2010 Ninth International Conference on Grid and Cloud Computing

Ontology Automatic Constructing Based on Relational Database

Peng Liu[1], Xiaoying Wang[1,2], Aihua Bao[1], Xiaoxuan Wang[3]

(1. Command Automation Institute, PLA University of Science & Technology, Nanjing, China)
(2. The 202th Hospital, Joint Logistics Department of Shenyang Military Area Command, Shenyang, China)
(3. Department of Surveying and Geo-informatics, Tongji University, Shanghai, China)
gloud@126.com, monica.wxy@163.com, wxx1012@163.com

***Abstract*—Ontologies are often used to achieve interoperability between the nodes in the grid and the cloud. At present, most ontologies are built manually by the experts in the field. A new method aiming at ontology learning based on relational database is proposed. By analyzing the schema of relational database, a group of rules of transition from relational schema to OWL are established. Then a framework for extracting ontology from relational database is proposed. Based on the above work, a heterogeneous data integration based semantic prototype system is designed and implemented. Finally, strong and firm support to applications has been provided.**

Keywords—Ontology; Relational Database; Extraction; Ontology Learning

I. INTRODUCTION

Ontology is famous for Gruber's definition as "Ontology is a defined specification of conceptual model"[1].The importance of Ontology is demonstrated in many aspects and getting generally accepted. Nowadays, Ontology is widely used in semantic web, information intelligent retrieval System, and digital library, etc[2]. Ontology has become vital ways to build semantic knowledge models.

In the past ten years, many tools of ontology constructing have appeared, but most of them can only edit ontology and only support manual method. Thus, ontology constructing is an extremely difficult task. And it makes sense if we pay our attention on reducing the costs of ontology constructing using the technology of ontology learning[3].Nowadays, many scholars have done much work.Du Xiaoyong and Li Man[3] divides ontology learning into learning of structured, unstructured and semi-structured data.

Structured data mainly refers to relational database or object-oriented database. Relational database uses relational models which is a classical model of domain specific modeling[4]. The main task of capture ontology from relational database is to analyze the relational models and try to extract more semantic information inside and map it into the corresponding part of ontology.

II. RESEARCH ON ONTOLOGY EXTRACTING FROM RDB

So far we have many studies on ontology extracting from RDB.

The method employed by tojanovic[5] only analyzes the situation of datasets with equal or include relations.

The method put forwarded by Kashyap[6] requires users to do the semantic tagging，therefore, it is not automatic enough to extract semantic information to generate axiom.

Dogan&Islamaj[7] also have their own way, but their way omits the succession relations and fails to optimize the structure of relational database, so the structure of ontology turns out to be too simple.

Astrova[8] did have a more comprehensive way and corresponding tools, but the method requires using SQL script as input and limits the usage in actual applications.

Protégé has two plugins to extract ontology from relational database named DataMaster[9] and DataGenie[10]. DataMaster tries to build a ontology that can descript the source data structure instead of exploring the semantic relations inside the data, while DataGenie is only responsible for access to database Oracle and fails to meet the demand of extract from heterogeneous database. What's more, this way can only find limited semantic information and cannot explore succession relations, object properties and restriction relations.etc.

III. ONTOLOGY AUTOMATIC CONSTRUCTING BASED ON RELATIONAL DATABASE

The semantic feature makes it possible to solve heterogeneous data integration and achieve interoperation between heterogeneous systems. Now the main method of ontology constructing is still in manual way. Manually-constructed ontology is rich in semantic meaning, but it costs too much time and energy and makes ontology constructing a huge and complicated task and finally becomes bottleneck for knowledge obtaining. As relational database is constructed on the basis of demand analysis by system engineers and it includes many conceptional models

978-0-7695-4313-0/10 $26.00 © 2010 IEEE
DOI 10.1109/GCC.2010.86

of specific fields. Therefore, automatically extracting semantic information from mainstream database and reversely obtaining the ontology is of prime importance. And this is the key point the essay focused on.

IV. FORMAL DEFINITION OF RELATIONAL DATABASE AND ONTOLOGY

Relational model contains a limited relationship set *R*, a limited property set *A* and a group of primary keys and foreign keys .etc. Besides, there are also some other common functions[11] as follows:

• ***Function*** $dom(A_i)$, it sets the limits of the values attribute A_i can take, wherein $A_i \in A$.

• ***Function*** $Attr(R_i)$, it sets the attribute of relation R_i, wherein $R_i \in R$, $Attr(R_i) \subseteq A$.

