

A SURVEY OF MODEL MAPPING APPROACHES FOR STORING XML DOCUMENTS IN RELATIONAL DATABASES

¹PUSHPA SURI, ²DIVYESH SHARMA

^{1,2}Department of Computer science and applications .Kurukshetra University , Kurukshetra
E-mail: ¹ pushpa .suri@yahoo.co.in, ²divyeshsharma89@gmail.com

Abstract— XML documents are used for data exchange over the internet. Relational databases provides a repository for store and query these documents. Storage of XML documents can be categorized in two areas: Structured mapping approaches and Model mapping approaches . In this paper we provide a survey of most cited model mapping approaches. We describe each approach and its structure in this paper.

Index Terms— XML, DATABASES, MODEL MAPPING.

I. INTRODUCTION

Extensible Markup Language (XML) is a standard for data representation over the internet. Database management system is used to manage large amount of data efficiently. For storage of XML data, Three approaches can be used:- File system, native databases and relational databases. Relational databases provides a mature way to store and query xml documents. We use relational databases to store xml documents. XML databases can be classified into two categories: - XML enabled database and native databases. We work on XML enable databases , in XML databases , XML documents are decomposed into graph structure and then store them into databases. While in native XML databases there is no need of mapping of XML documents in databases.

II. STRUCTURAL AND MODEL MAPPING APPROACHES

In structural mapping of XML documents in relational databases, XML schema or DTD is activated with XML document. In model mapping approaches, XML schema or DTD is associated with xml document. In DTD based approaches, firstly we have to resolve DTD operators after then we make DTD graphs then we make relational schema.[8][9][10] While in model mappings , DTD is not associated with documents. In these mappings, firstly xml documents is converted into trees and then nodes of the tree are to be stored into databases. Several model mapping approaches deals with this area and pervious [1][2][3][4][5][6][7]. In this paper we survey most cited model mapping approaches and briefly explains all these approaches.

III .OVERVIEW OF MODEL MAPING APPROACHES

1. EDGE APPROACH: - Edge approach was proposed in 1999.This approach use edge

information of XML tree for storing into relational databases. The structure of this approach is:- Edge (source, ordinal, target, label, flag, value). This table stores the identifiers of source and target nodes of each edge in the graph. Flag indicates whether the node is internal node or it contains any value. Ordinal number reflects the ordered edges.

2. XREL:- It is path based approach. It use Xpath data model. Therefore XML document is represented a tree structure of nodes. It use start and end position of nodes rather than the node identifier. The basic schema is as follows:- Path (Path id, Path Expr) Element (Doc id , Path id, Start , end, index, reindex) Text (Doc id , Path id, Start , end, value) Attribute (Doc id , Path id, Start , end, value) Here Doc id represents document identifier. Path expression represents simple path expressions, start represents start position of a region, end represents end position of region and value represents a string value.

3. X PARENT:- XPARENT approach has the following schema. Label path (id, len, path) Data path (Pid, Cid) Element (Path id, Did, Ordinal) Data (Path id, Did, Ordinal, Value) Label path table preserves the attribute id, len and path. Id attributes represents a unique label path identifier, len represents path and path attribute represents the simple path expression respectively. In data path table, attribute PID and CID represent the parent and child node id of an edge. In the element in data table attributes such as path id is foreign key of the id the label path table. DID is data path identifier. Ordinal number represents the order of the attribute and value represents the value.

4. XPEV:- This approach represents the three table schema. Path (Pid, Path, Expression) Edge (Pid, source, target, label, ordinal, flag) Value (Pid, source, target, label, ordinal, value) This

approach is edge oriented approach. PID represents the id of distinct path. Path express attribute is a simple path expression. Ordinal attribute the order of nodes. Flag attribute is used to represent whether it is internal node or a value.

5. LNV:- LNV consists of following table. Path (doc id, path of signatures, path of nodes, values, type of nodes, position) Signature (Label, signature id) Doc id represents the document identifier.

