Chapter 19
Conclusions and Further Research

Ibrahim Dweib
Sultan Qaboos University, Oman

Joan Lu
University of Huddersfield, UK

ABSTRACT

In this chapter, the authors characterize a new model for mapping XML documents into relational database. The model examines the problem of solving the structural hole between ordered hierarchical XML and unordered tabular relational database to enable use of the relational database systems for storing, updating, and querying XML data. The authors introduce and implement a mapping system called MAXDOR to solve the problem.

CONTRIBUTIONS

The following are the main contributions presented throughout this thesis:

- **XML Document Mapping into Relational Database**: A novel method is introduced to partition XML document into tokens (i.e., element and attributes). It relies on assigning a tuple in a relational table for each token information and relations with its neighbours. The method works efficient and performs well for large XML documents.

- **Building XML Document from Relational Database**: A novel method is introduced to build original XML document or update one from relational database. It relies on retrieving document contents depending on token links and token levels which formulate XML document as a group of subtrees.

DOI: 10.4018/978-1-4666-1975-3.ch019

Copyright © 2013, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Conclusions and Further Research

- **Updating XML Document Contents:** A novel method is used to update (i.e. insert new token or modify its name or value) XML document contents stored in relational database. It is based on creating links for each token with its neighbours to maintain document structure without a need to relabel or re-index document contents.

- **Querying and Retrieving Many XPath Axes of Xml Document:** A novel method is introduced to access most of XPath axes preceding-sibling, following-sibling and descendant without storing all possible XPath information for document contents (Tatarinov et al., 2002; O’Neil et al., 2004). It relies on dynamically creating result subtree(s) on the fly using a temporary table “XPathQuery table” for the desired XPath expression storing all interested tokens.

**ADVANTAGES**

- **High Flexibility of Updating:** MAXDOR approach performed updating processes of inserting new tokens in any location in the document and at any level of relevance to the candidate element (i.e. parent, child, left-sibling and right-sibling), updating token name and value at constant cost of execution time since there is no need to relabel following tokens IDs or overwrite tokens paths.

- **Stability:** The approach worked fine in both directions; mapping and rebuilding for large documents: “Auction” document with 600MB size and 9244050 tokens can be processed without trouble.

**RECOMMENDATIONS**

1. Our model is strongly recommended for a system where XML document contents needs to be updated very frequently.

2. Our model is strongly recommended for a system where maintaining document structure is important as in document-centric documents.

**DRAWBACKS AND LIMITATIONS**

- **Loss of Information:** Our mapping algorithm does not consider some information in the original XML document such as processing instructions, comments, CDATA sections and external entities. Furthermore, it needs an enhancement to consider multiple occurrences of texts in one element.

- Since XPath query expression is used for retrieving information from XML document, it ascribes two limitations to our approach:
  1. Only one query upon one document will be applied at the time.
  2. XPath language doesn’t have commands to insert or update an XML document content that enforces us to add an editor to manage updating process. The editor can manage small documents only.

- Our approach uses fixed schema in relational database and one table “tokens table” is used to store document contents. In addition, maximum table size in Microsoft Access is limited to 2GB including System Objects and indexes. These limitations restrict the maximum XML document size to be mapped in our approach to 600MB approximately.
Related Content

Design of a Novel Search Engine for Prospective Question Answering
[www.igi-global.com/article/design-of-a-novel-search-engine-for-prospective-question-answering/126327?camid=4v1a](www.igi-global.com/article/design-of-a-novel-search-engine-for-prospective-question-answering/126327?camid=4v1a)

Ranking Algorithm for Semantic Document Annotations
[www.igi-global.com/article/ranking-algorithm-semantic-document-annotations/72703?camid=4v1a](www.igi-global.com/article/ranking-algorithm-semantic-document-annotations/72703?camid=4v1a)

Information Retrieval Model Using Uncertain Confidence's Network
[www.igi-global.com/chapter/information-retrieval-model-using-uncertain-confidences-network/198562?camid=4v1a](www.igi-global.com/chapter/information-retrieval-model-using-uncertain-confidences-network/198562?camid=4v1a)

The Use of Mixed Methods in Organizational Communication Research
[www.igi-global.com/chapter/use-mixed-methods-organizational-communication/69732?camid=4v1a](www.igi-global.com/chapter/use-mixed-methods-organizational-communication/69732?camid=4v1a)