Volume Editors

Heng Tao Shen
The University of Queensland, Brisbane, QLD, Australia
E-mail: shenht@itee.uq.edu.au

Jian Pei
Simon Fraser University, Burnaby BC, Canada,
E-mail: jpei@cs.sfu.ca

M. Tamer Özsu
University of Waterloo, Canada
E-mail: tamer.ozsu@cs.uwaterloo.ca

Lei Zou
Peking University, China
E-mail: zoulei@icst.pku.edu.cn

Jiaheng Lu
Renmin University of China, China
E-mail: jiahenglu@gmail.com

Tok-Wang Ling
National University of Singapore, Singapore
E-mail: lingtw@comp.nus.edu.sg

Ge Yu
North-East University, Shenyang, China
E-mail: yuge@mail.neu.edu.cn

Yi Zhuang
Zhejiang University, Hangzhou, 310058, China
E-mail: zhuang@zjgsu.edu.cn

Jie Shao
University of Melbourne, Australia
E-mail: jsh@unimelb.edu.au

Library of Congress Control Number: 2010937439


LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-642-16719-5 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, re-printing, 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.com
© Springer-Verlag Berlin Heidelberg 2010
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180
WAIM 2010 Workshop Chair’s Message

WAIM is a leading international conference for researchers, practitioners, developers and users to share and exchange cutting-edge ideas, results, experience, techniques and tools in connection with all aspects of Web data management. The conference invites original research and industrial papers on the theory, design and implementation of Web-based information systems, as well as proposals for demonstrations, tutorials and panels. Previous WAIM conferences were held in Shanghai (2000), Xian (2001), Beijing (2002), Chengdu (2003), Dalian (2004), Hangzhou (2005), Hong Kong (2006), Huangshan (2007), Zhangjiajie (2008) and Suzhou (2009). Along with the main conference, WAIM workshops are intended to provide an international group of researchers with forum for the discussion and exchange of research results related to the conference topics.

This WAIM 2010 workshop volume comprises papers from three workshops, which were

1. The First International Workshop on Graph Database
2. International Workshop on Advanced Techniques on XML Data Management, and
3. The Second International Workshop on Web-Based Contents Management Technologies

The contents of these three workshops were selected from a public call-for-proposals process. The workshop organizers put a tremendous amount of effort into soliciting and selecting research papers with a balance of high quality and new ideas and new applications. We asked all workshops to follow a rigid paper selection process, including the procedure to ensure that any Program Committee members (including workshop Program Committee Chairs) be excluded from the paper review process of any papers they are involved with. A requirement about the overall paper acceptance ratio was also imposed to all the workshops.

We are very grateful to the main conference organizers. We would also like to take this opportunity to thank all workshop organizers and Program Committee members for their great effort in putting together the workshop program of WAIM 2010.

May 2010

Heng Tao Shen
Jian Pei
The First International Workshop on Graph Database (IWGD 2010) Chairs’ Message

The growing popularity of graph databases has generated interesting data management problems, such as indexing techniques, query algorithms and graph mining. This workshop focused on issues related to graph databases. IWGD 2010 was held in conjunction with WAIM 2010 in JiuZhai Valley, China. IWGD 2010 aimed at bringing together researchers in different fields related to graph databases who have common interests in interdisciplinary research. The workshop provided a forum where researchers and practitioners could share and exchange their knowledge and experience.

April 2010

M. Tamer Özsu
Lei Zou
This was the First International Workshop on XML Data Management. The workshop focused on the convergence of database technology with XML technology, and brought together academics, practitioners, users and vendors to discuss the use and synergy between these technologies.

This workshop attracted 18 submissions from many different schools and countries. The Program Committee accepted 10 papers, among which there are 6 long papers and 4 short papers. The topics of accepted papers include XML applications in semantic Web, XML data storage and indexing, XML query languages and optimization and so on. These proceedings will serve as a valuable reference for XML data management researchers and developers.

The paper “Effective XML Keyword Search Through Valid Lowest Information Unit” mainly discusses concepts and usages of the proposed concepts of LIU and VLIU, which can be used to resolve the problem that LCA-semantics may be incomplete or redundant in the results of keyword search in some cases. Ying Lou et al. first present related work, then present the concept of “information unit” and then introduce the notion of VLIU. Finally, an approach based on VLIU is also outlined in this paper.

The paper “Reducing Redundancy of Xpath Query over Networks by Transmitting XML Views” studies the problem of answering Xpath queries over networks. Xu et al. presented analysis on how to minimize communication cost by rewriting queries. This paper proposes keeping a balance between the network traffic and the complexity of the query, which is a good point.

