Advances in Computer Science and Information Technology (ACSIT)
Print ISSN: 2393-9907; Online ISSN: 2393-9915; Volume 2, Number 2; January-March, 2015 pp. 162-166

http://www.krishisanskriti.org/acsit.html

Survey of XML to Relational Database Mapping Techniques

Maya Mourya[1] and Preeti Saxena[2]

[1]School of Computer Science & IT DAVV, Indore, M.P., INDIA
[2]School of Computer Science & IT DAVV, Indore, M.P., INDIA
E-mail: [1]maya22mourya@gmail.com, [2]preeti_ms@rediffmail.com

Abstract—*Extensible markup language (XML) recommended by W3C (World Web Consortium) has become known standard for representing and exchanging data over the Internet. Because of simplicity and generality of XML, it is widely used in World Wide Web. To store and querying XML document three technologies are commonly in use. They are Native XML database, Object Oriented Database and Relational Database. Relational database technology is popular among all other due to its important features such as, effective storage and efficient retrieval techniques. Thus, the mapping of XML data into relational database is required for data centric applications.*

Mapping of XML to relational database used both database power capabilities. Mapping approach provides interoperability between data source. In XML database, mapping main challenges is mapping of hierarchical ordered nature of XML data model to unordered nature of relational data model. This paper presents review of XML-to-Relational database mapping approaches in the categories of schema aware and schema less techniques. It is presented in the form of a comparative study of these mapping approaches

1. INTRODUCTION

XML is richly structured document used to store, represent and exchange data over World Wide Web [1]. A large amount of XML data needed to be stored in digital copy and exchanged over the Internet. Storage and retrieval of the vast amount of web data is popular area of interest for the researchers and database vendors. However, the important issue is retrieval and querying these data in an efficient manner. For this task many solution have been proposed in past. They use different technologies. Native XML database, Object Oriented Database and Relational Database [2] are common technologies in practice. Relational database is most popular among them. The aim of mapping XML documents into relational database is to utilize relational database power capabilities such as storage and efficient retrieval technique. XML to Relational database are dividing into three categories: Model mapping, Structure mapping and Constraints mapping [3].

In model mapping approach, fixed XML data model is used and translation of XML does not consider structure information, which is associated with XML document. This approach does not dependent on the complexity of the XML schema. Implementation of modeling approach is present many research paper [14] [15] [16] [18] [19]. In this approach, the nodes and sides of the XML document tree are mapped to the relationship model. The relationship model creation does not depend on the complexity of XML Document structure. This approached method used a fixed relationship model to store all of the XML document structure.

In structure, mapping approach XML document is associated with DTD and XML schema information. This information is used to translate XML schema. To solve complex DTDs, inlining technique is used which apply simplification rules to XML documents. There are several inlining techniques available such as basic inlining, shared inlining and hybrid inlining [4]. After simplification of DTD, inlining DTD graph is drawn and then database schema is generating according to inlined DTD graph.

In the constraint-based approach, it is required to map XML data to relations based on the integrity constraints. In this method constraints are investigated in XML data model and relational schema generated with constraints. Constraints are use to optimized semantic query, which minimize the processing cost of the query.

Many papers [7] [9] [10] [12] [13] used structure and constraints approach to generate relational database from XML documents.

The remainder of this paper is organized as follows: Section 2 describes overview of mapping techniques. Section 3 explains comparative study. Section 4 is the conclusion.

2. OVERVIEW OF MAPPING TECHNIQUES

Mapping of XML to Relational database differs in terms of metadata, information to be preserved and creation of relational schema. According to metadata used, mapping

strategies are broadly divided into two categories schema aware and schema less. Techniques that store XML documents in pre-defined relational tables are called schema less. Model mapping approach uses this technique. Schema aware techniques used schema information and focuses on structure and constraints mapping. In this section, related work of these techniques is discussed.

3. SCHEMA AWARE TECHNIQUES

Normalized relational storage for XML [22] presented an algorithm for generating an optimal design for XML in relational setting. Algorithm based on computing a set of minimum covers for all functional dependencies on a universal relation for given XML. According to author, this approach was efficient in terms of reducing data redundancy, preserving semantic expression Functional Dependencies (XDs) and the schema information did not preserved order of element in XML document.

