Model of knowledge representation about materials in the form of a relational database for CAPCAST system

B. Mrzygłód*, K. Regulski
Department of Industrial Computer Science, Faculty of Metals Engineering and Industrial Computer Science, AGH, Mickiewicza 30, Krakow, Poland
*Corresponding author. E-mail address: bmrzyglod@tempus.metal.agh.edu.pl

Received 29-04-2011; accepted in revised form 04-05-2011

Abstract

This study is related with one of the components of a hybrid decision support system called CAPCAST, implemented under a research project conducted by WIMIIP AGH. It is a model of knowledge representation about materials in the form of a relational database, designed to allow the transformation of its records into components of a rule-based knowledge base.

Keywords: Application Of Information Technology In Foundry Industry, Alloys, Knowledge Integration, Knowledge Codification, Attribute Table, Databases, Methods Of Reasoning

1. Introduction

Competitiveness is one of the key attributes in modern market economy. It can be evaluated in terms of both technology and economy. Most often, quality improvement (e.g. raising the functional properties of a product) will entail an increase in production costs. Yet, good quality does not necessarily mean achieving the best possible (extreme) values of the controlled product parameters, but getting the values required by the customer, or by appropriate standards. Thus, at the design stage of any product, one of the important problems is to determine an optimum compromise between the quality and price.

For each product, a decision space is defined, which contains the admissible materials, manufacturing technologies, possibly also the upgrading technologies (e.g. heat treatment, surface treatment, plastic forming, etc.) with the associated costs. Manufacturers can choose between materials guaranteeing at the very beginning the required properties of the final product and materials that are much cheaper but capable of achieving the desired properties after specific treatment (e.g. toughening). Taking proper decision requires multifaceted analysis of the data on materials and technical knowledge. The assistance that cannot be overestimated in this situation are computer systems for decision support in specific areas. This study is related with one of the components of a hybrid decision support system CAPCAST, implemented under a research project conducted by WIMIIP AGH. It is a model of knowledge representation about materials in the form of a relational database, designed to allow the transformation of its records into a rule-based knowledge base.

2. Design assumptions for CAPCAST system

The first step in building such a model is to determine the object of the reasoning and standard entries into the system. Actually, the reasoning process begins at the stage of placing an order. It
is the customer who gives the technical specification of the product and it is the customer who has to decide whether he can or cannot accept the price dictated by the manufacturer. Based on these restrictions, it is now the manufacturer who must decide whether he is able to execute the order (within the deadline appointed and using the available facilities) and what will be the price of the product as dictated by the cost of production. Improving the procedure of production costs estimation should improve the contract negotiations and make them more efficient. On the other hand, estimating the cost of production involves, among others, also the need to determine the type of material and treatment. The manufacturer can choose what materials he will use for the product and at what price, providing he can check which materials among those that he is capable of producing will guarantee the required technical parameters of the product (and if they require any treatment or not).

So, the system is expected to “prompt” the manufacturer a list of materials that will guarantee the compliance with the required technical conditions. These will probably be the materials that can meet the requirements when still in the as-cast condition, as well as those that meet the requirements after heat treatment. The list must also include the cost of material and of the heat treatment, if the latter one is required. A model of this type will provide the manufacturer with a tool to assist the decision-making process (Fig. 1).

The diagram in Figure 1 shows the basic requirements for a material-related module in a CAPCAST system. The system user (manufacturer or his employees) introduces to the system the customer's technical requirements concerning the material, and also gives the current price/cost of materials and processing. Entering current prices is necessary to keep the calculations updated all the time against the ongoing market price fluctuations. However, this operation does not have to be performed each time when the type of material is determined; it is enough to provide current prices in the system keeping pace with the rate of changes in the market and introducing these changes with the same frequency with which the fluctuations occur in the market (e.g. once a week, once a month, etc.). The user will have at his disposal an intuitive and easy to use interface, which should correspond in its layout to the, already known to him, schemes or entry forms of the used documents, such as specifications or standards.

Using the data entered, the materials characteristics built in the database system, and the knowledge base on changes in properties due to additional treatment, the system will carry out the reasoning process. Creating a list of the acceptable materials will be an iterative process. The successive materials will be checked in terms of their ability to satisfy the technical requirements (the algorithm used in this test is the subject of separate publication). If the requirements are met, the material is on the list and its costs are estimated. If the raw material does not meet the requirements, the possibilities of its improvement are analysed (simulation of this analysis is beyond the framework of this study). If improvements can provide an adequate level of performance, the material
is added to the list along with the cost of both the material and its processing.

In this way, a list of materials is made which is then returned to the user. Based on this list, the manufacturer can decide which materials (among those technologically acceptable) will provide the most economical production. He may also consider other options (such as experience in production of a given type, production capacity, etc.).

