

Generating XML Schema from Relational Database and Object Oriented Database

Myint Myint Lwin, Thi Thi Soe Nyunt, Yuzana
University of Computer Studies, Yangon

lwin.myintmyint@gmail.com, thithisn@gmail.com, yuzana.yzn@gmail.com

Abstract

The number of XML documents is growing rapidly on the Web because Extensible Markup Language (XML) becomes a standard format for exchanging and representing data over the Internet. But most of the data are stored in structured databases such as relational and object-oriented databases. Therefore, data model heterogeneity and schema heterogeneity makes the main challenging tasks for data transferring on the Web and to get the standard method. Fortunately, XML schema provides the converting the physical data to XML format is more accurately and efficiently. This paper proposes a framework for generating semantic XML schema from relational and object-oriented databases schema through the concept of Extended Entity Relationship (EER) model and Unified Modeling Language (UML class diagram). To avoid the loss of information, schema matching step is used to compare the XML schemas generated by each model and merge them to provide good result of schemas.

1. Introduction

Nowadays, Web application plays as an important role on the Internet for providing the accessing and transferring data between different sites efficiently. But it faces many challenges for data transferring on the Web such as data model heterogeneity and schema heterogeneity. Therefore, many researchers investigate and find the best solution to solve this problem.

Fortunately, XML is the excellent method for exchanging data on the Internet. Then it is widely employed on the Web application because it is defined as the accepted standard language for data exchange and carries on the Web. In real world, most of data are stored in the traditional databases such as relational database (RDB) or object oriented database (OODB). Users spend a lot of time to access or manage the heterogeneous databases without the common schema format. In practice, the traditional data model and data sources are impossible to change immediately into the XML format and apply XML technology. Although XML provide many benefits for the data carry on the Web; it cannot competent with the existing database technology such as relational and object oriented database technologies because they reached the mature stages such as concurrency, persistence and consistency etc.

In the current situation, the traditional data source necessary to change into XML format but the developer indispensable the existing database technologies and find alternative solution for this issue. As a result, schema conversion becomes a new trend in the research areas and we develop schema conversion mechanism over the relational and object-oriented databases by converting a standard XML schema from the various schemas such as RDB schema and OODB schema. In our study, we investigated the EER and UML class diagram have the more common features to the XML format and they are more applicable for converting the XML format. Moreover, both models provide the semantic concepts for the XML schema. As a result, we apply the concepts of EER and UML class for translating the real database schema to XML schema. But EER and

UML models are used by many designers to present the design of the database or software. But, our proposed system used these models as a basis concept for generating the semantic XML schema.

The proposed system generated the mapping rules that are based on the concepts of these models and converted the XML schema from RDB and OODB schemas. Many researchers used only one model from these models to convert the XML schema. Therefore, some semantic are loss. In the proposed system used both models to avoid this problem. And then schema matching step enhanced the XML schema result by comparing the resulted schemas generated by each model and combining them to obtain more accurate XML schema.

This paper is organized as follows: in the following section we discuss the related works of the proposed system. In section 3 we present the nature of EER model, UML model and relationship of these models to XML format. The proposed framework is described in section 4. Finally, we conclude our proposed system in section 5.

2. Related Works

Many researchers have proposed various schemas conversion methods and mapping rules for converting from relational database to XML schema. The paper [8] proposed the translation method for relational schema with semantic constraints into XML schema through an Extended Entity Relationship (EER) model and showed data conversion from relational data to the XML documents. In [7], they proposed a method of translating the conceptual schema of a relational database into XML schema through EER model and then physical data are transferred from relational table to XML document using stepwise procedures. Another proposed algorithm [9] is Query pattern-based relational schema to XML schema and based on a concept that columns related to equi-join in relational schema can have referential integrity. The papers [12, 13] presented the schema conversion methods between XML and relational models with three schema conversion