• ***Function*** $PKey(R_i)$, it sets the primary key of relation R_i, wherein $R_i \in R$, $PKey(R_i) \subseteq Attr(R_i)$.

• ***Function*** $FKey(R_i)$, it sets the foreign key of relation R_i, wherein $R_i \in R$, $FKey(R_i) \subseteq Attr(R_i)$.

In addition to the above-mentioned entities, relational model also includes some restrictions, such as PRIMARY KEY,NOT NULL,UNIQUE and other dependence. All these factors are called Relation Schema which is used to explain the structure and relations inside. The tuple in the relation shows the value of Relation Schema and makes the content of the database.

Every relation is corresponding with one two-dimension table. We usually call it $R_i(A_1, A_2, ..., A_n)$ if we descript it in Relation Schema: R_i means relational table, A_j stands for properties of the line. These semantic meaning will play an important part in otology constructing.

In this essay, relations are divided into correlative relations and basic relations. Correlative relations do not have nonkey attribute and only refer to the relations between entilities.

Definition 1. Correlative Relation: Suppose $m = |PK(R_i)|$, $n = |A(R_i)|$, $q = |FK(R_i)|$, if $m = n = q \geq 2$, the relation Ri is called correlative relation。We can mark the set of correlative relations in Relation Schema as R_C.

Definition 2. Basic Relation: $\forall R_i \in R$, if $R_i \notin R_C$, that is to say R_i is not correlative relation，then R_i is Basic Relation。We can mark the set of basic relations as RB.

Ontology Structure is a quintuple [12] : $O = \{C, R, H_C, Rel, A_O\}$ .Here C & R are two totally different sets.Where the element of C is called Concept,and the element of R is called Relation; H_C denotes concept level,namely Taxonomy Relation; Rel shows non-Taxonomy Relation among concepts; A_O is ontology axiom.

Through the analysis of relational database, this essay explains the construction of rule base, and finally puts forward the regulations of relational models to ontology model.

Rule 1. If $R_i \in R_B$ R_i will be converted to a class of ontology,

Rule 2. If $R_i \in R_C$, R_i will be converted to $2C_n^2$ object properties，and $n = |A(R_i)|$。

Rule 3. As for the R_i relation property of $A_j \in A(R_i)$, if $A_j \notin F(R_i)$, then A_j will be converted to the property of data type .Its domain will be the class of relation R, and the property should be corresponding with OWL data type.

Rule 3. Handles the property line other than the foreign keys. These lines of property depict own characteristics of relations. As foreign keys emphasize on the depending relations between relations, we will discuss it in next chapter.

Rule 4. For $\forall R_i \in R_B$, if $|PK(R_i)| = 1$, it will be converted to the data type of functional with the restriction Cardinality=1。

Rule 5. For $\forall R_i \in R_B$, if $|PK(R_i)| > 1$, Cardinality=1 will be added to property of every primary keys..

Rule 6. For $\forall R_i \in R_C$, if $A_j \in PK(R_i)$, and $|PK(R_i)| = 1$, it will be converted into succession relations，and categories established by relation R_i will belongs to categories which are established by referenced relation R_i'。

Rule 7. In other conditions within Rule 6, foreign key properties of basic relation R_i will be totally converted to properties within ontology. The definition domain will be the class established by R_i, and value domain will be the class established by referenced relations.

Rule 8. If the property A_j of relation R_i is claimed to be U, the restriction will be changed into features of function property.

Rule 9. If the property of relation R_i is claimed to be N，the restriction will be changed to cardinality restriction of minCardinality=1.

V. ONTOLOGY AUTOMATICAL EXTRACTING ALGORITHM

We assume that unconverted relational database is obtained from concrete E-R model and follows the 3NF regulations, i.e. specific knowledge of the field has been stored in the chosen database mode and its data record. Every time we scan the relational database, we will get

