Chapter 6

Fuzzy XQuery:
A Real Implementation

José Ángel Labbad
Universidad Simón Bolívar, Venezuela

Ricardo R. Monascal
Universidad Simón Bolívar, Venezuela

Leonid Tineo
Universidad Simón Bolívar, Venezuela

ABSTRACT

Traditional database systems and languages are very rigid. XML data and query languages are not the exception. Fuzzy set theory is an appropriate tool for solving this problem. In this sense, Fuzzy XQuery was proposed as an extension of the XQUERY standard. This language defines the xs:truth datatype, the xml:truth attribute and allows the definition and use of fuzzy terms in queries. The main goal of this chapter is to show a high coupling implementation of Fuzzy XQuery within eXist-db, an open source XML DBMS. This extension strategy could also be used with other similar tools. This chapter also presents a statistical performance analysis of the extended fuzzy query engine using the XMark benchmark with user defined fuzzy terms. The study presents promising results.

INTRODUCTION

The Web has been become a popular tool for services such as travel agencies, shopping stores, car rental, encyclopedia, and so on. Thus, the Web plays an essential role in many online companies and it has made available an exorbitant amount of data from several websites. Many of these websites contain engines that query data from different existing sites. Most of these websites use XML (Extensible Markup Language) format (W3C, 2008) to interchange data, because it is the standard for this purpose.

XML documents may be queried through declarative query languages such as XPath (W3C, 2014) and XQuery (W3C, 2010). Both languages are XML-centric, i.e., their data model and type system are

based on XML. XQuery is an extension of XPath conceived to integrate multiple XML sources and it is the W3C standard language for XML data. Several database engines support XQuery either as native language or as an alternative language.

As several authors have adverted, XQuery is not accurate to handle search criteria based on user's preferences (Buche et al 2006) (Calmès et al 2007) (Goncalves and Tineo 2007) (Thomsom and Radhameni 2011) (Ueng 2012). XQuery is not able to discriminate query answers according to user's criteria. This weakness is often referred to as rigidity problem of query languages and it is due to query conditions are based on Boolean logic (Bordogna and Psaila, 2008).

As a motivating example, suppose researchers who want to attend a conference. They want to query a travel company website searching for the best flight trip according their own preferences. Someone would like a trip that were very cheap and made few connections. Another person might prefer a direct flight whose destination is a near city reaching the conference city by train.

Preference criteria in this example involve linguistic terms of vague nature. They are the natural language terms: very, cheap, few, and near. In general, semantics of such terms is context-dependent and may vary according to user's preference. For giving answers to user requests, in this case, many optional trips might exist, it would be helpful to discriminate them in terms of compatibility with user's criteria.

Fuzzy sets theory is a possible theoretical solution to this kind of needs. System might allow defining user's criteria and ranking query answers using a membership function; a membership function quantifies the satisfaction degree of each answer with respect to user's criteria and induces a total order of the dataset.

In order to give a solution to described problem, some proposals had arisen (Buche et al 2006) (Gaurav and Alhajj 2006) (Calmès et al 2007) (Goncalves and Tineo 2007) (Campi et al 2009) (Thomsom and Radhameni 2011) (Jin, Y. and Veerappan 2010) (Ma et al 2010) (Goncalves and Tineo 2010) (Ueng 2012) (Panic et al 2014). In particular, in a previous work Fuzzy XQuery has been defined (Goncalves and Tineo, 2010). At present time, fuzzy logic extensions introduced by Fuzzy XQuery are not included in the standard definition. Some efforts have been made in implementing such features, but resulting products are not wide available and there is still work to do.

This chapter shows the development of an implementation of Fuzzy XQuery, a language for fuzzy queries over XML. Fuzzy XQuery was proposed by Goncalves and Tineo (2010). Usual query languages fail at expressing user preferences and context sensitive search conditions. They are often very rigid because there based on Boolean logic. This implies that elements are either completely included or not included at all in a query's result set. This problem often keeps users from easily expressing their preferences in a natural language, which would help them obtain better results adapted to their needs. XQuery suffers of this problem as other query languages do. Fuzzy sets theory gives a mathematical and computational framework that allows the definition of fuzzy terms. This has been proposed as a solution for the rigidity of query languages.

This chapter addresses an implementation of Fuzzy XQuery, with three main objectives in mind. First, showing that the implementation is possible, such that running the queries in an appropriate way will produce the desired results. Second, checking if there is any mistake or inconsistency in the language definition. Third, conducting several performance analyses, checking that performance is not significantly affected by any additional computations needed for supporting fuzzy queries and to showing that, in the implemented extension, queries without fuzzy terms maintains the same performance as can be observed in the original DBMS.

Related Content

Knowledge Discovery From Massive Data Streams
Sushil Kumar Narang, Sushil Kumarand Vishal Verma (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications (pp. 1508-1534).*
www.irma-international.org/chapter/knowledge-discovery-from-massive-data-streams/198612

Predicting Users' Acceptance of E-Library from the Perspective of Technology Acceptance Model
Adeyinka Tella (2013). *Modern Library Technologies for Data Storage, Retrieval, and Use (pp. 240-251).*
www.irma-international.org/chapter/predicting-users-acceptance-library-perspective/73780

Life Insurance-Based Recommendation System for Effective Information Computing
Asha Rani, Kavita Tanejaand Harmunish Taneja (2021). *International Journal of Information Retrieval Research (pp. 1-14).*
www.irma-international.org/article/life-insurance-based-recommendation-system-for-effective-information-computing/274037

Delay Optimization Using Genetic Algorithm at the Road Intersection
Bharti Sharmaand Sachin Kumar (2019). *International Journal of Information Retrieval Research (pp. 1-10).*
www.irma-international.org/article/delay-optimization-using-genetic-algorithm-at-the-road-intersection/222764

XML Documents Normalization Using GN-DTD
Zurinahni Zainoland Bing Wang (2013). *Information Retrieval Methods for Multidisciplinary Applications (pp. 54-77).*
www.irma-international.org/chapter/xml-documents-normalization-using-dtd/75901