algorithms such as Constraints-preserving Inlining (CPI), Nesting-based Translation (NeT) and Constraints-based Translation (CoT). The paper [16] proposed the techniques to identify superclass/subclass relationship in XML Schema and proposed guidelines to map XML Schema to ORDB schema utilizing such semantic information. In [3], it presented the concept and methodology of using EER model to design an object-oriented database schema. The schema conversion approaches from OODB to XML schema are presented by some researchers. The presentation of object oriented database with XML-DTD is described by paper [4]. N. Taher el all [15] presented a novel approach for mapping an existing object-oriented database into XML and vice versa. Some researchers presented mapping rules of the UML class diagram to XML schema. K. Tobias and K. Thomas [10,11] presented a method of how to automatically extract and XML document structure from a conceptual data model (UML class diagram). B. Grady el all [2] suggested a graphical notation in UML for designing XML Schema. Many issues involved in representing XML Schema using UML and presented a transformation approach which is both comprehensive and thorough are discussed in paper [6]. D. Eladio [5] presented the current transformation of UML models into XML schema.

The above approaches converted relational or object-oriented database into the XML schema based on either EER model or UML class diagram. The proposed framework differs from the above approaches because it generates the semantic XML schema using the mapping rules that are based on both the concepts of EER model and UML class diagram. Therefore, it provides more semantic information in the schema.

3. Theory Background

The increasing amount of XML documents over the Internet and the benefits of XML are accepted as a standard for data representation and exchanging between different data model. However, in real world, data are stored in traditional database such as relational

databases and object oriented databases. The owners of the business want their data to present with XML format for data exchanging. Their desires are fully complete if these data are stored in XML databases. Unfortunately, all data cannot be transformed into XML format immediately because XML is immature state for storing and retrieving data than relational database system. Therefore, the conversion of relational data into XML document and mapping to XML schema becomes the challenging tasks for the research area. Many approaches for storing XML documents in relational databases are classified into three categories: model-mapping, structure-mapping and semantic-preserving mapping. The model-mapping approaches used fixed relational schema to store XML documents. Most of these approaches based on the tree or graph structure to store XML data. In the structure-mapping approach, schemas are generated from XML documents and use them as a mapping of attributes and values in XML documents into the relational tables. The latest XML trend is leading to the semantic. These trends provide the data semantic in the relational schema are translated and preserved in the XML document. To get the semantic XML, researchers try to find the efficient model to convert the relational or object oriented data to XML documents. The entity relationship model and UML class diagrams can provide these semantic to the XML schema.

In this section describes the concepts of EER, UML class and the relationship with XML schema by considering the semantic constraints.

3.1. Extended Entity Relation (EER)

Extended Entity Relationship (EER) model is the model that has resulted from extending the original entity relationship (ER) model with semantic data modeling concepts such as superclass/subclass, specialization/generation, inheritance, categorization, higher-degree relationships, aggregation etc. It is widely used as a language for definition of structuring of database or information system. In the current database environment, designer employ EER model as a conceptual schema to describe the

data structure and relationships of a relational database.

3.2. Unified Modeling Language (UML)

The Unified Modeling Language (UML) is a language for visualizing, constructing and documenting object-oriented software systems. It had been widely accepted as a standard for modeling software systems. However, UML class diagrams are more popular and used to model the static design view of a system. Moreover, it can support semantic information for the database design. The semantic of the class diagram is considered from the multiplicity on their attributes and inheritance association.

3.3. XML Schema

XML schema is a more powerful way to describe the structure of XML documents. It provides the syntax and defines a way in which elements and attributes can be represented in a XML document. It can also support not only inheritance as the opportunity for re-usability but also provides the ability to specify data types for both elements and attributes. They provide semantics of XML documents. XML schemas are a richer and more powerful of describing information than what is possible with DTDs.

3.4 The relationship between XML schema, EER and UML class diagram

Both EER and UML class have many common features of XML schema. They provide the semantic information in the XML schema. The nature of XML document are nested and inheritance. These features are support by both models. The entity and attributes same as the element and attributes in XML schema. The class and attributes also same as the elements and attributes in the XML schema. The associations of both models can support the semantic constraints for the XML schema.

