iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5398

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

Mapping of XML Document and Relational Database (using Structural Queries)

Pallavi Chikhale[1], Shital Harihar[2], Vandna Adhude[3], Prof. Prashant Raut[4].
[1, 2, 3, 4]*Department of Computer Science, SIT Lonavala, Maharashtra, India*

Abstract: ***Now a day's, XML files are most vital facet of code trade. As a core normal, XML provides a solid foundation around that different standards could grow. Making DTD's is possibly what the creators of XML had in mind after they known as it extensile terminology. XML may be a file extension for XML file format accustomed produce common info formats and share each the format and also the information on the globe Wide internet, intranets, et al. victimization normal code text. XML provides regular thanks to describe semi-structured information. We tend to propose the propose methodology uses stemming stopping process in conjunction with tree structure. Performing stemming, stopping and parsing on that file, as per the tag file makes columns in relational database and insert tag information in the respective columns, all files saves in one folder. These embody consistency problems for mappings and schema, additionally as imposing tighter restrictions on mappings and queries to attain tractable question responsive in information exchange. The propose method can achieve better performance in decreasing of time and space complexities.***

Keywords*:* ***Time and Space Complexity, Batch Stream Processing, Tree Data Structure***

I. INTRODUCTION

Today is an information era, most enterprises and individuals are deploying their businesses on the internet including e-learning, e-commerce and e-library. This will result in an exponential growth in data, which finally forming the so called BIGDATA. XML is an Extensible Markup Language that defines the set of rules for encoding document in both human and machine-readable formats. XML is widely used in web services, especially data transmission or data conversion and storing. XML data can be stored in disks or else in Relational database. A relational database is a collection of data items organized as a set of formally described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables. The standard user and application program interface to a relational database is the structured query language (SQL). SQL statements are used both for interactive queries for information from a relational database and for gathering data for reports. Batch Stream processing can be used to fasten the process, so overall system performance in case of time and space complexity can be reduced. Batch system have an option to allows users to set task processing in automated sequence.

II. PROBLEM STATEMENT

Propose methodology is employed to convert complicated XML structure into electronic database victimization structural queries, conjointly reduces the time complexness victimization Batch Stream process and area complexness victimization tree organization.

III. LITERATURE REVIEW

A. XRecursive

A Storage Method for XML Document Based on Relational Database

1) Authors: M.A. Ibrahim Fakharaldien, Jasni Mohamed Zain, and Norrozila Sulaiman

The protractible terminology (XML) is quickly changing into the actual normal for information exchange over the web and currently it plays a central role in information management, transformation, and exchange. Since its introduction to business within the Late Nineties, XML has achieved widespread support and adoption among all the leading code tools, server, and info vendors. As significantly, XML has become the interlanguage for information by lowering the price of process, searching, exchanging, and re-using data. XML provides a uniform, self-describing means that for expressing data during a manner that's decipherable by humans and simply verified, reworked, and printed, the recent topic is to hunt the simplest manner for storing XML documents so as to urge high question process potency. Additionally, information may be transmitted to remote services anyplace on the web mistreatment XML-based internet services to require advantage of the new presence of connected code applications. Today, the dominant storage mechanism for structured enterprise information is that the computer database, that has tested itself associate

©IJRASET: All Rights are Reserved

economical tool for storing, looking for, and retrieving data from huge collections of information. Relative information bases specialise in relating individual data records sorted by kind in tables. Developers will be a part of records along PRN mistreatment SQL and gift one or a lot of records to end-users as substantive data. The relational information base model revolutionized enterprise data storage with its simplicity, efficiency, and value effectiveness.

B. Mini XML

An efficient mapping approach between XML and Relational Database

1) *Authors:* Huchao Zhu, Huiqun Yu(Corresponding Author) In order to store numerous XML document data and offer faster response to the user, an efficient approach named mini-XML is proposed based on model-mapping approach. The major contributions of this paper are as follows: A new mapping schema named the mini-XML is proposed, which adopts the path-based technique as the among the non-leaf nodes. The reorganization of the XML documents tree is achived via labelling each leaf node with a sequence number. By comparing the time and space spent mapping XML data into a relational database, an effective evaluation is made and the baseline data is expanded to enhance the authenticity and efficiency of the experiment.

