Indonesian Journal of Electrical Engineering and Computer Science
Vol. 14, No. 1, April 2019, pp. 450~454
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v14.i1.pp450-454

Effective XQuery keyword using XML query processing

E. Seshatheri[1], T. Bhuvaneswari[2]
[1]Computer Science and Engineering, Manonmaniam Sundaranar University, India
[2]Department of Computer Applications, Queen Mary's College (Autonomous), India

Article Info

Article history:

Received Aug 9, 2018
Revised Oct 28, 2018
Accepted Jan 21, 2019

Keywords:

Data mining
Linear search algorithm
Tree based association rules (TAR)
Wild card search
XML document

ABSTRACT

The data has structured is determined using the standard is known as XML whereaslarge amount of data has consumed through internet consist of the both structural data format as well as semi structural data format which gets stored and processed whereas XML allow the data of semi-structured and hierarchical data representation not only consist of concept with individual items from various kind of database but also have relationship among data items. The utilized knowledge bed is provided with concise ideas for both structured and semi structured data files, XML document contents and rapid with exact solutions for the queries required at any time. The user can search their resources with the help of queries. Searching the resources with the help of queries is not a simple task, where inaccurate result and complexity would occur. Hence it is not a better way for searching the resources. This paper proposes the query answering system of Linear search usingwild card searchfor extracting the frequent pattern to maximize your search results in library database on XML document to extract the most relevant feeds from the large file directly. It will help the user to find his resources completely.

Corresponding Author:

E. Seshatheri,
Computer Science and Engineering,
Manonmaniam Sundaranar University,
Tirunelveli – 627 012, India.
Email: esesha3@gmail.com

1. INTRODUCTION

XML is a standard for describing how information is structured. It has become a popular format for storing and sharing data across heterogeneous platforms. The representation of XML is flexible and interoperable which is frequently used in application and able to create in various platforms. In order to know the structure of the XML file user needs to know the semantics before querying the document which needs to forming the query. In this research, it is proposed a method for retrieving more efficient more accurate results or the queries made by the users on the XML document [1].The original XML document is interpreted to Modified Tree based Association Rules (TAR) files which were shaped by frequent patterns on the original document. It provides concise representation of Xml document based on the content and structure of Xml file [2]. An approach for Tar used as mined rules which takes RSS feeds as input which provide the more suitable and standard data get stored in the format of XML in both the XML content as well as structure in the document [3]. A novel frequent-pattern tree (FP-tree) structure; our performance study shows that the FP-growth method is efficient and scalable for mining the frequent patterns of both long and short and also order of magnitude is faster than Aprioralgorithm [4]. In [5] proposes the algorithm of Maximal Frequent Itemset (MFI) and improvised frequent pattern tree for association rule mining. This algorithm generates frequent item sets without using candidate sets and Complexity Parameter (CP) trees. In [6] discusses the approach of Tree Based Association Rules (TAR) plays an important role for reducing the retrieval time of query. In [7] in their paper the study highlight the analysis of large scale dataset processing, handling challenges and its systematic review is comprehensive. In [8] has illustrated a method as mine Tree-based association rules in

XML documents whereas this ruleofferdata in XML document with content as well as structure.In this work [9] offers more suitable and standard data hasstored as Xml format in both the structure as well as content of Xml document based on the TAR. In [10] provide concise representation of XML document and also to provide fast, approximate answers to the queries whenever required. In [11] has proposed apriori algorithm is used to finding the usage patterns by modified version called apriori graph. These rules are used to assist for predicting the suitable web pages for the user to visit feasibly in further as a service provider. In [12]. The performance of this method is good in ease XML document but doesn't perform woth XML document with complex and irregular structure in tool is said to be Xquery, the language to identifying and element extraction, attributes from the XML document. In [13] have represented an algorithm namely CMtree Miner which is efficient in computational have determined all nearest and more repeated sub tree in the rooted unordered trees database. The DRYADEPARENT is represented from [14] is the recent quick tree mining algorithm. Hence, it has extracted the sub tree which is embedded with trees maintained with ancestor relationship among the nodes and between the ancestor descendent pairs even in parent-child nodes. This paper proposes the Multinomial Naïve Bayesian (MNB) Classifier, Artificial Neural Network (ANN) and Support Vector Machine (SVM) for mining emotion from text. In our setup, SVM outperformed other classifiers with promising accuracy [15]. This study has illustrated the disadvantages of the above mentioned technique get incorporated and discovery of the new technique. Attribute Oriented Induction High level Emerging Pattern (AOI-HEP) has been proven can mine frequent and similar patterns and the finding AOI-HEP patterns with confidence mining pattern [16]. This research work proposed improved algorithm for stemming Indonesian text. The result of the research shows that the proposed algorithm was the best for Indonesian text processing purpose with score of 0.648 [17].