3. Decision tables and databases

3.1. Relational database model

As it has already been mentioned, the system must use the internal resources. These will be material-related data, including mechanical properties, functional properties, and chemical composition, the knowledge base on changes in these properties due to an appropriate treatment and, of course, also knowledge about the treatment processes as such.

Proper selection of formalisms will determine the functionality of the system. Taking into account the iterative model of the design and implementation of a CAPCAST system, used in the determination of formalisms, it is also necessary to allow for the fact that the successive modules implemented in the future may require the data representation changed to some degree. It is therefore necessary to confer the greatest possible flexibility to each of the internal knowledge resources, so that the integrated form will allow their maximum reuse in other modules.

Currently, the most popular form of data recording and collection are relational databases based on SQL. The concept of relational databases comes from the set theory, which is one of the main branches of mathematical logic. Wherever we are dealing with relational databases, we are in fact operating on sets of elements. The database is presented in the form of tables for entities, relationships and their attributes. Tables, and thus the entire database, can be interpreted as relations in the mathematical sense. The main data structure is the relation which is a subset of Cartesian product of two selected sets representing the admissible values. Operations in the database are operations on relations.

Let $A_1 = \{a,b,c\}$, $A_2 = \{x,y\}$, then $A_1 \times A_2 = \{(a,x), (a,y), (b,x), (b,y), (c,x), (c,y)\}$. Examples of relations which are subsets of Cartesian product $A_1 \times A_2$ are:

$r_1 = \{(a,x), (b,x), (c,x)\}$

$r_2 = \{(a,x), (a,y), (b,y)\}$

Just as in the Cartesian product, the elements of the relation are called tuples. Relation is a set of tuples having the same structure, but different values. Each tuple has at least one attribute [1].

In the case of data on cast materials for example:

$\text{alloy-name} = \{\text{ductile iron, cast steel, aluminium bronze BA1032}\}$

$\text{alloy-group} = \{\text{ferroalloys, bronzes, aluminium alloys}\}$

relation:

$r = \{\{\text{ductile iron, ferroalloys}\}, \{\text{cast steel, ferroalloys}\}, \{\text{aluminium bronze BA1032, bronzes}\}\}$

is the relation determined with $\text{alloy-name} \times \text{alloy-group}$. In the relational data model, $A_1, A_2, \ldots, A_n$ denote attributes. $R = (A_1, A_2, \ldots, A_n)$ is schematic relation $R$.

To facilitate visualisation, it is customary to present a model of database – a schematic relationship - with ER (entity relationship) models. For example, a model database that was created for the purpose of collecting the data from experiments carried out on materials [2] may take the form as shown in Figure 2.
3.2. Integration of databases for the system

At the stage when the requirements for a module of decision support for the choice of materials are specified, a number of attributes describing the properties of materials are selected. Some of these attributes are defined by the standards (e.g. strength in the annealed condition, density, impact resistance KC, KCV, KCU, modulus of elasticity and tensile strength), some of them result from common technical specifications (e.g. abrasion resistance, ferromagnetic properties, damping coefficient). Still others result from the available data sources (databases from experiments, pictures of structures, and TTT, and TTT, diagrams.) In the case of data derived from photographs and diagrams, they are subject to previous digitisation to a tabular form to make them suitable for the process of reasoning. At each of these stages of defining attributes it is necessary to upgrade the work with the knowledge of experts and information comprised in literature to create a database model corresponding to the structure of domain knowledge (Fig 3).

Fig. 3. Schematic decomposition of attributes and data acquisition system for the CAPCAST knowledge base

Because the obtained knowledge structure is more extensive than the data structure in the available sources (in each separately), it is necessary to solve the problem of data integration.

The resulting, internal database system should allow the storage of all available data on materials, not just of those specified attributes that are necessary for the currently implemented model of reasoning, as at later stages of the system execution, other attributes may be needed as well. Hence, another optimisation problem appears: how to plan the structure of a database to, on the one hand, store as much as possible of the available information while, on the other, preserve data integrity and consistency, without generation of empty arrays requiring disk space allocation.

The solution here is to create multiple tables in a database storing different groups of attributes from different sources. And so, a separate database will contain, e.g. the data on materials characteristics, a separate group of tables will hold the data from the digitised images, and in still another place the results of experiments will be stored.

To effectively use so distributed information, it is necessary to develop proper mechanics of the data flow and reasoning algorithms. At the initial phase of project execution it has been assumed that the knowledge base of the system should correspond to the modern mechanisms of reasoning and knowledge integration, and enable their future integration with the use of ontology [3-9].

To this end, it was decided to allow the use of decision tables to store the knowledge in the form of both attributes and conditions.

3.3. Decision tables

Adopting the attribute convention of the domain description allows using the data written in the form of decision tables. Having developed a database describing the materials characteristics, the technological processes and product quality parameters, we can use those data, as well as an additional expert knowledge and literature to establish the relationships and dependencies between different attributes. The discovery of such relationships allows for the construction of rules. But before we proceed to the formalisation of rules in selected logic, they can be saved in a versatile form of attribute tables.

