___

Survey on Managing XML Search through Personalization

B. Kurinchi Meenakshi
Dept of CSE,
Bharath University
Chennai, India
kurinchimv@gmail.com

Dr. C. Nalini
Dept of CSE,
Bharath University
Chennai, India
drnalinichidambaram@gmail.com

Abstract - Information Retrieval IR based on keyword on web has become very significant and XML has become the widely used format for sharing of information. The number of resources of information has increased significantly and retrieval of correct data according to user preference may not be achieved efficiently. In order to improve the search of XML documents according to the user requirement and preference we use personalized search based on user preference stored as an XML. The problem in personalized search is in identifying the correct preferences based on the search text. In this paper we have done a survey of papers related to XML search and personalization. From this survey we have come up with a proposed solution, to store the user preferences with keywords and sub preferences as an XML document and related text as keywords. To identify the user preferences based on the query the user preference XML is loaded. The user preference nodes related to the keyword is identified and ranked based ranking function of top k algorithm. The documents will now be listed based on keyword and also based on keyword and preference node combination. The relevance status value of the resulting documents will be identified in both the search and the final search result will be listed by considering the RSV value using the re-ranking strategies.

Keywords – Relevance Status Value, XML, Ranking, top k algorithm, Re ranking, Personalization.

_____*****_____

I. INTRODUCTION

Information search has become very essential and an important part of day today life. There are wide range of information available in different formats. Handling wide range of information in different format will be difficult to handle.XML helps in overcoming this issue as it is the standard and universally accepted format. The XML documents can be retrieved and displayed based on the user search term.

As there are wide collection of information available the resultant set of information which are retrieved may be huge. The user may be interested only in some of the resultant documents out of the huge set of retrieved information. In order to make the user search more effective user personalization has been considered.

The user preferences can be stored and can be used in retrieving the result documents according to the user query. The issue in using user preference for retrieval is that not all user preferences are required in retrieving the result documents .So the issue is in identifying the user preference parameters according to the given query and deriving the weightage of these parameters according to the given query. In this paper have proposed an approach for identifying the user preference parameters based on user search text.

We will discussing on the related work, proposed system and conclusion for implementing personalization in XML search in this paper.

II. RELATED WORK

The works which had been carried out in the areas related to user personalized search have been discussed below.

Luis M. de Campos, Juan M. Fern´andez-Luna, Juan F.

Huete, and Eduardo Vicente-L´opez have proposed an approach for making the user search more effective by using the concept of weighted personalized parameters. The results of original query and expanded query have been compared to identify the documents based on user's interest based on ReRanking approach [1].

Harshal R Aher, Anupkumar Bongale have presented an approach of identifying the XML nodes according to the keyword and ranking them based on score of relevance. The problem of fuzzy typ-ahead search and the efficiency of methods such as LCA Based interactive search and top-k algorithm have been discussed to complete the search and ranking of XML documents more efficiently [2].

Varun Varma Sangaraju has proposed a system of using Adaptive Search for searching and ranking XML based documents. In this system users can select the search algorithm based on its benefits. The search algorithms like Boolean Retrieval algorithm and LCA based algorithm have been discussed along with ranking algorithm [3].

Jovan Pehcevski, Jaap Kamps, Gabriella Kazai, Mounia Lalmas, Paul Ogilvie, Benjamin Piwowarski, and Stephen Robertson have proposed evaluation measures which makes use of score of a document calculated based on highlighted text and full text of the document which will be a value between 0 and 1.This helps to identify the document parts based on relevance and content [4].

Laxman Dethe, Prof. R. M. Goudar, Prof. Sunita Barve have presented the advantages of using Fuzzy search techniques in XML search over Xpath and Xquery and the efficiency of Minimum cost tree and LCA based search algorithms. The user need not know about the XML data when using this search [5].

Pradeep Kumar Reddy Gade, N Prasanna Balaji, U

_____

Sreenivasulu have presented on ranking the XML query results according to the user search intention and relevance using XML TF*IDF and KWSearch algorithm. Using XML TF and XML IDF the confidence level of each node is computed for it to be searched. They have also proposed a ranking scheme based on XML TF and XML IDF to arrive at the hierarchical structure of XML data [6].

