Lecture Notes in Computer Science 7419

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Zhifeng Bao Yunjun Gao Yu Gu
Longjiang Guo Yingshu Li Jiaheng Lu
Zujie Ren Chaokun Wang Xiao Zhang (Eds.)

Web-Age Information Management

WAIM 2012 International Workshops:
GDMM, IWSN, MDSP, USDM, and XMLDM
Harbin, China, August 18-20, 2012
Proceedings

Springer

Volume Editors

Zhifeng Bao
National University of Singapore, Singapore, E-mail: baozhife@comp.nus.edu.sg

Yunjun Gao
Zhejiang University, Hangzhou, China, E-mail: gaoyj@zju.edu.cn

Yu Gu
Northeastern University, Shenyang, China, E-mail: guyu@ise.neu.edu.cn

Longjiang Guo
Heilongjiang University, Harbin, China, E-mail: longjiangguo@gmail.com

Yingshu Li
Georgia State University, Atlanta, GA, USA, E-mail: yli@cs.gsu.edu

Jiaheng Lu
Renmin University of China, Beijing, China, E-mail: jiahenglu@ruc.edu.cn

Zujie Ren
Hangzhou Dianzi University, Hangzhou, China, E-mail: renzju@gmail.com

Chaokun Wang
Tsinghua University, Beijing, China, E-mail: chaokun@tsinghua.edu.cn

Xiao Zhang
Renmin University of China, Beijing, China, E-mail: zhangxiao@ruc.edu.cn

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-33049-0 e-ISBN 978-3-642-33050-6
DOI 10.1007/978-3-642-33050-6

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012945067

CR Subject Classification (1998): H.3, H.4, H.2.8, H.2.4, C.2.1, E.1, F.2.2

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

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

WAIM 2012 Workshop Chairs' Message

Web-Age Information Management (WAIM) is an annual international conference for researchers, practitioners, developers, and users to share and exchange their cutting-edge ideas, results, experiences, techniques, and tools in connection with all aspects of Web data management. The conference invites original research papers on the theory, design, and implementation of Web-based information systems. WAIM 2012, the 13th in the series, was held in Harbin during August 18–20, 2012.

Along with the main conference, WAIM workshops intend to provide the international group of researchers with a forum for the discussion and exchange of research results contributing to the main themes of the WAIM conference. This WAIM 2012 workshop volume contains the papers accepted for the following five workshops that were held in conjunction with WAIM 2012. These five workshops were selected after a public call-for-proposals process, each of which focuses on a specific area that contributes to the main themes of the WAIM conference. The five workshops were:

- The First International Workshop on Graph Data Management and Mining (GDMM 2012)
- The Second International Wireless Sensor Networks Workshop (IWSN 2012)
- The First International Workshop on Massive Data Storage and Processing (MDSP 2012)
- The Third International Workshop on Unstructured Data Management (USDM 2012)
- The 4th International Workshop on XML Data Management (XMLDM 2012)

All the organizers of the previous WAIM workshops and conferences have made WAIM a valuable trademark and we are proud to pursue their work. We would like to express our thanks and acknowledgement to all the workshop organizers and Program Committee members who contributed to making the workshop program such a success. They put a tremendous amount of effort into soliciting and selecting research papers with a balance of high quality, novelty, and applications. They also followed a vigorous review process. A total of 34 papers were accepted. We are very grateful to the main conference organizers and the local Organizing Committee for their great support and wonderful arrangements.

August 2012

Xiaochun Yang
Hongzhi Wang

Preface of the 4th International Workshop on XML Data Management

It is our great pleasure to welcome you to the proceedings of the 4th International Workshop on XML Data Management (XMLDM 2012).

XML has gained much attention from database and Web researchers who are actively working in one or more of the emerging XML areas. XML data are self-describing and provide a platform-independent means to describe data and, therefore, can transport data from one platform to another. XML documents can be mapped to one more of the existing data models such as relational and object-relational models, and XML views can be produced dynamically from the pre-existing data models. XML queries can be mapped to the queries of the underlying models and can use their optimization features. XML data integration is useful for E-commerce applications such as comparison-shopping, which requires further study in the domain of data-, schema- and query-based integration. XML change management is another important area that has attracted attention in the context of Web warehouses. XML has been in use in upcoming areas such Web services, sensors and biological data management. The Third International Workshop on XML Data Management 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.