2. SEMI STRUCTURED DATA TECHNIQUES USING XML QUERY PROCESSING

2.1. Association Rules

The focus of data mining community is based on the advancement technique for general structure extraction from heterogeneous XML data is said to be mining semi-structured data. The default approach from XML data for association rule mining whereas it help to record the document of XML into the model of rational data and finally it get stored in a relational database. Hence, these standard tools get applied using this method in the relational database to perform rule mining. The time consumption and involvement of manual intrusion due to mapping process are available in this method. Therefore, this approach is not appropriate for streaming XML data. XQuery is a language of XML Query which has addresses the capable requirement for querying intelligently the source of XML data. Hence, it is highly adoptable in order to query a wide spectrum of source in XML data which is inclusive of both documents and databases. Thus, the XQuery has managed to perform mining with association rule from XML documents as straight forward approach. The XML Query language has developed the XQuery which is used for usual functions for searching and extracting of both elements and attributes from the XML documents whereas the implementation of complex algorithm is frequently hard in XQuery. The major issue in association rule mining has proposed initially and several algorithm implementations have developed in the literature database. XQuery has used various methods for extracting association rules from ease XML documents. The set of functions from XQuery has developed and get implemented in Apriori algorithm in order to show a better perform only in ease XML document.

2.2. Clustering

In data mining, clustering is one of the important technique used to discover pattern and also for data distribution from the original data. The categorization of World Wide Web documents, array of proteins with same kind of functions, group of genes and the seismic fault detections using catlog of earthquake with the entries which are grouped can able to processed by clustering. These samples have in general that clustering algorithm quality is good then the benefit o recognized higher. The researcher [18] has represented this method in according to two language uses with class description for semi-structural data in automated clustering. The first class language has classes lattice which is created as a model for enveloping the entire dataset. The second class language is the base for interpretation of lattice part in which the user needs to be addressed. One significant XML concepts is Document Type Definition (DTD) whereas the complete advantages are notconsidered in the present application. The researcher [19] has illustrated cluster algorithm for extraction of semi-structural data from the original data whereas clustering novel method with DTDs is presented which can be used for clustering the XML document. This approach has two level cluster approaches namely

a. Clustering the element in DTDs: The first level method with element clustering that provided with dement clusters which has appropriate elements for initiation.

b. Clusters DTDs separately: This is a second level in the entire clustering process in which the DTD clustering has utilized the generalized data.

2.3. Classification

In most of the cases, behavior of classification in XML document is concealed with structure information presented in the document. In some cases, the classifier of informational retrieval has probably progress to be ineffectual for XML documents which has focused on the rule based classifier uses as an efficient tool for data classification. The motive technique is rule based classifiers which have integrated the issues of both classification and associations. The structural rules with their created problems are discussed using XRULE [20] which is to perform the classification task. The structures which are firmly associated to the respective class variables are identified in the training phase. Once the training phase get completed, the testing phase start performing these rule which are utilized to predict the unknown XML document classes whereas the XML documents can be modeled as rooted trees which is ordered and labeled. In [21] provided the form of XML document defines pattern of subtrees in the XML document. In [22] introduced XMINERULE for enrich XQuery with knowledge discovery and datamining capabilities. In [23] described the simple XML document has illustrated the proposed technique that perform good only in simple XML document but not in the complex XML document which has irregular structure.

The limitation of this method is a huge number of rules are produced by rule generator algorithm, and it is very difficult to store the rules, retrieve the related rules, and set the rules. In most cases, XRULE achieves high-classification accuracy by using considerably large number of rules in the classifier, which successively might cause overfitting, particularly for small training datasets.

