

# A Query Driven Method of Mapping from Global Ontology to Local Ontology in Ontology-based Data Integration

Haifei Zhang

Mechanical and Electrical Engineering School of Nantong Textile Vocational Technology College, Nantong, China  
Email: zhanghaifei@nttec.edu.cn

**Abstract**—At present, the mediator/wrapper integration methods are widely used in ontology based data integration because they solve the data update problems of data warehouse method. The key of this method is building of mapping from the global ontology in mediator to the local ontology in wrapper. This article analyzes the general mapping methods and designs a SPARQL query driven Global Local as View (GLAV) mapping method, it is proved in the experiment that this method can effectively support query decomposition and makes the ontology-based integration system has very good extensibility.

**Index Terms**—Ontology based Data Integration, Global Ontology, Local Ontology, GLAV mapping, SPARQL driven

## I. INTRODUCTION

As information technology develops, enterprises in order to improve the economic benefits and market competitiveness, control and integratedly manage various information in production operation process by making use of computer, network, and database technologies in certain depth and breadth on the basis of business flow optimization and reconfiguration, so as to realize information sharing inside and outside the enterprise as well as use the information effectively.

However, such information serving different applications is stored in many separate data sources. In order to effectively utilize the information, a public and integrated environment shall be built for integrating data from many sources that are separate, heterogeneous, and autonomous, so as to provide the user with a unified and transparent data access interface.

Ontology is a formal and explicit specification of a shared conceptualization[1]. In Web environment, ontology provides a consensus for the given domain, which can eliminate the difference of terms between various data sources and realize semantic interoperability. Therefore, ontology-based data integration has become into a hot topic for research in recent years.

At present, the mediator/wrapper integration methods[2] are widely used in ontology-based data integration because they solve the data update problems of data warehouse method[3]. This method first describes

the schemas of heterogeneous data sources into an ontology specification language (such as RDF(S)[4][5], OWL[6]) which according to the mediator schema use the ontology wrapper, then it builds the semantic mapping from the local ontology in the ontology wrapper to the global ontology in the mediator.

Mainly there are two methods for building semantic mapping from the local ontology in the ontology wrapper to the global ontology in the mediator: One is Global-as-View (GAV)[7][8], the other is Local-as-View (LAV)[8][9]. The currently emerging Global-Local-as-view (GLAV)[10] mapping method is a combination of GAV and LAV, formed by combination of the views in global schema and those in the related data sources. The method of GLAV can comprehensively make use of the advantages of GAV and LAV, capable of providing more expressive semantic mapping to the data integration system.

This article analyzes the above mapping methods as well as designs a query driven GLAV mapping method, the query is described as an ontology query language which named SPARQL [11]. It is proved in the experiment that this method can effectively support query decomposition and makes the ontology-based integration system has very good extensibility.

## II. ONTOLOGY-BASED DATA INTEGRATION

Ontology-based data integration mainly adopts Mediator/Wrapper integration method. Local data source schema is wrapped into a local ontology by wrapper, the local data is still stored in the original database (see the work of the author in the previous phases [12]). The mediating schema in mediator is global ontology formed by domain ontology. Mappings will be built between global ontology and local ontology; the system architecture diagram is as shown in Figure 1.

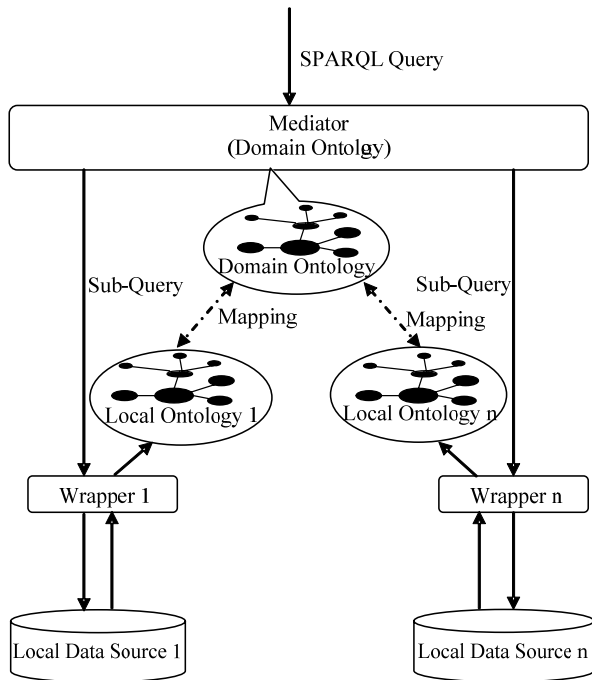


Figure 2. The architecture of Ontology-Based Data Integration

III. SEMANTIC MAPPING

There are three main methods for mapping of global ontology and local ontology, which are Global-as-View (GAV), Local-as-View (LAV), and Global-Local-as-View (GLAV). The followings are the definitions and formal descriptions of the three basic mapping methods with examples. The global ontology is as shown in Figure 2; the two local ontologies are shown in Figure 3 and Figure 4.

