Chapter 15
State of the Art Technology

Ibrahim Dweib
Sultan Qaboos University, Oman

Joan Lu
University of Huddersfield, UK

ABSTRACT

This chapter presents the state of the art approaches for storing and retrieving the XML documents from relational databases. Approaches are classified into schema-based mapping and schemaless-based mapping. It also discusses the solutions which are included in Database Management Systems such as SQL Server, Oracle, and DB2. The discussion addresses the issues of: rebuilding XML from RDBMS approaches, comparison of mapping approaches, and their advantages and disadvantages. The chapter concludes with the issues addressed.

APPROACHES FOR STORING AND QUERYING XML

A number of different techniques for storing XML documents in a RDB have been established. These techniques can be divided into two groups: the schemaless-centric technique and the schema-centric technique (Dweib et al., 2008). The first one makes use of XML document structure to manage the mapping process (Tatarinov et al., 2002; Dweib et al., 2008; Soltan and Rahgozar, 2006; Zhang and Tompa, 2004b; Jiang et al., 2002; Yoshikawa et al., 2001). The second one depends on schema information to develop a relational schema for XML documents (Fujimoto et al., 2005; Shanmugasundaram et al., 1999; Amer-Yahia et al., 2004; Atay et al., 2007b; Xing et al., 2007b; Knudsen et al., 2005; Lee et al., 2006).

The aim of mapping XML documents into relational database is to make use of the capabilities of the relational database which are: indexes, triggers, data integrity, security, multi-user access, and query optimization by SQL query language. In the meanwhile XML technology is trying to gain the above-mentioned capabilities, developed for RDBs, and efficiently store, retrieve, and rebuild XML data from RDBs.
The studies that address the problem of mapping XML document into RDB take care of the above issues, and attempt to translate users’ XML queries, either XPath expression (Berglund et al., 2007) or W3C’s recommendation XQuery expression (Boag et al., 2007), into SQL queries (Oracle, n. a.). XQuery gives power to the translation method since XQuery comprises XPath, and it is recommended by W3C, while XPath is not. The translation method should also consider its ability to rebuild, the stored XML document without losing information, and retrieve it in an acceptable time. Many studies have tried to address translation and restore constructing labelling methods. Labelling methods aim to reserve nodes order, parent-child and ancestor-descendant relationships, and document structure (Tatarinov et al., 2002; Chung and Jesurajaiah, 2005; Soltan and Rahgozar, 2006; Li and Moon, 2001; O’Neil et al., 2004; Wu et al., 2004; Kobayashi et al., 2005).

**Schema-Based Mapping**

One of the early studies in this area was conducted by (Shanmugasundaram et al., 1999) from the University of Wisconsin-Madison. They proposed three mapping techniques: Basic, Shared, and Hybrid Inlining. These are proposed to map DTDs into relational schemas. Basic Inlining proposed building a separated table for each element in the DTD while in the Shared Inlining each element is represented in one table. The Hybrid Inlining technique inlines shares an element which is not repeated or recursively related. These techniques are different from one another in the degree of redundancy; they vary from being highly redundant in Basic Inlining, to containing no redundancy in Hybrid Inlining.

The above approach offers limited structures to represent the features of XML data, such as nested relationships, ordering of XML documents, and the DBMS schema representations. Querying these structures is usually complex since the end users are not familiar with them.

Mapping algorithms for XML DTDs to relational schemas were proposed by Atay et al. (2007b) from Wayne State University. They attempted to enhance the shared-inlining algorithm (Shanmugasundaram et al., 1999), in away to overcome its incompleteness and eliminate redundancies caused by the shared elements. They claimed that the algorithm can deal with any DTDs including arbitrary cyclic DTDs, but shared-inlining algorithm deals merely with two mutually recursive elements. Dealing with cycles which involve more than two elements in a DTDs is not clear. Figure 1 shows the three cases they considered in their inlining procedure. In case 1, a node $a$ is connected to a node $b$ by a normal edge, and $b$ has no other incoming edges. In this case, node $b$ is inlined into its parent node $a$, and the parent-child relationships are maintained between $b$ and its children. In case 2, node $a$ is connected to node $b$ by a normal edge where $b$ has other incoming edges (i.e. $b$ is a shared node). In this case node $b$ is not inlined into its parent node $a$ since $b$ has multiple parents. In case 3, node $a$ is connected to a node $b$ by a star edge, such that every node of $a$ can contain multiple occurrences of $b$. In this case, the node $b$ is not combined into its parent node $a$ in order to avoid redundancy.

Figure 2 gives an example of the idea of the inlining procedure clear. Figures 2.A and 2.C show the DTD graphs, where the inlining results are shown in Figures 2.B and 2.D after applying the inlining algorithm. It could be noted from figures that nodes which are connected by, -edge or *-edge and, -edge must point to a shared node.

Redundancy reduction XML storage in relations (RRXS) within XML Functional Dependency (XFD) was proposed by (Chen et al., 2003). They defined constraints to capture the structural constraints as well as semantic information. It makes use of XML schema semantic constraints. Using the semantics of a document could reduce the redundancy since node identifiers can be removed where value based keys are still available for particular elements. Unfortunately the sug-
Related Content

On the Updating of Domain OWL Models at Runtime in Factory Automation Systems
www.igi-global.com/chapter/on-the-updating-of-domain-owl-models-at-runtime-in-factory-automation-systems/198619?camid=4v1a

State-of-the-Art Information Retrieval Tools for Biological Resources
www.igi-global.com/chapter/state-of-the-art-information-retrieval-tools-for-biological-resources/198546?camid=4v1a

Service-Oriented Architectures for Context-Aware Information Retrieval and Access
www.igi-global.com/chapter/service-oriented-architectures-context-aware/28075?camid=4v1a

Automatic Text Document Summarization Using Graph Based Centrality Measures on Lexical Network
www.igi-global.com/article/automatic-text-document-summarization-using-graph-based-centrality-measures-on-lexical-network/204464?camid=4v1a