

*Commenced Publication in 1973*

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

## Editorial Board

David Hutchison

*Lancaster University, UK*

Takeo Kanade

*Carnegie Mellon University, Pittsburgh, PA, USA*

Josef Kittler

*University of Surrey, Guildford, UK*

Jon M. Kleinberg

*Cornell University, Ithaca, NY, USA*

Friedemann Mattern

*ETH Zurich, Switzerland*

John C. Mitchell

*Stanford University, CA, USA*

Moni Naor

*Weizmann Institute of Science, Rehovot, Israel*

Oscar Nierstrasz

*University of Bern, Switzerland*

C. Pandu Rangan

*Indian Institute of Technology, Madras, India*

Bernhard Steffen

*University of Dortmund, Germany*

Madhu Sudan

*Massachusetts Institute of Technology, MA, USA*

Demetri Terzopoulos

*New York University, NY, USA*

Doug Tygar

*University of California, Berkeley, CA, USA*

Moshe Y. Vardi

*Rice University, Houston, TX, USA*

Gerhard Weikum

*Max-Planck Institute of Computer Science, Saarbruecken, Germany*

Zohra Bellahsène Tova Milo  
Michael Rys Dan Suciu  
Rainer Unland (Eds.)

# Database and XML Technologies

Second International XML Database Symposium, XSym 2004  
Toronto, Canada, August 29-30, 2004  
Proceedings



Springer

## Volume Editors

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

Tova Milo  
Tel Aviv University, Computer Science Department  
Tel Aviv 69978, Israel  
E-mail: milo@post.tau.ac.il

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

Dan Suciu  
University of Washington, Computer Science and Engineering  
Box 352350, Seattle, WA 98195-2350, USA  
E-mail: suciu@cs.washington.edu

Rainer Unland  
University of Duisburg-Essen  
Institute for Computer Science and Business Informations Systems (ICB)  
Practical Computer Science, especially Data  
Management Systems and Knowledge Representation  
Schützenbahn 70, 45117 Essen, Germany  
E-mail: UnlandR@informatik.uni-essen.de

Library of Congress Control Number: 2004110717

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

ISSN 0302-9743  
ISBN 3-540-22969-8 Springer 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. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media  
springeronline.com

© Springer-Verlag Berlin Heidelberg 2004  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Protago-TeX-Production GmbH  
Printed on acid-free paper SPIN: 11317616 06/3142 5 4 3 2 1 0

# Preface

Modern database systems enhance the capabilities of traditional database systems by their ability to handle any kind of data, including text, image, audio, and video. Today, database systems are particularly relevant to the Web, as they can provide input to content generators for Web pages, and can handle queries issued over the Internet.

The eXtensible Markup Language (XML) is used in applications running the gamut from content management through publishing to Web services and e-commerce. It is used as the universal communication language for exchanging music and graphics as well as purchase orders and technical documentation.

As database systems increasingly talk to each other over the Web, there is a fast-growing desire to use XML as the standard exchange format. As a result, many relational database systems can export data as XML documents and import data from XML documents and provide query and update capabilities for XML data. In addition, so called native XML database and integration systems are appearing on the database market, whose claim is to be especially tailored to storing, maintaining, and easily accessing XML documents.

After the huge success of the first XML Database Symposium (XSym 2003) last year in Berlin (already then in conjunction with VLDB) it was decided to establish this symposium as an annual event that is supposed to take place as an integral part of VLDB. 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 about 60 submissions from all over the world. After a careful reviewing process, the international program committee accepted 15 high-quality papers of particular relevance and quality. The selected contributions cover a wide range of exciting topics, in particular XQuery processing, searching, ranking, and mapping XML documents, XML constraints checking and correcting, and XML processing. An exciting highlight of the symposium was the keynote by Mary Fernandez from AT&T Research. Her talk "*Building an Extensible XQuery Engine: Experiences with Galax*" was a perfect start for this symposium.

As the editors of this volume, we would like to thank all the program committee members and external reviewers who sacrificed their valuable time to review the papers and helped in putting together a truly convincing program.

We would also like to thank the organizers of VLDB 2004 (and of XSym), especially Ilju Kiringa, and the general chair of VLDB 2004, John Mylopoulos, without whom this symposium would not have been possible. Our special thanks also go to Akmal Chaudhri. He not only was always available when a helping hand was needed but he also did an excellent job with implementing and maintaining the XSym homepage. Finally,

we would like to thank Alfred Hofmann from Springer for his friendly cooperation and help in putting this volume together.