The structure of relational databases, and possible operations performed on sets (set union, intersection and difference, Cartesian product) enable creating the decision-making tables, which also take the form of a relationship. The rows in decision-making tables define the decision rules, which can be expressed as:
where $X = x_1 \land x_2 \land \ldots \land x_n$ is the conditional part of rule, and $Y$ is the decision-making part. Each decision rule sets the decisions to be taken if the conditions given in the table are satisfied.

It is easy to note that the decision table represents an information system, where columns correspond to attributes and rows correspond to objects.

Checking the dependency of attributes is done to omit the unnecessary attributes, the step which can be of crucial importance in optimising the decision-making process. A smaller number of attributes means less of a dialogue with the user and quicker search of the rule base looking for adequate procedure of reasoning. In the case of decision tables that contain very large sets of redundant attributes (created during the operations associated with data mining), the possibilities of reduction can become critical elements in building of a knowledge base. A totally different situation occurs when the decision table is created by knowledge engineers in a controlled manner, based on e.g. literature, expert knowledge, and/or standards, when the set of attributes is authoritatively created basing on the available knowledge about the phenomena. In this case, the reduction of attributes is not necessary, as it can be assumed that the number of unnecessary attributes (if any) shall not deteriorate the model classificability.

In the concept proposed in this article, the attributes will be selected from the existing databases and from our knowledge of the processes of materials treatment, due to which the table will contain only the carefully selected attributes, which are indispensable in the process of reasoning and are operating as premises, i.e. facts possible to exist.

For example, in the case of heat treatment, we can collect data on the values of various mechanical properties, depending on the conducted processes of improvement (Fig. 4).

For example, in the case of heat treatment, we can collect data on the values of various mechanical properties, depending on the conducted processes of improvement (Fig. 4).

<table>
<thead>
<tr>
<th>Mat symbol</th>
<th>Quenching 950°C → cooling rapid (water) H1</th>
<th>medium (oil) H2</th>
<th>slow (air) H3</th>
<th>Aging temperature (100–500°C)</th>
<th>Aging rate [with furnace/ in air]</th>
<th>$R_m$</th>
<th>$R_{0.2}$</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>H3</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>442</td>
<td>227.94</td>
<td>7.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>H3</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>446</td>
<td>232.99</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>H3</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>447</td>
<td>227.3</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>H1</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>465</td>
<td>212.93</td>
<td>23.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>H3</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>472</td>
<td>233</td>
<td>10.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>H3</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>462</td>
<td>229.95</td>
<td>12.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>H3</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>489</td>
<td>235.03</td>
<td>11.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>H3</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>510</td>
<td>256</td>
<td>20.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>H3</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>513</td>
<td>253.38</td>
<td>14.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>H3</td>
<td>$\delta$</td>
<td>$\delta$</td>
<td>516</td>
<td>177.43</td>
<td>18.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reading out the rules from this table will be very simple: "if we apply modifier G and quenching at 950°C and then cooling in air (H3), and additionally aging at 500°C followed by cooling with furnace, then we can expect the mechanical properties of the following values: $R_m = 442$, $R_{0.2} = 238$, $A = 7.5$. In this case we get

$$IF \ldots THEN \ldots: X \rightarrow Y$$

(1)

can see that columns 1-4 are conditional attributes, while columns 5-7 are the decision-making attributes.

Physical data recording does not enforce the use of tables, since individual data can be saved in textual form, though proper meta-description has to be used. This recording method is possible, e.g. in XML language (Extensible Markup Language), designed to represent different data in a structured way. Owing to this solution we can use the informational nature of decision tables without their costly, in terms of disk space, counterpart. This solution should enable the creation of a universal knowledge base which can be used in reasoning applying virtually any language of logic, whether it be a rule-based reasoning, approximate reasoning, or fuzzy reasoning.

4. Summary

This article presents a methodology used in creation of a knowledge base for CAPCAST system, based on different types of input data under the conditions specified by the problem domain, that is, selection of materials for the designed cast products. The database structure and, as a consequence, also the structure of the knowledge base are strongly determined by the availability of data on this particular subject and by the adopted relational model. Under these circumstances, and bearing in mind the need for expansion of the system in subsequent iterations, which cannot be predicted in advance, the decision tables were chosen as a formalism for recording the conditional knowledge of the toughening processes. It also gives prospects for creation of a universal, from the viewpoint of the reasoning systems, tool to generate rules in a particular language of logic based on rules encoded in an attribute decision-making table.

Acknowledgments

Commissioned Research Project No. 7924/B/T02/2011/40 financed from the funds for science by Ministry of Scientific Research and Information Technology.

References


