Computer and Information Science
www.ccsenet.org/cis
Vol. 3, No. 1
February 2010

Ontologies Acquisition from Relational Databases

Shufeng Zhou (Corresponding author)
College of Mathematics and Science, Liaocheng University
No.34 Wenhua Road, Liaocheng 252059, Shandong, China
E-mail: zhousflc@sohu.com

Guangwu Meng
College of Mathematics and Science, Liaocheng University
No.34 Wenhua Road, Liaocheng 252059, Shandong, China

Haiyun Ling
College of Computer Science, Liaocheng University
No.34 Wenhua Road, Liaocheng 252059, Shandong, China

The research is financed by Natural Science Foundation of Liaocheng University(X051034).

Abstract

Ontologies play an important role in solving the problem of semantic heterogeneity of heterogeneous data sources. The ontology acquisition from the relational database is one of the fundamental technologies on the information integration field. Ontology acquisition from relational database (OARDB) and transformation rules used are presented and discussed. The key technologies and model are given.

Keywords: Ontology, Relational Database, Rules, Semantic Web

1. Introduction

The Semantic Web proposed by Tim Berners-Lee has been viewed as the next generation of the current Web. And with the development of Semantic Web research, people have realized that the success of Semantic Web depends on the proliferation of ontologies and pay more attention to the construction of ontologies. The popularity of ontologies is rapidly growing. However, most of the world's data today are still stored in relational databases. Therefore, it is necessary to acquire ontologies from relational databases.

Though ontology construction tools have become mature over the years, the manual development of ontologies still remains a tedious and cumbersome task. So this paper proposes how to generate ontology automatically or semi-automatically from relational databases.

The paper is organized as follows. Section 2 introduces related works. Section 3 analyzes the ontology acquisition from relation database. Section 4 introduces the implementation of OARDB in detail. And section 5 gives a summary and the future works.

2. Related works

A number of researchers have made contribution to this problem, such as defining semantics in database schema, extracting semantics out of database schema, and serve as a foundation for our work. But there does not exist hierarchies and cardinality about properties in those previous approaches. So these approaches cannot be used to describe the ontology from relational database directly and correctly. In this paper, we analyze the database schema information firstly. Then systematically present how relational database can be transformed into ontologies.

3. Ontologies Acquisition from Relational Databases

In this section, the paper explains the ontologies acquisition from relational databases. First, the paper lists the fundamental condition to express the ontologies acquisition. Then, explains steps to acquire the ontologies from relational databases.

3.1 Fundamental condition

In order to acquire ontologies from relational databases, the paper makes the assumption: The relational database schema is normalized, at least up to third normal form. In fact, the third normal form is the most common normal form in relational database schema. So if some databases might not be well normalized, it is possible to automate the process of finding functional dependencies within data and to algorithmically transform a relational database schema to third normal form.

3.2 Steps to acquire the ontologies

The process of acquiring ontologies can be divided into two stages: Extracting Relational Database Schema Information and Acquiring Ontology. In the first stage, the paper adopts the reverse engineering methods to extract relational database schema information. It is the key stage to acquire the ontology, and then based on the relational database schema information ontological structure can be constructed. After that, the relational database tuples can be retrieved and formed as the ontological instances. So the OARDB generally consists of four steps.

Step 1: Extracting relational database schema information. Such as relation names, attribute names, primary keys, foreign keys and integrity constraints.

Step 2: Analysis of primary keys, foreign keys and attributes information, and then construct ontology and ontological concept.

Step 3: Retrieving tuples from relational database.

Step 4: Mapping the tuples to ontological instances and forms knowledge base.

In the four steps, the step 1 and step 2 are the transformation of relational schema, which can be extracted using Java API. Based on the relational schema information, the ontology and ontological concept are constructed. And the step 3 and step 4 are the transformation of relational data. After step 1 and step 2, an ontological structure is formed. Then we can use the relational tuples as the ontological instances according the transformation rules.

4. Implementation of OARDB

According to schema information extracted from relational database, the paper firstly analyzes the relationship between relational schemas, and then classifies the relational schema. Finally, the rules for acquiring ontology are given below.

4.1 Transformation from relational database schema information to ontology

Schema 1: For relation R, suppose that there is no foreign key. Such as, Course(courseId, courseName, credits, creditHours). This is the most basic relational schema, the corresponding transformation rules are as follows.

Rule 1: A Class can be created, using the relation name as the name of Class.

Rule 2: All attributes domain in relational schema are mapped into xsdDatatype.

