Mapping a Database Schema to the Structure of an Existing Ontology

Anahita Nafissi*, Fabio Fiorani and Björn Usadel
Plant Sciences (IBG-2), Forschungszentrum Jülich, D-52425 Jülich, Germany
* a.nafissi@fz-juelich.de

In this work, we describe an approach for mapping the structure of a database schema to the structure of existing ontologies. The database contains plant traits values and plant experimental history. The goal is to identify the semantic correspondences between databases and ontologies and provide a tool that can be more broadly adopted by the community. The approach presented in this work is a semi-automatic approach.

In the literature, there are several approaches which map the database schema to an ontology. The underlying assumption by all approaches is that the chosen ontologies model the same domain as the one modelled by the relational database schema. Some mapping approaches are R2O (Barrasa et al., 2004), DartGrid (Chen et al., 2006), Linked Data Mapper (Zhou et al., 2008), RDOTE (Vavliakis et al., 2010), RDB2OWL [Bumans & Cerans, 2010], MAPONTO [An et al., 2006]. The difference between the above approaches is that some approaches are manual and some are semi-manual. Furthermore, for some approaches a human expert gives the correspondences between database terms and ontology terms.

The database schema of our Phenomis database considered for mapping contains plant phenotyping information and environmental information. The ontologies considered for mapping are plant ontology, phenotypic quality ontology, plant trait ontology, plant environmental conditions, and environment ontology. The mentioned ontologies are very large and contain over 1000 concepts. Unlike the number of concepts, the number of roles is very small (less than 10). Note that the roles denote the relations between domain objects.
In order to map the database to the ontology, we first consider the schema of the database and extract relation names and attributes of each relation. Note that the relations in a relational schema are classified into two categories, namely entity relations and relationship relations (Hu & Qu, 2007). Furthermore, an attribute is also classified into two categories, namely foreign key attribute and non foreign key attribute (Hu & Qu, 2007). A relationship relation is used to connect two other relations and contains foreign key attributes. Unlike a relationship relation which contains only foreign key attributes, an entity relation contains non foreign key attributes. For the mapping process, we do not consider relationship relations and all their attributes. Similarly, we extract concept - and role names of the ontology.

For the mapping process, we have to discover the correspondences between the terms of the database and the terms of the ontology. For this purpose, we compare the relation - and attribute names of the database with the concept names of the ontology. The comparison is performed according to the similarity matches. This means that we find similar matches among the relation and attribute names of the database and concepts of the ontology. Then, the results should be evaluated by a human expert (plant biologist) who is familiar with both the terms used in the database and in the ontologies. Thus, this approach is a semi-automatic approach. For some ontologies the mapping results are more than the others. Furthermore, the human involvement required for mapping varies across different ontologies.
The softwares used for this work are Java, Protégé, SQL.

ACKNOWLEDGEMENTS

This work is performed within the German-Plant-Phenotyping Network which is funded by the German Federal Ministry of Education and Research (project identification number: 031A053)

REFERENCES

An, Y., Borgida A. and Mylopoulos, J. (2006). Discovering the Semantics of Relational Tables through Mappings, Journal on Data Semantics, VII, pp. 1-32.

Barrasa, J., Corcho O., and Gomez-Perez A (2004). R2O, an Extensible and Semantically Based Database-to-Ontology Mapping Language, in Second International Workshop on Semantic Web and Databases (SWDB 2004).

Bumans, G., and Cerans, K. (2010). RDB2OWL: a Practical Approach for Transforming RDB Data into RDF/OWL, in A. Paschke, N. Henze and T. Pellegrini, eds. Proceedings of the 6th International Conference on Semantic Systems (I-SEMANTICS 2010), ACM.

Chen, H., Wu, Z., Mao, Y., and Zheng, G. (2006). DartGrid: a Semantic Infrastructure for Building Database Grid Applications, Concurrency and Computation: Practice and Experience, 18(14), pp. 1811-1828.

Hu, W., and Qu, Y. (2007). Discovering Simple Mappings between Relational Database Schemas and Ontologies, Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), Busan, South Korea, Volume 4825 of LNCS, page 225—238, Berlin, Heidelberg, Springe Verlag.

Vavliakis, K.N., Grollios, T.K., and Mitkas, P.A. (2010). RDOTE-Transforming Relational Databases into Semantic Web Data, in A. Polleres and H. Chen, eds., Proceedings of the ISWC 2010 Posters & Demonstrations Track: Collected Abstracts, pp.121-124.

Zhou, C., Xu, C., Chen, H. and Idehen, K. (2008). Browser-based Semantic Mapping Tool for Linked Data in Semantic Web, in C. Bizer, T. Heath. K. Idehen and T. Berners-Lee, eds., Proceedings of the WWW 2008 Workshop on Linked Data on the Web (LDOW 2008).