2.4. Construction of TAR based XQuery Search

The most flexible architecture is XML documents which can be preprocessed whereas the XML pre-processing is done by XML parser. The DOM (Document Object Model) parser is used here which is used to construct the tree from the XML document. According to the XML document, DOM has created a structure of tree within the internal memory whereas DOM can able to store the entire documents in to the internal memory before processing the accessible XML documents which get loaded as an object of XML DOM. It allows the users to traverse the document using wild card approach XML trees, access, insert, update the content, style and structure of the document and also to delete the nodes from the tree. Therefore XML document forms a tree structure. Also the XML document should be validated (i.e) the tags should be started and ended correctly without leaving any tag without its pair.

2.4.1. Frequent Pattern Extraction

The frequent event of datasets with huge amount of collected data is determined in the association rules.The two data items considered are X and Y and it is represented in term ofX ΩY. Support and Confidence are the factors used to measure the association rule whereas the Support is represented with frequency of the set namely X and Y which is available in the data set and Confidence is representedwith conditional probability about finding Y, having got X. The interesting patterns among the subtrees of the given XML document can be identified. In the XML document, the subtree pattern has defined the XML document in the set of TAR whereas the whole document of XML is accessed in order to provide support and confidence standards.

According to TAR mining there are two stage of process involved is mentioned below:

Stage 1: Mining frequent sub trees

Stage 2: Computing interesting rules

The data considered in all the files are merged in term of one XML document, after acquiring the set of files from the proposed model.The step next to this is to obtain the TAR of all the files. Once it is done, the proposed model of CMTree Miner algorithm will give the most frequent feeds of all the files whereas this process is completed then feed search has performed which are provided with filtered result. Searching the resources with the help of queries is not a simple task, where inaccurate result and complexity would occur.

3. PROPOSED METHOD

3.1. Effective XQuery Keyword Search

The major challenge of this research is to rank all these queries based on the individual matches. Tree based association rule where it is mainly query based system. The user can search their resources with the help of queries. Searching the resources with the help of queries is not a simple task, where inaccurate result and complexity would occur. It is not a better way for searching the resources. Hence, this research has focused to resolve the above limitations and also in our research, the above all disadvantages is also

incorporated and determined the novel technique. The association rule has specific mining ideas for providing summarized representation in XML documents have investigated through several proposals either using language like Xquery and advanced technique in XML content or using implementation of linear search algorithm. Therefore, an advanced search technique is wildcard which has used for maximizing the search result in the database. In order to search the represented one or more character in the word, wildcard is the most effective technique whereas the representation of a single character is mentioned in the form of question mark (??) which is very essential while there are several spellings for a word and it has to search for all variation at once.More efficiency because if the user forget the exact resources that the user want. In this case also Wildcard search will help the user to find his resources completely.
For example, Searching for Java would return java.
The forms of wildcard syntax specified by this XML document are:

a. A single period, without any qualifiers: Matches a single arbitrary character.
b. A period immediately followed by a single question mark, "?": Matches either no characters or one character.
c. A period immediately followed by a single asterisk, "*": Matches zero or more characters.
d. A period immediately followed by a single plus sign, "+": Matches one or more characters.
e. A period immediately followed by a sequence of characters that matches the regular expression {[0-9]+,[0-9]+}: Matches a number of characters, where the number is no less than the number represented by the series of digits before the comma, and no greater than the number represented by the series of digits following the comma.

In XQuery, contains wildcard search option consists of multiple search keywords namely *,?, full stop,+ that are alternatively followed by a qualifier. Every wildcard search has matched zero or many characters with a XQuery token in the text are being searched. The number of characters that can be matched depends on the qualifier. This search used to improve the performance and retrieve relevant information from the XML document.

4. CONCLUSION

The proposed model of linear search algorithm is provided with most frequent feeds of all the files while the feed search has been performed by the wildcard based search on XQuery does provided with the filtered result. Therefore, the proposed wildcard search is an advance search technique that can be utilized for maximizingthe search results in library databases with less time consumptions in order to find the resource completely for the users.