Rule 3: All attributes in relational schema are mapped into DatatypeProperty.

Schema 2: For relations R_1, R_2 in database, suppose that the primary key of R_1 consists of only one foreign key referring to R_2, such as PhdStudent(phdId, memo), phdId referring to Student(id, name, birthday,…). Then, in addition to abide by the first schema, we must express the subclass relationship between the two relations. So the rule is as follows.

Rule 4: The class corresponding to R_1 is a subclass of the class corresponding to R_2.

Schema 3: For relations R_1, R_2, R_3, if the primary key of R_1 is consisted of only two foreign keys A_{11}, A_{12} referring to R_2, R_3 respectively, and there is no other attribute except the primary key of R_1, such as Teaching(teacherId, courseId). In this case, the relation R_1 is used to describe the many-to-many relationship between two relations. Then, two inverse object properties can be created based on semantics of the relation. The rule of transformation is as follows.

Rule 5: Two object properties P_{23} and P_{23}' are created based on the semantics of R_1. Suppose that the classes corresponding to R_2, R_3 are C_2, C_3 respectively. The domain and range of P_{23} are C_2 and C_3, and the domain and range of P_{23}' are C_3 and C_2. At the same time, P_{23} and P_{23}' are two inverse object properties.

Schema 4: All schemas with the exception of above schemas. These relational schemas are divided into four cases.

Case 1: For a relation R, the number of the foreign key is equal or greater than 3.

Case 2: For relations R_1, R_2, R_3, if the primary key of R_1 is consisted of only two foreign keys A_{11}, A_{12} referring to R_2, R_3 respectively, and there are other attributes except the primary key of R_1.

Case 3: For a relation R, there is a foreign key. But the foreign key is not integral part of the primary key.

Case 4: For a relation R, there is a foreign key. And the primary key of R is consisted of the foreign key and other attributes of R.

In above four cases, all relations can be transformed into classes according Schema 1. In addition to describe the relationship between relations, inverse object properties can be created applying rule 5.

In addition, the following rules can be created to describe the cardinality constraint on properties.

Rule 6: For a relation, if attribute A (except foreign key) is declared as NOT NULL, then the cardinality of the property corresponding to A is 1.

Rule 7: For a relation, if attribute A (except foreign key) is declared as UNIQUE, then the maximal cardinality of the property corresponding to A is 1.

Rule 8: For a relation, if foreign key F is declared as NOT NULL, then the minimal cardinality of the object property corresponding to F is 1.

Rule 9: For a relation, if foreign key F is declared as NULL sometimes, then the minimal cardinality of the object property corresponding to F is 0.

4.2 Transformation from relational database tuples to ontological instances

Applying above rules, an ontological structure can be extracted from a relation schema. Then, the process of transformation from relation database tuples to ontological instances can start. The rules are as follows.

Rule 10: For a relation R, suppose that class C is corresponding to R, then every tuple t can be mapped to a ontological instance associated with unique identifier by appending the value of the primary key to the name of the relational name.

Rule 11: For a relation R, the values of the tuple can be mapped to the values of the corresponding property of ontological instance.

Rule 12: For a relation R, suppose that there a foreign key F, then the value of F can be mapped to ontological instance I, and the I can be mapped to the value of the object property corresponding to the foreign key.

5. Summary and Future Works

In summary, the main contributions of this paper are listed as follows. Firstly, we have presented a new approach to acquire ontology from relational database. It captures semantic information contained in the structures of the entities. Secondly, we have experimentally evaluated our approach on several data sets from real world domains. The results demonstrate that our approach performs well as compared to some existing approaches.

In the future work, we look forward to comparing our approach with some intermediate approaches. We also hope to consider some machine learning techniques for mining some other interesting and useful semantic mappings.

References

An, Y., Borgida, A., Mylopoulos, J. (2005). Inferring complex semantic mappings between relation tables and ontologies from simple correspondences. In Proceedings of International Conference on Ontologies, Databases and Applications of Semantics (ODBASE'05).1152-1169.

Astrova, Irina. (2004). Reverse Engineering of Relational Databases to Ontologies. *The Semantic Web: Research and Applications*, 327-341.

Berners-Lee, T., Hall, W., Hendler, J., (2006). Creating a science of the Web. Science.769-771.

Chen,H., Wu, Z., Wang, H., Mao, Y. (2006). RDF/RDFS-based relational database integration. In Proceedings of the 22nd International conference on Data Engineering (ICDE'06). 94.