Mapping Relational Database Schemas to Ontologies: an Approach

Yassaman Zand-Moghaddam

Supervised by: Dr. J. D. Horton

Faculty of Computer Science, University of New Brunswick, Fredericton, NB

April 2012

Motivation

"A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities" (Berners-Lee, Hendler, and Lassila, 2001).

The Semantic Web that is a new form or an extension of the current web follows the main objective of bringing structure to the content of web pages in order to make the information on the web machine-understandable and machine-processable. To facilitate this goal and represent knowledge in an explicit form, general or domain specific ontologies are created.

The fact that majority of the data found in the current web is stored in relational databases brings up the problem of data migration. How can data stored in a relational database be mapped and migrated to an ontology?

Problem Statement

Mapping relational database schemas to ontologies shows how a database and an ontology are semantically related and correspondences are established between their components.

Database to ontology mapping approaches are classified in two main categories (Fig. 1). Discovering mappings usually has two phases. First simple mappings between entities of the schema and the ontology are found. Second complex compositions based on simple mappings are constructed.

We investigate the problem of finding mappings between an existing relational database schema and an existing ontology and propose a new approach based on the problem of finding isomorphic sub-graphs.

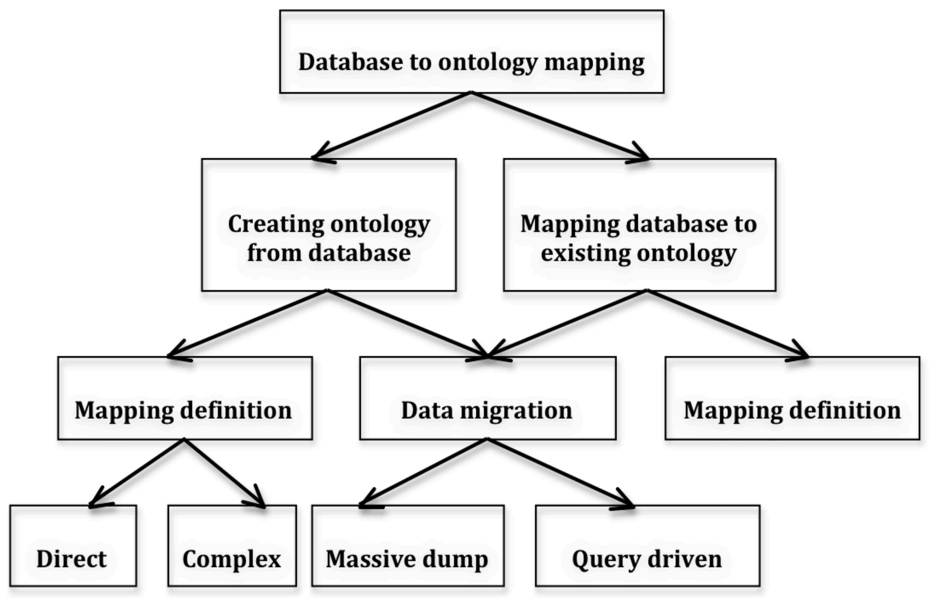

Figure 1. Classification of database to ontology mapping approaches (Ghawi & Cullot, 2007).

The Method

We propose a method in which structural similarities are found first then name similarities, in the matched structures, are investigated. Fig. 2 shows an example of finding similar subtrees.

The inputs of the method are a validated ontology in OWL DL and a relational database schema in SQL DDL; the output is a list of mappings.

The proposed method has the following steps:

1) Create intermediate graphs for the schema and the ontology so that they can be compared together.
2) Ignore labels and compare the structure of the created graphs to find isomorphic subgraphs.
3) Apply a set of name-based checks by following predefined rules. As an example, in two isomorphic subgraphs found after step 2, find the similarity between the names of the nodes with the same number of children. WordNet can be useful in measuring the relatedness of words.
4) Based on the similarities found in step 3 list the mappings.

Problems in the implementation of the method:

- *Time complexity of the subgraph isomorphism algorithms:* For general graphs, the subgraph isomorphism problem is known to be NP-complete that means the matching time can increase exponentially in terms of the size of the graphs; therefore, it causes a problem when dealing with large size general graphs.
- *inheritance modeling in relational database schemas:* Finding IS-A relations in relational databases is not easy since, in contrast to ontologies, in relational databases inheritance is not expressed explicitly; however, in their schemas it can be modeled in different ways(S.H. Tirmizi, J. Sequeda, D. Miranker, 2008).
- *naming styles:* Different naming styles may be used. In building complex words, one developer may use the underscore character as the word separator while other may concatenate simple words together and start each new word with a capital letter. An example: Social_Event, SocialEvent.

