A Database Trigger Strategy to Maintain Knowledge Bases Developed Via Data Migration

Olivier Curé and Raphaël Squelbut

Université de Marne-la-Vallée, Laboratoire ISIS,
77454 Champs-sur-Marne, France

Abstract. The mapping between databases and ontologies is an issue of importance for the creation of the Semantic Web. This is mainly due to the large amount of web data stored in databases. Our approach tackles the consideration of the dynamic aspects of relational databases in knowledge bases. This solution is of particular interest for "ontology-driven" information systems equipped with inference functionality and which require synchronization with legacy database.

1 Introduction

The role of ontologies is foundational in providing semantics to vocabularies used to access and exchange data within the Semantic Web. The creation and maintenance of large scale, concrete and useful ontologies is a major challenge which will drive the success of the next generation Web [6]. Several researches in this direction aim to exploit the databases underlying the "Deep Web" [7] to develop knowledge bases. Different efficient mapping techniques are already available (see section 2) and usually involve a reverse engineering processing. This task can be defined as the analysis of a "legacy" system in order to identify the system's components and their inter-relationships [10]. In the context of the "Deep Web", legacy data is generally stored in (object-) relational database management systems.

The DBOM (DataBase Ontology Mapping) system uses this approach and enables the creation, population and maintenance of expressive and richly axiomatized OWL ontologies from relational database schemata. This paper does not present the global architecture of the DBOM system, which has already been described in [12], but instead focuses on developing maintenance functionalities for knowledge bases. This set of services, called the trigger strategy, handles dynamic aspects of relational databases (triggers, referential actions).

The knowledge bases created and maintained with DBOM belong to the family of Description Logics (DL) systems [4] and we thus make the usual distinction between the description of a domain terminology (called the Tbox) and the actual assertions that make use of the terminology (called the Abox). The trigger strategy is designed to automatically maintain the knowledge base assertional box synchronized as closely as possible to the states of the mapping source database. This approach implies that whenever a database tuple related

C. Bento, A. Cardoso, and G. Dias (Eds.): EPIA 2005, LNAI 3808, pp. 206–217, 2005.

to a DL entity (concept, property) is updated (via a conjunctive query, e.g. SQL query), the Abox is similarly modified to ensure that it remains in a consistent state wrt to its source. With such characteristics, DBOM is designed to be a valuable tool for information-driven and inference enabled applications that are highly dependant on frequently updated data stored in databases. This paper is organized as follows :

- Section 2 presents a survey of the most influencial solutions designed in the field of data migration between databases and ontologies. This section also presents the main characteristics of the DBOM solution.
- Section 3 explains the motivations behind DBOM's trigger strategy and presents its general principle.
- Sections 4 and 5 respectively provide a formal definition of the concepts used in the trigger strategy and detail the implementation, concentrating particularly on the heuristics of this strategy.
- Section 6 concludes this paper with a discussion and hints on DBOM-related future works.

2 Survey of Existing Solutions

This survey is broadened to the study of solutions addressing data migration between databases and ontologies. We have selected eight influencial solutions which pertain to three approaches :

- creation of a knowledge base (Tbox and Abox) from an existing database [3,19,17,8,9].
- creation of a database schema from an existing knowledge base [18].
- creation of a mapping between an existing ontology and database schemata [2,5,16], in order to enable information integration. In this approach, an ontology schema corresponding to the database schema has been manually designed and a mapping is required to enable interoperability.

An evaluation framework has been designed to highlight the main features of these influencial solutions as well as to emphasize the distinctive features of DBOM. These features are :

- an automaticity creation feature which can take two different values:
  - automatic solutions which do not involve user interactions but are limited in terms of providing semantics.
  - semi-automatic solutions which require interactions with human designers, for example via the creation of a mapping file (i.e. [19]) or via adjustments of the system results [17]. This approach is time-consuming and requires a knowledge of both the source schema and the functionalities of the target application. But these solutions also enable the design of expressive and usually richly axiomatized ontologies.
- data coupling which proposes two solutions: a loose coupling, where the Abox is populated at mapping processing time, and a tight coupling, where the Abox is instanciated on-demand, whenever requested by a query.