We can store more than one document in same relational schema. The attribute path of signature represents the list of signature that exists on each path. The attribute value represents the value of path. The attribute type of node represents whether the node is element attribute or text. The position represents the position of nodes among its siblings. The attribute is signature id is signature given for each distinct label in the XML tree.

6. SCHEMA LESS APPROACH:- Dewib proposed a model mapping approach which consist of two tables. Documents (doc id, doc structure) tokens (doc id, token id , token name , token value) This approach used document table as a master table. In this table doc id attributes used to describe the unique id of each document. Doc structure is big text field containing a coded sting describing each document structure, any changing on the document structure should be reflected in this field. In token table, token id is primary key generated for each token. Doc is foreign key linking the token table to a document table. Token. Token value is text value of the xml tag properly.

7. SMX/R APPROACH:- SMX/R is mapping approach which use path based concept to store XML documents. The schema is as follows Path Table (doc id, Pid, start position, end position, node id, node type, node values). Doc id is unique id of the path. Start pos is the start location of the node. End pos is the end location of the node. Node lvl is the nesting depth. Node type is the type of the node whether it is an element, text or attribute .The second table is as follows:- Path index (Pid, path, node name). Pid is the id of the unique path. Path is path string. Node name is the name of the node .This approach also provides an algorithm for storing XML documents in relational databases.

8. XRECURSIVE

In XRECURSIVE approach, the storage of XML documents in relational databases done by two tables. The structure of two tables is as follows:- Tag structure (tag id , tag name) Tag value (tag id , tag value, tag type) In tag structure table, element name is represented by tag name field . ID represents the id of the parent of the element. In tag value table, Type attribute is used t represent the type of the element. whether it is an element or an attribute.

CONCLUSION

In this paper we provide a survey of most cited model mapping approaches. In future research, we will compare model mapping approaches by its database sizes.

REFERENCES

- [1] M. YoshiKawa.,T.Amagasa,T.Shtimura, "XREL: A path based approach to storage and retrieval of XML documents using relational databases".ACM Transactions on Internet Technology,1(1) ,pp:110-141, August 2001.
- [2] J. Qin, S.Zhao, S. Yang, W. Dau, " XPEV : A Storage Approach for Well-Formed XML Documents", FKSD, LNAI 3613,2005 pp.360-369.
- [3] D.Florescu, D.Kossman, " A Performance Evaluation of Alternative Mapping Schemes for storing XML Data in a Relational Database",Rapport de Recherche No. 3680 INRIA, Rocquencourt, France,1999.
- [4] A.Salminen,F.Wm, "Requirements for XML Document Database Systems. First ACM Symposium on Document Engineering", Atlanta.2001 pp:85-94.
- [5] H. Zafari, K. Hasami, M.Ebrahim Shiri, " Xlight,an Efficient relational schema to store and query XML data", In the proceedings of the IEEE International conference in Data Store and Data Engineering 2011, pp:254-257.
- [6] M.Ibrahim Fakhraldien, J.Mohamed Zain, N. Sulaiman, "XRecursive: An efficient method to store and query XML documents",Australian Journal of basic and Applied Sciences,5(12) 2011 pp: 2910-2916.
- [7] M.Sharkawi, N. Tazi, " LNV : Relational database Storage structure for XML documents", The 3rd ACS/IEEE International Conference On Copmputer Systems And Applications, 2005, PP:49-56.
- [8] J.Shanmugasundaram, K. Tufte, C. Zhang, G.He, D. Dewitt,J.Naughton, "Relational Databases for Querying XML Documents:Limitations and opportunities,"VLDB 1999 ,pp : 302-314.
- [9] M. Atay, A Chebotko, D. L iu, S. Lu, F. Fotoubi , "Efficient schema based XML to relational data mapping", Information systems ,Elsevier 2005.
- [10] S.Lu,Y. Sun, M.Atay,F. Fotouhi, " A New inlining algorithm for mapping XML DTDS to relational schema" In Proceedings of the First International Work-shop on XML Schema and Data Management , in conjunction with the 22nd ACM International Conference on Conceptual Modeling , Chicago, IL, October 2003.

★★★