

# Lecture Notes in Computer Science

2824

Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

**Springer**

*Berlin*

*Heidelberg*

*New York*

*Hong Kong*

*London*

*Milan*

*Paris*

*Tokyo*

Zohra Bellahsène Akmal B. Chaudhri  
Erhard Rahm Michael Rys  
Rainer Unland (Eds.)

# Database and XML Technologies

First International XML Database Symposium, XSym 2003  
Berlin, Germany, September 8, 2003  
Proceedings



Springer

## Volume Editors

Zohra Bellahsène  
LIRMM UMR 5506 CNRS/Université Montpellier II  
161 Rue Ada, 34392 Montpellier, France  
E-mail: bella@lirmm.fr

Akmal B. Chaudhri  
IBM developerWorks  
6 New Square, Bedfont Lakes, Feltham, Middlesex TW14 8HA, UK  
E-mail: akmal.b.chaudhri@uk.ibm.com

Erhard Rahm  
University of Leipzig  
Augustusplatz 10-11, 04109 Leipzig, Germany  
E-mail: rahm@informatik.uni-leipzig.de

Michael Rys  
Microsoft Corporation  
One Microsoft Way, Redmond, WA 98052, USA  
E-mail: rys@acm.org, mrys@microsoft.com

Rainer Unland  
University of Duisburg-Essen  
Institute for Computer Science and Business Information Systems  
Schützenbahn 70, 45117 Essen, Germany  
E-mail: UnlandR@informatik.uni-essen.de

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek  
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;  
detailed bibliographic data is available in the Internet at <<http://dnb.ddb.de>>.

CR Subject Classification (1998): H.2, H.3, H.4, D.2, C.2.4

ISSN 0302-9743

ISBN 3-540-20055-X Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York  
a member of BertelsmannSpringer Science+Business Media GmbH

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2003  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin GmbH  
Printed on acid-free paper SPIN: 10953624 06/3142 5 4 3 2 1 0

# PREFACE

The Extensible Markup Language (XML) is playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. The database community is interested in XML because it can be used to represent a variety of data formats originating in different kinds of data repositories while providing structure and the possibility to add type information.

The theme of this symposium is the combination of database and XML technologies. Today, we see growing interest in using these technologies together for many Web-based and database-centric applications. XML is being used to publish data from database systems on the Web by providing input to content generators for Web pages, and database systems are increasingly being used to store and query XML data, often by handling queries issued over the Internet. As database systems increasingly start talking to each other over the Web, there is a fast-growing interest in using XML as the standard exchange format for distributed query processing. As a result, many relational database systems export data as XML documents, import data from XML documents, provide query and update capabilities for XML data. In addition, so-called native XML database and integration systems are appearing on the database market, and it's claimed that they are especially tailored to store, maintain and easily access XML documents.

The first XML Database Symposium, XSym 2003, is a new forum on the combination of database and XML technologies. It is built on several previous XML, Web and database-related workshops that were held at the CAiSE 2002, EDBT 2002, NODe 2002 and VLDB 2002 conferences. The goal of this symposium is to provide a high-quality platform for the presentation and discussion of new research results and system developments. It is targeted at scientists, practitioners, vendors, and users of XML and database technologies.

The call-for-papers attracted 65 submissions from all over the world. After a careful reviewing process, the international program committee accepted 18 high-quality papers of particular relevance and quality. The selected contributions cover a wide range of exciting topics, in particular XML query processing, stream processing, XML-relational mappings, index structures, change management, and new prototypes. Another highlight of the symposium was the keynote by Mike Franklin, University of Berkeley.

As editors of this volume, we would like to thank once again all program committee members and all the external referees who gave up their valuable time to review the papers and helped in putting together an exciting program. We would also like to thank the invited speaker, authors and other individuals, without whom this

symposium would not have been possible. Moreover, our thanks go out to the local organizing committee who fulfilled with a lot of patience all our wishes. Finally, we would like to thank Alfred Hofmann from Springer-Verlag for his friendly cooperation and help in putting this volume together.