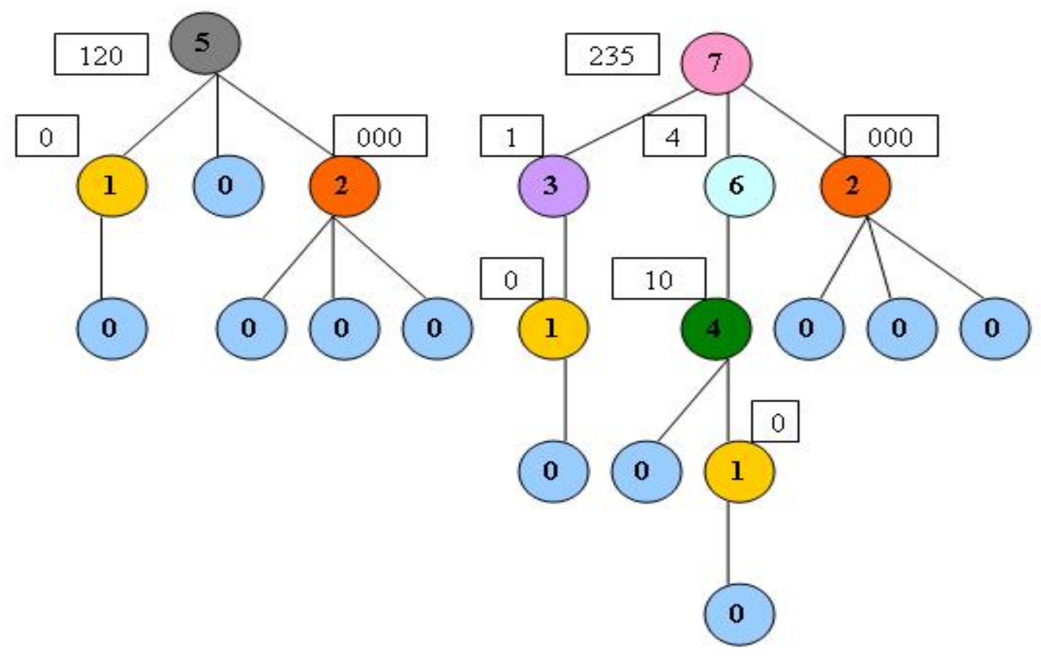

Figure 2. Finding similar subtrees between two concept hierarchy models. Based on the complexity of subtrees in both hierarchy models, a distinct integer is assigned to the root of each subtree.

Evaluation

- Accuracy: With small data sets such as the one used in (Y. An, A. Borgida, and J. Mylopoulos, 2005) it is possible to compare the results with the mappings found by an expert.
- Computation time: It is compared with other approaches in which the linguistic checks are done prior to the structure check.

Real-world Applications

Finding mappings between relational database schemas and ontologies facilitates integration and migration of the data between databases and ontologies. One of the prototype tools that can be used as a Protégé plugin is MapOnto (Y. An, A. Borgida, and J. Mylopoulos, 2005). It is a system for constructing complex mappings between ontologies and relational or XML schemas.

Selected Bibliography

T. Bernrs-Lee, J. Hendler, and O. Lassila, *The Semantic Web*, Scientific American, Vol. 284, 2001, pp. 34-43.

Y. An, A. Borgida, and J. Mylopoulos, *Inferring Complex Semantic Mappings Between Relational Tables and Ontologies from Simple Correspondences*, OTM Conferences, Cyprus, 2005, pp. 1152-1169.

R. Ghawi and N. Cullot, *Database-to-Ontology Mapping Generation for Semantic Interoperability*, 3rd International Workshop on Database Interoperability, Austria, 2007.

S.H. Tirmizi, J. Sequeda, D. Miranker, *Translating SQL Applications to the Semantic Web*, Database and Expert Systems Applications, LNCS, vol. 5181, pp. 450-464, 2008.

S.S. Sahoo *et al.*, *A Survey of Current Approaches for Mapping of Relational Databases to RDF*, W3C RDB2RDF XG Report, January 2009.

Contact

yassaman.zand@unb.ca