Query semantic data from relational database: an on-demand mapping approach

Haifei Zhang*

School of Mechanical and Electrical Engineering, Nantong Textile Vocational Technology College, CN-226007 Nantong, China

Received 1 March 2014, www.cmnt.lv

Abstract

One of the tasks for semantic web is to integrate large amounts of current information in relational database, which behind Web into machine-understandable RDF data model to form a "web of data". So relational database semantic query namely RDF access to relational database is an important issue in semantic web research. To realize the query is to build mapping relation between relational database schema and ontology. However, there is natural isomerism between them. The traditional method to eliminate the isomerism is to convert relational database schema into a similar ontology form and then to build all concepts and attributes mappings between conversion ontology and input ontology. This paper realized an on-demanding mapping method when users request query, avoided building all concepts and attributes mappings between conversion ontology and input ontology and improved mapping efficiency.

Keywords: relational database, semantic data, conversion ontology, input ontology, on-demanding mapping

1 Introduction

Relational database semantic query namely RDF [1] access to relational database is an important issue in semantic Web research. There are two access approaches: one is to build mapping between relational database schema and known ontology [2], the other is to issue or transform existing relation schema into RDF ontology [3, 4]. With the development of semantic Web, the number of ontology increases, thus we can make full use of existing known ontology to realize the RDF access to relational database that is the first access method is discussed in above. However, there is natural isomerism between relational database schema and ontology. The isomerism can be eliminated by converting relational database schema into similar ontology (we call the conversion ontology) and then building mapping between conversion ontology and input ontology (that is known ontology mentioned above). The basic graph pattern of input ontology SPARQL query reflects ontology classes and attributes that users need to query. This paper realized an on–demanding mapping method when users request query, avoided building all concepts and attributes mapping between conversion ontology and input ontology and improved mapping efficiency.

2 Relational database schema

Definition 1: A relational database schema can be described as $D=(N, attr, pk, fk, datatype)$, and is pentad, where name set $N=ET\cup RT\cup DT$ is a finite set composed of pairwise disjoint sets, including set ET of entity relation name, set RT of relationship relation name, and set DT of data type name, each data type name is data type name predefined by RDBMS. For $\forall T\in ET\cup RT$, the $Rel(T)$ denotes the T is a relational table.

In $\forall T\in ET\cup RT$, T has a non-empty set of attributes $attr(T)$ and each attribute $A\in attr(T)$ has a relevant predefined data type $datatype(A)\in DT$ as its range, $datatype(*)$ denotes the predefined data type of "$*$".

In $\forall T\in ET\cup RT$, T has only a single attribute or attribute set determining its tuple, called primary key $pk(T)$ of T. If $pk(T)\in attr(T)$, $pk(T)$ is called as single primary key, while if $pk(T)\subseteq attr(T)$, $pk(T)$ is composite primary key.

In $\forall T\in ET\cup RT$, if there is relation attribute quotes the primary key $pk(G)$ of other entity relation $G\in ET$ in T, then this attribute is called foreign key, satisfying $fk(T,G)\subseteq attr(T)$ and:

$$value(fk(T,G))\subseteq value(pk(G))\cup\{null\},$$

$value(*)$ denotes the range of "*", and "*null*" denotes null value. T may have $n\ (n\geq 0)$ foreign keys.

3 Ontology and semantic modelling

An ontology is an explicit specification of a conceptualization [5], which can be formalized as O(C, R, F, A, I), where C is the set of concepts/classes; R is the set of relations over elements of C; F is the set of functions; A is a set of axioms; and I is a set of instances.

* ***Corresponding author*** e-mail: zhanghaifei@nttec.edu.cn

Concepts can be used to describe anything, concrete or abstract. The ontology engineer analyzes relevant entities and organizes them into concepts. The backbone of ontology consists of a generalization/specialization hierarchy of concepts, i.e., taxonomy. Relations between concepts of a domain are a subset of the Cartesian product: C×C, which describe the relationships between two entities of the entities in C; Functions are special relations, which presents a functional property on pair entities. For example, mother(x, y) presents that x has a mother y, so, x determines y, like a function: y=mother(x); A is a set of axioms which are tautology assertions; and I is a set if instances of the concepts in C.

In the perspective of Semantic Web, the ontology can be used to model the knowledge about a specific domain, which requires a terminology to represent the concepts and the relations between these concepts. What's more, these concepts and the relations are organized as a hierarchy form. Ontology provides the Semantic Web a formal model of domain knowledge, on which we can use some logic to inference. In Semantic Web, the logical foundation of ontology is the Description Logic (DL) [6].