July 2003

Montpellier, Bedford Lakes, Redmond, Leipzig, Essen,

Zohra Bellahsene (General Chair)

Akmal Chaudhri (Program Committee Co-chair)

Michael Rys (Program Committee Co-chair)

Erhard Rahm (Publicity and Communications Chair)

Rainer Unland (Publications Chair)

## Program Committee

Bernd Amann, CNAM and INRIA (France)  
Valeria De Antonellis, Politecnico di Milano (Italy)  
Zohra Bellahsene, LIRMM (France)  
Elisa Bertino, University of Milan (Italy)  
Timo Böhme, University of Leipzig (Germany)  
Akmal B. Chaudhri, IBM developerWorks (USA)  
Sophie Cluet, INRIA (France)  
Istvan Cseri, Microsoft (USA)  
Gillian Dobbie, University of Auckland (New Zealand)  
Mary F. Fernandez, AT&T Research (USA)  
Daniela Florescu, BEA (USA)  
Irin Fundulaki, Bell Labs/Lucent Technologies (USA)  
Udo Kelter, University of Siegen (Germany)  
Donald Kossmann, Technical University of Munich (Germany)  
Mong Li Lee, National University of Singapore (Singapore)  
Eng Wah Lee, Gintic (Singapore)  
Stuart Madnick, MIT (USA)  
Ioana Manolescu, INRIA (France)  
Jim Melton, Oracle (USA)  
Alberto Mendelzon, University of Toronto (Canada)  
Laurent Mignet, University of Toronto (Canada)  
Tova Milo, Tel Aviv University (Israel)  
Guido Moerkotte, Universität Mannheim (Germany)  
Allen Moulton, MIT (USA)  
M. Tamer Özsu, University of Waterloo (Canada)  
Shankar Pal, Microsoft (USA)  
Erhard Rahm, University of Leipzig (Germany)  
Michael Rys, Microsoft (USA)  
Jérôme Siméon, Bell Labs (USA)  
Zahir Tari, RMIT (Australia)  
Frank Tompa, University of Waterloo (Canada)  
Hiroshi Tsuji, Osaka Prefecture University (Japan)  
Can Türker, Swiss Federal Institute of Technology, Zurich (Switzerland)  
Rainer Unland, University of Essen (Germany)  
Agnes Voisard, Fraunhofer ISST and Freie Universität Berlin (Germany)  
Osamu Yoshie, Waseda University (Japan)  
Jeffrey Xu Yu, Chinese University of Hong Kong (Hong Kong)

## External Reviewers

Sharon Adler, IBM (USA)  
Marcelo Arenas, University of Toronto (Canada)  
Denilson Barbosa, University of Toronto (Canada)  
Omar Benjelloun, INRIA (France)  
David Bianchini, University of Brescia (Italy)  
Ronald Bourret, Independent Consultant (USA)  
Lei Chen, University of Waterloo (Canada)  
Grégory Cobena, INRIA (France)  
David DeHaan, University of Waterloo (Canada)  
Hai Hong Do, University of Leipzig (Germany)  
Rainer Eckstein, Humboldt-Universität zu Berlin (Germany)  
Elena Ferrari, Università dell'Insubria, Como (Italy)  
Leo Giakoumakis, Microsoft (USA)  
Giovanna Guerrini, Università di Pisa (Italy)  
Jim Kleewein, IBM (USA)  
Sailesh Krishnamurthy, IBM (USA)  
Allen Luniewski, IBM (USA)  
Ingo Macherius, Infonbyte (Germany)  
Susan Malaika, IBM (USA)  
Florent Masegla, INRIA (France)  
Michele Melchiori, University of Brescia (Italy)  
Rosa Meo, Università di Torino (Italy)  
Benjamin Nguyen, INRIA-FUTURS (France)  
Yong Piao, University of Siegen (Germany)  
Awais Rashid, Lancaster University (UK)  
Mark Roantree, Dublin City University (Ireland)  
Kenneth Salem, University of Waterloo (Canada)  
Torsten Schlieder, Freie Universität Berlin (Germany)  
Bob Schloss, IBM (USA)  
Soumitra Sengupta, Microsoft (USA)  
Xuerong Tang, University of Waterloo (Canada)  
Alejandro Vaisman, University of Toronto (Canada)  
Pierangelo Veltri, INRIA (France)  
Brian Vickery, IBM (USA)  
Sabrina De Capitani di Vimercati, Università degli Studi di Milano (Italy)  
Florian Waas, Microsoft (USA)  
Norbert Weissenberg, Fraunhofer ISST (Germany)  
Liang-Huai Yang, National University of Singapore (Singapore)  
Hui Zhang, University of Waterloo (Canada)  
Ning Zhang, University of Waterloo (Canada)  
Hongwei Zhu, MIT (USA)

