

Studies in Fuzziness and Soft Computing

Volume 311

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

For further volumes:
<http://www.springer.com/series/2941>

About this Series

The series “Studies in Fuzziness and Soft Computing” contains publications on various topics in the area of soft computing, which include fuzzy sets, rough sets, neural networks, evolutionary computation, probabilistic and evidential reasoning, multi-valued logic, and related fields. The publications within “Studies in Fuzziness and Soft Computing” are primarily monographs and edited volumes. They cover significant recent developments in the field, both of a foundational and applicable character. An important feature of the series is its short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Li Yan · Zongmin Ma · Fu Zhang

Fuzzy XML Data Management

Li Yan
School of Software
Northeastern University
Shenyang
People's Republic of China

Fu Zhang
College of Information Science and
Engineering
Northeastern University
Shenyang
People's Republic of China

Zongmin Ma
College of Information Science and
Engineering
Northeastern University
Shenyang
People's Republic of China

ISSN 1434-9922
ISBN 978-3-642-44898-0
DOI 10.1007/978-3-642-44899-7
Springer Heidelberg New York Dordrecht London

ISSN 1860-0808 (electronic)
ISBN 978-3-642-44899-7 (eBook)

Library of Congress Control Number: 2013954019

© Springer-Verlag Berlin Heidelberg 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The Web is a huge information resource depository all around the world and the huge amount of information on the Web is getting larger and larger every day. Nowadays the Web is the most important means for people to acquire and publish information. Against this context, it has become very crucial to exchange and share data on the Web. Being the de facto standard for data representation and exchange over the Web, Extensible Markup Language (XML) has emerged and extensively applied in many business, service, and multimedia applications. As a result, a large volume of data is managed today directly in XML format.

XML and related standards allow the easy development of applications that exchange data over the Web. This creates a new set of data management requirements involving XML, such as the needs of constructing, storing, querying, reasoning, and integrating XML documents. For the purpose of XML data management, it is necessary to integrate XML and databases. Various database models, including relational, object-oriented, and object-relational database models as well as conceptual data models, have been used for mapping to and from XML model so that XML model can be extracted from the database models or/and reengineered into the database models. Through reengineering XML model into the relational or object-oriented database model, XML documents can be stored in databases for XML data processing based on database technologies. Through extracting XML model from the relational or object-oriented database model, XML documents can automatically be constructed from databases for database data exchange and share on the Web. Note that XML lacks sufficient power in modeling real-world data and their complex interrelationships in semantics, and the conceptual data models with powerful data abstraction contain clear and rich semantics and do not have data type limitation. Through reengineering XML model into the conceptual data models, XML document integration can be carried out based on the conceptual data models. Through extracting XML model from the conceptual data models, XML documents can conceptually be designed with the conceptual data models.

With the wide and in-depth utilization of XML in diverse application domains, some particularities of data management in concrete applications emerge, which challenge current XML technology. In data- and knowledge-intensive applications, one of the challenges can be generalized as the need to handle imprecise and uncertain information in XML data management. Imprecise and uncertain data can

be found, for example, in the integration of data sources and data generation with nontraditional means (e.g., automatic information extraction and data acquirement by sensor and RFID). So it is crucial for Web-based intelligent information systems to explicitly represent and process imprecise and uncertain XML data.

Fuzzy logic has been applied in a large number and in a wide variety of applications and has been a crucial means of implementing machine intelligence. So, in order to bridge the gap between human-understandable soft logic and machine-readable hard logic, fuzzy logic cannot be ignored because none of the usual logical requirements can be guaranteed: there is no centrally defined format for data, no guarantee of truth for assertions made, and no guarantee for consistency. Fuzzy logic has been introduced into databases for fuzzy data management. It can be believed that fuzzy logic can play an important and positive role in XML data management. Currently the researches of fuzzy logic in XML data management are attracting increased attention.

