Preface

Since its first edition in 2003, the XML Database Symposium series (XSym) has been a forum for academics, practitioners, users and vendors, allowing all to discuss the use of and synergy between database management systems and XML. The symposia have provided many opportunities for timely discussions on a broad range of topics pertaining to the theory and practice of XML data management and its applications. XSym 2010 continued this XSym tradition with a program consisting of 11 papers and a keynote shared with the 36th International Conference on Very Large Data Bases (VLDB 2010). We received 20 paper submissions, out of which 8 papers were accepted as full papers, and 3 as short papers. Each submitted paper underwent a rigorous and careful review by four referees.

The contributions in these proceedings are a fine sample of the current research in XML query processing, including XPath satisfiability, approximate joins, pattern matching, linear index construction for trees, dynamic labeling, and XQuery update translation based on schema. The papers focus on recent advances in detecting functional dependencies, modeling complex XML twig pattern output, promoting semantics capability of XML keys, and searchable compression of Microsoft office documents. In addition, we include a paper that shares lessons learned from real XML database development.

The organizers would like to express their gratitude to the authors, for submitting their work, and to the Program Committee, for providing very thorough evaluations of the submitted papers and for the discussions that followed under significant time constraints. We also would like to thank the invited keynote speaker, Prof. M. Tamer Özsu, for the challenging and thought-provoking contribution. Finally, we are also grateful to Microsoft and Michael Rys for their generous sponsorship, Andrei Voronkov and other contributors for the EasyChair conference management system, and the local organizers for their efforts in making XSym 2010 a pleasant and successful event. Finally, we would also like to thank Alfred Hofmann and his great team from Springer for their support and cooperation in putting this volume together.

July 2010

Mong Li Lee
Jeffrey Xu Yu
Zohra Bellahsene
Rainer Unland
Organization

Steering Committee

Zohra Bellahsene  LIRMM-CNRS/University Montpellier 2 (France)
Ela Hunt       University of Strathclyde (UK)
Michael Rys   Microsoft (USA)
Rainer Unland University of Duisburg-Essen (Germany)

Program Co-chairs

Mong Li Lee  National University of Singapore (Singapore)
Jeffrey Xu Yu Chinese University of Hong Kong (China)

International Program Committee

Bernd Amann  Université Paris 6 (France)
Veronique Benzaken Université Paris-Sud (France)
Sourav S. Bhowmick Nanyang Technological University (Singapore)
Stéphane Bressan National University of Singapore (Singapore)
Chee Yong Chan National University of Singapore (Singapore)
Yi Chen      Arizona State University (USA)
Minos Garofalakis Technical University of Crete (Greece)
Giorgio Ghelli Università di Pisa (Italy)
Torsten Grust Universität Tübingen (Germany)
Giovanna Guerrini Università di Genova (Italy)
H.V. Jagadish University of Michigan (USA)
Yaron Kanza  Technion Israel Institute of Technology (Israel)
RagHAV Kaushik Microsoft Research (USA)
Jiaheng Lu    Renmin University of China (China)
Murali Mani   Worcester Polytechnic Institute (USA)
Peter McBrien Imperial College - London (UK)
Atsuyuki Morishima University of Tsukuba (Japan)
Tadeusz Pankowski Poznan University of Technology (Poland)
Prakash Ramanan Wichita State University (USA)
Pierre Senellart Télécom ParisTech (France)
Jerome Simeon  IBM Research (USA)
Martin Theobald Max-Planck-Institut für Informatik (Germany)
Wee Hyong Tok Microsoft (China)
VIII Organization

Vasilis Vassalos  Athens University of Economics and Business (Greece)
Yuqing Wu  Indiana University (USA)
Ni Yuan  IBM (China)
Xiaofang Zhou  University of Queensland (Australia)

External Reviewers

Pantelis Aravogliadis
Federico Cavalieri
Vassilis Christophides
Pierre Genevès
Françoise Gire
Mirian Halfeld-Ferrari.
# Table of Contents

## Keynote Address

Distributed XML Query Processing (Extended Abstract) .................... 1

*M. Tamer Özsu and Patrick Kling*

## XML Query Processing

Approximate Joins for XML Using g-String .......................... 3

*Fei Li, Hongzhi Wang, Cheng Zhang, Liang Hao, Jianzhong Li, and Hong Gao*

Linear Computation of the Maximum Simultaneous Forward and Backward Bisimulation for Node-Labeled Trees .......................... 18

*Nils Grimsmo, Truls Amundsen Bjørklund, and Magnus Lie Hetland*

Extending the Tractability Results on XPath Satisfiability with Sibling Axes ......................................................... 33

*Yasunori Ishihara, Shogo Shimizu, and Toru Fujiwara*

Extending XQuery with a Pattern Matching Facility ................. 48

*Peter M. Fischer, Aayush Garg, and Kyumars Sheykh Esmaili*

## XML Update and Applications

A Schema-Based Translation of XQuery Updates ..................... 58

*Leonidas Fegaras*

EBSL: Supporting Deleted Node Label Reuse in XML .................. 73

*Martin F. O’Connor and Mark Roantree*

Lessons Learned from DB2 pureXML Applications: A Practitioner’s Perspective .......................................................... 88

*Matthias Nicola*

Searchable Compression of Office Documents by XML Schema Subtraction ............................................................. 103

*Stefan Böttcher, Rita Hartel, and Christian Messinger*

## XML Modeling

Fast Detection of Functional Dependencies in XML Data .......... 113

*Hang Shi, Toshiyuki Amagasa, and Hiroyuki Kitagawa*
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP+Output: Modeling Complex Output Information in XML Twig</td>
<td>128</td>
</tr>
<tr>
<td>Huayu Wu, Tok Wang Ling, and Gillian Dobbie</td>
<td></td>
</tr>
<tr>
<td>Promoting the Semantic Capability of XML Keys</td>
<td>144</td>
</tr>
<tr>
<td>Flavio Ferrarotti, Sven Hartmann, Sebastian Link, and Jing Wang</td>
<td></td>
</tr>
<tr>
<td><strong>Author Index</strong></td>
<td>155</td>
</tr>
</tbody>
</table>