XMLDM attracted 16 submissions from Asia, Europe, and Singapore. The Program Committee accepted seven full papers. These papers cover a variety of topics, including XML keyword search, XML concurrency control protocols, indexing, XPath, uncertain XML dataset, classification and so on. We hope that they will serve as a valuable starting point for much brilliant thinking in XML data management.

The paper "Effective Keyword Search with Synonym Rules over XML Document" introduces a novel XML keyword search that can find the semantic information behind user input queries. The authors use synonyms, acronyms, and abbreviations that define the equality between strings. Finally, they have devised a transformation matching-based IL algorithm (TMIL) with synonym rules to improve the effectiveness of SLCA-based keyword search over XML documents.

In the paper "XML Concurrency Control Protocols: A Survey," Shan et al. present an overview of some of the most important XML concurrency control protocols, such as locking-based, timestamp-based, and optimistic XML concurrency control protocols. In addition, a summary and comparison are given for each protocol. While indexing XML documents for research purposes can be a complex task especially when we consider content and structure, the paper "Using Conceptual Scaling for Indexing XML Native Databases" proposes using conceptual scaling-based formal concept analysis for indexing both content and

structure. It aims to provide a combined structure while assuring hierarchical levels of data content and structure representation.

The paper "Indexing Compressed XML Documents" consists of studies and analyzes some suitable compressors to improve the indexing compressed XML documents process in order to exploit the compressed data for querying and information retrieval. The authors propose a new indexing process which leads to compressed XML data by re-indexing compressed XML data under an XMill compressor.

In the paper "Path-Based XML Stream Compression with XPath Query Support," Qian et al. present a compression for XML stream technology which divides XML streams into structure and context, and then encodes them respectively. They also present experimental results that demonstrate the effectiveness and efficiency of the methods proposed.

The paper "Uncertain XML Functional Dependencies Based on Tree Tuple Models" studies the functional dependencies and their applications in uncertain XML datasets. In this paper, Lv et al. propose three new kinds of functional dependencies based on tree tuple models for uncertain XML datasets. Finally, they also provide a sound and complete set of inference rules as well as two applications.

In "XML Document Classification Using Closed Frequent Subtree," Wang et al. propose an efficient SVM- and SLVM-based classification approach for XML documents that combines the content with the structure of XML documents to compute the similarity between the categories and documents. The experimental results show that this approach performs better than any other competitor's approach on XML classification.

Making XMLDM 2012 possible was a team effort. First of all, we would like to thank the authors and panelists for providing the content of the program. We would like to express our gratitude to the Program Committee and external reviewers, who worked very hard in reviewing papers and providing suggestions for their improvement. In particular we extend our special thanks to Linlin Zhang for maintaining the XMLDM 2012 website and for his effort in organizing the workshop.

We hope that you will find these proceedings interesting and thought-provoking.

Zhifeng Bao
Jiaheng Lu
Talel Abdessalem

XMLDM 2012 Workshop Organization

Workshop General Chairs

Tok Wang Ling	National University of Singpaore, Singapore
Ge Yu	North-East University, China

Workshop Co-chairs

Talel Abdessalem	Telecom Paristech Institute, France
Zhifeng Bao	National University of Singapore, Singapore
Jiaheng Lu	Renmin University of China, China

Program Committee

Stephane Bressan	National University of Singapore, Singapore
Mongli Lee	National University of Singapore, Singapore
Guoliang Li	Tsinghua University, China
Jianxin Li	Swinburne University of Technology, Australia
Zhanhuai Li	Northwestern Polytechnical University, China
Chengfei Liu	Swiburne University of Technology, Australia
Lu Qin	Chinese University of Hong Kong, Singapore
Hongzhi Wang	Harbin Institute of Technology, China
Junhu Wang	Griffith University, Australia
Wei Wang	University of New South Wales, Australia
Huayu Wu	Institute for Infocomm Research, Singapore
Xiaochun Yang	North-East University, China
Junfeng Zhou	Yanshan University, China
Rui Zhou	Swinburne University of Technology, Australia

It Is Time to Exploit Unstructured Data
Preface to the Third International Workshop on USDM

The management of unstructured data has become a hot topic in academia, industry, and government. Most data today, such as Web data, media data, sensor data, are generated by a mass with free will, or automatically by software or hardware, and lack explicit, predefined schema. In contrast to the existing relational data, these data are called unstructured data. The large volume, high change ratio, implicit and heterogeneous structures of these unstructured data pose great challenges to database researchers and engineers. So far, we are encountering the challenges of universal data models, highly flexible storage organization, metadata management, content understanding and so on. Fortunately, there are some initial and enlightened ideas.

