International Journal of Multimedia and Ubiquitous Engineering
Vol. 10, No. 3 (2015), pp. 389-396
http://dx.doi.org/10.14257/ijmue.2015.10.3.35

Application of the Semantic Annotation Technology Based on Ontology

Haiming Jing

School of Informaion Science and Technology
Shijiazhuang Tiedao University
Shijiazhuang, 050043, China
jimperlov@yeah.net

Abstract

To resolve the semantic conflicts in the process of data exchange, the ontology is made full use to describe the characteristics of the relation between concept, adding semantics to XML Schema. In this paper, the semantic conflict resolution model is introduced, the indirect method is applied and the ER model is applied to realize the semantic mapping from relational schema to XML Schema, this is the so-called semantic annotation technology. In the field of railway, there exist some semantic conflict in the data exchange, by the Ontology OWL fragment extracted from the experiment, semantic conflict will be reduced and even eliminated.

Keywords: *semantic conflicts, data exchange, ontology, semantic annotation*

1. Introduction

For the heterogeneous system, if the semantic information integration must be done, some systems map database schema to Ontology[1], then rewrite the semantic query to the SQL statement; others are adppted stratified ontology and vocabulary to achieve semantic integration. But these methods can only solve the local semantic conflict, can not meet the needs of practical application. Yu[2] put forward a heterogeneous data integration method based on hybrid ontology, by building the mapping relationship between global ontology and local ontology, data sources, so as to solve the problem of semantic heterogeneity in heterogeneous data integration.

Semantic conflict is the conflict caused by using different ways in heterogeneous systems to express the same entity in reality[3].Semantic conflict is mainly caused by heterogeneous data sources in different table name, column name and data type [4]. For example, Use the ‘author’ to represent the author of a book in one data source, in another system, use ‘writer’ to represent homonymy heterogeneous caused by writer; Another example, The data type conflict in one data source ‘datetime’ to represent the date time with in another data source ‘date’ to represent date time. Ontology provides shared concept system for system design in specific domain application.

The Research on the data integration mainly focused in XML and ontology to solve the heterogeneous data conflict, because the data format of XML is cross platform and good scalability, the XML data integration system can well solve the operating system heterogeneity and data source heterogeneity and some other syntax heterogeneity. But the XML data format can not handle the semantic heterogeneity. And the ontology can shield the inconsistence of the local semantics of data source, provide a unified global concept for the users, in order to solve the semantic conflict in integration [5-6].

ISSN: 1975-0080 IJMUE

2. Related Work

Semantic annotation is the annotation of the original data, it make original data have semantic information. not only understood by people, but the machine. In the semantic web it is generally annotated with XML annotation language for data, taking RDF/XML as data description model, combined with Ontology, the annotated data has a clear meaning, so that the machine can understand[7-8].

Current research about the semantic annotation roughly are the same, all is in the background of the semantic web, using ontologies to study various semantic annotation, and the results of the image annotation and video annotation fields are more abundant.

A semantic conflict resolution scheme is proposed, it adopt semantic mapping from relational schema to XML schema and the introduction of ontology mapping file to solve the problems of semantic annotation. Each XML Schema document of heterogeneous data source can be annotated for the corresponding elements through mapping and with ontology. In this way, the XML Schema documents with semantic annotation will make mapping relationship, heterogeneous data sources also make a mapping relationship with semantic information.

The process of semantic annotation of information resource is a difficult step. On the one hand, the semantic annotation may extract concept from information resource and describe with ontology base, and the information resource is an instance of the ontology base. On the other hand, the semantic annotation is the step which can enrich the ontology base, when the process occurred to find a new concept to the ontology base, the process can learn the new concept by the knowledge base and describe the concept with the current concept in ontology, and then, add the concept into the ontology base, so the semantic annotation is an enriched process of the ontology base. The sketch map of semantic annotation is showed as follow.