In their paper “Structure and Content Similarity for Clustering XML Documents,” Zhang et al. proposed a new model to represent the structure and content information in XML documents. Based on the model, they define similarity measures that can be used to cluster XML documents.

In the paper “pq-Hash: An Efficient Method for Approximate XML Joins,” the authors propose a data structure called pq-array to support searching for approximate tree set matching. In addition to that, an efficient algorithm, pq-hash, is proposed to avoid nested loop joins where possible.

In their paper “SeCCX: Semantics-Based Fine Granular Concurrency Control for XML Data,” Rong et al. provide a new locking protocol, named SeCCX, for the concurrent transactions of XML data. First, the authors introduce the SeCCX, including its lock modes, lock compatibility matrix, protocol rules and conflict detections and then the authors analyze the isolation of SeCCX.
In the paper “XRCJ: Supporting Keyword Search in XML and Relation Co-occurrence,” Zhang et al. propose a new method that combines XML and RDB in the same platform.

In “Building Web Application with XQuery,” the authors explore a XQuery-based framework for developing Web applications. This paper proposes that the suggested framework can simplify Web applications development and provides a flexible and scalable architecture.

“Functional Dependencies for XML” studies how to express the functional dependency for XML data based on the current standard XML schema language, i.e. XML Schema. In particular, the authors shows how to support complex values such as list and set, and propose the “match tree” to judge the value equality for complex typed elements, which is further used to decide the satisfaction of the given functional dependency constraints.

In “TwigLinkedList: Improvement of TwigList,” the authors introduce the concept of “level list,” a linked list, into TwigList to check the parent-child relationship between node and solve the out-of-orderness problem. The paper also conducts several tests to evaluate its performance.

The paper “Compression Algorithms for Structural Query Results on XML Data” provides a compression technique for the results of an XML query. Wang et al. propose to compress XML data in order to reduce XML query-result sizes. At the end, this paper extends bitmap indexes for compression.

These papers cover a variety of topics. We believe that they will provide researchers and developers with a brief glimpse into this exciting new technology, specifically from the perspective of XML databases. We hope you enjoy reading the papers.

April 2010

Jiaheng Lu
Tok Wang Ling
Ge Yu
The Web is ever-increasing in size and involves quite a broad range of technologies such as databases, XML and digital contents, data and Web mining, Web services, the Semantic Web and ontology, information retrieval, and others. The Second International Workshop on Web-Based Contents Management Technologies (WCMT 2010) was held in conjunction with the WAIM 2010 conference and invited original research contributions on Web and Internet content technologies. WCMT 2010 aimed at bringing together researchers in different fields related to Web information processing who have a common interest in interdisciplinary research. The workshop provided a forum where researchers and practitioners could share and exchange their knowledge and experience.

April 2010

Yi Zhuang  
Jie Shao
The First International Workshop on Graph Database (IWGD 2010)

Honorary Chair

Jianguo Xiao  
Peking University, China

Program Co-chairs

M. Tamer Özsu  
University of Waterloo, Canada
Lei Zou  
Peking University, China

Program Committee

K. Singh Ambuj  
University of California at Santa Barbara, USA
Gutierrez Claudio  
Universidad de Chile, Chile
Zhao Dongyan  
Peking University, China
Francis Ilyas Ihab  
University of Waterloo, Canada
Cheng James  
Nanyang Technological University, Singapore
Xu Yu Jeffery  
The Chinese University of Hong Kong, China
Cheng Jiefeng  
The University of Hong Kong, China
Zaki Mohammed  
Rensselaer Polytechnic Institute, USA
Varadarajan Ramakrishna  
University of Wisconsin-Madison, USA
Giugno Rosalba  
University of Catania, Italy
Jin Wei  
North Dakota State University, USA
Li Wen-Syan  
SAP Technology Lab, China
Lian Xiang  
Hong Kong University of Science and Technology, China
Wan Xiaojun  
Peking University, China
Yan Xifeng  
University of California at Santa Barbara, USA
Xiao Yanghua  
Fudan University, China
Zhuang Yi  
Zhejiang Gongshang University, China
Peng Zhiyong  
Wuhan University, China
International Workshop on Advanced Techniques on XML Data Management (XMLDM 2010)

Program Co-chairs

Tok Wang Ling National University of Singapore, Singapore
Ge Yu North-East University, China
Jiaheng Lu Renmin University of China, China