The Third International Workshop on Unstructured Data Management (USDM 2012) aimed at bringing together researchers, developers, and users to discuss and present current technical developments in this area. The first Workshop on Unstructured Data Management (USDM 2010) was held with APWeb 2010 (April 6, 2010, Busan, Korea) and the Second Workshop on Unstructured Data Management (USDM 2011) was held with APWeb 2011 (April 20, 2011, Beijing, China), which provided a successful international forum for the dissemination of research achievements in unstructured data management. This year, we received 24 submissions on diverse topics of interest, and selected nine of them through a rigorous review progress and extensive discussions. These accepted papers handle issues on unstructured data storing, querying, retrieval, analysis, mining, and applications. The Program Committee composed a diverse and exciting program for USDM 2012.

The workshop was a forum for both presenting new research results and discussing practical experiences, which can help shape and solve critical problems in unstructured data management. We believe it provided participants with a chance to gain more knowledge in the field. We had two accepted papers on the storage and indexing of unstructured data, three on image data retrieval and processing, two on text search, and two on data mining of unstructured data.

This workshop was partially supported by the Unstructured Data Management System Projects in the HGJ program of China. We would like to thank all the people for their help in making the workshop successful. We thank the Steering Committee Chairs (Xiaoyong Du, Jianmin Wang, and Tengjiao Wang)

and the Steering Committee members (Dianfu Ma and Yueting Zhuang) for their suggestions and important instructions. We would like to thank all PC members, especially Zhenying He, Jinchuan Chen, and YueguoChen. Finally, we would like to thank all the speakers and presenters at the workshop, and all the participants at the workshop, for their engaged and fruitful contributions.

Xiao Zhang
Chaokun Wang
Jun Gao

Organization

Program Committee Co-chairs

Xiao Zhang	Renmin University of China
Chaokun Wang	Tsinghua University
Jun Gao	Peking University

Program Committee Members

Jinchuan Chen	Renmin University of China
Yueguo Chen	Renmin University of China
Zhenying He	Fudan University
Jun Gao	Peking University
Chaokun Wang	Tsinghua University
Xiao Zhang	Renmin University of China

Steering Committee Co-chairs

Xiaoyong Du	Renmin University of China
Jianmin Wang	Tsinghua University
Tengjiao Wang	Peking University

Steering Committee Members

Xiaoyong Du	Renmin University of China
Dianfu Ma	Beihang University
Jianmin Wang	Tsinghua University
Tengjiao Wang	Peking University
Yueting Zhuang	Zhejiang University

Preface to the 1st International Workshop on MDSP

On behalf of the Program Chairs for MDSP 2012, consisting of two General Co-chairs and two Program Co-chairs, we are pleased to present you with this volume. It contains the papers accepted for presentation in the workshop program of the 13th International Conference on Web-Age Information Management held in Harbin, China, during August 18–20, 2012.

This is the First International Conference on Massive Data Storage and Processing (MDSP). Twenty papers were submitted to the MDSP program, from which eight were accepted for presentation and inclusion in the conference proceedings. An acceptance rate of 40% makes MDSP one of the most selective workshops of WAIM 2012.

We would like to thank all the authors of submitted papers for choosing MDSP 2012 for the presentation of their research results. Owing to the high quality of the submitted papers, selecting the eight papers for the main conference was a very difficult task. We are deeply indebted to the four Program Chairs and 16 Program Committee members for their conscientious and impartial judgment and for the time and effort they contributed in preparation of this year's conference. All Area Chairs and reviewers are listed on the following pages.

The organizers of the conference are very happy with the response to our call for papers, noticing the interest of the data storage and processing community in this field. The workshop is composed of eight papers selected for presentation, covering a wide range of topics and showing interesting experiences. A brief summary of all the contributions, classified in three main areas, is presented below.