This paper describes XRel, a completely unique approach to storage and retrieval of XML documents exploitation relative databases. During this approach, associate XML document is rotten into nodes supported its tree structure, and keep in relative tables in step with the node sort, with path data from the foundation to every node. XRel allows America to store XML documents employing a mounted relative schema with none data regarding DTDs and part varieties, and conjointly allows America to utilize indices like the B+-tree and therefore the R-tree supported by management systems. For the process of XML queries, we tend to gift associate algorithmic program for translating a core set of XPath expressions into SQL queries. Thus, XRel doesn't impose any extension of relative databases for storage of XML documents, and question retrieval supported XPath expressions are often accomplished in terms of a preprocessor for info source language. Finally, they demonstrate the effectiveness of this approach through many experiments exploitation actual XML documents.

System S → { I, P, O }
Input = I
XML file = X
X = {x1, x2, x3,……xn}
Tag = T
Column = C
Value = V
Value – {Tv1, Tv2,….Tvn}
RDBMS file = R
Process = P
Step 1
Upload XML file
X = {x1, x2, x3,……xn}
Step 2
Convert file tag into columns
T → C
Step 3
Insert tag value into respective columns
{Tv1, Tv2,….Tvn} → C
Step 4
Update File
Update[X]
Output O
File saved in RDBMS
R – C[X = {x1, x2, x3,……xn}]

©IJRASET: All Rights are Reserved

IV. ALGORITHM

A. *Algorithm*

K-means Algorithm

The process to achieve the result sets of classified data is quite simple. It basically consists on several iterations of a specific process, designed to get a optimal minimum solution for all data points.

Let's look this process in detail.

First, we need to establish a function of what we want to minimize, in our case the distance between every data point and the correspondent centroid.

So, what we want is:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} \|x_j - c_i\|^2$$

With this function well defined, we can split the process in several steps, in order to achieve the wanted result. Our starting point is

$$c_i = \frac{1}{|k_i|} \sum_{x_{j \in k}}^{k} x_j$$

5 – If the cluster centers change, repeat the process from 2. Otherwise you have successfully computed the k means clustering algorithm and got the partition's members and centroids.

The achieved result is the minimum configuration for the selected start points. It is possible that this output isn't the optimal minimum of the selected set of data, but instead a local minimum of the function. To mitigate this problem, we can run process more than one time in order to get the optimal solution.

It is important for you to know that there are some variations of the initial center choice method. Depending on the problem you want to solve, some initial processes might benefit your implementations.

1) Algorithm: Sentence Level Clustering

Step 1: Enter the user query.

(e.g., What is java.)

Step 2: Apply Steaming and stopping

(e.g., Remove stop words. Here in this example, 'what' and 'is' words will be removed)

Step 3: Remaining words will be search in to another document file in database.

Step 4: result will be displayed as sentences where the keyword is present.