Based on the concept of ontology, W3C recommends a set of specifications to model the information on the Web of data. RDF (Resource Description Framework) [1] is a lightweight ontology language, which originally designed as a metadata data model. RDF is based upon the idea of making statements about Web resources in the form of triple (subject, predicate, object) expressions. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object. Therefore, RDF can be used to describe the entities and the relationships between these entities. However, RDF has limited weak semantics, and we need a rich-semantic ontology language. OWL (Web Ontology Language) [7] is another proposal which is recommended by W3C. OWL is designed for use by applications that need to process the content of information instead of just presenting information to humans. It can facilitate greater machine interpretability of Web content than that supported by RDF and RDFS by providing additional vocabulary along with a formal semantics. OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full.

Semantic data modelled with OWL described in RDF is the foundation of Semantic Web applications, how to get the RDF data is an emergency question for many environments. So, lots of researchers have focused on the fields of mapping the existing traditional data to the RDF data [8]. W3C also recommends a language to map the relational data to RDF data, the R2RML (RDB to RDF Mapping Language) [9]. It can be used to express customized mappings from relational databases to RDF datasets. With R2RML, we can view existing relational data in the RDF data model. The target RDF data is modelled in an existing ontology and different distributed relational data can be mapped to the unified semantic data which implements a data integration application.

4 Relational database semantic query

Traditional database semantic query is as follows: First generating conversion ontology from relational database schema by semantic wrapper, then building mapping relation from input ontology to conversion ontology. So SPARQL query on input ontology can be transformed into conversion ontology according to the mapping relation. Lastly submitting conversion ontology to semantic wrapper for transforming into SQL query of relational database. The construction of semantic wrapper refers to the author's previous work [10]. And the key function is to generate conversion ontology and rewrite from conversion query to SQL query and SQL query results transform into RDF data. Relational database semantic query is shown as Figure 1. This paper mainly describes mapping on query, so query results are not discussed here and there is no hint of query results in the graph. All these are realized in the author's previous work [10].

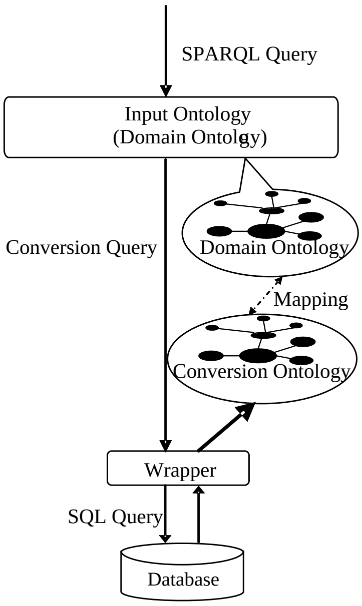

FIGURE 1 The structural diagram of relational database semantic query system

5 SPARQL basic graph pattern

SPARQL is the query language of semantic Web and is matched on graph patterns. Complex graph patterns can be obtained by combining simple patterns. Combining methods include: 1. Basic graph pattern; 2. Group graph pattern; 3. Optional graph pattern; 4. Alternative graph pattern.

Definition 2: A SPARQL query graph pattern GP is defined as the following expression:

$$GP \rightarrow TP \mid GP\ AND\ GP \mid GP\ OPTIONAL\ GP \mid GP\ UNION\ GP \mid GP\ FILTER\ \exp$$, where

$GP\ AND\ GP$ is namely group graph pattern, which could be the combination of any other graph patterns.

TP namely basic triple graph pattern, which could be the set consists of many RDF triples.

GP OPTIONAL GP stands for alternative graph pattern, the latter *GP* of which is the range of optional graph pattern.

GP UNION GP means that query graph pattern contains alternative graph pattern.

GP FILTER exp stands for value constraint, in which exp is the expression of value constraint.

Basic graph patterns (abbreviated to BGP) are sets of triple patterns. SPARQL graph pattern matching is defined in terms of combining the results from matching basic graph patterns. A sequence of triple patterns interrupted by a filter comprises a single basic graph pattern. Any graph pattern terminates a basic graph pattern [11].