This book goes to great depth concerning the fast growing topic of technologies and approaches of fuzzy XML data management. The topics of this book include representation of fuzzy XML, query of fuzzy XML, fuzzy database models, extraction of fuzzy XML from fuzzy database models, reengineering of fuzzy XML into fuzzy database models, and reasoning of fuzzy XML. Concerning the representation of fuzzy XML, the fuzziness in XML documents, fuzzy XML representation model, and fuzzy XML algebraic operations are discussed. Concerning the query of fuzzy XML, querying fuzzy XML with AND, OR, and NOT predicates is proposed, respectively, and building the index on fuzzy XML query is investigated. Concerning the fuzzy database models, three kinds of fuzzy database models are introduced, which are the fuzzy UML data models, fuzzy relational database models, and fuzzy object-oriented database models. Concerning the extraction of fuzzy XML, extracting fuzzy XML from the fuzzy UML data models, fuzzy relational database models, and fuzzy object-oriented database models is proposed, respectively. Concerning the reengineering of fuzzy XML, reengineering fuzzy XML into the fuzzy UML data models, fuzzy relational database models, and fuzzy object-oriented database models is presented, respectively. Concerning the reasoning of fuzzy XML, reasoning on fuzzy XML with fuzzy Description Logic and fuzzy ontology are investigated.

This book aims to provide a single record of current research in the fuzzy data management with XML. The objective of the book is to provide the state-of-the-art information to researchers, practitioners, and graduate students of the Web intelligence and at the same time serve the data and knowledge engineering professional faced with nontraditional applications that make the application of conventional approaches difficult or impossible. Researchers, graduate students, and information technology professionals interested in XML and fuzzy data processing will find this book a starting point and a reference for their study, research, and development.

We would like to acknowledge all of the researchers in the area of databases, XML, and fuzzy databases. Based on both their publications and the many discussions with some of them, their influence on this book is profound. The materials

in this book are the outgrowth of research conducted by the authors in recent years. The initial research work was supported by the *National Natural Science Foundation of China* (60873010, 61370075, 61073139, 61202260), and in part by the *Program for New Century Excellent Talents in University* (NCET-05-0288). We are grateful for the financial support from the *National Natural Science Foundation of China* and the *Ministry of Education of China* through research grant funds. Additionally, the assistance and facilities of Northeastern University, China, are deemed important and highly appreciated. Special thanks go to Janusz Kacprzyk, the series editor of *Studies in Fuzziness and Soft Computing*, and Thomas Ditzinger, the Senior Editor of *Applied Sciences and Engineering* of Springer-Verlag, for their advice and help to propose, prepare, and publish this book. This book would not be completed without the support from them.

Shenyang, September 2013

Li Yan
Zongmin Ma
Fu Zhang

Contents

1	Databases and XML for Data Management	1
1.1	Introduction	1
1.2	Logical Database Models	3
1.2.1	The Relational Database Model	4
1.2.2	The Nested Relational Database Model	11
1.2.3	The Object-Oriented Database Model	12
1.3	Conceptual Data Models	14
1.3.1	Entity-Relationship and Enhanced ER Models	15
1.3.2	UML Class Model	19
1.4	XML Models	22
1.4.1	XML Documents	22
1.4.2	XML DTD Constructs	23
1.4.3	XML Databases	26
1.5	Summary	27
	References	28
2	Fuzzy Sets and Fuzzy Database Models	31
2.1	Introduction	31
2.2	Imperfect Information and Fuzzy Sets	32
2.2.1	Imperfect Information	33
2.2.2	Fuzzy Sets	34
2.3	Fuzzy UML Data Models	36
2.3.1	Fuzzy Class	37
2.3.2	Fuzzy Association	38
2.3.3	Fuzzy Generalization	39
2.3.4	Fuzzy Aggregation	42
2.3.5	Fuzzy Dependency	44
2.3.6	An Example of Fuzzy UML Data Model	45
2.4	Fuzzy Relational Database Models	46
2.4.1	Fuzzy Relational Database Models	47
2.4.2	Fuzzy Relational Algebra Operations	48

