

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

Stéphane Bressan Stefano Ceri
Ela Hunt Zachary G. Ives
Zohra Bellahsene Michael Rys
Rainer Unland (Eds.)

Database and XML Technologies

Third International XML Database Symposium, XSym 2005
Trondheim, Norway, August 28-29, 2005
Proceedings

Volume Editors

Stéphane Bressan

National University of Singapore, Department of Computer Science

School of Computing

3 Science drive 2, 117543 Singapore, Republic of Singapore

E-mail: steph@nus.edu.sg

Stefano Ceri

Politecnico di Milano, Dipartimento di Elettronica e Informazione

Via Ponzio, 34/5, 20133 Milano, Italy

E-mail: ceri@elet.polimi.it

Ela Hunt

University of Glasgow, Department of Computing Science

Lilybank Gardens 8-17, Glasgow G12 8QQ, UK

E-mail: ela@dcs.gla.ac.uk

Zachary G. Ives

University of Pennsylvania, Computer and Information Science Department

3330 Walnut Street, Philadelphia, PA 19104-6389, USA

E-mail: zives@atcis.upenn.edu

Zohra Bellahsène

LIRMM UMR 5506 CNRS/Université Montpellier II

161 Rue Ada, 34392 Montpellier, France

E-mail: bella@lirmm.fr

Michael Rys

Microsoft Corporation, One Microsoft Way, Redmond, WA 98052, USA

E-mail: 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

Library of Congress Control Number: 2005931472

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

ISSN 0302-9743

ISBN-10 3-540-28583-0 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-28583-0 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 2005

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik

Printed on acid-free paper SPIN: 11547273 06/3142 5 4 3 2 1 0

Preface

This year marks an exciting time in the XML-database space: XQuery is moving closer to becoming a full W3C Recommendation, and the “Big 3” database vendors (IBM, Oracle, Microsoft) are expected to release XQuery support in their relational DBMSs, joining a number of existing open source and commercial products. Thus, we are very pleased to feature an industrial paper (describing the XML-specific features of Microsoft SQL Server) as well as 14 research papers. XSym’s focus this year was on building XML repositories, and papers discussed the following topics: indexing support for the evaluation of XPath and XQuery; benchmarks and algorithms for XQuery and XPath evaluation; algorithms for constraint satisfaction checking, information extraction, and subtree matching; and applications of XML in information systems.

This year, XSym also coordinated its efforts with the Database and Programming Languages Symposium, DBPL 2005. The resulting program included not only presentations of the papers in this proceedings, but also a joint DBPL-XSym keynote talk by Giuseppe Castagna, developer of the CDuce language for XML processing, and a joint panel on open XML research problems and challenges.

The organizers would like to express their gratitude to the XSym Program Committee and external reviewers for their efforts in providing very thorough evaluations of the submitted papers under significant time constraints and to Microsoft for their sponsorship and for the use of the Microsoft Conference Management Toolkit. We would also like to thank Gavin Bierman and Christoph Koch, the organizers of DBPL, for their efforts and their willingness to coordinate with us.

These proceedings are dedicated to Alberto Mendelzon who sadly passed away this year. As a strong supporter of and active contributor to this symposium series he will always remain in our memory.

Singapore, Milan, Montpellier, Glasgow
Philadelphia, Essen, Redmond

July 2005

Stéphane Bressan
Stefano Ceri
Zohra Bellahsene
Ela Hunt
Zachary Ives
Rainer Unland
Michael Rys

General Chair

Stéphane Bressan, National University of Singapore (Singapore)

General Co-chair

Stefano Ceri, Politecnico di Milano (Italy)

Organizing Chair

Zohra Bellahsene, LIRMM (France)

Program Committee Chairs

Ela Hunt, University of Glasgow (UK)

Zachary Ives, University of Pennsylvania (USA)

Proceedings

Rainer Unland, University of Duisburg-Essen (Germany)

Sponsorship

Michael Rys, Microsoft (USA)

Communications

Akmal B. Chaudhri, IBM ISV & Developer Relations (USA)