Figure 2 shows the SPARQL query statement pattern, in which ?x, ?y and ?z respectively stand for instances of Class A, Class B and Class C. Class A has value property A1 and relationship property has B; Class B has value property B1 and B2 and relationship property has A and has C. Among them, A1 is value property variable with value constraint Exp(?A1); B2 is value property variable under OPTIONAL pattern; B1 is value property variable and takes the Value(B1); C1 is value property variable. From this SPARQL query statement pattern, it is known that this query covers basic graph pattern, optional graph pattern, value constraint and so on; for alternative graph pattern (UNION), its query rewriting result can be regarded as union of two SQL statements; SPARQL query statement of UNION graph pattern is not given.

```
PREFIX ex:<http://example.org/schemas/university#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?A1 ?B2 ?C1
WHERE{
    ?x rdf:type ex:ClassA.
    ?x ex:A1 ?A1.
    ?x ex:hasB ?y.              Basic Graph
    ?y rdf:type ex:ClassB.      Pattern
    ?y ex:hasA ?x.
    ?y ex:B1 Value(B1).
    OPTIONAL{                   Optional Graph      Graph Pattern
    ?y ex:B2 ?B2. }             Pattern
    ?y ex :hasC ?z.             Basic Graph
    ?z rdf:type ex :ClassC.     Pattern
    ?z ex :C1 ?C1.
    FILTER(?A1=constant)        Value Constraint
}
```

FIGURE 2 SPARQL query statement pattern

6 On-demanding mapping on query request

In addition to the above described SPARQL basic graph patterns that are composed of triples, ontology is also composed of triples. Inputting the basic graph patterns of ontology SPARQL query in relational database semantic query reflects ontology classes and attributes that users need to query. This chapter presents how to realize the on-demanding mapping method when users request query by examples to illustrate how to build the mapping between the basic graph patterns in SPARQL query and conversion ontology.

Assuming conversion ontology generated from relational database schema is shown as Figure 3.

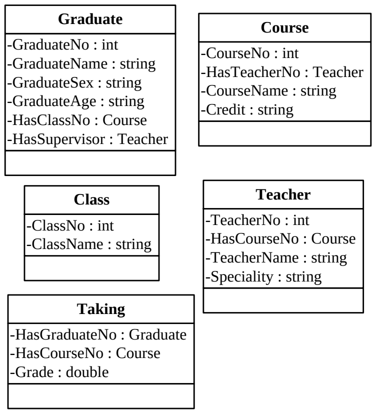

FIGURE 3 Graduate ontology

Figure 5 shows SPARQL query that users input into input ontology shown as Figure 4.

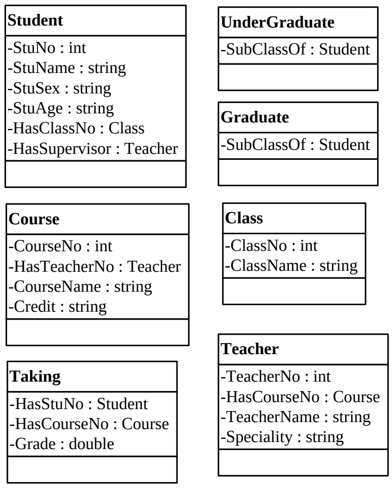

FIGURE 4 Student ontology

```
PREFIX ex:<http://example.org/schemas/university#>
PREFIX  rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
  SELECT ?cid ?cname ?tid ?specia
    WHERE {?x rdf:type ex:Course.
            ?x ex:CourseNo ?cid.
            ?x ex: CourseName ?cname.
            ?x ex: Credit ?Credit.
            ?x ex: hasTeacherNo ?y.
            ?y rdf:type  ex:Teacher.
            ?y ex:TeacherNo ?tid.
            ?y ex:hasCourseNo ?x.
            OPTIONAL{?y ex:Speciality ?specia.}
            FILTER(?credit=3)}
```

FIGURE 5 SPARQL query

The mapping relation between triples of basic and optional graph patterns and conversion ontology Graduate Ontology (as shown in Figure 3) is shown as Figure 6.

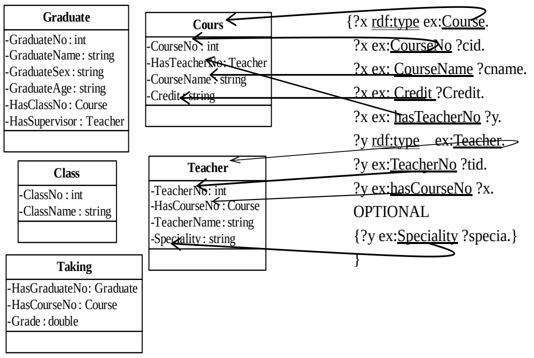

FIGURE 6 Mappings from BGP of SPARQL to conversion ontology

ALGORITHM 1 (the mapping algorithm): Triples to conversion ontology mapping algorithm TP2OMap

Algorithm TP2OMap (TP_{BGP}, TP_{OPT})

