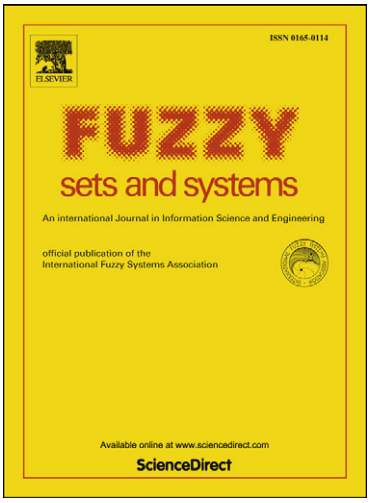
Accepted Manuscript



Modeling fuzzy data with XML: a survey

Zongmin Ma, Li Yan

PII: S0165-0114(15)00446-7
DOI: http://dx.doi.org/10.1016/j.fss.2015.09.016
Reference: FSS 6911

To appear in: *Fuzzy Sets and Systems*

Received date: 31 January 2015
Revised date: 12 August 2015
Accepted date: 22 September 2015

Please cite this article in press as: Z. Ma, L. Yan, Modeling fuzzy data with XML: a survey, *Fuzzy Sets and Systems* (2015), http://dx.doi.org/10.1016/j.fss.2015.09.016

Modeling Fuzzy Data with XML: A Survey

Zongmin Ma* and Li Yan
College of Computer Science & Technology, Nanjing University of Aeronautics and Astronautics
Nanjing, 211106, China
** e-mail: zongminma@gmail.com*

Abstract. Uncertain information extensively exists in data and knowledge intensive applications, where fuzzy data play an import role in nature. Fuzzy set theory has been extensively applied to extend various database models and resulted in numerous contributions. This paper concentrates on a crucial issue in fuzzy data management: fuzzy data modeling in XML. An up-to-date overview of the current state of the art in fuzzy XML data modeling is provided in the paper. The paper serves as identifying possible research opportunities in the area of fuzzy XML data management in addition to providing a generic overview of the approaches proposed to modeling fuzzy XML data.

***Keywords*:** XML; fuzzy sets; fuzzy XML data; data modeling

1. Introduction

One of the major areas of database research has been the continuous effort to enrich existing database models with a more extensive collection of semantic concepts. Database models have developed from hierarchical and network database models to the relational database model. Also to process complex objects and semantic relationships in many data- and knowledge-intensive applications, some non-traditional data models have been proposed for databases such as the object-oriented database model and the object-relational database model. Moreover, with the prompt development of the Internet and the popularity of Web-based applications, the requirement of managing information based on the Web has attracted much attention both from academia and industry. Being the de-facto standard for data representation and exchange over the Web, XML (eXtensible Markup Language) allows the easy development of applications that exchange data over the Web (Bray, Paoli and Sperberg-McQueen, 1998). This creates a set of data management requirements involving XML. XML and related standards have been extensively applied in many business, service, and multimedia applications. As a result, a large volume of data is managed today directly in XML format.

While traditional data models can provide efficient data management capabilities, they often suffer from some inadequacy of necessary semantics. One of these inadequacies can be generalized as the inability to handle imprecise and uncertain information. In real-world applications, information is often imperfect. One of the semantic needs not adequately addressed by the traditional data models is that of uncertainty. Traditional data models assume that the models are a correct reflection of the world and further assume that the stored data is known, accurate and complete. It is rarely the case in real life that all or most of these assumptions are met.

Management of uncertain data typically involves two primary technical issues: storage and queries. These two issues are actually closely related. Efficient querying of uncertain data is supported by the storage structure. Uncertain data modeling provides the infrastructure for uncertain data management and uncertain data querying is one major goal of uncertain data management. In order to represent and manipulate uncertain data, two major foundations, which are *probability theory* and *fuzzy set theory*, have been developed and applied to extend various database models (Parsons, 1996). Probabilistic database models developed in the literature mainly include the probabilistic relational database model (e.g., (Barbara,