B. *Block Diagram Of System*

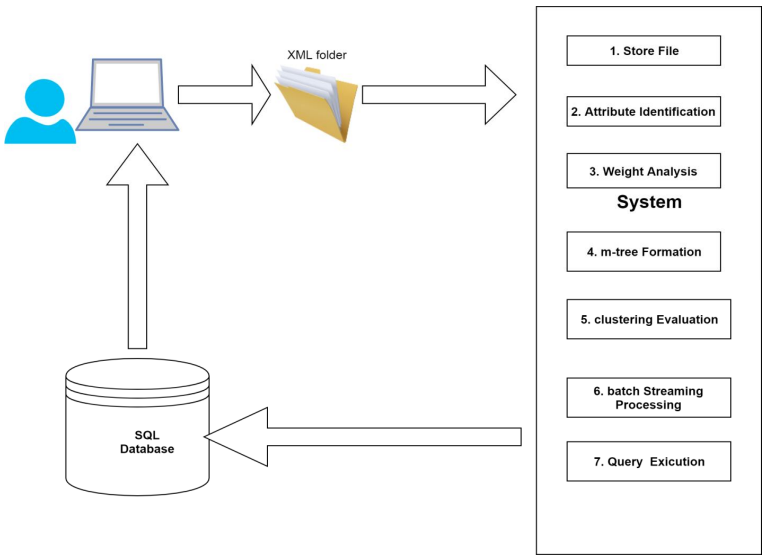


Figure 6.1. Block diagram of Identity base

©IJRASET: All Rights are Reserved

1) Collect the XML document from customers
2) Store to system for conversion
3) Create the XML folder
4) Then parse the XML document
5) Apply the propose algorithm
6) Apply the SQL queries and
7) Convert the document into Relational Database.

