Common Data Mapping System Between XML and Relational Database

Hong Xin

College of Computer Science and Technology,
HuaQiao University
Xiamen, China 361021
xinhong@hqu.edu.cn

Chen Weibin

College of Computer Science and Technology,
HuaQiao University
Xiamen, China 361021
chenwb@hqu.edu.cn

Abstract—**Although most of current databases support XML format data, the different interfaces they applied are lack of commonality. An XML exchange system is build by java and the standard SQL. The system includes two virtual platforms. Through the platform of virtual relational database system and the platform of virtual XML database system, the XML-document and the relational-data can be exchanged by ignoring the system platform.**

Keywords- XML, Relation, Data Mapping

I. INTRODUCTION

XML (Extensible Markup Language) like the HTML, are the reduced version of SGML (Standard Generalized Markup Language). XML is platform independenct, but relies on content. XML is a powerful tool to deal with the structured data in the Internet environment [1].The XML is widely used in the internet as a data description language because of its self-descriptive Character. We developed a data exchange system to realize two direction of data mapping between XML data and Relational data [2-4].

Nowadays, different database systems deal with the data format providing by themselves. They do not support data format to each other. This cause many problems when people want to exchange data between different systems or use data from deferent database systems. By using our platform, the data can be exchanged between different databases. For example, the data in the Oracle can be output as XML data, and then the XML data can be imported into SQL Server. Our platform also supports the data exchange between the pure XML file and the relational database systems. It meets a widely demands for data exchange requirement.

In this article, firstly we discuss about the lack of XML support of database systems now. Secondly, we introduce the mapping algorithm between the XML and the relational database system. Thirdly, we show the mapping platform of the two direction mapping. Fourthly, we point out the principle of the system in data mapping. Lastly, we make a conclusion of our works.

II. THE XML INTERFACE IN DATABASE SYSTEM

For the widely use of the XML in the network, many database applications nowadays provide the XML interfaces, ie, SQL Server and Oracle. The architecture of SQL Server is shown in fig 1. The SQL Server 2005 provides the FOR XML and OPEN XML command for XML. The architecture of Oracle is shown in fig 2.Oracle provide the Java, C and C++ parser for XML.
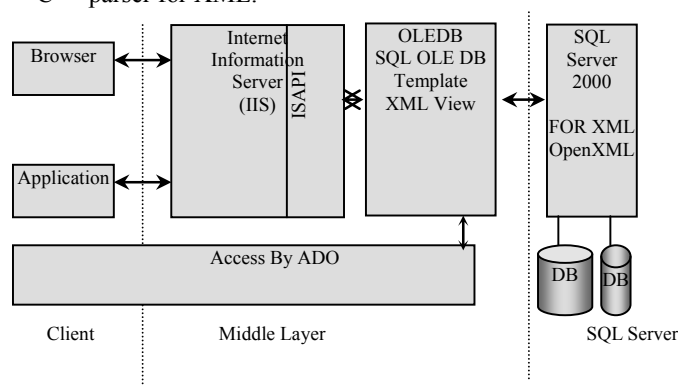

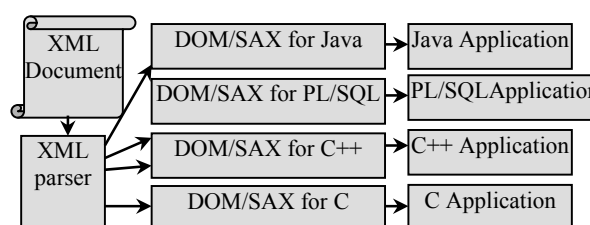Figure 1. SQL Server architecture

Figure 2. XML parser in Oracle for Java , C and C++

As mentioned before, although many large database systems provide the XML interfaces, the interfaces are different to each other. The lack of commonality cause great difficulties for the heterogeneous database to process the XML data. So it is important to establish a platform independent method to support XML data exchange between different database systems.

The XML interfaces providing by the database systems treat the XML data as simple file. They can not deal with the relations beneath the XML node. Our platform can extract the relation when the XML data are mapped into relational database system. At the same time, the data in the database systems can be mapped into XML data and preserved the relation.