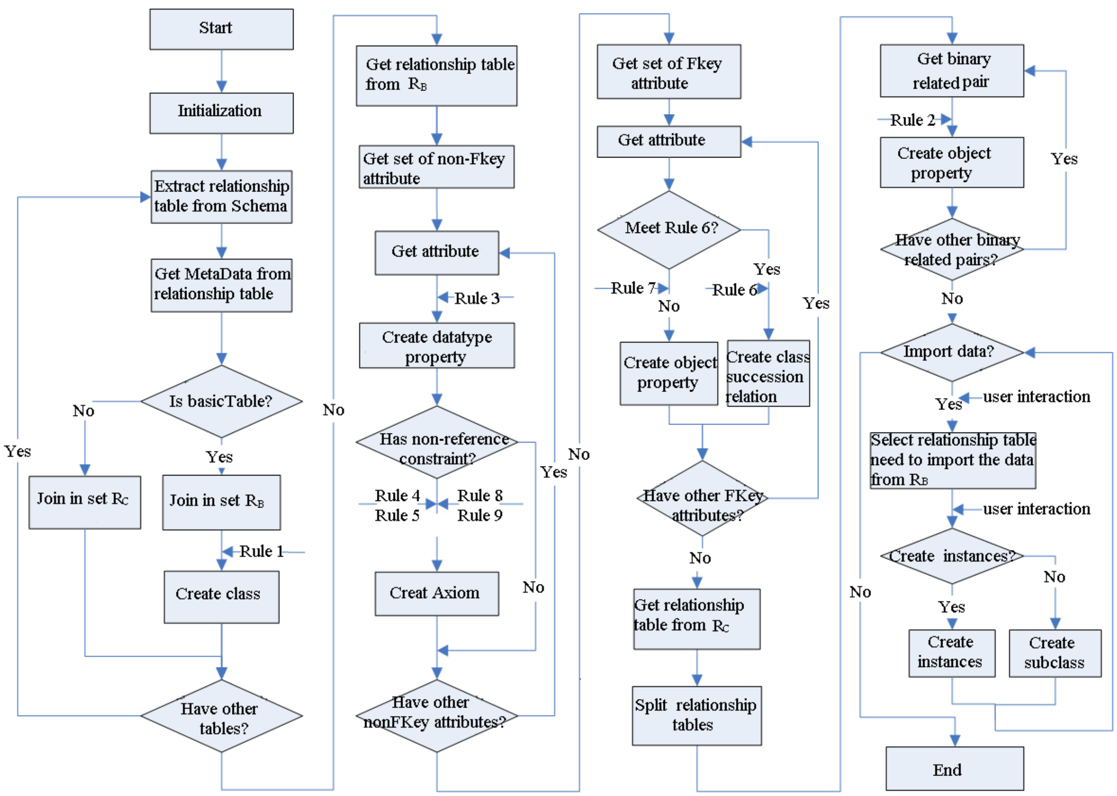
MetaData of every table and judge whether this two-dimension table shows the basic relations or correlative relations; and then we will go on convert it into factors of ontology according to preset rules. Algorithm design points are shown in Algorithm 1 and the process is shown as below.

Algorithm 1:OntoExtraction

Input:DBConnectInfo.

Output:file OWL.

1. // initialization
   Collection<TreeNodeInfo>nodeInfos=getTreeNodeInfosForTableNames(tableNames);
   ArrayList<TreeNodeInfo>basicTables=new ArrayList<TreeNodeInfo>();
   ArrayList<TreeNodeInfo>relationshipTables=new ArrayList<TreeNodeInfo>();
2. *Classification(nodeInfo,dbMetadata,strTableName)*;// Classification of the relationship
3. *for(TreeNodeInfo nodeInfo: basicTables){* // Operate on the basic relationship
4. *CreateClassForTables(dbMetadata,strTableName,nodeInfo)*;//Create class
5. *CreateDatatypePropertyForTable(dbMetadata,strTableName,nodeInfo)*;
6. *CreateObjectPropertyForTable(dbMetadata,strTableName,nodeInfo)*;
7. *CreateCardinalityForColumns(dbMetadata,strTableName,nodeInfo)*;
8. *CreateInstancesForTable(dbMetadata,strTableName,nodeInfo,TRUE)*;//Create instance
9. *CreateSubclassForTable(dbMetadata,strTableName,nodeInfo,TRUE)*;// Create subclass
   }
10. *for(TreeNodeInfo nodeInfo: relationshipTables){* // Operate on the Correlative Relation
11. *CreateInversePropertyForTable(dbMetadata,strTableName,nodeInfo)*;
    }



1. Figure 1. Flowchart of ontology automatical extracting

VI. EXPERIMENTS

This essay carry out experiments on Oracle 10g、SQL Server 2000 and MySQL Server 5.0 database, we can see some fragments of experimental database mode in Table Ⅰ.

