Semantic Conflict Resolution Scheme Based on Ontology

Ning ZHAO

School of Information Economic and Management Shijiazhuang Tiedao University
Shijiazhuang, 050043, China
zz_ning860@sohu.com

Keywords: Semantic Heterogeneity, Semantic Conflict, Conflict Resolution, Ontology, Semantic Annotation.

Abstract. In this paper, the ontology is made full use to describe the characteristics of the relation between concept and concept, adding semantics to XML Schema. This is the so-called semantic annotation technology. To solve the problem of semantic conflict. In the field of railway, there are many such semantic conflicts, by using this method can well solve the problem of semantic heterogeneity among them.

Introduction

XML(Extensible Markup Language) is meta annotation language, is a simplified subset of SGML, it combine the powerful function of SGML and the usability of HTML to Web applications, which denotes the relationship between data.

The adoption of the XML[1] language, designed by the World Wide Web Consortium (W3C), for representing semi-structured data on the Web, has lead to a rapid growth in the amount of XML data available over the internet. Taking advantage of the huge amount of implicit and distributed information on the Web is a significant challenge.

But Schema is mainly used to determine the structure of XML documents, cannot be used to determine the specific meaning and semantic relation between elements. Although it organization elements with a hierarchical way, but this way does not provide the relationship between the elements, but only provides a grammar, reuse some simple structure to construct more complex structure, cannot express the semantic relationship between these elements. In order to add the semantics for the Schema, its structure and the carrier that can express the semantic information must link together. Ontology can be the carrier, because ontology can provide semantic description for some domain concepts.

Semantic conflict is the conflict caused by using different ways in heterogeneous systems to express the same entity in reality [2].Ontology is a new specification business data in recent years, which can accurately describe the data semantics, and infer implicit data semantic relations [3]. Ontology has become a new generation of data transmission and data description language format. It is mainly applied in the agent system, knowledge management system, digital library, electronic commerce platform, information retrieval, natural language processing, etc. However, the semantic description of knowledge expression ability and accuracy is insufficient.

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.

It highlights the structure description in an open and self description way, which reflects the relationship between data.

**Ontology**

**Semantic Conflict**

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 cannot 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].

**The OWL Technology**

The most used and well-known language to describe ontologies is Ontology Web Language (OWL). This technology allows for a common, formally defined and description logics supported data-format to be specified. Effectively, it models the world around the systems in a graph model. This graph contains the concepts present in the modeled domain, the relations between those concepts and potentially classification axioms which specify generic knowledge about the domain and can be exploited by the business logic of an application at a generic level of abstraction, by means of generic reasoning mechanisms. This common, agreed data-format can then be used to exchange the information and its attached domain model beyond the undertakings’ system boundaries, thus facilitating an integrated view of the asset’s conditions. Moreover, the ontology technology not only allows to model the data in a formalized manner, but also to reflect the semantics of that data, thus the information and the knowledge of those systems.

An example of such domain knowledge modeled in the ontology instead of in the application is graphically illustrated in Fig 1. The example shows how the transfer of domain-specific rules from the application to the model can facilitate a reusable and generic end-user application. As the domain logic is incorporated inside the model, the application does not need to be changed when the characteristics of the logic change. The first approach, illustrated at the top of the diagram, has the domain rules, i.e. that a fault is a measurement with a value of 10 in the first system, and a measurement with a value of 20 in the second measurement system, included in the application. Thus, with every other deployment with yet another type of measurement systems, the end-user application would have to be altered as well. In the second approach, illustrated at the bottom of the diagram, the logic is included in the model. Together with a reasoner, the end-user application would not have to be changed.
for new deployments with other types of measurement systems. This work is shifted towards the model engineering process, a task of domain experts. As a result, the application should only ask for Faults. The reasoner can then automatically infer for every domain, using the ontology model, which measurements are actually faults for that specific domain.

Figure 1. Example Illustrating the Use of Descriptions and Restrictions in an Ontology.

**Semantic Conflict Resolution Scheme**

**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 Fig. 2.

Figure 2. The Translation Process from Relational Schema to XML Schema.

**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 [7].

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 annotation [8]. In addition, the mature technology in pattern matching is also used [9]. The annotation result is added to the XML Schema by using namespace of XML Schema extension mechanism.

Implementation

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

```xml
<xs:element name="railway bridge" type="xs:string" semantic="&ontol;# railway bridge"/>
<xs:element name="railway tunnel" type="xs:string" semantic="&ontol;# railway tunnel"/>
<xs:element name="rail" type="xs:string" semantic="&ontol;# rail"/>
<xs:element name="sleeper" type="xs:string" semantic="&ontol;# railroad tie"/>
```

Relation Schema Railway Infrastructure(railway bridge, railway tunnel, track, railroad tie, locomotive, carriage, electric system)

```xml
<xs:element name="locomotive" type="xs:string" semantic="&ontol;# train"/>
<xs:element name="carriage" type="xs:string" semantic="&ontol;# train"/>
<xs:element name="railway tunnel" type="xs:string" minOccurs="0" semantic="&ontol;# railway tunnel"/>
<xs:element name="sleeper" type="xs:string" semantic="&ontol;# railroad tie"/>
```

**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 Fig. 3.

Figure 3. 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.

Acknowledgement
This work was supported by the follows: The Young Foundation of Shijiazhuang Tiedao University. (No. 20153011)

Reference

[1] World Wide Web Consortium (W3C). Extensible Markup Language (XML) 1.0, second edition, 1999. Available at: http://www.w3.org.

[2] T. Gruber, A translation approach to portable ontology specifications, Technical Report KSL 92-71, Knowledge Systems Laboratory, Stanford University, Stanford, CA, 1992.

[3] Dietz J L G. Enterprise ontology: Theory and methodology[M]. Berlin: Springer, 2010: 35-80.

[4] Halevy AY, Ashishy N, Bittonz D. Enterprise information integration: successes, challenges and controversies[C]/SIGMOD, 2005: 778-787.

[5] Y Chen, S Li and H Li. A data integration system based on Ontology. Computer Engineering, 2007, 33(23): 90-93.

[6] Z. X, Wang, T. R, Wang, C. Q. Ye. Research on Ontology Integration. Computer Engineering, 2007, 33(2): 4-6.

[7] Semantic annotation, indexing, and retrieval. Web Semantics: Science, Services and Agents on the World Wide Web 2 (2004) 49-79.

[8] Miller G A. WordNet: A Lexical Database for the English Language[Z]. http://www.cogsci.princeton.edu/~wn.

[9] Rahm E, Bernstein P A. A Survey of Approaches to Automatic Schema Matching[J]. The VLDB Journal, 2001, 10 (4).