

A Bootstrapping Architecture for Integration of Relational Databases to the Semantic Web

Juan F. Sequeda

The University of Texas at Austin
Department of Computer Sciences
Austin, Texas, USA
+1 512 471 9540

jsequeda@cs.utexas.edu

Syed H. Tirmizi

The University of Texas at Austin
Department of Computer Sciences
Austin, Texas, USA
+1 512 471 9540

hamid@cs.utexas.edu

Daniel P. Miranker

The University of Texas at Austin
Department of Computer Sciences
Austin, Texas, USA
+1 512 471 9541

miranker@cs.utexas.edu

ABSTRACT

The vision of the Semantic Web is to create a web of data with well-defined meaning. Most data in the current web is managed by relational databases. Thus, it is imperative for the Semantic Web community to offer easily implemented solutions to bridging relational database content and RDF. Direct mappings means to use the SQL schema to create an OWL ontology and use it to represent the data in RDF. Direct mapping methods have an advantage that they are, intrinsically, automated. If a SQL-schema was created using contemporary model-driven software engineering tools, the resulting OWL ontology can be semantically rich. However, few SQL databases are so developed and therefore semantically weak. We suggest that direct-mapping methods can be integrated by a refinement process. We propose a two step bootstrapping architecture of integrating relational databases with the Semantic Web by first generating a “database-derived putative ontology” and second, refining the putative ontology with a domain ontology and database individuals.

Categories and Subject Descriptors

H.2.4 [Database Management]: Systems

General Terms

Standardization

Keywords

Semantic Web, Relational Databases, Direct Mapping, Ontology

1. INTRODUCTION

It is rare that integrating a database with the Semantic Web accrues obvious benefit to the owners of a database. Thus, organizationally, integrating a database with the Semantic Web is rarely a priority. Hence, it is imperative for the community to make it as easy as possible to bridge relational database content and the Semantic Web. The W3C RDB2RDF Incubator Group is working towards this effort¹.

A widely used approach is the Relational Databases and Ontology mapping methods such as D2RQ, R2O, etc [1]. An advantage of this approach is that existing domain ontologies can be widely reused. A downside of this approach is that the mapping between

the relational schema and the existing ontology is a programming process, which is not attractive for users who want a simple way of exposing their relational content on the Semantic Web.

On the other hand, Direct Mapping methods offer an automatic method of exposing relational content to the Semantic Web by generating a “*database-derived putative ontology*” from the relational database schema. In principle if the relational database schema was successfully created with rich semantics, a direct-mapping method would generate a rich equivalent putative ontology. But in practice, relational database schemas do not appreciate the value of encoding domain semantics. Therefore it would be foolish to think that direct-mapping by itself would be the “easy” way of integrating relational database content with the Semantic Web.

As a result, we propose that by combining direct-mapping with a bootstrapping method, the putative ontology can be refined into a semantically richer putative ontology and matched to a domain ontology. We define bootstrapping as a deterministic mapping of information into an axiom system suitable for more sophisticated learning methods.

Direct mapping shifts the mapping problem from a relational database to ontology mapping problem to an ontology to ontology matching problem. We consider different ontology matching efforts as horizontal and vertical matching. Horizontal ontology matching can be understood as matching different domain ontologies that are “semantically rich”. Vertical ontology matching emphasizes on matching a putative ontology (poor semantics) with a domain ontology (rich semantics). Therefore, our bootstrapping proposal involves vertical ontology matching.

2. FIRST STEP: DIRECT MAPPING

Direct mapping concerns the automatic transformation of database content and schema to a Semantic Web representation. In simplified form this often means that a database’s domain semantics as encoded in its SQL-database schema are identified and translated into equivalent OWL expressions [2], generating a “*database-derived putative ontology*”. The database contents can then be translated into consistent RDF triple representation.

Even though the putative ontology is not semantically equivalent as a domain ontology, it is not semantically incorrect. We consider direct mapping the first step because it automatically generates a putative ontology.

¹ <http://www.w3.org/2005/Incubator/rdb2rdf/>

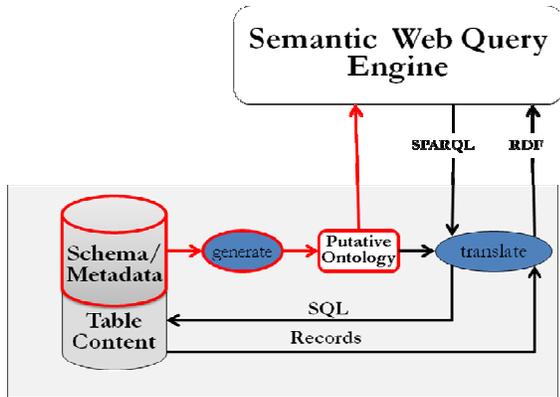


Figure 1. Direct Mapping architecture

Furthermore, we believe that the putative ontology can be the input of a SPARQL to SQL translator. This would allow to (semi)-automatically create SPARQL endpoints to RDB when the putative ontology is generated, refined and matched to a domain ontology, which would facilitate retrieving relational content in RDF on the fly.