C. Hardware Requirement

1) System Processors: Core2Duo
2) Speed: 2.4 GHz
3) Hard Disk :150 GB

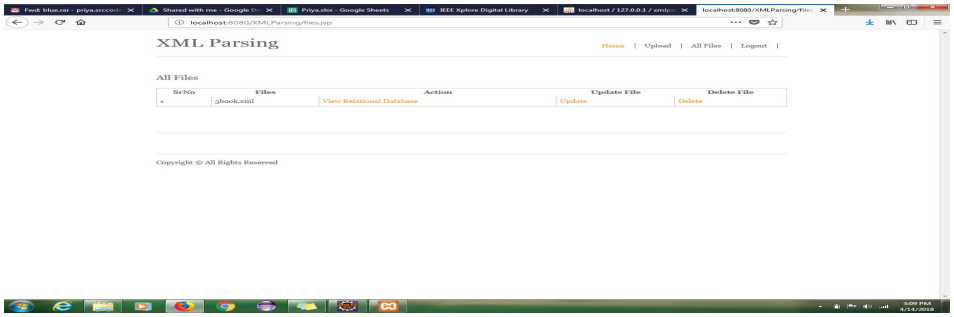
V. ADVANTAGES

A. Cost effective
B. Effective result
C. Time saving

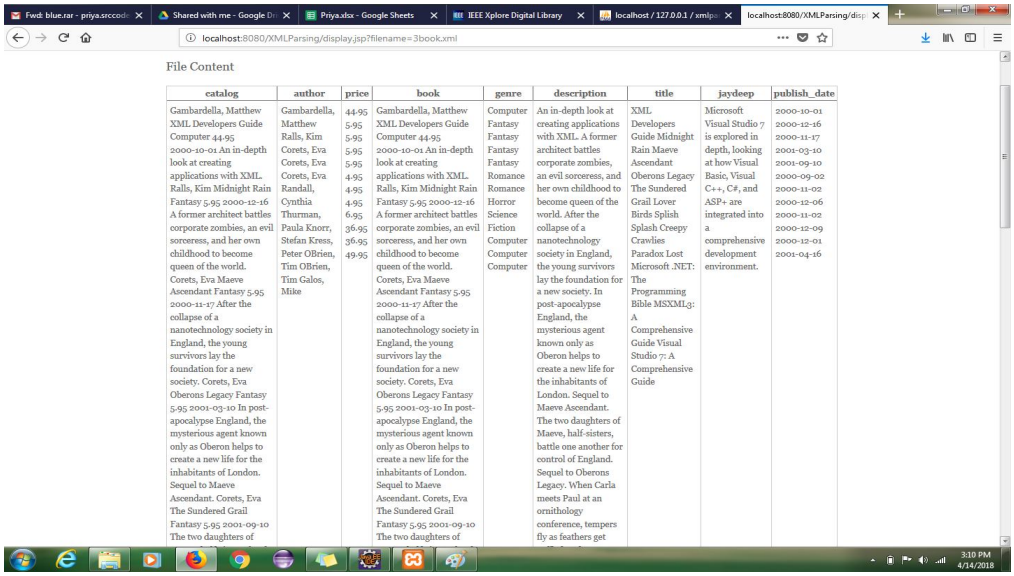
VI. APPLICATION

A. Web log mining
B. Organization data ensembling like new agencies etc.

VII. RESULT

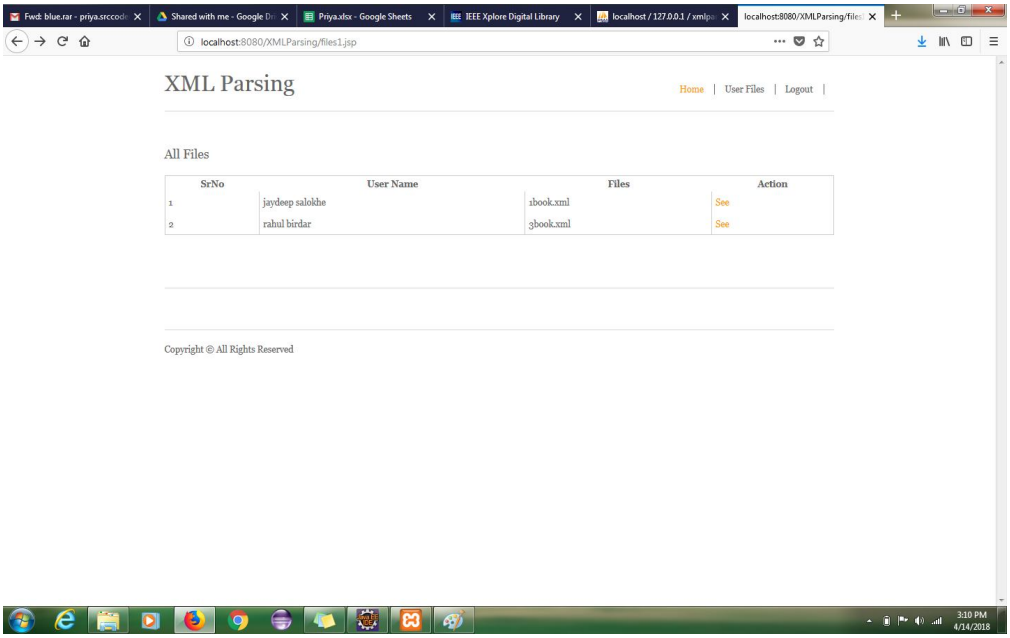


View all files

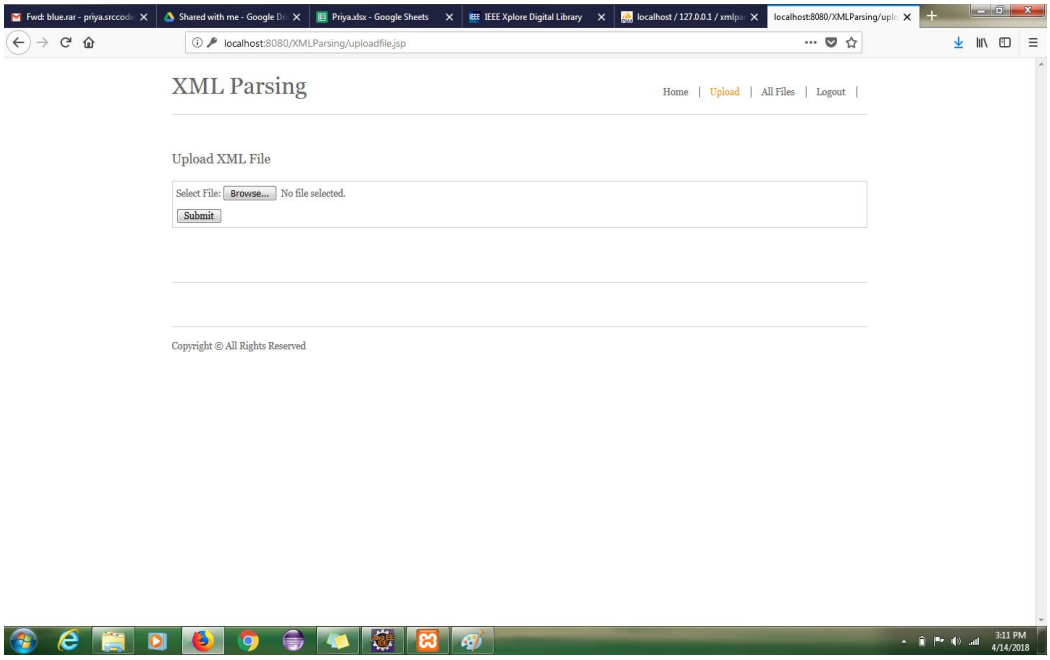


View relational database files

©IJRASET: All Rights are Reserved



Admin view all user files



User Upload XML file

VIII. CONCLUSION AND FUTURE SCOPE

XML is Associate in nursing extensile language and may be used for storing and transportation of data. A knowledgebase could also be a collective set of multiple data sets organized by tables, records and columns. It establishes a well-defined relationship between info tables. Tables communicate and share knowledge, that facilitates data search ability, organization and coverage. It uses structured search language (SQL) which will be a commonplace user application that features easy programming interface for info interaction. Knowledgebase information service springs from the relation plan of mapping data sets and was developed by Edgar F. Codd. For fasten method, there is would really like of correct conversion of XML to database. We tend to propose system that's helpful for the user to store Xml file into info and once we tend to build changes in xml file changes get reflected to info to boot. We've a bent to propose Associate in nursing economical mapping approach, the mini-XML, to mapping XML into the database.

©IJRASET: All Rights are Reserved

Propose work concludes that the proper parsing of XML data. Batch stream method beside tree structure is commonly accustomed fasten the tactic. Once creating tree, attribute objects square measure usually combined into clusters therefore, the time quality of Associate in nursing algorithmic rule square measure usually reduced.