- PTL: Partitioned Logging for Database Storage on Flash Solid State Drives by Jun Yang and Qiong Luo from Hong Kong University of Science and Technology. The authors describe a storage scheme for databases on flash solid state drives.
- Adaptation Mechanism of iSCSI Protocol for NAS Storage Solution in Wireless Environments by Shamim Ripon and Sung Park from East West University. This paper presents an architecture to adapt iSCSI protocols with traditional network attached-storage cluster systems with error recovery methods.
- Band Selection for Hyperspectral Imagery with PCA-MIG by Kitti Koonsanit and Chuleerat Jaruskulchai from Kasetart University, Thailand. In this paper, an integrated PCA, maxima–minima functional method and information gain is proposed for hyperspectral band selection.
- NestedCube: Toward Online Analytical Processing on Information-Enhanced Multidimensional Networks by Jing Zhang, Xiaoguang Hong, and Qingzhong Li from Shandong University. This paper presents Nested Cube, a new

data warehousing model, which can support OLAP queries on information-enhanced multidimensional networks.
- MRFM:An Efficient Approach to Spatial Join Aggregates by Yi Liu, Luo Chen, Ning Jing, and Wei Xiong from University of Defense Technology. In this paper, the authors study the problem of answering spatial join aggregate queries under the MapReduce framework.
- A Distributed Inverted Indexing Scheme for Large-scale RDF Data by Xu Li and Xin Wang from. This paper presents a distributed inverted indexing scheme for large-scale RDF data. A scalable inverted index is built using the underlying data structure of Cassandra, which is a distributed key-value storage system.
- MSMapper: An Adaptive Split Assignment Scheme for MapReduce by Wei Pan, Zhanhuai Li, Qun Chen, Shanglian Peng, Suo Bo, and Jiang Xu from Northwestern Polytechnical University. This paper introduces the MSMapper (Multi-Split Mapper), a modified self-tuning mapper in which multiple splits can be assigned to one mapper.
- Driving Environment Reconstruction and Analysis Systems on Multi-sensor Networks by Chunyu Zhang, Yong Su, Jiyang Chen, and Wen Wang from the Research Institute of Highway. The authors construct a driving environment reconstruction and analysis system based on multi-sensors network onboard and some functional subsystems.

We would like to thank everyone who helped us. We greatly appreciate the advice and support by the WAIM 2012 General Co-chairs, Jianzhong Li (Harbin Institute of Technology, China) and Qing Li (City University of Hong Kong, China), Program Co-chairs, Hong Gao (Harbin Institute of Technology, China) and Local Organization Chair, Jizhou Luo (Harbin Institute of Technology, China), Workshops Chairs, Xiaochun Yang (Northeast University, China) and Hongzhi Wang (Harbin Institute of Technology, China).

Weisong Shi
Yunjun Gao
Weiping Wan
Zujie Ren

Organization

General Co-chairs

Weisong Shi	Wayne State University, USA
Yunjun Gao	Zhejiang University, China

Program Co-chairs

Weiping Wang	Chinese Academy of Sciences, China
Zujie Ren	Hangzhou Dianzi University, China

Program Committee

Guoray Cai	Pennsylvania State University, USA
Yong Woo LEE	University of Seoul, Korea
Hung Keng Pung	National University of Singapore
Xiaofei Liao	Huazhong University of Science and Technology, China
Yijun Bei	Zhejiang University, China
Yi Zhuang	Zhejiang GongShang University, China
Tao Jiang	Jiaxing University, China
Weiwei Sun	Fudan University, China
Xiaokui Xiao	Nanyang Technological University, Singapore
Yimin Lin	Singapore Management University, Singapore
Bin Yao	Shanghai Jiaotong University, China
Shaojie Qiao	Southwest Jiaotong University, China
Congfeng Jiang	Hangzhou Dianzi University, China
Jilin Zhang	Hangzhou Dianzi University, China
Qi Qiang	Alibaba Corp., China
Yongjian Ren	Infocore Corp., China

Preface to the 2nd International Workshop on IWSN

Over the past decade, signficant advances in wireless communication and computing technologies have led to the proliferation of reliable and ubiquitous infrastructure and infrastructureless wireless sensor networks all over the world, as well as a diverse range of new applications, such as the surveillance and protection of critical infrastructures and environment monitoring. Wireless sensor networks collect sensing measurements or detect special events, perform node-level processing, and export the combined data from their sensing nodes to the outside world. Sensing, processing, and communication are three key elements whose combination in one small device is instrumental to pervasive computing and gives rise to countless applications. These applications have raised new challenges ranging from the theoretical foundations of these systems, algorithms and protocol design, security and privacy to rigorous and systematic design and evaluation methodologies and new architectures for next-generation wireless sensor networks.