TABLE I. STUDY OF RELATED WORK

S.No	Title	Author	ISSN Number/Volume	Technique	Advantage/Disadvantage
1	Using Personalization to Improve XML Retrieval	Luis M. de Campos, Juan M. Fern´andez-Luna, Juan F. Huete, and Eduardo Vicente-L´opez	Vol 26 No.5 May 2014.	Bayesian network model and Reranking strategy	**Advantage** XML documents can be retrieved according to user interest which will make the search more effective **Disadvantage** Delay in document retrieval as the weightage of each parameter has to be calculated and then follow the reranking strategy
2	Create XML Document and Efficient Interactive Keyword Search Technique over XML Data	Harshal R Aher , Anupkumar Bongale	ISSN: 2277 128X/ Volume 4, Issue 5, May 2014	Fuzzy typ-ahead ,top-k Algorithm,LCA-based method ,Minimum Cost Tree	**Advantage** Effective search of XML with relevance ranking
3	Ranking Of XML Documents by Using Adaptive Keyword Search	Varun Varma Sangaraju	Vol. 5 (2) , 2014, 1619-1621/ISSN:0975-9646	Boolean Retrieval algorithm and LCA based algorithm with ranking algorithm like XRank and XReal	**Advantage** User based selection of search algorithm Time saving and increased performance
4	INEX 2007 Evaluation Measures.	Jovan Pehcevski1, Jaap Kamps2, Gabriella Kazai3, Mounia Lalmas4, Paul Ogilvie5,5 Benjamin Piwowarski6, and Stephen Robertson3	Volume 4862, 2008, pp 24-33	INEX 2007 Evaluation Framework	**Advantage** Improves the effectiveness of XML document retrieval based on relevance and best of content.
5	Performance Evolution of XML Data Searching by Using Fuzzy Type a head Search	Laxman Dethe, Prof. R. M. Goudar, Prof. Sunita Barve	Vol. 2, Issue 11, November 2014	Minimum Cost Tree,ELCA,Fuzzy type-ahead	**Advantage** Minimum cost tree: Retrieves a node even if there is no descendant nodes that include all the keyword ELCA based method: Retrieves more relevant than the LCA based method **Disadvantage** LCA-Based interactive search: It is provide low result quality and answers are not relevant
6	An Effective XML Keyword Search with User Search Intention over XML Documents	Pradeep Kumar Reddy Gade, N Prasanna Balaji, U Sreenivasulu	Volume 11 Issue 16 Version 1.0 September 2011	XML TF*IDF and KWSearch algorithm	**Advantage** Ranking of query results according to the user objective and relevance to user search intention

III. PROPOSED SYSTEM

The approach which is followed in identifying the related preferences involves parsing the user preference XML using DOM (Document Object model). Identify the nodes which contain the search text using DOM model. The rank of these nodes will be calculated using the ranking function in top-k algorithm. The ranking function will calculate the rank based on the below calculation

If node n contain keyword ki
The score or relevance of node n and keyword ki is calculated by

$$\text{SCORE1 }(n, ki) = \frac{\ln(1+tf(ki,n))*\ln(idf(ki))}{(1-s)*s*ntl(n)}$$

Where

tf (ki, n)- no of occurrence of ki in sub tree rooted n
idf (ki)- ratio between number of node in XML to number of nodes that contain keyword ki
ntl (n)-length of n/length of nmax
nmax - node with max terms
s -constant set to 0.2

If there are more number of keywords then we first evaluate the relevance between node n and each input keyword, and then the overall score can be calculated by combining the individual relevance scores. We then rank the nodes based on the score.

Now we try to get the XML documents based on two sets of queries.one set will retrieve the documents based on the entered text and the next set will retrieve the documents based on the entered text and also the node text in the ranking order taking one at a time.

For the retrieved XML documents in both the cases the relevance status value has to be identified which can be calculated as ratio of total number of selected text in the document and total text in the document.

The documents matching both the list will be now reranked

considering HRR, SRR and IRR reranking approaches [1].

This will be repeated for the next node in the rank list.

IV. CONCLUSION

From the survey of the papers mentioned in the related work we have identified the approach which will make the retrieval of XML documents more effective by making use of the personalized information. The approach involves calculating the weightage of the user preferences according to the user query by using the ranking function and using those user preferences in retrieving the documents and reranking the documents based on HRR, SRR and IRR.

V. REFERENCE

[1] Luis M. de Campos, Juan M. Fern´andez-Luna, Juan F. Huete, and Eduardo Vicente-L´opez "Using Personalization to Improve XML Retrieval", May 2014,Vol 26 No.5

[2] Harshal R Aher , Anupkumar Bongale "Create XML Document and Efficient Interactive Keyword Search Technique over XML Data",May 2014,ISSN: 2277 128X/ Volume 4, Issue 5

[3] Varun Varma Sangaraju "Ranking Of XML Documents by Using Adaptive Keyword Search",2014,Vol. 5 (2) , 1619-1621/ISSN:0975-9646

[4] Jovan Pehcevski1, Jaap Kamps2, Gabriella Kazai3, Mounia Lalmas4, Paul Ogilvie5,5 Benjamin Piwowarski6 and Stephen Robertson3 "INEX 2007 Evaluation Measures", 2008, Volume 4862, pp 24-33

[5] Laxman Dethe, Prof. R. M. Goudar, Prof. Sunita Barve "Performance Evolution of XML Data Searching by Using Fuzzy Type a head Search",November 2014 ,Vol. 2, Issue 11

[6] Pradeep Kumar Reddy Gade, N Prasanna Balaji, U Sreenivasulu " An Effective XML Keyword Search with User Search Intention over XML Documents",September 2011,Volume 11 Issue 16 Version 1.0