4. Proposed Framework

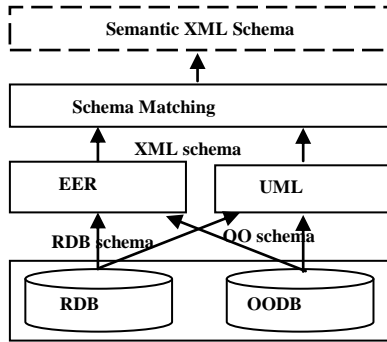


Figure 1: Architecture of the proposed system

The proposed framework is shown in Figure 1. It consists of two mapping flows to produce semantic XML schema. They are mapping from relational database schema to XML schema and another is object oriented database schema to XML schema. Each mapping flows includes following steps.

Step 1: Analyze the real database schema to map the concept of EER and UML class such as relationship, inheritance, cardinality or multiplicity, etc.

Step 2: Map the semantic constraints of EER and UML class into XML schema

Step 3: Generate the semantic XML schema by merging the XML schema generated by each model to avoid the loss of semantic information.

To generate XML schema, we already defined the semantic constraints for associated mapping rules with each database model and the XML schema. The proposed system applied the concepts of EER model and UML class as the intermediate mapping step. The mapping rules are considered based on the nature of each databases such as relational and object oriented database and then mapped with concepts of EER model and UML class.

To determine inheritance constraint, in the relational database schema, the inheritance is based on the primary key and foreign key constraints. In EER model, to define the strong and weak entity is very important as well as to define the superclass and subclass in UML class. This concept is viewed as the complex type element in XML schema. Because the element is

defined as complex type if an element contains one or more child elements. Another important factor is association. In EER model, cardinality such as one-to-one, one-to-many, many-to-many are investigated using their association. On the other hand, multiplicity constraint is considered in UML class diagram. These cardinality or multiplicity constraints are mapped into the XML schema as the attribute of minOccurs and maxOccurs. The mapping constraints described above are general. The intermediate semantic XML schemas are generated by each model after mapping steps. These schemas can lack some semantic information due to the structure of each model. The final step is schema matching step. Schema matching is the process of developing semantic matches between two or more schemas. In this step, identifying semantic mappings, or correspondences, between two or more schemas are needed. But the proposed system needs to match only two XML schemas. They have similar element and attributes. This step is added in the process because to fill the semantic lack from each model and to generate fully semantic schema.

Now, we will present some mapping rules. The following tables show the most general feature of mapping rules.

Table 1. Basic mapping rules for relational database schema

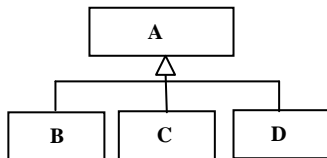
Relational Database	EER	UML	XML schema
Table	Entity	Class	Element
Column	Attribute	Attribute	Attribute
Key	Relationship	Association	Nested element

Table 2. Basic mapping rules for object oriented database schema

Object-Oriented Database	EER	UML	XML Schema
Object	Entity	Class	Element
Attribute	Attribute	Attribute	Attribute
Association	Relationship	Association	Nested Element

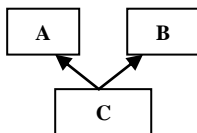
The semantic information is essential for XML schema because the structures of XML documents are relied on it. To get the semantic information from database schema, the proposed system analyzes and defined the most complex constraints for the mapping rules. Therefore, it analyzed the relationship between these models and used the detail concepts of each model such as inheritance and cardinality to capture the semantic constraint of the real database schema. In the analyzing step, we consider the mapping concepts of two models (EER and UML class) for each schema to provide fully semantic constraints. The proposed system need to provide valuable mapping rules to express these semantic constraints in the XML schema. The important features of EER and UML class diagram are collected and find the associated mapping constraints are defined. The following examples are some of complex constraints defined in the proposed system.

Example 1: Generalization and specialization



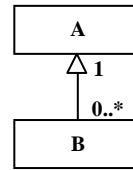
```
<xs:element name="A" type="xs:string"/>
<xs:element name="B" type="xs:string"
  substitutionGroup="A"/>
<xs:element name="C" type="xs:string"
  substitutionGroup="A"/>
<xs:element name="D" type="xs:string"
  substitutionGroup="A"/>
```

Example 2: Multiple generalizations



```
<xs:element name="C" substitutionGroup =
  "A, B"/>
```

Example 3: Cardinality or Multiplicity



```
<xs:element name="A" type="xs:string"
  abstract="true" />
<xs:complexType name="xs:string">
  <xs:sequence>
    <xs:element name="B" minOccurs="0"
      maxOccurs="unbounded" />
  </xs:sequence>
</xs:complexType>
```

The above illustrations are general constraints of mapping rules into XML schema. Other mapping constraints are needed to consider and generate the complete semantic XML schema. Final result of the proposed system is semantic XML schema.

