Use of an Object-based system with reasoning capabilities to integrate relational databases

Goñi A., Illarramendi A., Blanco J.M., Mena E.

Facultad de Informática, Universidad del País Vasco. Apdo. 649,
20.080 San Sebastián Basque Country
e-mail: jipileca@si.ehu.es
phone: + 34 43 218000

Abstract

The integration of heterogeneous and autonomous information sources is a requirement for the new type of cooperative information systems. In this paper we show the advantages of using a terminological system for integrating pre-existing relational databases. From the resulting integrated schema point of view, using a terminological system allows for the definition of semantically richer integrated schema. From the integrated schema generation process point of view, the use of a terminological system permits the definition of a more consistent, broad and automatic process. Last, from the query processing point of view, terminological systems provide interesting features for incorporating semantic and caching query optimization techniques. The advantages are presented in detail for each main step of the integration process: translation, integration and query processing.

1

1   Introduction

Within most organizations there exists many heterogeneous databases, that have been defined independently, which store data that are somehow related. A global and uniform treatment of all these data, maintaining at the same time the autonomy of the component systems, will bring the organizations the opportunity of improving the management of their information.

In the literature several attempts to solve the interoperability problem and in particular the database interoperability can be found. However, a standard solution that gives an answer to all requirements has not been proposed yet. In [?] an overview of different proposals is presented.

In this paper we wish to present our experience of building a system that allows the integration of heterogeneous and autonomous relational databases using a terminological system. This last type of systems, more recently termed description logics (DL), allow for the specification of classes using intensional descriptions phrased in terms of the necessary and sufficient properties that must be satisfied by their instances. The key difference between such descriptions and the earlier class specifications is that such a definition can be used to recognize instances of the class, or to infer answers to queries in situation where there is incomplete knowledge [?]. Indeed the main focus of the paper is on presenting the advantages of using a terminological system for the main steps of the integration process: a *translation* step for obtaining a uniform and richer representation of the schemata that must be integrated; an *integration* step for creating an integrated schema and a *query processing* step for giving answers to the queries formulated over the integrated schema.

Among the related work that also use a terminological system for the system integration process we can mention [?], [?], [?] and [?]. In [?] the use of taxonomic reasoning techniques to support the conceptual design of schemata is proposed. With this aim they have incorporated taxonomic reasoning techniques to some well-known semantic data models. They maintain that the designed schemata will be more consistent. In [?] an existing terminological system (LOOM) is used as a model to describe database schemata and, a system that uses LOOM to provide efficient access to a relational database is described. However, the schemata integration task is somehow limited; they do not allow relations between data elements that belong to different schemata. Our proposal is richer in this sense. In [?] it is argued that the exploitation of pre-existing databases using a terminological system (CLASSIC) allows the generation of new information sources. Nevertheless,

they assume that the integrated schema is defined previously and so the integration task is reduced to the mapping process among the local and the integrated schemata. Finally, in [**?**] CANDIDE, a DBMS based on description logics, is used as a tool to automate a significant part of the schemata integration process.

The main advantages that the use of a terminological system provides for the integration process and that are explained in this paper are the follo