Method", "Microstructure" and "Application". Metadata on each group are stored in a correspondent subdictionary. Each subdictionary consists of a series of relational tables organized in a hierarchical structure.

The complex relations between different materials technical terms are also defined and represented "step by step" due to the wide variety of existing information and our imperfect understanding on materials and their utilization. Relations most commonly used are KIND_OF, PART_OF, and DESCRIBED_BY. By using these relations, hierarchical relation of each group of materials information can be obtained. The definitions and representations of the hierarchies are to be very general and natural for most users but to be flexible for easier improvements of the system for data base administrators. Fig.1 shows a portion of materials groups hierarchy.

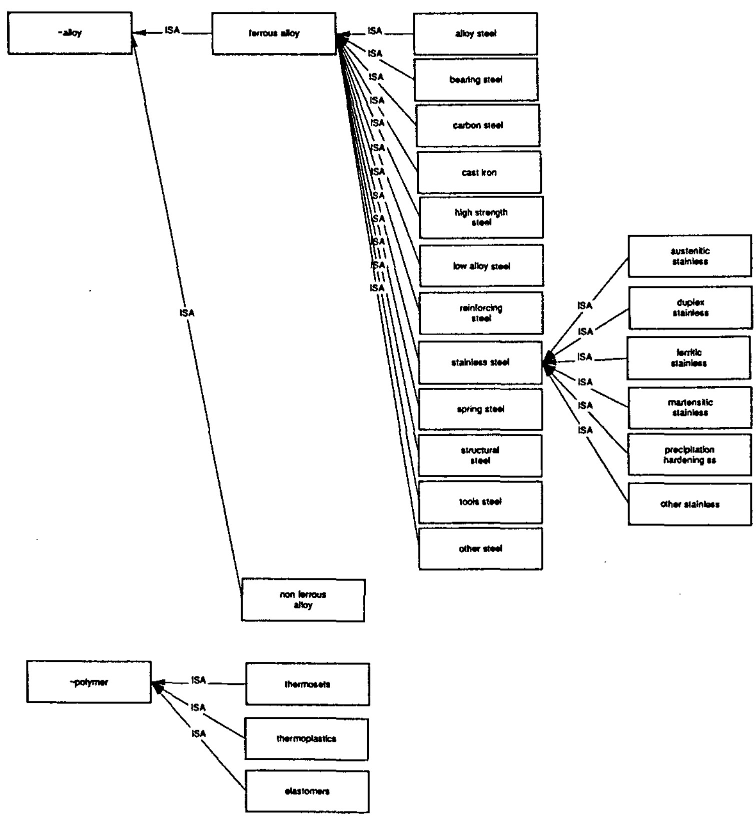

Fig.1 Examples of hierarchy of materials groups.

In addition to the dictionary used partly to generate graceful user interface, such directory information of data systems as data base names, table names and field names, are also included in the metadata to manage networking.

Data Model of Prototype Dictionary System

A prototype dictionary system CDSMM is implemented in our integration system using a tool INGRES/Windows4GL. Pop-up menus are made for the communication between users and the integrated system. When a user logs in the system, s/he sees the root menu of the menu tree of the system. The selection of any item from this menu invokes some other menus and/or programs.

The selection of item "Metadata Manipulation" from the root menu starts the schema of CDSMM and the user sees the directory screen as shown in Fig.2. By selecting from this menu, user can start exploration of materials data.