International Programme Committee

Ashraf Aboulnaga, University of Waterloo (Canada)
Sihem Amer-Yahia, AT&T Research (USA)
Ricardo Baeza-Yates, Universidad de Chile (Chile)
Veronique Benzaken, LRI – Université Paris XI (France)
Tiziana Catarci, University of Roma “La Sapienza” (Italy)
Yi Chen, University of Pennsylvania (USA)
Giovanna Guerrini, Università di Pisa (Italy)
Ashish Kumar Gupta, University of Washington (USA)
Raghav Kaushik, Microsoft Research (USA)
Qiong Luo, Hong Kong University of Science and Technology (China)
Ioana Manolescu, INRIA (France)
Peter McBrien, Imperial College London (UK)
Guido Moerkotte, University of Mannheim (Germany)
Felix Naumann, Humboldt University Berlin (Germany)
Werner Nutt, Heriot-Watt University (UK)
Beng Chin Ooi, National University of Singapore (Singapore)
M. Tamer Ozsu, University of Waterloo (Canada)
Tadeusz Pankowski, Poznan University of Technology (Poland)
Alexandra Poulovassilis, Birkbeck College, University of London (UK)
Prakash Ramanan, Wichita State University (USA)
Elke A. Rundensteiner, Worcester Polytechnic Institute (USA)
Arnaud Sahuguet, Bell Laboratories – Lucent Technologies (USA)
Monica Scannapieco, University of Roma “La Sapienza” (Italy)
Jayavel Shanmugasundaram, Cornell University (USA)
Jerome Simeon, IBM Research (USA)
Wang-Chiew Tan, University of California, Santa Cruz (USA)
Yannis Velegrakis, AT&T Research (USA)
Stratis Viglas, University of Edinburgh (UK)
Peter Wood, Birkbeck College, University of London (UK)
Yuqing Melanie Wu, Indiana University (USA)
Jun Yang, Duke University (USA)
Jeffrey Xu Yu, Chinese University of Hong Kong (China)

External Reviewers

Andrei Arion	LRI – Université Paris XI (France)
Patrick Bosc	LRI – Université Paris XI (France)
Chavdar Botev	Cornell University (USA)
Giuseppe Castagna	LRI – Université Paris XI (France)
Laura Chiticariu	University of California, Santa Cruz (USA)
David DeHaan	University of Waterloo (Canada)
Maged El-Sayed	Worcester Polytechnic Institute (USA)
Fan Yang	Cornell University (USA)
Mirian Halfeld-Ferrari	LRI – Université Paris XI (France)
Ming Jiang	Worcester Polytechnic Institute (USA)
Ming Lee	Worcester Polytechnic Institute (USA)
Diego Milano	University of Roma (Italy)
	University of Edinburgh (UK)
Feng Shao	Cornell University (USA)
Frank Tompa	University of Waterloo (Canada)
Song Wang	Worcester Polytechnic Institute (USA)
Rui Yang	National University of Singapore (Singapore)
Ning Zhang	University of Waterloo (Canada)

Table of Contents

Invited Talk (Shared with DBPL)

Patterns and Types for Querying XML Documents	1
<i>Giuseppe Castagna</i>	

Constraints and Views

Checking Functional Dependency Satisfaction in XML	4
<i>Millist W. Vincent and Jixue Liu</i>	
A Theoretic Framework for Answering XPath Queries Using Views	18
<i>Jian Tang and Shuigeng Zhou</i>	

Labeling and Path Evaluation

A Path-Based Labeling Scheme for Efficient Structural Join	34
<i>Hanyu Li, Mong Li Lee, and Wynne Hsu</i>	
The BIRD Numbering Scheme for XML and Tree Databases – Deciding and Reconstructing Tree Relations Using Efficient Arithmetic Operations	49
<i>Felix Weigel, Klaus U. Schulz, and Holger Meuss</i>	
Efficient Handling of Positional Predicates Within XML Query Processing	68
<i>Zografoula Vagena, Nick Koudas, Divesh Srivastava, and Vassilis J. Tsotras</i>	

Indexing

Relational Index Support for XPath Axes	84
<i>Leo Yuen and Chung Keung Poon</i>	
Supporting XPath Axes with Relational Databases Using a Proxy Index ..	99
<i>Olli Luoma</i>	
An Extended Preorder Index for Optimising XPath Expressions	114
<i>Martin F. O'Connor, Zohra Bellahsene, and Mark Roantree</i>	

Benchmarking and Query Processing

XPathMark: An XPath Benchmark for the XMark Generated Data	129
<i>Massimo Franceschet</i>	

MemBeR: A Micro-benchmark Repository for XQuery	144
<i>Loredana Afanasiev, Ioana Manolescu, and Philippe Michiels</i>	
Main Memory Implementations for Binary Grouping	162
<i>Norman May and Guido Moerkotte</i>	

Documents and Biometrical Applications

Logic Wrappers and XSLT Transformations for Tuples Extraction from HTML	177
<i>Costin Bădică and Amelia Bădică</i>	
Approximate Subtree Identification in Heterogeneous XML Documents Collections	192
<i>Ismael Sanz, Marco Mesiti, Giovanna Guerrini, and Rafael Berlanga Llavori</i>	
A Framework for XML-Based Integration of Data, Visualization and Analysis in a Biomedical Domain	207
<i>Nathan Bales, James Brinkley, E. Sally Lee, Shobhit Mathur, Christopher Re, and Dan Suciu</i>	

Industrial Session

Optimizing Runtime XML Processing in Relational Databases	222
<i>Eugene Kogan, Gideon Schaller, Michael Rys, Hanh Huynh Huu, and Babu Krishnaswamy</i>	

Panel (Together with DBPL)

Panel: “ <i>Whither XML, ca. 2005?</i> ”	237
--	-----

Author Index	239
--------------------	-----