III. THE EXCHANGE ALGORITHM BETWEEN THE XML DATA AND THE RELATIONAL DATA

We know that the formats of the XML data and the relational data are different. Data mapping from XML to the relational data will be loss. The XML is organized by tree structure. And the relational data is organized by two-dimensional table. The key point is building a mapping algorithm between them.

A. The mapping algorithm from the XML data to the relational data

First of all, the XML is not strict in schema. One XML schema may allow different kinds of data organization. At the same time, not all the XML data have the schema file. So before mapping the XML data to the relational data, we need to establish a XML schema. The main algorithm of XML schema modeling is reading the full XML data file and training a model of the XML schema corresponding to the file[5].

Secondly, we create a XML2DB algorithm to realize the mapping from XML schema to the relational schema. The attribute of the XML schema is mapping to the column of a table. The node in different layer of XML schema is mapping into different table. The parent and the son node are represented by the relation of the table[6].

Lastly, all the data in the XML will be mapped into relational database system under the rule of the XML2DB algorithm[7].

The XML2DB algorithm is shown in Figure3.

Algorithm1：XML 2DB
1. **read XML data and build XML schema**
2. **XML schema normalize**
3. **mapping XML schema to relational schema**
4. **mapping XML data to relational data**

Figure 3.　DB2 XMLrecursion－ algorithm

B. The mapping algorithm from the relational data to the XML data

The relational data are restricted by the table schema. It is easy to map the relational data to the XML data. We create a DB2XML algorithm to realize the mapping from relational database to XML. The column is mapped into XML node. And the relation of table is mapped into relation of node, for example, the parent node and the son node. The DB2XML algorithm is shown in Figure4.

Algorithm2:DB2XML
1. **mapping DB relation to XML node structure**
2. **mapping table schema to XML schema**
3. **mapping relational data to XML data**

Figure 4.　DB2XML

C. Discussion

For the different format between the XML data and the relational data, there are some special conditions we should consider.

Firstly, the recursion structure is allowed in the XML data, but forbidden in the relational data. The XML2DB-recursion－ mapping algorithm can map the recursive structure in the XML schema into the relational schema , and form a relational schema in normal formulas,by adding a new node in the recursive structure.The algorithm is shown in Figure5.

Algorithm 2XML2DB-recursion－normalize
If the element definition reference a farther element
then the two elements make an indirect recursion
else{**If** element reference itself
　　　then the element make a direct recursion
　　　else there is no recursion}
If there is a recursion
Then{ create relation between two elements ;
　　　delete the reference; Output the SQL to create table}

Figure 5.　XML2DB-recursion－ algorithm

Secondly, the XML data usually have redundant structure. Such structure can not be stored in the relational database. So before mapping the XML data to the relational database system, we should normalize the XML data at first. The redundant mapping algorithm can normalize the XML data and form the non-redundant schema of XML. So that the XML data can be mapping as normal structure[8].The algorithm is shown in Figure 6.

Algorithm 2:XML2DB- redundant－normalize
1. If there is attribute redundant
　　then normalize the attribute "maxOccurs" tag
2. If there is group redundant
　　then normalize the group "maxOccurs" tag
3. If there is element redundant
　　then normalize the element "maxOccurs" tag
4. If "maxOccurs" tag is "*"
　　then create table relation

Figure 6.　XML2DB- redundant－normalize algorithm

Lastly, there are something more we can do to preserve the different structure between the XML data and the relational data. For example, we can add some columns in a table to record the attributes and notions of the XML data[9].

IV. THE SYSTEM DESIGN

The platform-independent database exchange system is created through two virtual platforms by java: the virtual relational database platform base on XML, and the virtual XML platform base on the relational database. The platform processes the data by using the middle standard language, so both of the virtual systems are independent of any system. The different XML interface of the heterogeneous database

system is no more a problem by using the java parser and the standard SQL language. The structure of the system is shown in Figure7.

A. The virtual relational database platform base on XML

The virtual relational database platform enables the XML document to be shown as relational data without relational database system, for example, the XML data can be shown as relational table, and can be added，updated，deleted in the virtual table as we do in the real database system. At the same time, the XML data can be easily mapped into any relational database through the virtual relational database platform.