# Table of Contents

## XML–Relational DBMS

XML-to-SQL Query Translation Literature: The State of the Art and Open Problems .....	1
<i>Rajasekar Krishnamurthy, Raghav Kaushik, Jeffrey F. Naughton</i>	
Searching for Efficient XML-to-Relational Mappings .....	19
<i>Maya Ramanath, Juliana Freire, Jayant R. Haritsa, Prasan Roy</i>	
A Virtual XML Database Engine for Relational Databases .....	37
<i>Chengfei Liu, Millist W. Vincent, Jixue Liu, Minyi Guo</i>	

## XML Query Processing

Cursor Management for XML Data .....	52
<i>Ning Li, Joshua Hui, Hui-I Hsiao, Parag Tijare</i>	
Three Cases for Query Decorrelation in XQuery .....	70
<i>Norman May, Sven Helmer, Guido Moerkotte</i>	
A DTD Graph Based XPath Query Subsumption Test .....	85
<i>Stefan Böttcher, Rita Steinmetz</i>	

## Systems and Tools

PowerDB-XML: A Platform for Data-Centric and Document-Centric XML Processing .....	100
<i>Torsten Grabs, Hans-Jörg Schek</i>	
An XML Repository Manager for Software Maintenance and Adaptation .....	118
<i>Elaine Isnard, Radu Bercaru, Alexandra Galatescu, Vladimir Florian, Laura Costea, Dan Conescu</i>	
XViz: A Tool for Visualizing XPath Expressions .....	134
<i>Ben Handy, Dan Suciu</i>	

## XML Access Structures

Tree Signatures for XML Querying and Navigation .....	149
<i>Pavel Zezula, Giuseppe Amato, Franca Debole, Fausto Rabitti</i>	
The <i>Collection Index</i> to Support Complex Approximate Queries .....	164
<i>Paolo Ciaccia, Wilma Penzo</i>	

Finding ID Attributes in XML Documents .....	180
<i>Denilson Barbosa, Alberto Mendelzon</i>	

## Stream Processing and Updates

XML Stream Processing Quality .....	195
<i>Sven Schmidt, Rainer Gemulla, Wolfgang Lehner</i>	

Representing Changes in XML Documents Using Dimensions .....	208
<i>Manolis Gergatsoulis, Yannis Stavrakas</i>	

Updating XQuery Views Published over Relational Data: A Round-Trip Case Study.....	223
<i>Ling Wang, Mukesh Mulchandani, Elke A. Rundensteiner</i>	

## Design Issues

Repairs and Consistent Answers for XML Data with Functional Dependencies .....	238
<i>S. Flesca, F. Furfaro, S. Greco, E. Zumpano</i>	

A Redundancy Free 4NF for XML .....	254
<i>Millist W. Vincent, Jixue Liu, Chengfei Liu</i>	

Supporting XML Security Models Using Relational Databases: A Vision .....	267
<i>Dongwon Lee, Wang-Chien Lee, Peng Liu</i>	

Author Index .....	283
--------------------	-----