Program Committee

Bao Zhifeng National University of Singapore, Singapore
Bressan Stephane National University of Singapore (NUS), Singapore
Boncz Peter Centrum Wiskunde & Informatica, The Netherlands
Yong Chan Chee National University of Singapore, Singapore
Du Xiaoyong Renmin University of China, China
Feng Jianhua Tsinghua University, China
Gao Jun Peking University, China
Kitsuregawa Masaru Tokyo University, Japan
Li Guoliang Tsinghua University, China
Lin Xuemin University of New South Wales, Australia
Meng Xiaofeng Renmin University of China, China
Xu Liang National University of Singapore, Singapore
Xu Yu Jeffrey The Chinese University of Hong Kong, China
Vagena Zografoula Microsoft Research, USA
Wang Bin North-East University, China
Wang XiaoLing Fudan University, China
Wang Hongzhi Harbin Institute of Technology, China
Wood Peter University of London, UK
Zhou Aoying East China Normal University, China
Zhou Yonghuan University of Southern Denmark, Denmark
Zhang Rui University of Melbourne, Australia
Zhang Xiao Renmin University of China, China
The Second International Workshop on Web-Based Contents Management Technologies (WCMT 2010)

General Co-chairs
Dickson K.W. Chiu Dickson Computer System, China
Hua Hu Hangzhou Dianzi University, China

Program Co-chairs
Zhuang Yi Zhejiang Gongshang University, China
Shao Jie University of Melbourne, Australia

Program Committee
Yang Jun Google, USA
Qian Weining East China Normal University, China
Kalnis Panos National University of Singapore, Singapore
Li Cuiping Renmin University, China
Chen Yi Arizona State University, USA
Wu Weili University of Texas at Dallas, USA
Wang Wei University of New South Wales, Australia
Yu Yi New Jersey Institute of Technology, USA
Xu Jianliang Hong Kong Baptist University, China
Chi-Wing Wong Raymond Hong Kong University of Science and Technology, China
Hu Haiyang Hangzhou Dianzi University, China
Gao Yunjun Zhejiang University, China
Zou Lei Peking University, China
Huang Zi University of Queensland, Australia
Wu Ziang Nanjing University of Science and Technology, China
He Bingsheng Microsoft Research Asia, China
Cong Gao Aalborg University, Denmark
Mehrotra Sharad University of California, Irvine, USA
Li Feifei Florida State University, USA
Shen Hong University of Adelaide, Australia
Choi Byron Hong Kong Baptist University, China
Wang Wei University of New South Wales, Australia
Zhang Jia Northern Illinois University, USA
Jeung Hoyoung EPFL, Switzerland
Zhou Mingqi East China Normal University, China
Zhang Jun Nanyang Technological University, Singapore
He Jing Victoria University, Australia
Liu Xiaoyan University of Melbourne, Australia
Zhou Xiangmin CSIRO, Australia
# Table of Contents

## The First International Workshop on Graph Database (IWGD 2010)

- Mining Graphs with Constraints on Symmetry and Diameter ............. 1  
  *Natalia Vanetik*

- Graph Partitioning Strategies for Efficient BFS in Shared-Nothing Parallel Systems .................................................. 13  
  *Victor Muntés-Mulero, Norbert Martínez-Bazán, Josep-Lluís Larriba-Pey, Esther Pacitti, and Patrick Valduriez*

- HyperGraphDB: A Generalized Graph Database ....................... 25  
  *Borislav Iordanov*

- Survey of Graph Database Performance on the HPC Scalable Graph Analysis Benchmark .................................................. 37  
  *D. Dominguez-Sal, P. Urbón-Bayes, A. Giménez-Vaño, S. Gómez-Villamor, N. Martínez-Bazán, and J.L. Larriba-Pey*

- A Model for Automatic Generation of Multi-partite Graphs from Arbitrary Data .................................................. 49  
  *Ricardo Baeza-Yates, Nieves Brisaboa, and Josep Larriba-Pey*

- A Fast Two-Stage Algorithm for Computing SimRank and Its Extensions .................................................. 61  
  *Xu Jia, Hongyan Liu, Li Zou, Jun He, and Xiaoyong Du*

- Hotspot District Trajectory Prediction ..................................... 74  
  *Hongjun Li, Changjie Tang, Shaojie Qiao, Yue Wang, Ning Yang, and Chuan Li*