Joost Visser et al. [7] introduce an algebraic approach to schema transformation that is constraint-aware. Constraints are preserved from source to target schemas and that new constraints are presented. This paper proposes an improved approach rather than labeling. Approach used refinement theory and point-free program transformation. The presented work is unique in many ways as compare to work presented in [8-10] [13]. This approach Expresses constraints as point free functions, which can be converted automatically and from structure programs including path expressions. Employs strongly typed Boolean function to capture constraints this has advantage of being more expressive and allowing a fully compositional treatment. Their method is not limited to hierarchical-to-relational mappings. It can be used for schema transformation in general and deal with arbitrary constraints.

Davidson et al [9] presented an alternative constraint-preserving approach, using path expressions. They presented a technique for refining the design of relational storage. Three algorithms are presented, First check whether a given functional dependency is propagated from XML keys via a predefined view. In addition, others two algorithms compute a minimum cover for all functional dependencies. The experimental result demonstrates that the algorithms are efficient in practice. The proposed framework that improves the consumer relational database design is the major contribution of their work. The presented work is first approach, which gives framework and algorithms to map XML constraints in relational view.

Designing information-preserving mapping schemes [10] discuss generation of constraints on relational schemas that make XML to Relational mappings information preserving. Nonstructural constraints on the initial XML schema are not taken into result. Constraints and conversion functions are expressed in variations on Datalog, which can be manually rewritten to normal form in a mechanical way [7]. This paper addresses the problem of designing information-preserving mapping schemes. The Contributions of this paper is proposed a sound and extensible framework for designing information-preserving mapping schemes. The generated relational schema of this framework is equivalent to original XML schema.

H. Pacheco et al. [12] presents Coupled Schema Transformation and Data Conversion for XML and SQL, which used two-level data transformation. In two-level data transformation, type-level transformation of a data format is couple with value-level transformations of data instances. The proposed system is consists of a combinatory library for composing type-changing rewrite rules that preserves structural information and referential constraints. The presented paper, discusses practical application Haskell based two-level transformation support. It has contributed in two manners. First, it elaborates the rules for structure and constraint information. Second, it inserts the general transformation kernel into a language-specific transformation framework. Limitation of presented work is to capture constraints method, which employs a type annotation mechanism. As a result, a smaller class of possible constraints is covered.

Teng et al. [13] presented concept and the formal definition of XML functional dependencies. This paper presents the method map constraints, such as functional dependencies, cardinality, domain, choice, reference constraints. These constraints also preserve the structure of DTDs in relational schema. The definition of XML functional dependencies proposed in this paper overcomes the shortcomings of the papers [9] [11] [17]. Definitions presented in the paper differ in several aspects. The method captures the characteristics of XML structure and differentiates between global and local functional dependencies for XML. In XML document, it considers string values of attributes and element themselves to represent functional dependencies. This method does not map multi-valued functional dependency.

4. SCHEMA LESS TECHNIQUES

.A Model Mapping Approach for storing XML documents in Relational database[14] proposes a method to store XML data using two relational table, Node table and data table. Here Note table stores all node id's along with node names. Data table stores corresponding node values in it. Proposed algorithm does not use path concept. It f Firstly assign a unique identifier to each node. Then the node value and the node id are stored in a relational table. It stores parent node id along with every node id to maintain parent child relationships. The structure of this approach is as follows:

Node (Node id, Node name)

Data (Doc id, Node id, Parent id, Node value, Node type, Node pos)

Node table stores node ids with their names. In Data table stores ids, values, Node type attribute is use to indicate whether the node is an element or an attribute or a text, Node Pos attribute is a position of the node among its siblings in the XML data graph. The paper shows that it requires less space to store data using this approach as compared to XRel[19]. In addition, it requires less number of join operations in query processing.

Edge [15] stores XML document in a single table using edge label information. It is simple and straightforward approach in terms of loading each edge into a tuple of single relation. Edge approach has a limitation that it requires multiple self joins to create path. To complete query it requires a large number of joins, which make it inefficient. Some information is also lost during simplification process thus reconstruction of XML document is not possible.