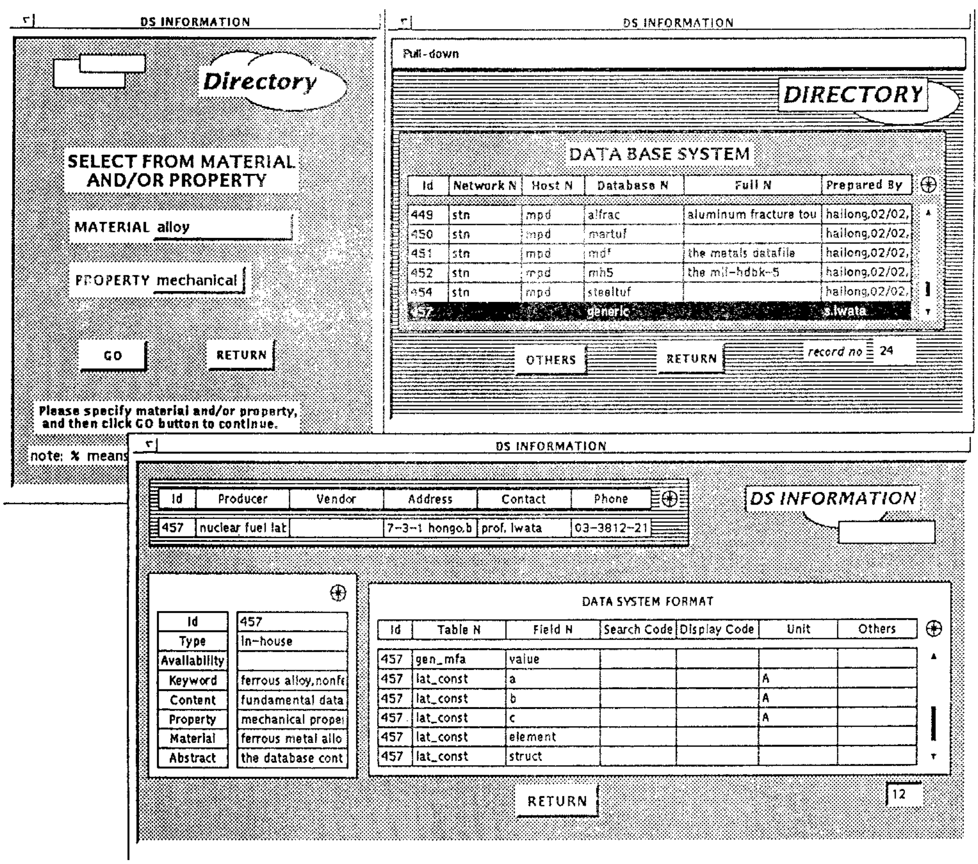

Fig.2 Examples of user guides by the directory system.

INTEGRATION EXAMPLES

One of the ways to cope with the complexity is to approach it from generic information, so that an interactive interface is structured on this criterion, namely, from generic to specific, from whole to part, from macroscopic to microscopic, from set to element, from global to detail. As design procedures are usually continuous searching for better solutions through trials and errors, an interactive interface for integration of multiple data base components is developed by taking advantage of the CDSMM.

When logging in the integration system, a user can access any data base component through a series of menus or through the interactive interface that connects users and the data base components. This connection is to be transparent to the users. By displaying menus and prompting commands, the system makes queries for users to retrieve necessary information. CDSMM helps to rewrite the generic query at first prepared by users into a set of executable queries which meet data models of connected materials data systems. Some rewriting examples are listed below.

```
Ex.1. Query:      SELECT           [corrosion]
                       FROM        [integrated files]
                       WHERE       [condition]='BWR-7'
                       AND         [materials]='Zircaloy-2';
      Rewriting using DESCRIBED_BY:
           [corrosion] -> [environments^ oxide thickness v
                       solute element  diffusion] -> ---
           [integrated files] -> [table A, table B, ---] -> ---
           ['BWR-7'] -> ['composition of water' ^ 'heat flux' ^
                       'irradiation condition' ---] -> ---
      Rewriting using PART_OF:
           ['Zircaloy-2'] -> ['liner' ^ 'substrate alloy' ^ 'surface oxide'] -> ---
      Rewriting using KIND_OF:
           ['surface oxide'] -> ['semiconductor'] -> ---
```

By taking advantage of the predefined hierarchical relations of material information, the system can navigate to retrieve relevant data. If stored information is incomplete, CDSMM evokes to search alternative information through DESCRIBED_BY, PART_OF and KIND_OF relations as shown in the above example. Besides, the user can partly modify the output formats as s/he likes regardless of the exact format of each data base component. As an example, Fig.3 shows an output example of a subset of zirconium alloys in all alloys defined in Japanese Industrial Standards and their grouping by crystal structures with respect to elongation and ultimate tensile strength.