B. The virtual XML platform base on relational database system

The virtual XML platform enables the relational data to be shown as an XML document without XML database. The relational data can be added, updated, deleted in this virtual platform as process the data in a real XML database. And the data in the relational database can be easily mapped into XML document format through the virtual XML platform.
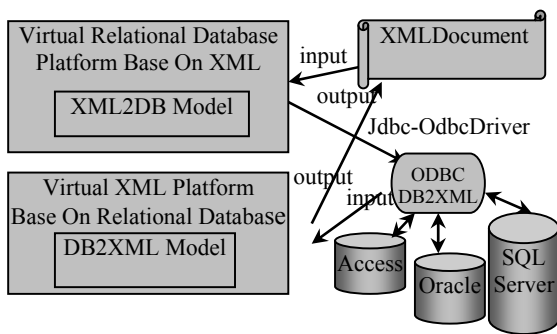

Figure 7.   the common data mapping system between XML and relational database

V.   THE PRINCIPLE OF THE SYSTEM OPERATION

A.   The mapping subsystem from XML document to relational data

In the Virtual Relational Database Platform, the XML document can be shown as relational database without data mapping and be mapped into relational data by the standard SQL through the XML2DB model. The XML2DB model use java parser to read the XML document and create the corresponding table in relational database through ODBC and also create the XSL (Extensible Stylesheet Language) file. The XSL file can show the XML document as a relational table without any relational database management behind. The XML document can be modified as in a real relational database, and be output into any relational database in the Virtual XML Platform. The different XML interfaces in different database are no more problems by using the java parser and the standard SQL language.

The flow chart of the XML2DB model is shown in Figure8. Schema is not necessary for the XML document,

the mapping from XML document to relational database is divided into two parts: the mapping without XDR model and the mapping with XDR model. For the commonality of the program, our experiment system realize the mapping in three steps: firstly, it creates the XDR model; secondary, applied the mapping algorithm from XML model to relational model, the mapping creates the table and the XSL file; finally, it outputs the XML document to the relational data.
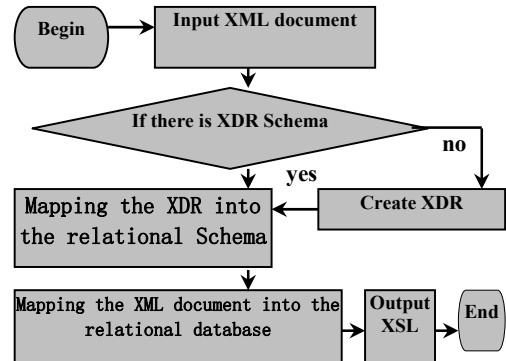

Figure 8.   The flow chart of the XML2DB model

B.   The mapping subsystem from relational data to XML document

In the process of mapping from XML data to relational data, the relational data is fetched by SQL and be mapped into XML document through ODBC, the data can be shown and processed as an XML database though there is not any XML database behind. Because all the process is using standard SQL language, so this platform is fit for any database management system.

The system can map the relational database to XML data in two ways: one way the system outputs tables individually, the other way the system output tables and their relations at the same time. The first way is simple, but will lose tables' relation. The second way considering the relations of the tables, can output all the tables and their relations. The working flow in the system is shown in Figure9.
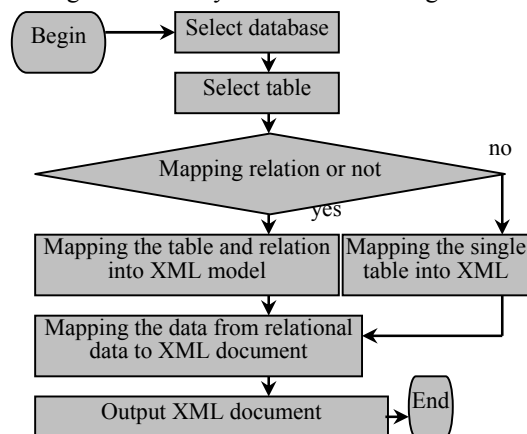

Figure 9.   The flow chart of the DB2XML model

VI. CONCLUSIONS

In order to solve the problem of multiple XML interfaces in the heterogeneous relational database system. We create a platform independent system by using the java XML parser and standard SQL language to realize the data exchange between XML document and any relational database system.