To overcome Edge approach limitations in XRel [19] a path-based approach is used. XRel schema contains four tables (Path, Element, Attribute and Text) to store XML document in relational table. In this approach XML document is presented as nodes of tree. Relational tables store each node according to node type with path information from the root to node. In this schema, node identifiers are replacing by start and end positions. Schema tables are as follows.

Path (PathID, Pathexp)

Element (DocID, PathID, Start, End, index, reindex)

Text (DocID, PathID, Strat, End, Value)

Attribute (DocID, PathID, Start, End, Value)

XRel schema requires high storage space to save large XML documents. Query response time is less then Edge but still poor because it uses large number of Θ-joins. Joins are required to determine ancestor-descendant relationships and check edge connections between elements.

Xlight, An Efficient Relational Schema to Store and Query XML Data [18] stores shredded XML documents in Relational database using five tables.

Document (DocId, Name)

Path (PathId, Path)

Data (DocId, PathId, LeafNo, LeafGroup, LinkLevel, LeafValue, hasAttrib)

Ancestor(DocId, LeafGroup, AncestorPre, AncestorLevel)

Attribute (Name, Val, id, pre)

The Document table stores the names of the documents and identifier (Docid) in the database. Docid is used as a reference in other schema. The path table stores unique id in pathid and existing root to leaf paths. Data table stored stores the related information of all leaf nodes. Ancestor information of each leaf node is stored in Ancestor table. Attribute table stores information of existing attribute.

Xlight path table implementation is similar to XRel method but difference is that, Xlight stores the information for only the leaf nodes. In these schema Θ-joins is replace by with equi-joins over the set of ancestors. As compare to Edge [15], XRel [19], Xparent [16] Xlight requires less storage space and required less query processing time.

5. COMPARATIVE STUDY

Table 3.1 presents the comparison of Model, Structure and Constraints mapping in terms of approached type, schema type, preserved order, constraints and experimental evaluation.

Table 3.1: Comparison of approaches

Research paper	Approached type	Schema aware/Schema less	Preserved ordered	Preserved constraints	Experiment Result Present	Comment
C-mapping [3], 2014	Structure and Constraints	Schema aware	Yes	Yes	Yes	Deal with N-to-one mapping constraints
Joost Visser et al. [7] , 2012	Constraints	Schema aware	No	Yes	Yes	Use algebraic approached for transformation, Deal with arbitrary constraints.
Lee et al[8], 2001	Constraints	Schema aware	No	Yes	Yes	First addressed constraints preserving
Davidson et al [9], 2003	Constraints	Schema aware	No	Yes	Yes	First framework and algorithm to mapped XML constraints to relational view.

Barbosa D. et al.[10], 2005	Constraints	Schema Aware	Yes	Yes	Yes	Addresses the problem of designing information-preserving mapping schemes
H. Pacheco et al. [12], 2007	Constraints	Schema aware	No	Yes	Yes	Used two-level data transformation for mapping
Research Paper	**Approached Type**	**Schema Aware/ Schema Less**	**Preserved Ordered**	**Preserved Constraints**	**Experiment Result Present**	**Comment**
Teng et al. [13], 2005	Constraints and structure	Schema aware	No	Yes	No	Presented formal definition of XML functional dependencies
[14], 2013	Model	Schema-less	No	No	Yes	It represents a compact structure for storing the whole information of an XML document in relational databases and it requires less memory space for storage
Edge[15], 1999	Model	Schema-less	Yes	No	Yes	First approach which store XML document in fixed relational database.
Xparent[16], 2002	Model	Schema-less	Yes	Yes	Yes	Stores multiple XML documents in relational schema, improvement of Edge and XRel method
Xlight[18], 2011	Model	Schema-less	Yes	No	Yes	Query performance is better than other model
Mustafa Atay et al. [20], 2007	Structure	Schema aware	Yes	No	Yes	A lossless mapping scheme for storing ordered XML data
A. Kannan et al. [21], 2012	Structure	Schema aware	Yes	Yes	No	Mapped XML document in relational table with Access control information.
Reduan Samad et al [22],2011	Structure	Schema aware	Yes	Yes	Yes	Introduced an algorithm, which is able to reduce data redundancy and preserve semantic constraints
MAXDOR [23] ,2009	Model	Schema less	Yes	No	Yes	No loss of information, Reconstruction of original XML documents, preserved order, search semantics
J. Wenny Rahayu et al [24], 2006	Structure and Constraints	Schema aware	Yes	Yes	Yes	Aim to map additional constraints
Zijing Tan et al. [25], 2011	Structure	Schema aware	Yes	Yes	Yes	Provide an algorithm to construct a canonical target instance