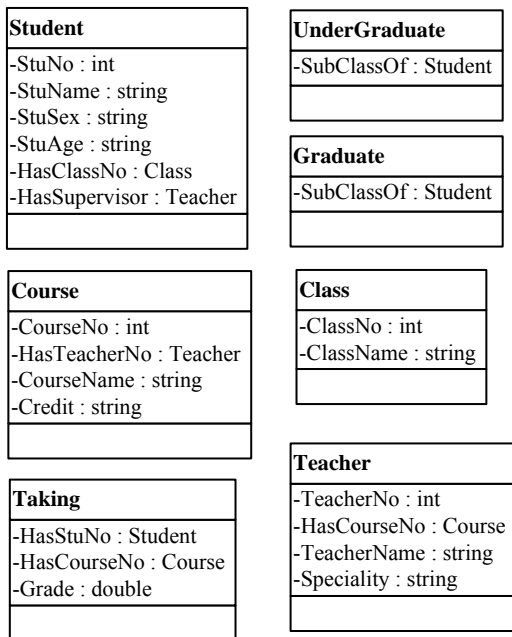


Figure 3. Global Ontology 1

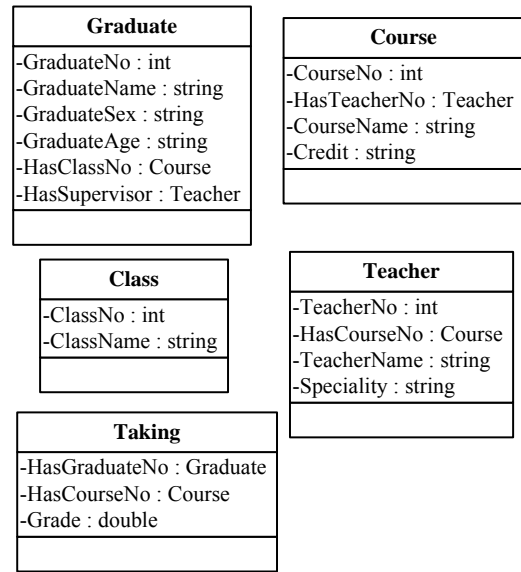


Figure 4. Local Ontology

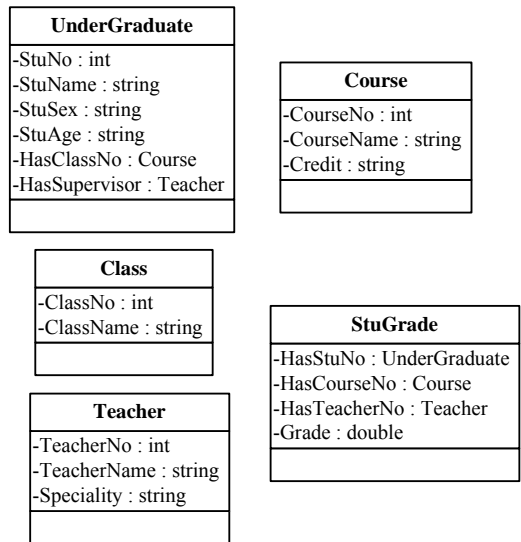


Figure 1. Local Ontology 2

A.. GAV Mapping

In GAV, term  $T_G$  in the global ontology is regarded as a view of the local ontology, such as query  $Q_L$  of the local ontology. In this case, GAV interrelates each term in the global ontology with the query of the local ontology. The form of the series of assertions in GAV mapping is as follows:  $T_G \leftarrow Q_L$ .