2.1 Ontology

Ontology is a new specification business data in recent years, which can accurately describe the data semantics, and infer implicit data semantic relations[7]. Usually, ontology is a formal description of the relationship between tuple contains concept, individual and concept, it can be used to express the attribute information, the quantity standard, inconsistent statements and logical relations between different objects [8]

With the popularity of ontology, the technologies related to ontology have emerged rapidly and gradually formed system. For example, after a long-term development and evolution, the ontology specification finally fixed two norm, the resource description framework (RDF) and Web ontology language (OWL). RDF defines three tuple form subject attribute and object knowledge representation, OWL introduces owl:Thing, RDF owl: Class, owl:Property and other words to make knowledge description more convenient [9].

OWL is also compatible with the web architecture. It has, amongst others, an XML-based encoding and it is backward compatible with RDF Schema, which can be seen in Figure 1. As ontologies are also tailored towards the distributed nature of the Web, OWL additionally provides constructs for (de-)composition, extension,adaptation, sharing and reuse.

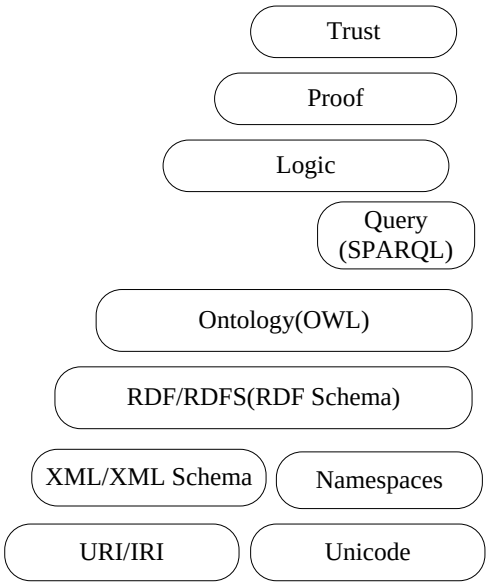

Figure1. Layered View on Semantic Web Enabling Technologies

2.2 Model Design

In order to further enhance the semantic expression ability of XML Schema, Ontology is constructed to realize knowledge sharing and reuse firstly. XML Schema files of each heterogeneous data source can be add semantic tag by mapping with Ontology. Therefore, the mapping relationship will occur between the XML Schema file with semantic annotation, Mapping relationship with semantic information will be produced among the heterogeneous data sources.

Semantic conflict resolution model is shown in Figure 2. When using the indirect mode to realize the conversion from relational model to XML mode, by the ER diagram of relational model, ER diagram can map to XML Schema diagram, by XML ER map, XML Schema diagram will map to XML Schema through semantic mapping.

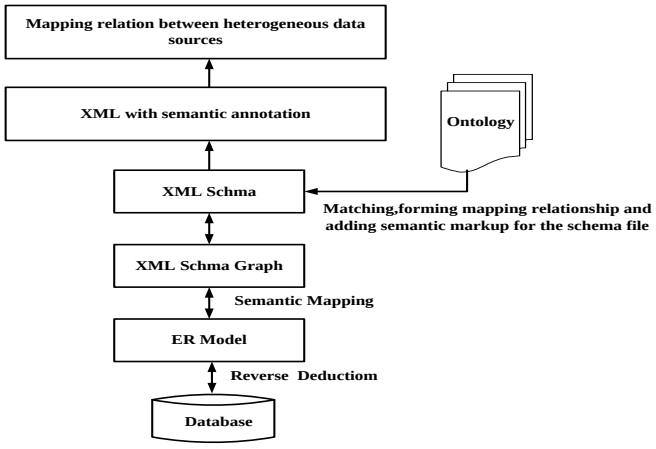

Figure 2. Semantic Conflict Resolution Model

3. Semantic Conflict Resolution Scheme

3.1. Mapping from Relational Schema to XML Schema

Since the relational database is not an efficient way for data explosion, electronic transfer of data, and electronic business on the Web, we introduce a methodology in which a relational schema will be translated to an Extensible Markup Language (XML) schema definition for creating an XML database that is a simple and efficient format on the Web. We apply the Indirect Schema Translation Method that is a semantic-based methodology in this project.