Input: TP_{BGP}, TP_{OPT}
Output: $Match(C_{ti}, C_{oi})$, $Match(P_{ti}, P_{oi})$
Steps:
1. FOR each triple T_i in TP_{BGP} and TP_{OPT}
2. String $S_i \leftarrow$obtain the subject in T_i;
3. String $P_i \leftarrow$obtain the predicate in T_i;
4. String $O_i \leftarrow$obtain the object in T_i;
5. IF "type" EQUALS P_i THEN
6. BEGIN
7. Obtain corresponding(matching) concept C_{oi} of O_i from conversion ontology;
8. $C_{ti} \leftarrow O_i$;
9. Return $Match(C_{ti}, C_{oi})$
10. END
11. ELSE
12. BEGIN
13. Obtain corresponding(matching) concept P_{oi} of O_i from conversion ontology;
14. $P_{ti} \leftarrow P_i$;
15. Return $Match(P_{ti}, P_{oi})$
16. END
17. END FOR

7 Conclusion

This paper describes an on-demanding method when users request query by building mapping between basic graph patterns in SPARQL query and conversion ontology. To compared with the method described in classic literature [2] of the first access approach in section 1, this method avoid constructing all concepts and properties mapping between input ontology and conversion ontology, thus the efficiency is obviously.

The match between basic graph pattern and conversion ontology is similar to directly regarding input ontology SPARQL query as query on conversion ontology. But this query term (class and attribute) is not an exact match and needs to establish matching relation by using lots of similarity calculation algorithm in existing mapping of basic graph patterns and terms on conversion ontology. The on-demanding mapping method using basic graph patterns of SPARQL query statement in relational database semantic query avoided building all concepts and attributes mapping between conversion ontology and input ontology and undoubtedly improved query efficiency.

Acknowledgment

This work is sponsored by the following grants: (i) a grant No. 61163057 from the National Natural Science Foundation of China, (ii) a grant No. BK2010520 from the Natural Science Foundation of Jiangsu Province of China, (iii) Excellent Youth Scholar of “Qing Lan Project” of Jiangsu Province in China (2012), and (iv) two grants No. FYKY/2012/2 and No. FYKY/2013/8 from Science and Technology Program of Nantong Textile Vocational Technology College.

References

[1] Klyne G, Carroll J eds 2004 Resource Description Framework (RDF): Concepts and Abstract Syntax *W3C Recommendation* 10 Feb 2004 http://www.w3.org/TR/rdf-concepts/
[2] Hu W, Qu Y 2007 Discovering simple mappings between relational database schemas and ontologies *Proceedings of ISWC/ASWC2007* Busan, South Korea **4825** 225-38
[3] Xu Z, Dong Y, Lu Y 2006 Semantic-preserving translation from ER schema to OWL DL ontology *Chinese Journal of Computers* **29**(10) 1786-96 *(in Chinese)*
[4] Bizer C, Cyganiak R 2006 D2R Server-Publishing Relational Databases on the Semantic Web *Poster at the 5th International Semantic Web Conference (ISWC2006)* Athens GA USA
[5] Perez A G, Benjamins V R 1999 Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods *Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods* Stockholm Sweden 1-15
[6] Baader D, Calvanese D, McGuinness D L, Nardi D, Patel-Schneider P F 2003 The Description Logic Handbook: Theory, Implementation, and Applications *Cambridge Cambridge University Press* 1-3
[7] McGuinness D L, van Harmelen F 2004 OWL Web Ontology Language Overview W3C Recommendation 10 February 2004 http://www.w3.org/TR/owl-features/
[8] Sequeda J F, Tirmizi S H, Corcho O, Miranker D P 2011 Survey of Directly Mapping SQL Databases to the Semantic Web *The Knowledge Engineering Review* **26** 445-86
[9] Das S, Sundara S, Cyganiak R 2012 R2RML: RDB to RDF Mapping Language *W3C Recommendation* 27 September 2012 http://www.w3.org/TR/r2rml/
[10] Zhang H, Lin L, Qiang B 2012 A Semantic Wrapper of Deep Web Data Source Based on RDF View *Journal of Computational Information Systems* **8**(9) 3527-39
[11] Prud'hommeaux E, Seaborne A 2008 SPARQL Query Language for *RDF W3C Recommendation* 15 January 2008 http://www.w3.org/TR/rdf-sparql-query/

Author

Haifei Zhang, born in November, 1980, Nantong, China

Current position, grades: lecturer of Mechanical and Electrical Engineering School of Nantong Textile Vocational Technology College. PhD candidate at Hohai University.
University studies: MS degree in Software Engineering from Soochow University in 2005.
Scientific interests: databases, GIS, ontological engineering, semantic web and web intelligence.
Publications: 1 patent, 2 software copyrights, 18 papers.