

## Chapter 7

# Probabilistic Ranking Method of XML Fuzzy Query Results

Wei Yan

Liaoning University, China

### ABSTRACT

*Fuzzy query processing for XML database systems is an important issue. Based on the fuzzy set theory, XML fuzzy query can be expressed exploiting fuzzy predicates. To deal with the ranking problem of XML fuzzy query results, this chapter proposes a novel ranking approach. Firstly, according to the workload of XML documents, this chapter speculates how much the users care about each attribute node and assign a corresponding weight to it. Then, a membership degree ranking method, which ranks the fuzzy query results according to corresponding membership degree, is presented. Furthermore, this chapter proposes the probabilistic ranking method, which improves the PIR method. The improved probabilistic ranking method considers the relevance between the nodes specified by fuzzy query and the nodes unspecified by fuzzy query. Finally, top-k ranking algorithm of XML fuzzy query results is presented. The efficiency and effectiveness of the approach are also demonstrated by experimental results.*

### INTRODUCTION

With the rapid development of the World Wide Web, the XML data has been recognized as a standard for data representation and transmission over the Internet. Many data have been produced and transformed into the XML data format, especially in scientific data and Web logs the information is kept in XML. Several query languages have been proposed until XPath and XQuery received a general consensus, becoming the standard query language. Traditional query processing method of XML data requires users to specify the precise queries, and the query processing systems return precise query results to the users. However, in many cases, the users can not express their query intentions accurately. Therefore, the traditional XML query processing systems lack flexibility to the users. The users would like to issue fuzzy query with fuzzy predicates, then XML database system is more convenient to the users.

DOI: 10.4018/978-1-4666-8767-7.ch007

For example, published information of the books is stored in an XML document. The user needs to search the books that were published in year around 2012 as well as the price is less than \$60. In the XML document, the user may request the following fuzzy query:

```
//book[ {@year close to 2012} ] [ {@price at most 60} ]
```

where, fuzzy predicates “close to” and “at most” express users’ fuzzy query intension.

In the practical application, the users don’t know exactly what they are looking for. When searching the XML document, the users often don’t know the XML Schema in detail. The XML data format has the structure and content information. The same data sometimes can be described using different structure. Therefore, the users’ query intensions are usually fuzzy or imprecise. How to extend XML query language as well as provide some flexibility to the users is an important issue. The XPath query language can select XML data nodes effectively through the expression of query traversal, then it able to formulate a query expression that accurately expresses users’ query intensions. This chapter uses fuzzy predicates to expend XPath query language, to make the fuzzy query expression to express users’ fuzzy query intension. Moreover, the users may propose the fuzzy query with fuzzy terms or fuzzy relations. Thus, the extensional fuzzy query expression can help users to improve their interaction with system in order to obtain more query results.

After fuzzy query of XML data, another problem faced by the users will be that how to use efficient ranking method to rank the fuzzy query results. This chapter uses the membership ranking method to rank the fuzzy query results according to their membership degree. For the nodes with the same membership degree, each node specified by fuzzy query is assigned a score according to its desirableness to the user. This chapter also considers the workload information of the XML documents, which is a log of past users’ queries. Based on the workload, the chapter speculates how much the user cares about each node and assigns a corresponding weight to the attribute nodes according to their importance.

This chapter presents an automatic estimative approach via workload as well as data analysis to resolve above problem. Fuzzy query and probabilistic ranking (FQPR) method developed the ranking functions, which are based on probabilistic information retrieval (PIR) model (Chaudhuri *et al.* 2004, 2006). The architecture of FQPR ranking model has a pre-processing component that collects workload statistics to determine the appropriate ranking scores. The extracted ranking scores are materialized in an intermediate knowledge representation layer, to be used later by a query processing component for ranking the fuzzy query results. Although FQPR ranking approach derives these quantities automatically, the architecture allows users or domain experts to tune these quantities further. Therefore, FQPR ranking model may customize the ranking functions for different applications.

The objectives of the chapter are summarized as follows:

- This chapter proposes an XML fuzzy query method, and use fuzzy predicates to expand XPath query language.
- In order to resolve the raking problem of fuzzy query results, the chapter proposes a novel membership degree ranking method, which can rank the fuzzy query results according to their membership degree.
- Based on workload statistics, the chapter presents probabilistic ranking method, which considers the relevance between the nodes specified by fuzzy query and the nodes unspecified by fuzzy query.

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/probabilistic-ranking-method-of-xml-fuzzy-query-results/138697](http://www.igi-global.com/chapter/probabilistic-ranking-method-of-xml-fuzzy-query-results/138697)

## Related Content

---

### Optimal Composition of Services for Intelligent Systems Using TOPSIS

Ashish Sethand Kirti Seth (2021). *International Journal of Information Retrieval Research* (pp. 49-64).  
[www.irma-international.org/article/optimal-composition-of-services-for-intelligent-systems-using-topsis/280526](http://www.irma-international.org/article/optimal-composition-of-services-for-intelligent-systems-using-topsis/280526)

### Decentralized Search and the Clustering Paradox in Large Scale Information Networks

Weimao Ke (2012). *Next Generation Search Engines: Advanced Models for Information Retrieval* (pp. 29-46).  
[www.irma-international.org/chapter/decentralized-search-clustering-paradox-large/64419](http://www.irma-international.org/chapter/decentralized-search-clustering-paradox-large/64419)

### On the Updating of Domain OWL Models at Runtime in Factory Automation Systems

Juha Puttonen, Andrei Lobovand José L. Martinez Lastra (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications* (pp. 1665-1685).  
[www.irma-international.org/chapter/on-the-updating-of-domain-owl-models-at-runtime-in-factory-automation-systems/198619](http://www.irma-international.org/chapter/on-the-updating-of-domain-owl-models-at-runtime-in-factory-automation-systems/198619)

### The Automatic Detection of Diabetes Based on Swarm of Fish

Reda Mohamed Hamou (2018). *Handbook of Research on Biomimicry in Information Retrieval and Knowledge Management* (pp. 154-174).  
[www.irma-international.org/chapter/the-automatic-detection-of-diabetes-based-on-swarm-of-fish/197701](http://www.irma-international.org/chapter/the-automatic-detection-of-diabetes-based-on-swarm-of-fish/197701)

### Generating and Adjusting Web Sub-Graph Displays for Web Navigation

Wei Lai, Maolin Huangand Kang Zhang (2004). *Intelligent Agents for Data Mining and Information Retrieval* (pp. 241-253).  
[www.irma-international.org/chapter/generating-adjusting-web-sub-graph/24167](http://www.irma-international.org/chapter/generating-adjusting-web-sub-graph/24167)