Through the Virtual XML Platform and the Virtual Database Platform, the data can be shown without the corresponding database system's support. The XML document can be shown as relational data without the data mapping by the Virtual Relational Database Platform. The relational database can be shown as XML document without the data mapping by the Virtual XML Platform.

The XML document can be mapped into any relational data by the XML2DB model in the Virtual Relational Database Platform. Any Relational Database can be mapped into XML by the DB2XML model in the Virtual XML Platform. In fact, the system is efficient in changing data between XML document and relational database system in the heterogeneous database systems.

We discuss the mapping algorithm between the XML data and the relational data. The different structures cause the mapping is loss. The XML2DB-recursion — mapping algorithm and XML2DB-redundant—normalize algorithm solve the main format different problems between the two kinds of data. There are also some detail differences between them. We usually ignore them when we exchange data. On one side, the mapping from relational data to the XML data is lossless. On the other sides, actually most mappings from the pure XML data to the relational data are loss. If we want to keep the full information of the XML data, we need to add some table columns to record additional data of the XML file.

REFERENCES

[1] xml. http://baike.baidu.com/view/63.htm,2009-5-12

[2] Kang Xiao-Bin,Zhang Er-Hu, Wu Xu-Yi.A mapping algorithm from XML Schema to Relational Schema[J]. Computer Applications. ,2004,24 (5):106-108

[3] WAN G ZhiOping.XML-Based Heterogeneous Relational Databases Integration System[J]. Journal of Henan University (Natural Science). 2007

[4] HONG Xin.The research of Mapping technology from XML Document to Relational Database Base on XDR[D].Fujian:HuaQiaoUniversity,2004.

[5] HONG Xin , CHEN Weibin , JIAN Chongjun .Mapping from XML Model to Relational Model[J]. Journal of Huaqiao University (Natural Science).2009.30(6). 646-649.

[6] HONG Xin , CHEN WeiBin , DU JiXiang.Extract Relation Schema from XML Schema[J]. Journal of Huaqiao University (Natural Science). 2010.31(3).288-291.

[7] HONG Xin,CHEN Weibin,JIAN Chong-jun.Mapping from XML-Data to Relational-Data[J]. 2011. 36（1）：39-42.

[8] HONG Xin, CHEN Miao , JIAN Chongjun.Normalization mapping from redundant XML schema to relational schema[J]. Journal of Yanzhou University (Natural Science). 2011.14(1):56-60.

[9] HONG Xin , CHEN Weibin. Lossless mapping research between XML schema and relational schema[J]. Journal of Shandong University of Technology(Natural Science Edition). 2010.24(5):23-30.

[10] AREMAS M, LIBKIN L. Normal form for XMLdocument[J].ACM Trans Database Syst.2004,29(1):195-232.

[11] VINCENT M, LIU Ji-xue, Liu Cheng-fei.Strong functional dependen-cies and their application to normal forms in XML [J]. ACM Trans Database Syst.2004, 29(3): 445-462.

[12] Buneman P, Davidson S, Fan Wen-fei, et al. Keys for XML[C]//Proc. of the 10th International WWW Conference. Hong Kong, China:[s. n.], 2001: 201-210.

[13] CHEN Yi, DAVIDSON S, HARAC, et al. RRXS: redundancy reducing XML storage in relations [C]//Proceedings of the 29th VLDB Conference. Berlin, Germany, Morgan Kaufmann Publishers Inc,2003: 189-200.

[14] TAN Zi-Jing,SHI Bai-Le.Propagating Functional Dependency and Normalization Between Relations and XML. Journal of Software. 2005,16(4):534-539.

[15] Writing by Paul j.Burke, Translating by Fang Li etc. Professional SQL Server 2000 XML [M].Beijing: China Electric Power Press.2003.1

[16] Writing by （American） Graeme Malcolm ; Translating by Lain studio.Programming Microsoft SQL Server 2000 with XML[M]:Beijing:PeKing University Press, 2001,11.

[17] The Guide for Oracle 9i XML network database Development[M].Beijing:Beijing Hope Electronic Press, 2002,4.