REFERENCES

[1] Sashatheri, E., and Dr. Bhuvaneshwari, T., "A Novel Method to Managing Semi Structured Data in Distributed Environment using Modified Tree based Association Rules(TAR)", *Australian Journal of Basic and Applied Sciences,* vol. 9, no. 35, pp. 277-286, 2015.
[2] Vikhe, P. B., and Gunjal, B. L., "Extracting Tree Based Association Rules from XML Document", *International Journal of Emerging Technology and Advanced Engineering,* vol. 3, no. 6, June 2013.
[3] AlfiyaIqbal, A. S., and Sanchika, B., "Frequent Pattern Mining for XML Query- Answering Support", *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* vol. 4, no. 2, July 2014.
[4] Neelesh, S., and Richa, K., "FP-Growth Tree Based Algorithms Analysis: CP-Tree and K Map", *Binary Journal of Data Mining & Networking,* vol.5, pp. 26-29, 2015.
[5] Vandit, A., Mandhani, K., and Dr. Preetham, K., "An Improvised Frequent Pattern Tree Based Association Rule Mining Technique with Mining Frequent Item Sets Algorithm and a Modified Header Table", *International Journal of Data Mining & Knowledge Management Process (IJDKP),* vol. 5, no. 2, 2015.
[6] Shambhu, K. S., et al., "Association Policy for XML Query Answering By Mining Tree", *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 4, no. 3, 2015.
[7] Seshatheri, E., and Dr. Bhuvaneshwari, T., "An Efficient distributed data processing method for smooth environment", *Journal of Engineering and Applied Sciences* vol. 11, no. 8, pp. 1855-1858, 2016.
[8] Swarupa, N. S., "TAR: Algorithm for Mining XML Query Answering", *International Journal of Advanced Research in Computer Science and Software Engineering,* vol. 6, no. 6, 2016.
[9] Poonam, R. M., "Answering XML Query Using Tree Based Association Rule", *IJCSMC,* vol. 6, no. 2, pp.75 – 80, 2017.
[10] Neha, H. N., and Kapil, H., "Data Mining for Intensional Query Answering Using Tree Based Association Rules", *IJEDR,* vol. 4, no. 2, 2016.
[11] Pritish, Y., and Suneetha, K. R., "Modified Apriori Graph Algorithm for Frequent Pattern Mining", *International Conference on Innovations in information Embedded and Communication Systems.* 2016.
[12] Wan, J. W., and Dobbie, G., "Extracting Association Rules fromXML Documents Using XQuery," *Proc. Fifth ACM Int'l WorkshopWeb Information and Data Management,* pp: 94-97, 2003.

[13] Chi, Y., Yang, Y., Xia, Y., and Muntz, R. R., "CMTreeMiner: Mining both Closed and Maximal Frequent Subtrees", *Proc. Eighth Pacific- Asia Conf. Knowledge Discovery and Data Mining,* pp: 63-73, 2004.
[14] Termier, A., Rousset, M., Sebag, M., Ohara, K., Washio, T., and Motoda, H., "DryadeParent, an Efficient and Robust Closed Attribute Tree Mining Algorithm", *IEEE Trans. Knowledge and Data Eng,* vol. 20, no. 3, pp. 300-320,
Mar, 2008.
[15] Muhammad, A. A., and Mahmudul, H. B., "Text to Emotion Extraction Using Supervised Machine Learning Techniques", *TELKOMNIKA,* vol.16, no.3, pp. 1394~1401, 2018.
[16] Harco, L. H. S. W., Agung, T., & Richard, R., "Confidence of AOI-HEP Mining Pattern", *TELKOMNIKA,* vol.16, no.3, pp. 1217-1225, 2018.
[17] Afian, S. R., Aris, T., and Rahmat, T., "Comparison of stemming algorithms and its effect on indonesian text processing", *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol-17, 2018.
[18] Nathalie, P., Marie-Christine, R., & Veronique, V., "Automatic Construction and refinement of a class hierarchy over semi-structured data",
[19] Svetlozar, N., Serge, A., and Rajeev, M., "Extracting schema from semi-structured data",
[20] Abiteboul, S., Buneman, P., Suciu, D., "Data on the Web', Morgan Kaufmann, 2000.
[21] Suganya, I., Velmurugan, N., and Ganeshkumar, P., "XML Query-Answering Support System using AssociationMining Technique", *IEEE conference on ICT.* 2013.
[22] Braga, D., Campi, A., Ceri, S., Klemettinen, M., and Lanzi, P., "Discovering Interesting Information in XMLData with Association Rules", *Proc. ACM SYmp. Applied Computing,* pp. 450-454, 2003.
[23] Wan, J. W. W., and Dobbie, G., "Extracting Association Rules from XML Documents Using XQuery," *Proc. Fifth ACM Int'l Workshop Web Information and Data Management,* pp. 94-97, 2003.