Mapping relational databases through ontology matching: A case study on information migration

Manuel Rodriguez-Mancha, Hector G. Ceballos, Francisco J. Cantu, and Aldo Diaz-Prado

Tecnologico de Monterrey, Campus Monterrey
Eugenio Garza Sada 2501, 64789, Monterrey, Mexico
{mjrodriguez,ceballos,fcantu,jadiaz}@itesm.mx

Abstract. In order to aid domain experts during data integration, several schema matching techniques have been proposed. Despite the facilities provided by these techniques, mappings between database schemas are still made manually. We propose a methodology for mapping two relational databases that uses ontology matching techniques and takes advantage of tools like D2R Server and AgreementMaker for automating mapping generation and for enabling unified access to information. We present the results obtained by some ontology matching algorithms in this context, demonstrating the feasibility of this approach.

Keywords: Relational DataBases, Ontology matching, Information Integration

1 Introduction

Specifying schema matches is a tedious, time-consuming, error-prone and therefore expensive process that requires the participation of one or more database experts. Current advances on the creation of *RDF views* of relational data[1] has provided an uniform interface to data that improves information integration and retrieval, as well as enables applying ontology matching algorithms to the database schema mapping problem.

2 A database mapping methodology based on ontology matching techniques

Counting with experts for both databases/systems during data migration is really exceptional, therefore we propose using training data for guiding this process. A domain expert, i.e. a regular user of the original system, is asked to capture in the new system some examples obtained from the original system (training data) in order to have equivalent information in both databases.

Next we apply an iterative process divided in three phases. In the first phase we generate an ontological representation of each database schema using D2R

[1] W3C RDB2RDF Working Group. `http://www.w3.org/2001/sw/rdb2rdf/`

[1]. The resulting D2RQ mapping is used for mounting a SPARQL end-point that provides unified access to database records as RDF instances and for generating a plain ontological representation of database schemas [3].

In the second phase we use matcher algorithms of AgreementMaker v.023 [2] for identifying correspondences between tables (classes) and fields (properties). In the third phase, the expert verifies mappings' correctness through the comparison of instances of both databases using the corresponding SPARQL end-point and schema equivalences.

Expert's feedback is persisted in a mapping ontology that is used in further iterations. The next iteration begins in the second phase whereas the entire process finishes when no new matching is found or the number of incorrect matches is greater than good ones.

3 First results on information migration

Our methodology was motivated by the migration from an information system in operation to a new one with analogous functioning. Both systems rely on normalized relational databases, the first in MySQL and the second in Oracle. Each database has about 180 tables and 1,400 fields in total.

From all matchers implemented in AgreementMaker, only ASM, DSI and the combination Anatomy identified correspondences. Resulting matches were evaluated by an expert, showing an average 22% of recall for class/table and property/field mappings. On the other hand, precision for class/table mappings was above 90% for these three matchers, but on property matching Anatomy obtained an 84% in comparison with 28% and 25% for ASM and DSI.

String matching techniques used by these matchers allowed detecting similarities between mnemonic names of tables and fields in the first place. Additionally, relations property-class allowed detecting mappings between equivalent fields across related tables.

In this way, mappings detected in the first iteration reduced in a 62% the number of comparisons required for mapping both databases. This means that the expert would have to perform manually only a 38% of the total comparisons.

As future work we will develop an automatic evaluation of mappings by comparing instances extracted from both databases and incorporating them in the ontological representation of schemas. This will enable matchers like IISM that uses mapped individuals for aligning classes as well.

References

1. Bauelos, E., Zrate, J., Almazn, S., Zubia, A.: D2RQ mappings incorporating semantic of domain. Journal of Environmental Studies 18, 59–62 (2009)
2. Cruz, I.F., Antonelli, F.P., Stroe, C.: Agreementmaker: efficient matching for large real-world schemas and ontologies. Proc. VLDB Endow. 2, 1586–1589 (August 2009)
3. Ghawi, R., Cullot, N.: Database-to-ontology mapping generation for semantic interoperability. In: Third International Workshop on Database Interoperability (InterDB 2007) (2007)