After getting a big picture on target materials, the user could go into the details of zirconium alloys navigated by part_of relation as shown in Fig.4.

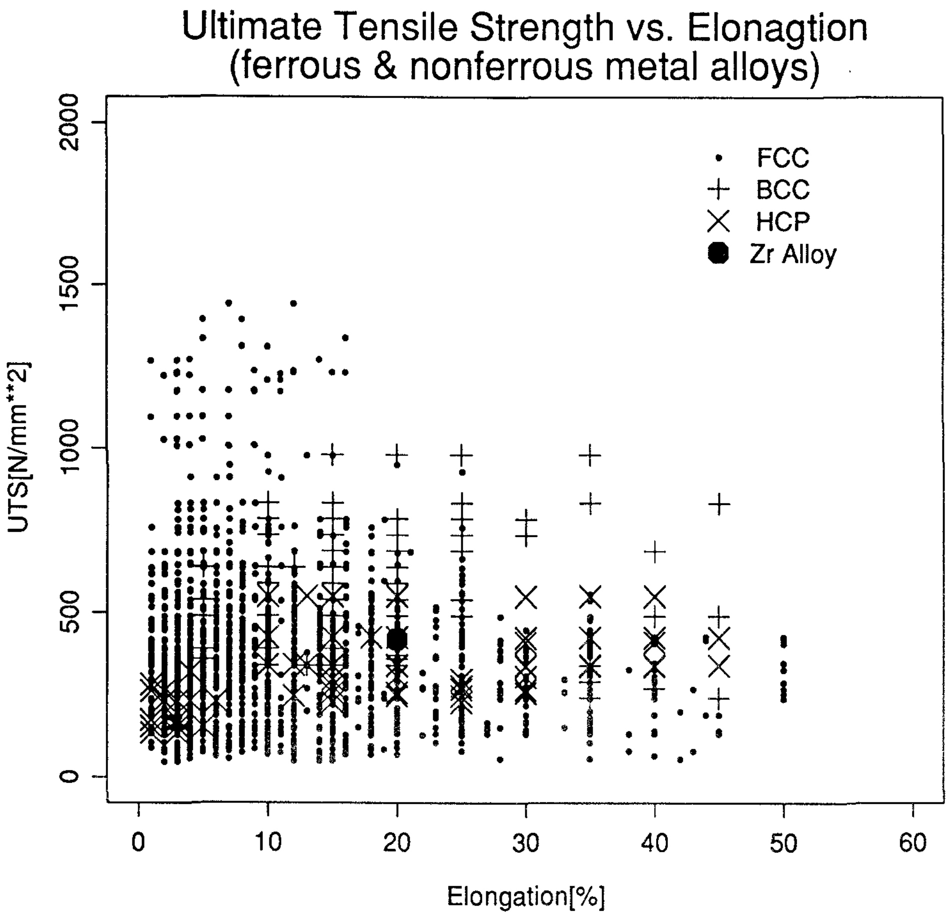

Fig.3 An output example of a subset of zirconium alloys in all alloys defined in Japanese Industrial Standards and their grouping by crystal structures with respect to elongation and ultimate tensile strength.

The communication between the user and the system is conducted by menus. The user can sets up his query conditions only by selecting suitable menus or answering the prompts in restricted natural language. This removes the memory burden from the user because s/he does not have to remember all the details about data formats of each data base component.

CONCLUDING REMARKS

As one of important peripherals to integrate relevant systems, utilization of computerized dictionary is proposed by using sample data bases. In order to make the dynamic features of materials information more realistic, integrations with other methods as qualitative and quantitative simulations, and visualization tools are required.

