Semantic Mapping of Relational Databases to Ontologies in a Mediation Setup Environment

Pieter Verheyden
Scientific researcher, Ph.D. student
Vrije Universiteit Brussel – STARLab
pverheyd@vub.ac.be

Presentation Abstract

The integration and interoperation of dispersed, heterogeneous (relational) databases is not a new problem. In the last decade, several solutions have been proposed and developed. A more recent approach is the specification of database interoperability at the semantics level, rather than at the “flat” syntactic level, by using so-called ontologies or formal, agreed representations of the semantics of some domain of interest. An ontology not only specifies the domain concepts and their relationships in some language, but also includes the manner in which applications or services are permitted to make use of these concepts. Therefore, ontologies also play a key role in making databases interoperate.

The DOGMA approach to ontology engineering (Developing Ontology-Guided Mediation for Agents) is specifically adapted to the classical model-theoretic view of (relational) databases. Notably, it rigorously separates an ontology base of elementary lexical fact types called lexons, from the rules and constraints governing the concepts referred to by the lexons in the ontology base. These rules are reified in so-called ontological commitments of applications to the ontology base. In this presentation the structure of the commitment layer is made precise by defining Ω-RIDL (Ontology Base Reference and Idea Language), a new type of so-called commitment language. In a nutshell, an application system and in particular its database schema can be assigned a formal semantics, also known as (first order) interpretation. Such semantics in our approach has two separate components, (a) a mapping from the database symbols and relationships to an ontology base, and (b) expressions, separate and “ontological”, of how database constraints restrict the use of, or precisely commit to, the concepts referred by the terms in this ontology base.

In the scope of the SCOP project (Semantic Connection of Ontologies to Patient data), VUB STARNLab and Language and Computing NV (L&C) collaborated in order to find a suitable and working solution for the semantic integration of dispersed, heterogeneous databases in a non-trivial domain, viz. the health sector. Throughout the years, L&C has built up and still maintains an extensive, multilingual medical ontology called LinKBase®. We used the National Drug Code (NDC) Directory of the U.S. Food and Drug Administration (FDA) as a case study. The commitment of the relational database of the NDC Directory to a DOGMA ontology base containing pharmaceutical knowledge imported from LinKBase® was defined in Ω-RIDL. Parts of this commitment definition will be used to explain the main components, the syntactic principles, and the novel aspects of Ω-RIDL.

Defining ontological commitments (i.e. semantic mappings) for relational databases must aim for some practical use. We demonstrate how ontological commitments specify mediators for translating conceptual queries (queries on ontology level) to correct logical queries (queries on database level). Further, we illustrate how some of Ω-RIDL’s key constructs can conveniently be reused in a conceptual query language. Finally, we briefly present the MaDBoKS system (Mapping Databases onto Knowledge Systems), a mediator software component developed by L&C as an extension to their ontology management system LinKFactory®.