The series of assertions in GAV mapping of the above example:

$$GO.Student \supseteq Q(LO1.Graduate) \tag{G1}$$

$$GO.Course \supseteq Q(LO1.Course) \tag{G2}$$

$GO.Class \supseteq Q(LO1.Class)$  G3

$GO.Teacher \supseteq Q(LO1.Teacher)$  G4

$GO.Taking \supseteq Q(LO1.Taking)$  G5

$GO.Graduate = Q(LO1.Graduate)$  G6

$GO.Student \supseteq Q(LO2.UnderGraduate)$  G7

$GO.Class \supseteq Q(LO2.Class)$  G8

$GO.Course \supseteq Q_{CourseNo,TeacherNo,CourseName,Credit}(LO2$   
 $.Course \infty_{CourseNo} LO2.StuGrade)$

G9

The SPARQL query statement of assertion G9 is briefly as follows:

```
Select ?CourseNo ?TeacherNo ?CourseName ?Credit
Where {
  ?x rdf:type Course.
  ?x rdf:CourseNo ?CourseNo.
  ?x rdf:CourseName ?CourseName.
  ?x rdf:Credit ?Credit.
  ?y rdf:type StuGrade.
  ?y rdf:HasCourseNo ?x.
  ?y rdf:HasTeacherNo ?TeacherNo.}
```

$GO.Teacher \supseteq Q_{TeacherNo,CourseNo,TeacherName,Speciality}(LO2$

G10

$.Teacher \infty_{TeacherNo} LO2.StuGrade)$

The SPARQL query statement of assertion G10 is briefly as follows:

```
Select ?TeacherNo ?CourseNo ?TeacherName ?Speciality
Where {
  ?x rdf:type Teacher.
  ?x rdf:TeacherNo ?TeacherNo.
  ?x rdf:TeacherName ?TeacherName.
  ?x rdf:Speciality ?Speciality.
  ?y rdf:type StuGrade.
  ?y rdf:HasTeacherNo ?x.
  ?y rdf:HasCourseNo ?CourseNo.}
```

$GO.UnderGraduate = Q(LO2.UnderGraduate)$  G11

$GO.Taking \supseteq Q(LO2.StuGrade)$  G12

“ $\supseteq$ ” stands for complete view, when a concept is complete, its extension provides any superset of the tuples satisfying the corresponding view. In other words, from the fact that a tuple is in this concept cannot conclude that such a tuple satisfies the corresponding view.

“=” stands for exact view, when a concept is exact, its extension is exactly the set of tuples of objects satisfying the corresponding view.

*B. LAV Mapping*

In LAV, term  $T_L$  in the local ontology is regarded as a view of the global ontology, such as query  $Q_G$  of the global ontology. In this case, LAV interrelates each term in the local ontology with the query of the global ontology. The form of the series of assertions in LAV mapping is as follows:  $T_L \leftarrow Q_G$ .

The series of assertions in LAV mapping of the above example:

$LO1.Graduate = Q(GO.Graduate)$  L1

$LO1.Course \subseteq Q(GO.Course)$  L2

$LO1.Class \subseteq Q(GO.Class)$  L3

$LO1.Teacher \subseteq Q(GO.Teacher)$  L4

$LO1.Taking \subseteq Q(GO.Taking)$  L5

$LO2.UnderGraduate = Q(GO.UnderGraduate)$  L6

$LO2.Class \subseteq Q(GO.Class)$  L7

$LO2.Course \subseteq Q_{CourseNo,CourseName,Credit}(GO.Course)$  L8

$LO2.Teacher \subseteq Q_{TeacherNo,TeacherName,Speciality}(GO.Teacher)$  L9

$LO2.StuGrade \subseteq Q_{StuNo,CourseNo,TeacherNo,Grade}(GO.Taking \infty_{CourseNo} GO.Teacher)$  L10

The SPARQL query statement of assertion L10 is briefly as follows:

```
Select ?StuNo ?CourseNo1 ?TeacherNo ?Grade
Where {
  ?x rdf:type Teacher.
  ?x rdf:TeacherNo ?TeacherNo.
  ?x rdf:HasCourseNo ?CourseNo1.
  ?y rdf:type Taking.
  ?y rdf:HasCourseNo ?CourseNo2.
  Filter (?CourseNo1=?CourseNo2)}
```

*C. GLAV Mapping*

GLAV is the combination of GAV and LAV as well as the extended form of LAV, with it, the view (query)  $Q_L$

of the local ontology can be the view (query)  $Q_G$  of the global ontology in the mapping rules. In this case, GLAV can export data with the view of the local ontology, which excels the expression of LAV, besides, it also allows the terms in the global ontology to connect with another, which excels the expression of GAV. The form of the series of assertions in GLAV mapping is as follows:

$$Q_L \leftarrow Q_G$$

The series of assertions in GLAV mapping include all the assertions of LAV mapping as well as the connection of the mapping definitions from the same data source.

