Managing data and knowledge in the industrial Unilever environment

Piet-Hein Speel
Advanced Manufacturing Technology Group
Unilever Research Laboratorium Vlaardingen
P.O. Box 114, 3130 AC Vlaardingen, the Netherlands
Email: Piet-Hein.Speel@unilever.com

Abstract. In data- and knowledge-intensive areas of the industrial Unilever environment, usually many database systems and knowledge-based systems exist and are being built. However, most systems do not interact which results in various problems. As a solution, one could strive for one huge, multi-purpose data-knowledge-based system having the advantages of both the database research and the knowledge representation research. However, due to both practical and technical problems, this option is not viable. Instead, the vision is to come to an intelligent software system that is build on top of the available database and knowledge-based systems. This software system should be able to manage all underlying systems and to present itself to users as one overall, intelligent system.

1 CURRENT SITUATION

In the Unilever business and research environment, both data and knowledge play a crucial role in decision making in general, and in research, innovation and business processing in particular. Huge amounts of data and knowledge of both empirical and theoretical nature are available. Within Unilever, more and more attention is being paid to an information culture, where data and knowledge in electronic format are shared and reused between various groups and disciplines.

In this information culture, empirical data are considered a treasure. This data is usually stored in relational databases (Oracle is the Unilever standard).

In addition, Unilever possesses a large amount of empirical knowledge in various disciplines, including processing experiences in the foods and detergents business. This knowledge is used in various tasks, including process assessment, monitoring, diagnosis, and design. For these tasks, a progressively increasing number of knowledge-based systems is being designed and implemented in object-oriented programming environments (G2 of Gensym is one of the Unilever standards).

Finally, knowledge about processes that are well understood is usually captured in physical models and implemented in mathematical tools like Matlab. Using these process simulators, all kinds of prediction tasks can be performed.

At the moment, database systems and knowledge-based systems (including mathematical models in the remainder of this paper) exist next to each other, without interaction. Various problems arise due to this situation, including the following ones:

1. redundant information is contained in these systems, which is difficult to maintain and may lead to inconsistencies;
2. user frustrations arise if data that is included in one system need to be re-entered in another system;
3. terminology differs in different systems focused on the same field, leading to confusion;
4. users don't know where to find relevant data and knowledge to accurately perform their tasks.

2 VISION: AN OVERALL INTELLIGENT SYSTEM

Decision makers within Unilever could increase the quality and efficiency of their work when an overall intelligent system would be available that manages all data and knowledge and includes an effective userinterface which deals with all user wishes, including query answering and storage and modification of data and knowledge.

3 RESEARCH ISSUES RELEVANT FOR THE DESIGN ON AN OVERALL INTELLIGENT SYSTEM

Given the current situation in the Unilever industry and research environment and the desire for a userfriendly system, a suitable software architecture needs to be developed. Two possibilities are considered: a huge, single knowledge-based system and an integrated hybrid system.

A software architecture in the form of a single, multi-purpose knowledge-based system normally includes a single database, a single knowledge base and an efficient and a rich reasoning mechanism. We have identified two problems for applying such a software architecture to the industrial Unilever environment.

Within the Unilever business, many independent database systems and knowledge-based systems of various sizes exist and will be constructed in the future. These systems are owned and secured by different (groups of) people. Using standardization approaches, attempts are being made to tune these systems. However, due to the time-consuming (and sometimes cumbersome) standardization process and the security problem, it is extremely unlikely that a huge, single knowledge-based system will be build in the (near) future which replaces all existing and future database and knowledge-based systems.

Besides this practical argument, it is also impossible from a technical point of view to construct a single knowledge-based system which includes the huge amount of data and knowledge that Unilever possesses and is able to operate effectively and efficiently (for some empirical runtime results of some existing knowledge representation systems, for example, see Chapter 6 of [3]).

Instead, a modular, hybrid approach is preferred. Modules correspond to database and knowledge-based systems having their own representation language, reasoning mechanism, and database/knowledge base management system. Most issues in the current database and knowledge representation research fields are focused on the design and construction of these modules.

On top of these modules, an overall coordinative system needs to be developed which manages these modules and which establishes an interface to the users. This approach is closely related to the trend of developing very large knowledge-based systems [1] where the modularity issue is favoured. Modularity leads to a transparent, maintainable, and manageable architecture that can be secured.

In order to realise a modular system, issues regarding to the interactions between the modules are important. First, in order to communicate between agents, research on shareable and reusable terminologies, often called ontologies, needs to be considered [workshops at the ECAI'94, IJCAI'95 and ECAI'96]. In addition, the Knowledge Sharing Effort [2] has initiated extensive research on various topics, resulting, among others, in a knowledge communication language (KQML) and a sharable knowledge specification language (Ontolingua/KIF). A European initiative is formed by the Euroknowledge Esprit Project [AI Watch, August 1995] in which recommendations for knowledge representation standardisation are established, which are relevant for knowledge technology users, theorists and tool vendors. Finally, research on agent architectures needs to be considered.

All these issues need to be elaborated in order to design a management system of the intelligent system that is able to communicate with the users on the one hand, and the various database and knowledge base management systems of the modules on the other hand. The resulting intelligent system will realise our vision and help decision makers to improve the quality and efficiency of their work.

References

[1] N.J.I. Mars. *Towards Very Large Knowledge Bases* IOS Press & Ohmsha, Amsterdam, the Netherlands, 1995

[2] R. Neches, R. E. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, W.R. Swartout. Enabling technology for knowledge sharing AI Magazine 12, 36-56, 1991.

[3] P.H. Speel. *Selecting Knowledge Representation Systems* Ph.D. thesis, University of Twente, Enschede, the Netherlands, 1995