When establishing database system, formal structure abstract is must be done firstly, then the frame of data express and data manipulation is constructed by the data model. ER model is a classical concept data model, it depict data in simple graph such as ER graph and translate the ER graph to relational model set, finally, the concept design of the database is accomplished.

The method of translating the relational schema to XML schema is an indirect method, namely, XML Schema is mapped by ER Graph. The process of mapping from relation schema to XML schema is depicted as Figure 3.

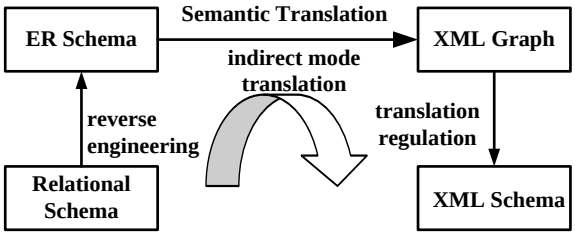

Figure 3. The Translation Process from Relational Schema to XML Schema

Although the database reverse engineering can extract the semantic structure relation model from a relational database, but more complicated implicit mode is very difficult to obtain, so the extracted OWL ontology fragments cannot fully express all the semantic information of relational database implied, it need experts in the field added for the OWL ontology fragment manually.

3.2 The Process of Adding Semantic for XML

Semantic annotation is a specific metadata generation and usage schema, aiming to enable new information access methods and to extend the existing ones[10].

The process of Adding semantic for XML Schema is as below, the connection between the element of ontology such as the concept, relationship and the individual in XML is established, which has a clear semantics. This process is also called semantic annotation based on ontology.

Semantic annotation using the concepts and properties defined in ontology to annotate the structure of XML in Schema. In the annotation process, WordNet is used to automatically process semantic annoation [11]. in addition, the mature technology in pattern matching is also used. The annotation results is added to the XML Schema by using namespace of XML Schema extension mechanism. The issue in the process of semantic annotation include:

(1) ComplexType, SimpleType were annotated by the concepts in ontology. when concept annotation was completed, the attribute of the concept must be annotated. The matching of the concept establish a context for the matching of its attributes, this context can automatically annotate its properties with high accuracy.

(2) Different XML schema was usually annotated by different granularity in the annotation process. If coarse granularity concept is adopted and there is no fine granularity distinction of certain elements in its structure, the coarse can be directly used for annotation. If The fine granularity is distinguished by other element, the classifier can be used to annotate the element.

(3) Restriction and Extension mechanism can be annotated by the parent-child class ,the primary key foreign key can be annotated by the property in the ontology.

4. Implementation

In order to describe the conflict resolution better, we choose the railway System as an example, ER model is derived from to relation model, ER diagram of database as shown in Figure 4. The ER model is mapped to the XML Schema, it is shown in Figure 5.