The series of assertions in GLAV mapping of the above examples include all the assertions of LAV mapping as well as the following ones:

$$Q_{GraduateNo, GraduateName, \\ ClassNo, CourseName, \\ TeacherName, Grade, Credit}$$

$$(LO1.Class \in LO1.Graduate \in \\ \text{ClassNo} \quad \text{GraduateNo}$$

$$LO1.Taking \in LO1.Course \in LO1.Teacher) \quad GL1$$

$$= Q_{StuNo, StuName, \\ ClassNo, CourseName, \\ TeacherName, Grade, Credit}$$

$$(GO.Class \in GO.Graduate \in \\ \text{ClassNo} \quad \text{StuNo}$$

$$GO.Taking \in GO.Course \in GO.Teacher) \quad GL2$$

$$Q_{StuNo, StuName, \\ ClassNo, CourseName, \\ TeacherName, Grade, Credit}$$

$$(LO2.Class \in LO2.UnderGraduate \in \\ \text{ClassNo} \quad \text{GraduateNo}$$

$$LO2.StuGrade \in LO2.Course \in LO2.Teacher) \quad GL2$$

$$= Q_{StuNo, StuName, \\ ClassNo, CourseName, \\ TeacherName, Grade, Credit}$$

$$(GO.Class \in GO.Graduate \in \\ \text{ClassNo} \quad \text{StuNo}$$

$$GO.Taking \in GO.Course \in GO.Teacher)$$

For querying the information of the courses for graduate students such as GraduateNo, GraduateName, CassName, CourseName, TeacherName, Grade, Credit, we just need the mapping of GL1. While for querying the information of the courses for undergraduate students such as StuNo, StuName, ClassName, CourseName, TeacherName, Grade, Credit, we just need the mapping of GL2, in which any of the sub-queries such as the query of student number, student name, and class name of the graduate student or the undergraduate student can adopt GL1 or GL2 mapping.

GLAV can provide an integration method that is more flexible, the building of the global schema and wrapping of the information source are separate. It is a combination of the views (queries) of the global ontology and the local ontology of each data source. The method of GLAV can comprehensively make use of the advantages of GAV

and LAV, capable of providing more expressive semantic mapping to the data integration system.

#### IV. CASE STUDY

Query  $Q_G$  of SPARQL of the global ontology in the integration system of the above examples is as follows:

```
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)}
```

According to the assertions in GLAV mapping, the two sub-queries as  $Q_{L1}$  and  $Q_{L2}$  of the two local ontologies univ1 and univ2 can be decomposed.

```
QL1:
PREFIX ex1:<http://example.org/schemas/univ1#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
SELECT ?cid ?cname ?tid ?specia
WHERE { ?x rdf:type ex1:Course.
?x ex1:CourseNo ?cid.
?x ex1: CourseName ?cname.
?x ex1: Credit ?Credit.
?x ex1: HasTeacherNo ?y.
?y rdf:type ex1:Teacher.
?y ex1:TeacherNo ?tid.
?y ex1:HasCourseNo ?x.
OPTIONAL{?y ex1:Speciality ?specia.}
FILTER(?credit=3)}
```

```
QL2:
PREFIX ex2:<http://example.org/schemas/univ2#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
SELECT ?cid ?cname ?tid ?specia
WHERE { ?z rdf:type ex2 :StuGrade.
?z ex2:HasCourseNo ?x.
?z ex2:HasTeacherNo ?y.
?x rdf:type ex2:Course.
?x ex2:CourseNo ?cid.
?x ex2: CourseName ?cname.
?x ex2: Credit ?Credit.
?y rdf:type ex2:Teacher.
?y ex2:TeacherNo ?tid.
OPTIONAL{?y ex2:Speciality ?specia.}
FILTER(?credit=3)}
```

$Q_{L1}$  and  $Q_{L2}$  query results output are as shwon in Figure 5 and Figure 6.