5. Conclusion

In this paper, we investigate and analyzed the nature of Extended Entity Relationship (EER) model and UML class diagram and used the concepts of both models to define the mapping rules. Then, we proposed a new framework for generating the semantic XML schema using these mapping rules. The resulted XML schemas are very applicable because it is converted with the semantic constraints of EER model and UML class diagram such as association, generation and specialization, inheritance. Therefore, this provides many benefits for the resource virtualization and exchanging data for web application. We will compare the accuracy of resulted schema and the benefit of proposed methods.

References

[1] B. Roger, "A Survey of Schema Matching Research", UMBCMWP 1031, September 2007.

- [2] B. Grady, C. Magnus, F. Matthew and K. Jari, "UML for XML Schema Mapping Specification", August 12, 1999.
- [3] Byung S. Lee, "OODB Design with EER", Journal of Object-Oriented Programming, March 1996, pp. 61-64.
- [4] D. Chitra, P. Manisha and S. Sahebrao, "Representing object oriented database using XML DTD", Advances Information Mining, ISSN: 0975, Volume 1, Issue 2, 2009, pp.11-13.
- [5] D. Dladio, L. Jorge, P. Beartriz, R. Aurea, Angel L. Rubio and Maria A. Zapata, "A Survey of UML Models to XML Schamas Transformations", Universidad de La Rioja. 26002 Logrono, Spain, 2007, Volume 4831, pp. 184-195.
- [6] Dilys S. Flora, P. Rosanne, K. Shonali and I. Maria, "UML Documentation support for XML Schema", Proceedings of the 2004 Australian Software Engineering Conference (ASWEC'04), IEEE, 2004.
- [7] F. Joseph and P. Francis, "Converting Relational Database into XML Document", Department of Computer Science, City University of Hong Kong, 152-41088/01 IEEE, 2001.
- [8] F. Joseph, F. Anthony, H. K. Wong and Y. Philip, "Translating Relational Schema with constraints into XML schema", City University of Hong Kong, International Journal of Software Engineering and Knowledge IJSEKE, 2006, Volume 16, Issue 2, pp.201-243.
- [9] K. Jinhyung, J. Dongwon and B. Doo-Kwon, "Query Pattern-based Relational Data to XML Data Translation Algorithm", Journal of Computer Science, 3 (4): pp. 212-217, 2007.
- [10] K. Tobias and K. Thomas, "Rule-Based Generation of XML Schema from UML Class Diagrams", Leipzig University of Applied Science, Department of Computer and Mathematics, D-04251 Leipzig, 2003.
- [11] K. Tobias and K. Thomas, "Rule-Based Generation of XML DTDs from UML class Diagrams", Leipzig University of Applied Sciences, Department of Computer Science and Mathematics, D-04251 Leipzig, 2003, Volume 2798/2003, pp. 339-354.
- [12] L. Dongwon, M. Murali and W. Chu, "Schema Conversion Methods between XML and Relational Models", Information and Software Technology, 2006, pp. 245-252.
- [13] L. Dongwon, M. Murali and Wesley W. Chu, "Effective Schema Conversions between XML and Relational Models", In European Conf. On Artificial Intelligence (ECAI), Knowledge Transformation Workshop, 2002.
- [14] N. Taher, A. Reda and Mick J. Ridley, "Two-Way Mapping between Object-Oriented Databases and XML", October 12, 2008.
- [15] S. Prasanna, Rukshan I. Athauda and G. Jeevani, "Utilizing Superclass/Subclass Relationship in Mapping XML Schema to ORDBs", Proceedings of the sixth IEEE International Conference on Computer and Information Technology (CIT'06), 2006.