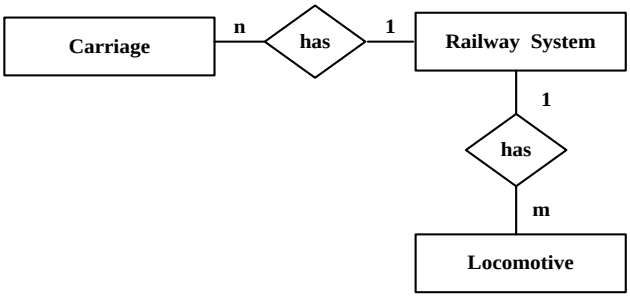

Figure 4. ER Graph of The Database Instance

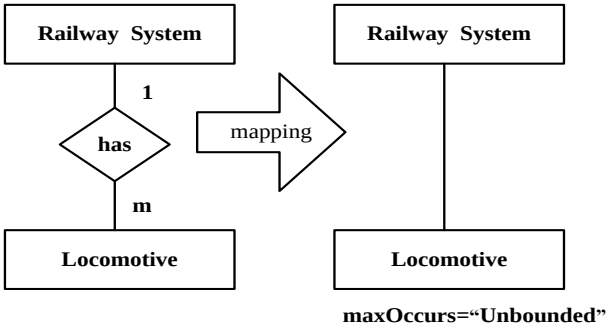

Figure 5. Process of ER Graph Mapping to XML Schema Graph

According to the rule of XML Schema diagram to XML Schema, The XML Schema code fragment of conversion is

```
<xsd: element name =“Railway System”>
< xsd: complex Type>
   <xsd: attribute name = “dno” type =“ ID”/>
 ...
   </xsd: complex Type>
</xsd: element>
<xsd: element name =“Locomotive”>
<xsd:complex Type >
<xsd: sequence>
< xsd: element ref =“ Railway System”>
</xsd: sequence >
< xsd: attribute name =“dno” type =“ ID”/>
```

< xsd: attribute name =“dno” type =“IDREF” />
</xsd: complex Type>
</xsd: element>

The relation model extracted from the railway system is as follow:

Railway System (railway bridge, railway tunnel, rail,sleeper, train, electric system)

There exist semantic conflict between relation mode Railway Infrastructure(railway bridge, railway tunnel, track, railroad tie, locomotive, carriage, electric system) and Railway System.The comparison of the two different model elements is as shown in table 1. The mapping relationship of the two heterogeneous model is as shown in table 2. by analysis we can draw a conclusion there exist heterogeneous conflict, data format conflict, data type conflict between them.

Table 1. Comparison of the Two Heterogeneous Model Elements

Railway System	Railway Infrastructure
railway bridge	railway bridge
railway tunnel	railway tunnel
rail	track
sleeper	railroad tie
train	connect (locomotive, carriage)
electric system	electric system

Table 2. Element Mapping Relation of the Two Heterogeneous Model Elements

Railway System	Railway Infrastructure
railway bridge	railway bridge
railway tunnel	railway tunnel
rail	track
sleeper	railroad tie
train	locomotive, carriage
electric system	electric system

The train attribute in Railway System is the combination of locomotive attribute and carriage attribute in railway infrastructure. In ontology knowledge fragment, on properties of locomotive, carriage and operation, namely owl:unionOf; The representation of ‘sleeper’ and ‘railroad tie’ is different, but they are equivalence relation. in OWL owl:equivalent Class is adopted to describe two identical example. Ontology OWL fragment extracted from the experiment is described below:

<owl:Classrdf:lD=” train” >
<owl:unionOfrdf: parseType=”Collection”>
<owl:Classrdf:about=”# locomotive”/>
<owl:Classrdf:about=”#carriage”/>
</owl:unionOf>
</owl:Class>
<owl:Classrdf:lD=”sleeper”>
<Owl:equiValentClassrdf：resource=”&railroad tie;sleeper”/>

</owl:Class>

Relation Schema railway system(railway bridge, railway tunnel, rail,sleeper, train, electric system) is annotated, the XML mode code is as follows:

<xs:elementname="railway bridge"type="xs:string"
semantic:type="&ontol; # railway bridge"/>
<xs:elementname="railway tunnel"type="xs:string"
semantic:type="&ontol;# railway tunnel"/>
<xs:elementname="rail"type="xs:string"
semantic:type="&ontol;# rail"/>
Relation Schema Railway Infrastructure(railway bridge, railway tunnel, track, railroad tie, locomotive, carriage, electric system)
<xs:elementname="locomotive"type="xs:string"
semantic:type="&ontol;# train"/>
<xs:elementname="carriage"type="xs:string"
semantic:type="&ontol;# train"/>
<xs:elementname="railway tunnel"type="xs:string"minOccurs="0"
semantic:type="&ontol;# railway tunnel"/>
<xs:elementname="sleeper"type="xs:string"
semantic:type="&ontol;# railroad tie"/>

5. Conclusion

When each data model is annotated, The elements and the structure of the model can achieve model mapping through semantic annotation matching and ontology applying. The mapping relationship is established by the matching between the two different data models and the ontology in matching technology. The mapping relationship between two XML Schema semantic with annotation is shown in Figure 6.