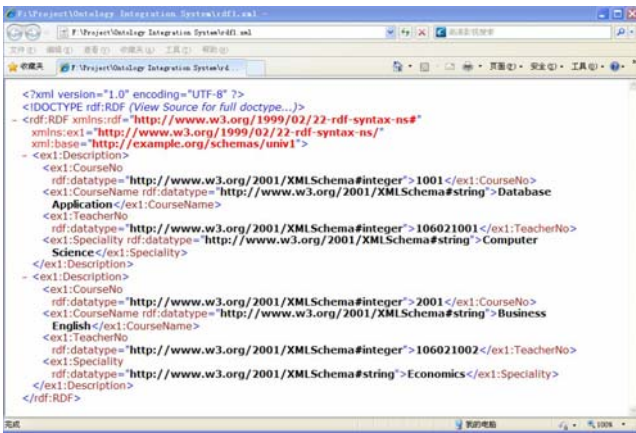


Figure 6. QL1 query result output

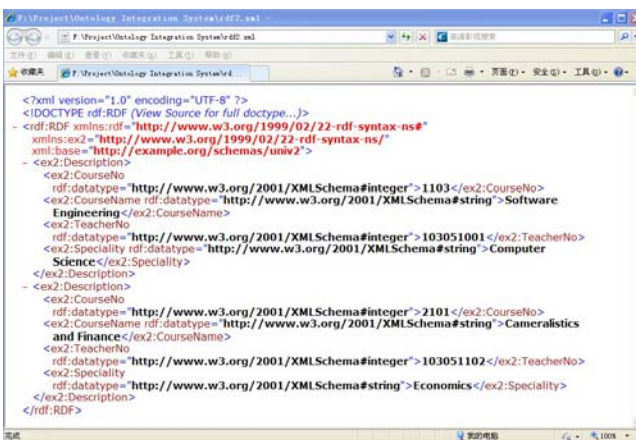


Figure 5. QL2 query result output

### V. CONCLUSION

This article analyzes mapping methods as GAV, LAV, and GLAV in the ontology integration framework based on Mediator/Wrapper. It's easy to see through the example in the article that the GLAV mapping method is more excellent in data adaptability and query capability mature than GAV and LAV mapping methods. The SPARQL query driven GLAV mapping method designed in this article effectively supports the decomposition from the global query on global ontology to the sub-queries on local ontologies and makes the Ontology-based integration system has very good extensibility.

In addition to the efficiency and capability of the mapping method itself, ontology-based data integration system also has some difficulties to be conquered completely, such as semantic heterogeneity. Therefore, elimination of semantic confliction and aggregation of the query results will be the key study direction of the author in the future.

### ACKNOWLEDGMENT

This work is sponsored by the following grants: (i) a grant No. FYKY/2012/2 from Nantong Textile Vocational Technology College Science and Technology Program, (ii) Qing Lan Project of Jiangsu Province, (iii) a

grant No. 61163057 from the National Natural Science Foundation of China and (iv) a grant No. BK2010520 from the Natural Science Foundation of Jiangsu Province of China.

### REFERENCES

- [1] Studer R. , Benjamins V. R. , Fensel D. . Knowledge engineering: principles and methods. Data and Knowledge Engineering, 25(1-2):161-197, 1998.
- [2] Gio Wiederhold. Mediators in the Architecture of Future Information Systems. IEEE Computer Society, 1992, 25 (3), pages 38-49, 1992.
- [3] Zhang Xiaoxiang et al. Encyclopedia of Computer Science Technology. Beijing: Tsinghua University Press, 2005. (in Chinese)
- [4] G. Klyne, J. Carroll, eds. Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/rdf-concepts/>
- [5] Dan Brickley, R.V. Guha, Brian McBride. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/rdf-schema/>
- [6] Deborah L. McGuinness, Frank van Harmelen. OWL Web Ontology Language Overview. W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/owl-features/>
- [7] A.Y. Halevy. Answering queries using view s: a survey. The VLDB Journal the International Journal on Very Large Data Bases, 10 (4): 270-294, 2001.
- [8] Maurizio Lenzerini. Data integration: A theoretical perspective. In Proceedings of 21st ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems. New York: ACM Press, pages 233-246, 2002.
- [9] Jeffrey D. Ullman. Information Integration Using Logical Views. 6th Int. Conf. on Database Theory ( ICDT'97 ) , 1186, 19-40, 1997.
- [10] Marc Friedman, Alon Levy, Todd Millstein. Navigational Plans for Data Integration. Proceedings of the 16th National Conference on Artificial Intelligence (AAAI' 99), 67-73, 1999.
- [11] Eric Prud'hommeaux, Andy Seaborne. SPARQL Query Language for RDF. W3C Recommendation, 15 January 2008. <http://www.w3.org/TR/rdf-sparql-query/>
- [12] Haifei Zhang, Lili Lin, Baohua Qiang. "A Semantic Wrapper of Deep Web Data Source Based on RDF View". Journal of Computational Information Systems, vol.8, no.9, pp. 3527-3539, 2012.



**Haifei Zhang** was born in Rudong Country, Nantong City, Jiangsu Province in 1980. He is a Ph.D. candidate of Hohai University. He received the MS degree in Software Engineering from Soochow University in 2005. He is a lecturer at Nantong Textile Vocational Technology College. His research interests include databases, GIS, ontological engineering, Semantic Web and Web intelligence.