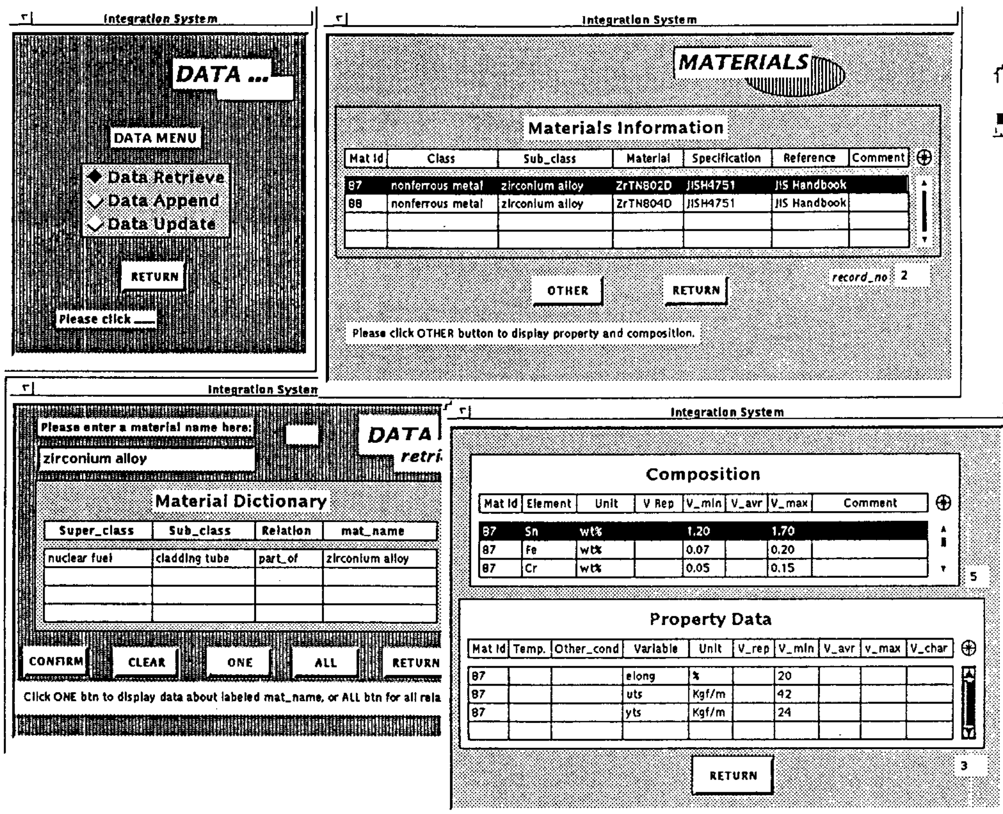

Fig.4 Retrieved example of zirconium alloy through part_of relation.

REFERENCES

[1] S. Iwata, T. Ashino and S. Ishino, Materials Database for Materials Design, Journal of Nuclear Materials 179-181(1991), 1135-1138.

[2] McCarthy, J. L., "Information Systems Design for Material Properties Data," Computerization and Networking of Materials Data Bases, ASTM STP 1017, J. S. Glazman and J. R. Rumble, Jr., Eds., American Society for Testing and Materials, Philadelphia, 1989, pp. 135-150.

[3] T. Ashino and S. Iwata, Integration of Materials Data System.(1) Prototype System Design and Extension of Data Types, Journal of Japan Society of Information and Knowledge, Vol.1,1990, No.1, pp75-91.

[4] H. Chen and S. Iwata, Interface Developments to Distributed Materials Data Systems(1), to be published in Journal of Japan Society of Information and Knowledge.

[5] Kaufman, J. G., "The National Materials Property Data Network: Response to a Critical Need," Chemical Engineering Data Sources, D. A. Jankowski and T. B. Selover, Eds., AIChE Symposium Series No. 247, Vol. 82, 1986.