TABLE I FRAGMENTS OF EXPERIMENTAL DATABASE MODE

No.	definitions of table
1	*Department(deptId int, deptName vchar, deptAddr vchar)*
2	*Student(stuId int, stuName vchar, sex vchar, deptId int)*
3	*Staff(staffId int, staffName vchar, email vchar)*
4	*GraduateStudent(stuId int, staffId int, researchArea vchar)*
5	*AcademicStaff(staffId int, researchArea vchar)*
6	*AdminStaff(staffId int, duty vchar)*
7	*Course(courseId int, courseName vchar, staffId int, deptId int)*
8	*ChooseCourse(courseId int, stuId int)*
9	*Scores(stuId int, courseId int, score float)*
10	*Project(projectId int, projectName vchar)*
11	*ProjectManagement(projectId int, deptId int, staffId int)*

The generated OWL can be opened in Protégé, and Figure 2 is the hierarchical structure of ontology. From that we can see the extracted ontology is in net structure which is more comprehensive in semantic meaning.

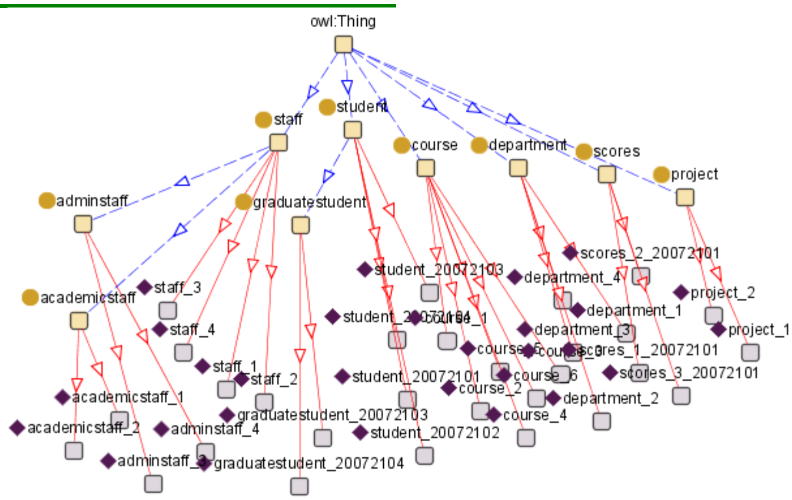


Figure 2. Hierarchical Structure of Ontology

VII . CONCLUSIONS

This essay brings up the formalized definitions of relational database and ontology, analyzes the semantic relations inside the relational database and put forward a series of converting regulations to automatically extract ontology from relational database. The converted ontology has enriched semantic meanings. Other than the simple class and property, the ontology can automatically build local ontology, keeping us free from manual construction and lay foundation for future steps on mapping between local ontologies.

REFERENCES

[1] Gruber TR. A translation approach to portable ontology specifications. Technical Report, KSL 92-71, Knowledge System Laboratory, 1993.

[2] Deng ZH, Tang SW, Zhang M, Yang DQ, Chen J. Overview of ontology. Acta Scientiarum Naturalium Universitatis Pekinensis, 2002, 38(5):730−738 (in Chinese with English abstract).

[3] DU XY, LI M, WANG SH. A Survey on Ontology Learning Research . Journal of Software, 2006,17(9): 1837-1847.

[4] Doan A, Madhavan J, Domingos P and Halevy A Y. Learning to Map between Ontologies on the Semantic Web[C]. In: Proceedings of the World Wide Web Conference, 2002.

[5] Stojanovic L, Stojanovic N, Volz R. Migrating data2intensive Web Sites into the Semantic Web. In: Proc. of the 17th ACM symposium on applied computing. 2002.

[6] Kashyap V. Design and creation of ontologies for environmental information ret- rieval. In: Proc. of the 12th Workshop on Knowledge Acquisition, Modeling and Management. 1999.

[7] Dogan G, Islamaj R. Importing Relational Databases into the Semantic Web. http://www.mindswap.org/webai/2002/fall/Importing-20Relational-20Databases-20into-20the-20Semantic-20Web.html (2002).

[8] Astrova, I., Korda, N., Kalja, A.Rule-Based Transformation of SQL Relational Databases to OWL Ontologies. In Proceedings of the 2nd International Conference on Metadata & Semantics Research, October 2007.

[9] DataMaster: http://protegewiki.stanford.edu/index.php/DataMaster.

[10] DataGenie: (2007) DataGenie. http://protege.cim3.net/cgi-bin/wiki.pl?DataGenie.

[11] ZHU JF, MA ZM, LU YH. Mapping OWL Ontologies to Relational Database Schema. Computer Science, 2008, 35(8):165-169.

[12] Maedche A. Ontology Learning for the Semantic Web. Boston: Kluwer Academic Publishers, 2002.