3. SECOND STEP: REFINEMENT

A criticism of direct mapping is that the putative ontology will still require integration with some global domain ontology. While this is true, in the larger context it is not a valid criticism because successful development of the Semantic Web will see the development of successful systems that implement ontology-to-ontology mappings [3]. In this context, refining the putative ontology with a domain ontology becomes the second step for successful RDB integration with the Semantic Web.

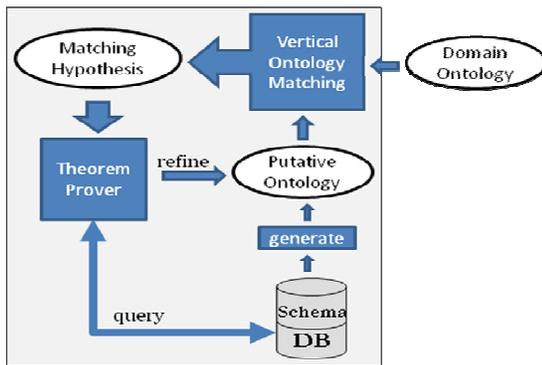


Figure 2. Refinement architecture

Due to the fact that the database-derived putative ontology does not contain all the semantics that are encoded in the RDB, by vertically matching it with a domain ontology, specific matching hypothesis can be determined that will enable the putative ontology to be refined and matched to the domain ontology.

An example of semantics that cannot be clearly represented in a database-derived putative ontology is inheritance. In a RDB, inheritance can be represented by several primary key and foreign key combinations: foreign key is also the primary key, foreign key

and primary key are disjoint or foreign key is a subset of the primary key. It is not possible to automatically identify inheritance in a RDB and represent it in an ontology. However by vertically matching the putative ontology with a domain ontology, matching hypothesis between the putative and domain ontology can be determined. Following the inheritance example, the matching hypothesis could suggest if two tables are in fact related by inheritance.

These matching hypothesis can now be processed by a theorem prover that can decide which matching hypotheses are correct by querying the data in the RDB and identifying if functional dependencies exist. The output of the theorem prover would be the correct matching hypotheses that would then refine the putative ontology and generate mappings between the putative ontology and the domain ontology.

4. RELATED WORK

During the past several years, several methods of direct mapping methods that generate database-derived putative ontologies have been published. Initial methods go back to when only RDFS existed while recent methods transform to OWL-DL or OWL Full. All previous published methods are based on expository examples and present no formal solution, which may lead to ambiguity [1]. In [4], we present a formal transformation system based on FOL and a notion of completeness for a direct mapping transformation system.

5. FUTURE WORK

Several disparities between relational databases and ontologies need to be bridged: inheritance model, identification of symmetric and transitive relationships, open/closed world, check constraint and triggers, and the use of owl:allValuesFrom and owl:someValuesFrom. Our future work entails tackling these problems and research vertical matching techniques that can produce matching hypotheses. Furthermore, our research involves how the putative ontology can be refined by the proven matching hypotheses and the relational database content.

6. REFERENCES

- [1] Sequeda, J.F. 2008. On Direct Mapping for Integrating SQL Databases with the Semantic Web. Technical Report #HR-08-09. The University of Texas at Austin. Department of Computer Sciences.
- [2] Sequeda, J. F., Tirmizi, S. H., and Miranker, D. P. 2007. SQL Databases are a Moving Target. Position Paper for W3C Workshop on RDF Access to Relational Databases (Cambridge, Massachusetts, USA, October 25 – 26, 2007)
- [3] Vrandečić, D., and Sure, Y. 2008. How to design better ontology metrics. In Proceedings of the 4th European Semantic Web Conference (Innsbruck, Austria, June 03 – 07, 2007)
- [4] Tirmizi, S. H., Sequeda, J. F., and Miranker, D. P. 2008. Translating SQL Applications to the Semantic Web. To appear in Proceedings of the 19th International Conference on Database and Expert Systems Applications (Turin, Italy, September 01 – 05, 2008)