GeM and WeBUSE: Towards a WWW-Database Interface

Madhav Sivadas and George Fernandez

Department of Computer Science,
Royal Melbourne Institute of Technology,
Melbourne, Australia
(madhav,george)@cs.rmit.edu.au

Abstract. This paper describes a generic metadatabase model to enable efficient browsing of structured databases by remote users in the World Wide Web environment. Existing Web-database interfaces rely on the premise that users know the database schema, and that they possess enough knowledge of the context of a schema for the correct interpretation of database semantics and query results. We propose a framework consisting of data structures, mechanisms and tools for representing a more complete description of database schemata. The Generic Metadatabase (GeM) model is capable of storing, as metadata, information about databases designed using most of the popular data modelling techniques. WeBUSE (Web-Based Uniform Schema-browsing Environment) is a suite of tools which enable remote users to browse the augmented database schemata using conventional Web browsers.

1 Introduction

Databases are accessed by users and applications through a set of interfaces, such as SQL, QBE, and graphical interfaces. Regardless of the interface, users are required to know the database schema to issue appropriate queries and to be able to correctly interpret query results. Although database metadata describe the structure of the database, not all the required information for query and result interpretation is available. Also, metadata are usually represented in formats which are better utilised by database management systems rather than users. Users and applications must use implicit contextual information, which is not part of the currently available metadata, to properly interpret data retrievals and updates. For example, if the attribute salary is retrieved with a query, the necessary contextual information to precisely determine its value should include currency, periodicity of pay, whether it is based on an hourly, or daily rate, whether it is gross or after tax, etc. Since database metadata was originally directed to support the DBMS in the management of data, they do not support structures to consistently store the complete context of an attribute, and their interfaces do not provide browsing or querying mechanisms to explore this context information.
In traditional database situations, this is reasonable since the contextual knowledge is embedded in the applications run by the organisation that owns the schema, or is common knowledge among the users within the organisation. However, advances in communications are making possible a different type of access by remote users, who are not in possession of this knowledge. Casual access and database interoperation via the Internet and the WWW require the storage of a more complete description of a database schema and data semantics, as well as adequate mechanisms to support browsing of contextual information.

Semantic modelling of a database schema is the process of formally describing the structure of the data to produce the Conceptual Schema, typically using methods such as Extended Entity/Relationship (EE/R), NIAM, or OMT. These methods attempt to capture the structure of the universe of discourse by describing the type of the relevant entities and the relationships between them. Many of the characteristics of these entities and relationships are described in the model itself. However, the experience of researchers and practitioners working in database interoperation highlights the shortcomings of modelling techniques to completely specify database schemata which in turn impedes the integration of databases. Their requirements for mediation, integration or interoperation are much more extensive than the detail supported by standard modelling techniques. Although this research does not deal with multidatabase integration or interoperation, our central concern is the same: how to provide a more complete description of database schemata, how can this information be stored, and how to provide mechanisms to retrieve the contextual knowledge.

The remainder of this paper is divided as follows. A review of related research is presented in Section 2. In Section 3 we describe a generic specification of the kinds of information stored in database schemata. Section 4 discusses GeM, a structural model for storage of schema information. A complete example of how to use the model is discussed in Section 5. Section 6 describes an architecture for the deployment of this model on the World Wide Web. We conclude in Section 7 with a summary and suggestions for further research.

2 Related Work

Research in the area of multidatabase interoperability has provided considerable insight about the different types of conflicts (incompatibilities) which may occur between two or more database systems [1, 12, 8, 13, 3]. Interoperability requirements ultimately require agreements on the meaning of data, otherwise interoperability is not possible [9]. The aim therefore should be to make information about the database as explicit as possible. If semantics were made explicit in the metadata, it would be possible to detect mismatched assumptions and to create mappings to overcome them [6]. Although this is not an easy task, it has been suggested that repository technology is the best available technique for achieving the goal [6]. A substantial research effort has been devoted to detecting and classifying conflicts between multiple databases. Batini et al provided the first compilation of research in the area of interoperability and schema integration [1].