Chapter 8.1
From Databases to Ontologies

Guntis Barzdins
University of Latvia, Latvia

Janis Barzdins
University of Latvia, Latvia

Karlis Cerans
University of Latvia, Latvia

ABSTRACT

This chapter introduces the UML profile for OWL as an essential instrument for bridging the gap between the legacy relational databases and OWL ontologies. We address one of the long-standing relational database design problems where initial conceptual model (a semantically clear domain conceptualization ontology) gets "lost" during conversion into the normalized database schema. The problem is that such "loss" makes database inaccessible for direct query by domain experts familiar with the conceptual model only. This problem can be avoided by exporting the database into RDF according to the original conceptual model (OWL ontology) and formulating semantically clear queries in SPARQL over the RDF database. Through a detailed example we show how UML/OWL profile is facilitating this new and promising approach.

INTRODUCTION

In this chapter we describe the role of Semantic Web languages, such as RDF and OWL, in transforming the field of traditional relational databases towards more open "world" based on shared ontologies. The purpose of this chapter is not to describe a novel theoretical result, but rather to gather and illustrate a broad range of techniques involved in what is nowadays called "Semantic Web". Special focus is devoted to the use of the UML profile for OWL (ODM, 2007) as an essential instrument for the described transformations. Although tools for some of these technologies are still rather infantile, they are sufficient to demonstrate the full spectrum of possibilities enabled by these new technologies compared to the traditional relational databases. The novelty of this chapter is that through a detailed "almost real life" example we illustrate how these theoretically known benefits can be implemented today with the currently available (though still largely experimental) tools and frameworks.

Semantic Web initially (in the seminal paper by Berners-Lee (2001)) was positioned as a meta-layer for adding meta-information to the unstructured documents stored on the traditional World Wide Web. However, OWL (2004), an ontology language developed for the Semantic Web, is lately emerging as the "lingua franca" for a wide variety of information exchange tasks, including the ones, which traditionally have been handled by the relational databases and their design frameworks, such as ER-models and UML. In this chapter we will consider only the later aspect of the Semantic Web – the applicability of OWL and its UML profile to the field of the traditional Information Systems.

The key idea of the Semantic Web is to unite the semantics of the data (metadata) and the actual data itself. For decades in the Information Systems based on the relational databases these two parts have been artificially separated – the conceptual model (metadata) often used in the design phase of the database was "lost" during the coding phase, were it got substituted by the normalized database schemas and low-level executable code of the user interfaces. Also data itself got buried in the database tables together with abundance of implementation-specific information that made this data hardly understandable to anyone but the programmers of the system. The key purpose of this chapter is to illustrate how semantic web technology is resolving this long standing database design problem and making data again easy accessible through the queries formulated in the terms of the high-level conceptual model (ontology).

Here we need to make an important note about terminology used in this chapter: OWL ontologies formally may contain both the concept definitions (referred to as "Tbox" in the underlying description logics theory) and the actual data ("Abox"). Meanwhile in the semantic web literature it is more common to use term "ontology" only for the concept definitions (Tbox) and to use term "RDF data" for the actual data (Abox). We will conform to this later terminology throughout this chapter.

The reminder of the chapter is structured as follows. In Section 2 we discuss the historic background of the techniques presented in the rest of the chapter and the overall motivation for the proposed approach. Section 3 introduces a legacy university enrollment Information System, which will serve as an example for the rest of the discussion. Section 4 starts the main part of the chapter illustrating in detail the use of the latest Semantic Web tools for re-engineering the university enrollment system to meet the basic Semantic Web requirements. Section 5 illustrates the database consistency constraint mapping to OWL reasoning. Finally, we conclude with the brief summary of the described methods and their potential future developments.

BACKGROUND

Systems for exporting relational database data to RDF have existed since the beginning of the Semantic Web (Berners-Lee, 1998). The need for such export initially was motivated purely by the web-related issues – that vast majority of the structured data on the web is currently buried in the traditional relational databases (so called "deep web") and is presented to the human viewer only through the dynamically generated web pages, where data and its semantics is mostly incomprehensible to the remote "software agents". In this approach relational databases are viewed as data-rich web nodes that need to be turned inside out (through the export to RDF), so that remote "software agents" could directly crawl and integrate structured data from across the web. Although technologies for partial recovering of the structured data from the deep web have been devised, such as "web scraping", "web services" and lately SPARQL, this semantic web vision has