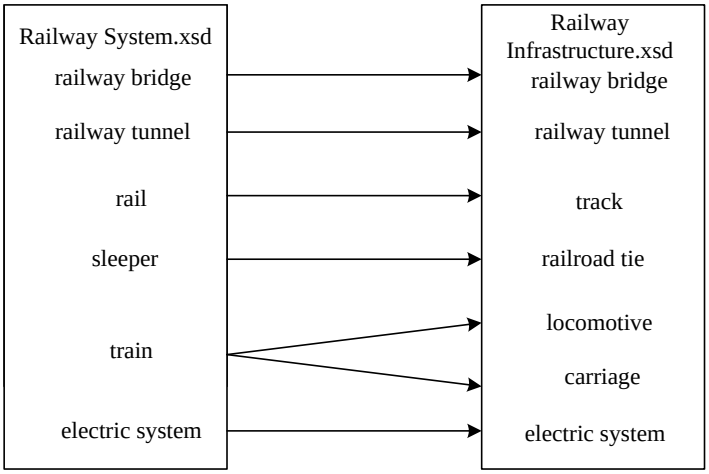

Figure 6. The XML Schema Mapping with Annotation

The XML Schema file concluded by the experiment, as attribute heterogeneous conflict, data format conflict, data type conflict semantic conflict are annotated, so the ambiguity will decreased by difference of natural language or symbols, so the semantic conflict will eliminate to some extent.

Acknowledgements

This work was supported by the follows: (1)The Young Foundation of Education Department of Hebei Province(No. 2011135) (2)The Young Foundation of Shijiazhuang Tiedao University. (No. 20133020).

Reference

[1] S. Ram, I. Park, "Semantic Conflict Resolution Ontology(SCROL):An ontology for detecting and resolving data and schema-level semantic conflicts", Knowledge and Data Engineering. IEEE Transactions , vol.16, no.2 **(2012)**, pp.189-202.
[2] J. C. Yu, L. L. Yang, "A heterogeneous data integration method based on Ontology", Information Research, vol.39, no.1, **(2013)**, pp.9-14.
[3] T. Gruber, "A translation approach to portable ontology specifications", Technical Report KSL 92-71, Knowledge Systems Laboratory, **(1992)**.
[4] A.Y. Halevy, N. Ashishy, D. Bittonz, "Enterprise information integration: successes, challenges and controversies", **(2005)**, pp. 778-787.
[5] Y. Chen, S. Li and H. Li, "A data integration system based on Ontology", Computer Engineering, vol.33, no.23, **(2013)**, pp. 90-93.
[6] Z. X. Wang, T. R.Wang, C. Q. Ye, "Research on Ontology Integration", Computer Engineering, vol. 33, no. 2 **(2007)**, pp. 4-6.
[7] Y. B. Wang, "Semantic Annotation Research Based on Ontology", Lanzhou University of Technology, **(2010)**
[8] R.P. Liang, H.W. Li , W. J. Li, "Progress in geography spatial semantics and the semantic web research", 2011 the national surveying and Mapping Science Doctoral Forum, Zhengzhou, **(2011)**.
[9] J .Z. Pan, "RDFS(FA): Connecting RDF(S) and OWL DL", IEEE Transactions on Knowledge and Data Engineering, vol.19, no.2, **(2007)**, pp. 192-206.
[10] S. Bechhofer, "The semantics of semantic annotation", **(2002)**.
[11] J. L. G. Dietz, "Enterprise ontology: Theory and methodology", Berlin: Springer, **(2010)**, pp. 35-80.

Author

Haiming Jing, he received the B.S. degree in Industrial Automation from the Harbin University of Science and Technology for Nationalities of China, Heilongjiang in 1999, and the M.S. degrees in Control Theory and Contrl Engineer, Electrical Engineering College, Yanshan university for Nationalities of China, Hebei in 2002. His current research interests include digital convergence, Ontology Computing. He is a vice professor of the Information Science and Technology