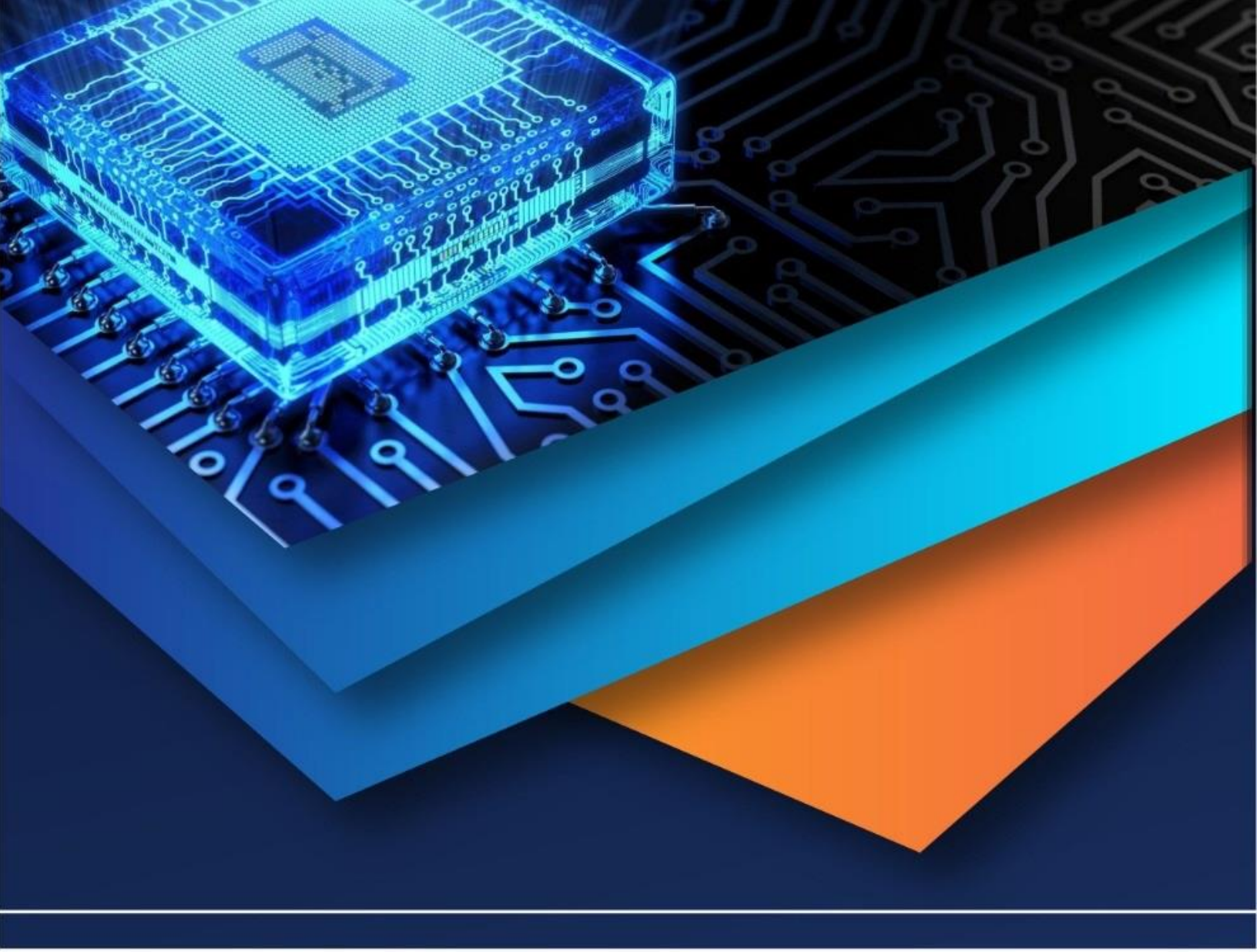
IX. ACKNOWLEDGMENT

Authors want to acknowledge Principal, Head of department and guide of their project for all the support and help rendered. To express profound feeling of appreciation to their regarded guardians for giving the motivation required to the finishing of paper.

REFERENCES

[1] Huchao Zhu,Huiqun Yu(Corresponding Author), “Mini XML: An efficient mapping approach,between XML and Relational Database”,2017.
[2] M.A. Ibrahim Fakharaldien, Jasni Mohamed Zain, and Norrozila Sulaiman,” XRecursive: AStorage Method for XML Document Based on Relational Database”,2011
[3] Bousalem, Zakaria, and Ilias Cherti. "XMap: A Novel Approach to Store and Retrieve XMl Document in Relational Databases." JSW 10.12 (2015): 1389-1401.
[4] Masatoshi Yoshikawa and Toshiyuki Amagasa, “XRel: A Path-Based Approach to Storage and Retrieval of XML Documents using Relational Databases”,2009
[5] Samini Subramaniam, Su-Cheng Haw, Poo Kuan Hoong,”s-XML: An efficient mapping scheme to bridge XML and relational database”
[6] Arenas M, Barcel P, Libkin L, et al. Foundations of Data Exchange[C]// Cambridge University Press, 2014.
[7]] Vaidya AV, Study of Optimal Indexing Schema for Xml Query Processing and Effective Storage Management 2009 to 2013, 2015
[8] Yoshikawa M, Amagasa T, Shimura T, et al. XRel: a path-based approach to storage and retrieval of XML documents using relational databases[J]. Acm Transactions on Internet Technology, 2001, 1(1):110- 141
[9] Jiang H, Lu H, Wang W, et al. Path materialization revisited: an efficient storage model for XML data[C]// Australasian Database Conference. Australian Computer Society, Inc. 2002:85-94.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)