## International Workshop on Advanced Techniques on XML Data Management (XMLDM 2010)

- Effective XML Keyword Search through Valid Lowest Information Unit .................................................. 85  
  *Ying Lou, Peng Wang, Zhanhuai Li, Qun Chen, and Xia Li*
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reducing Redundancy of XPath Query over Networks by Transmitting</td>
<td>95</td>
</tr>
<tr>
<td>XML Views</td>
<td></td>
</tr>
<tr>
<td>Yanjun Xu, Shicheng Xu, Chengbao Peng, and Zhang Xia</td>
<td></td>
</tr>
<tr>
<td>Building Web Application with XQuery</td>
<td>106</td>
</tr>
<tr>
<td>Zhiming Gui, Husheng Liao, and linlin Fu</td>
<td></td>
</tr>
<tr>
<td>Functional Dependencies for XML</td>
<td>110</td>
</tr>
<tr>
<td>Haitao Chen, Husheng Liao, and Zengqi Gao</td>
<td></td>
</tr>
<tr>
<td>Structure and Content Similarity for Clustering XML Documents</td>
<td>116</td>
</tr>
<tr>
<td>Lijun Zhang, Zhanhuai Li, Qun Chen, and Ning Li</td>
<td></td>
</tr>
<tr>
<td>pq-Hash: An Efficient Method for Approximate XML Joins</td>
<td>125</td>
</tr>
<tr>
<td>Fei Li, Hongzhi Wang, Liang Hao, Jianzhong Li, and Hong Gao</td>
<td></td>
</tr>
<tr>
<td>TwigLinkedList: Improvement of TwigList</td>
<td>135</td>
</tr>
<tr>
<td>Zengqi Gao, Husheng Liao, Hongyu Gao, and Kechao Yang</td>
<td></td>
</tr>
<tr>
<td>Compression Algorithms for Structural Query Results on XML Data</td>
<td>141</td>
</tr>
<tr>
<td>Qing Wang, Hongzhi Wang, Hong Gao, and Jianzhong Li</td>
<td></td>
</tr>
<tr>
<td>SeCCX: Semantics-Based Fine Granular Concurrency Control for XML</td>
<td>146</td>
</tr>
<tr>
<td>Data</td>
<td></td>
</tr>
<tr>
<td>Chuitian Rong, Wei Lu, Xiao Zhang, Zhen Liu, and Xiaoyong Du</td>
<td></td>
</tr>
<tr>
<td>XRCJ: Supporting Keyword Search in XML and Relation</td>
<td>156</td>
</tr>
<tr>
<td>Co-occurrence</td>
<td></td>
</tr>
<tr>
<td>Song Zhang and Xiaoyong Du</td>
<td></td>
</tr>
<tr>
<td><strong>The 2nd International Workshop on Web-Based Contents Management</strong></td>
<td></td>
</tr>
<tr>
<td>Technologies (WCMT 2010)</td>
<td></td>
</tr>
<tr>
<td>Formal Verification of Stochastic Timing Behavior in Web-Based</td>
<td>166</td>
</tr>
<tr>
<td>Business Process Collaboration</td>
<td></td>
</tr>
<tr>
<td>Haiyang Hu, Jianen Xie, and JiDong Ge</td>
<td></td>
</tr>
<tr>
<td>An Efficient Blind Ring Signature Scheme without Pairings</td>
<td>177</td>
</tr>
<tr>
<td>Jianhong Zhang, Hua Chen, Xue Liu, and Chenglian Liu</td>
<td></td>
</tr>
<tr>
<td>Improving the Throughput of Wireless Mesh Networks for Web Services</td>
<td>189</td>
</tr>
<tr>
<td>Hua Hu, Zheng Zhang, and Haiyang Hu</td>
<td></td>
</tr>
<tr>
<td>Author Name Disambiguation for Citations on the Deep Web</td>
<td>198</td>
</tr>
<tr>
<td>Rui Zhang, Derong Shen, Yue Kou, and Tiezheng Nie</td>
<td></td>
</tr>
<tr>
<td>Approximate Content Summary for Database Selection in Deep Web Data Integration</td>
<td>210</td>
</tr>
<tr>
<td>Fangjiao Jiang, Yukun Li, Jiping Zhao, and Nan Yang</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Detecting Comment Spam through Content Analysis</td>
<td>222</td>
</tr>
<tr>
<td>Congrui Huang, Qiancheng Jiang, and Yan Zhang</td>
<td></td>
</tr>
<tr>
<td>Enriching the Contents of Enterprises' Wiki Systems with Web</td>
<td>234</td>
</tr>
<tr>
<td>Information</td>
<td></td>
</tr>
<tr>
<td>Li Zhao, Yexin Wang, Congrui Huang, and Yan Zhang</td>
<td></td>
</tr>
<tr>
<td>Query Processing with Materialized Views in a Traceable P2P Record</td>
<td>246</td>
</tr>
<tr>
<td>Exchange Framework</td>
<td></td>
</tr>
<tr>
<td>Fengrong Li and Yoshiharu Ishikawa</td>
<td></td>
</tr>
<tr>
<td>Author Index</td>
<td>259</td>
</tr>
</tbody>
</table>