Tel Aviv and Orsay, Seattle, Essen, Montpellier, Redmond

July 2004

Tova Milo (Program Committee Co-chair)

Dan Suciu (Program Committee Co-chair)

Rainer Unland (General Chair)

Zohra Bellahsène (General Co-chair)

Michael Rys (General Co-Chair)

## Members of the International Programme Committee

Bernd Amann, CNAM Paris (France)  
Sihem Amer-Yahia, AT&T Research (USA)  
Michael Benedikt, Bell Labs (USA)  
Phil Bernstein, Microsoft Research (USA)  
Elisa Bertino, University of Milan (Italy)  
Angela Bonifati, CNR (Italy)  
Vassilis Christophides, ICS-FORTH & University of Crete (Greece)  
Gregory Cobena, Xyleme (France)  
Mariano Consens, University of Toronto (Canada)  
Alin Deutsch, University of California at San Diego (USA)  
Mary Fernandez, AT&T Research (USA)  
Dana Florescu, BEA (USA)  
Juliana Freire, OGI School of Science and Engineering (USA)  
H.V. Jagadish, University of Michigan (USA)  
Christoph Koch, Technische Universität Wien (Austria)  
Alberto Laender, Universidade Federal de Minas Gerais (Brazil)  
Laks Lakshmanan, University of British Columbia (Canada)  
Yossi Matias, Tel Aviv University (Israel) and HyperRoll (USA)  
Giansalvatore Mecca, Università della Basilicata (Italy)  
Hamid Pirahesh, IBM (USA)  
Evaggelia Pitoura, University of Ioannina (Greece)  
Neoklis (Alkis) Polyzotis, University of California at Santa Cruz (USA)  
Philippe Pucheral, INRIA (France)  
Elke A. Rundensteiner, Worcester Polytechnic Institute (USA)  
Arnaud Sahuguet, Bell Laboratories – Lucent Technologies (USA)  
Oded Shmueli, Technion (Israel)  
Val Tannen, University of Pennsylvania (USA)  
Vasilis Vassalos, New York University (USA)  
Stratis Viglas, University of Edinburgh (UK)  
Masatoshi Yoshikawa, Nagoya University (Japan)

## External Reviewers

Marcelo Arenas

Paolo Atzeni

Attila Barta

Marco Brambilla

Paolo Cappellari

Yi Chen

Valter Crescenzi

Francois Dang-Ngoc

Natasha Drukh

Maged El-Sayed

Vladimir Gapeyev

Georges Gardarin

Ashish Gupta

Vagelis Hristidis

Dao-Dinh Kha

Georgia Koloniari

Juliano Lage

Wei Luo

Marco Mesiti

Flavio Rizzolo

Stefanie Scherzinger

Toshiyuki Shimizu

Bernhard Stegmaier

Hong Su

Liyong Sui

Alberto Trombetta

Ling Wang

Nuwee Wiwatwattana

Yu Xu

Lingzhi Zhang

Yifeng Zheng

# Table of Contents

## Keynote Speech

Building an Extensible XQuery Engine: Experiences with Galax . . . . .	1
<i>Mary Fernández, Jérôme Siméon</i>	

## XQuery Processing

A Light but Formal Introduction to XQuery . . . . .	5
<i>Jan Hidders, Jan Paredaens, Roel Vercammen, Serge Demeyer</i>	
XML Query Processing Using a Schema-Based Numbering Scheme . . . . .	21
<i>Dao Dinh Kha, Masatoshi Yoshikawa</i>	
Implementing Memoization in a Streaming XQuery Processor . . . . .	35
<i>Yanlei Diao, Daniela Florescu, Donald Kossmann, Michael J. Carey, Michael J. Franklin</i>	

## Searching, Ranking, and Mapping XML Documents

XQuery Processing with Relevance Ranking . . . . .	51
<i>Leonidas Fegaras</i>	
Information Preservation in XML-to-Relational Mappings . . . . .	66
<i>Denilson Barbosa, Juliana Freire, Alberto O. Mendelzon</i>	
A Signature-Based Approach for Efficient Relationship Search on XML Data Collections . . . . .	82
<i>Giuseppe Amato, Franca Debole, Fausto Rabitti, Pasquale Savino, Pavel Zezula</i>	

## XML Constraints Checking and Correcting

Correctors for XML Data . . . . .	97
<i>Utsav Boobna, Michel de Rougemont</i>	
Incremental Constraint Checking for XML Documents . . . . .	112
<i>Maria Adriana Abrão, Béatrice Bouchou, Mirian Halfeld Ferrari, Dominique Laurent, Martin A. Musicante</i>	
EReX: A Conceptual Model for XML . . . . .	128
<i>Murali Mani</i>	



## XML Processing

A Runtime System for XML Transformations in Java .....	143
<i>Aske Simon Christensen, Christian Kirkegaard, Anders Møller</i>	
Teaching Relational Optimizers About XML Processing.....	158
<i>Siheem Amer-Yahia, Yannis Kotidis, Divesh Srivastava</i>	
Adjustable Transaction Isolation in XML Database Management Systems .....	173
<i>Michael P. Haustein, Theo Härder</i>	

## Clustering, Indexing, Statistics

Fractional XSKETCH Synopses for XML Databases .....	189
<i>Natasha Drukh, Neoklis Polyzotis, Minos Garofalakis, Yossi Matias</i>	
Flexible Workload-Aware Clustering of XML Documents.....	204
<i>Rajesh Bordawekar, Oded Shmueli</i>	
XIST: An XML Index Selection Tool .....	219
<i>Kanda Runapongsa, Jignesh M. Patel, Rajesh Bordawekar, Sriram Padmanabhan</i>	

<b>Author Index</b> .....	235
---------------------------	-----