6. CONCLUSION

The presented study provides overview of different mapping approaches that map XML documents to Relational database. The approaches are use to convert XML document into Relational database using both structure and semantic aspects. Mapping for XML to Relational database is proposed to use both schema power capabilities. The paper discusses the characteristics and limitation of different approaches. It also presents the comparison of different approaches in terms of order, constraints, experimental evaluation and schema techniques.

REFERENCES

[1] T. Bray, J. Paoli “Extensible Markup Language (XML) 3rd edition”, 2008.

[2] A.A.Abdel-aziz and H.Oakasha, "Mapping XML DTDs to Relational Schemas", 2009.

[3] Sosuke OTA Atsuyuki MORISHIMA, "C-Mapping: A Flexible XML-RDB Mapping Method based on Functional and Inclusion Dependencies", 2014.

[4] 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.

[5] Zijing Tan, Liyong Zhang, Wei Wang, Baile Shi, "XML data exchange with target constraints", 2011.

[6] Sihem AmerYahia, Fang Du, Juliana Freire," A Comprehensive Solution to the XML to Relational Mapping Problem", 2004.

[7] Tiago L. Alves,The Netherlands Paulo F. Silva, Portugal Joost Visser, "Constraint-aware Schema Transformation", 2012.

[8] Lee, D. and W. W. Chu, "Cpi: Constraints-preserving inlining algorithm for mapping xml dtd to relational schema", Data Knowl. Eng. 39, 2001.

[9] Davidson, S. et al., "Propagating XML constraints to relations", in: Proc. 19th Int. Conf. on Data Engineering, 2003

[10] Barbosa, D. et al., "Designing information-preserving mapping schemes for XML", in: VLDB'05: Int.Conf. Very Large Data Bases, 2005.

[11] M. Arenas, L. Libkin, A normal form for XML documents, Proceedings of Symposium on Principles of Database Systems (PODS'02), ACM press, Madison, WI, 2002, pp. 85–96.

[12] Berdaguer, P., A. Cunha, H. Pacheco and J. Visser, "Coupled schema transformation and data conversion for XML and SQL", 2007.

[13] Teng Lv, Ping Yana, Mapping DTDs to relational schemas with semantic constraints, 2005.

[14] Dr. Mrs. Pushpa Suri, Divyesh Sharma, "A Model Mapping Approach for storing XML documents in Relational database", 2013.

[15] D. Florescu and D. Kossmann, "Storing and Querying XML Data using an RDMBS",1999.

[16] H. Jiang, H. Lu, W. Wang, and J. X. Yu, "Path materialization revisited: an efficient storage model for XML data", 2002.

[17] M.L. Lee, T.W. Ling, W.L. Low, Designing functional dependencies for XML, Proceedings of VIII Conference on Extending Database Technology (EDBT'02), LNCS 2287, Springer, New York, 2002. pp124–141.

[18] 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.

[19] M. Yoshikawa, T. Amagasa, T. Shimura, and S. Uemura, "XRel: a path-based approach to storage and retrieval of XML documents using relational databases", 2001.

[20] M. Atay, A. Chebotko, D. Liu, S. Lu, and F. Fotouhi, "Efficient schema-based XML-to-Relational data mapping," Information Systems, vol. 32, pp. 458-476, 2007.

[21] A. A. Abd El-Aziz and A. Kannan "Storing XML documents and XML policies in Relational Databases", 2012.

[22] Kamsuriah A hmad and Reduan Samad Normalized Relational Storage for Extensible Markup Language (XML) Schema, 2011.

[23] Ibrahim Dweib*, Ayman Awadi and Joan Lu "MAXDOR: Mapping XML Document into Relational Database", 2009.

[24] Qiuju Lee, St´ephane Bressan, and J. Wenny Rahayu. "XShreX: Maintaining Integrity Constraints in the Mapping of XML Schema to Relational",2006.

[25] Zijing Tan, Liyong Zhang, Wei Wang, Baile Shi, "XML data exchange with target constraints", 2011.