2.5	Fuzzy Object-Oriented Database Models	56
2.5.1	Fuzzy Objects.	57
2.5.2	Fuzzy Classes.	57
2.5.3	Fuzzy Object-Class Relationships	58
2.5.4	Fuzzy Inheritance Hierarchies.	63
2.5.5	Algebraic Operations in Fuzzy Object-Oriented Databases.	65
2.6	Summary	77
	References	78
3	Fuzzy XML Data Models	83
3.1	Introduction	83
3.2	Fuzziness in XML Documents	84
3.3	Fuzzy XML Representation Models and Formalizations.	87
3.3.1	Representation of Fuzzy Data in the XML Document	87
3.3.2	Fuzzy DTD	88
3.3.3	Fuzzy XML Schema	90
3.3.4	Formalization of Fuzzy XML Models	94
3.4	Fuzzy XML Algebraic Operations.	97
3.5	Summary	99
	References	99
4	Fuzzy XML Queries and Index	101
4.1	Introduction	101
4.2	Fuzzy XML Queries with Predicates	103
4.2.1	Fuzzy Extended Dewey Encoding.	103
4.2.2	Fuzzy XML Twig Query with AND-Logic	105
4.2.3	Fuzzy XML Twig Query with AND/OR-Logic.	112
4.2.4	Fuzzy XML Twig Query with NOT-Logic.	118
4.3	Fuzzy XML Queries with Index	122
4.3.1	Fuzzy XML Twig Pattern Query Matching	122
4.3.2	Fuzzy XML Twig Pattern Query with Index	126
4.4	Summary	131
	References	131
5	Fuzzy XML Extraction from Fuzzy Database Models	133
5.1	Introduction	133
5.2	Fuzzy XML Extraction from Fuzzy UML Data Models	134
5.2.1	Extraction of Fuzzy Classes	134
5.2.2	Extraction of Fuzzy Generalizations	137
5.2.3	Extraction of Fuzzy Associations	138
5.3	Fuzzy XML Extraction from Fuzzy Relational Database Models	139

5.4	Fuzzy XML Extraction from Fuzzy Object-Oriented Database Models	145
5.4.1	Extraction of Fuzzy Classes	146
5.4.2	Extraction of Fuzzy Relationships	148
5.4.3	Extraction of Fuzzy Inheritance Hierarchies	149
5.5	Summary	150
	References	151
6	Reengineering Fuzzy XML into Fuzzy Database Models	153
6.1	Introduction	153
6.2	Reengineering Fuzzy XML into Fuzzy UML Data Models	155
6.3	Reengineering Fuzzy XML into Fuzzy Relational Database Models	159
6.3.1	DTD Tree and Mapping to the Relational Database Schema	159
6.3.2	Mapping the Fuzzy XML Model into the Fuzzy Relational Database Model.	161
6.4	Reengineering Fuzzy XML into Fuzzy Object-Oriented Database Models	163
6.5	Summary	166
	References	166
7	Fuzzy XML Reasoning.	169
7.1	Introduction	169
7.2	Fuzzy XML Reasoning with Fuzzy Description Logic	172
7.2.1	Fuzzy Description Logic $f\text{-ALCQ}_{\text{wf-reg}}$	172
7.2.2	Representing Fuzzy XML with Fuzzy Description Logic $f\text{-ALCQ}_{\text{wf-reg}}$	179
7.2.3	Reasoning on Fuzzy XML with Fuzzy Description Logic $f\text{-ALCQ}_{\text{wf-reg}}$	185
7.3	Fuzzy XML Reasoning with Fuzzy Ontology	189
7.3.1	Representing Fuzzy XML with Fuzzy Ontology	190
7.3.2	Reasoning on Fuzzy XML with Fuzzy Ontology	202
7.4	Summary	206
	References	206
	Index	209