The International Wireless Sensor Networks Workshop 2012 (IWSN 2012) provided a forum for researchers and practitioners worldwide to exchange ideas, share new findings, and discuss challenging issues for the current and next-generation wireless sensor networks.

IWSN 2012, co-located with the 13th International Conference on Web-Age Information Management (WAIM 2012), took place in Harbin during August 18, 2012. Each submission was reviewed by at least three Program Committee members. Following a rigorous review process, a total of six papers were selected for presentations at the workshop.

We thank all the authors for submitting their papers to the conference. Finally, many other people contributed to the success of IWSN 2012 directly and indirectly. Even though their names cannot be listed here because of space limitation, we owe them our gratitude.

August 2012

Yingshu Li
Jinbao Li
Longjiang Guo

Organization

Workshop Committee on International Wireless Sensor Networks Workshop (IWSN)

General Co-chairs

Yingshu Li	Georgia State University, USA
Jinbao Li	Heilongjiang University, China

Program Co-chair

Longjiang Guo	Heilongjiang University, China

Program Committee Members

Arif Selcuk Uluagac	Georgia Institute of Technology, USA
Donghyun Kim	North Carolina Central University, USA
Xiaoming Wang	Shaanxi Normal University, China
Fei Li	George Mason University, USA
Wenwei Li	Hunan University, China
Feng Wang	Arizona State University, USA
Kai Xing	University of Science and Technology of China, China
Yoora Kim	Ohio State University, USA
Chuanhe Huang	Wuhan University, China
Yan Yang	Heilongjiang University, China
Zenghua zhao	Tianjing, University, China
Chunyu Ai	Troy University, USA
Juan Luo	Hunan University, China
Jinghua Zhu	Heilongjiang University, China
Guilin Li	Xiamen University, China
Lei Yu	Clemson University, USA
Hongzi Zhu	Shanghai Jiao Tong University, China
Jing He	Kennesaw State University, USA
Meirui Ren	Heilongjiang University, China
Shouling Ji	Georgia State University, USA
Desheng Zhang	University of Minnesota, USA

GDMM 2012 Workshop Organizers' Message

Graph data have become a powerful tool for representing and understanding objects and their relationships. With the rapid growth of emerging applications like social network analysis, semantic Web analysis, bio-information network analysis and so on, there is an urgent need to support high-performance query processing and mining ability for various graph data structures. Current database researchers have been actively contributing to pressing problems on graph data management including storage for graph data, graph query processing, similarity measure and search, graph analysis and mining, graph query languages proposals, distributed graph data management, compressing large graph data, prototype systems for managing graph data, graph visualization, and browsing.

The First International Workshop on Graph Data Management and Mining (GDMM 2012) was held in August 2012 in Harbin, China, in conjunction with the 13th International Conference on Web-Age Information Management (WAIM 2012). The overall goal of the GDMM workshop is to bring researchers from different fields together, to exchange research ideas and results, share insights about how to provide efficient graph data management and mining techniques, and to understand the research challenges and solutions of this area.

The workshop attracted eight submissions from China and Japan, covering a broad range of interesting topics in graph data management. All submissions were peer reviewed by at least three Program Committee members to ensure that high-quality papers were selected. The Program Committee selected four papers for inclusion in the workshop proceedings (acceptance rate 50%). The accepted papers span exciting topics from graph extraction to graph data compressing, and query processing.

The Program Committee of the workshop consisted of 11 experienced researchers and experts in the area of data management. The workshop would not been successful without the help of many people and organizations. Firstly, we would like to acknowledge the valuable contribution of all the Program Committee members during the peer-review process. Secondly, we would also like to thank the WAIM 2012 workshop chairs for their great support in ensuring the success of GDMM 2012.

July 2012

Yu Gu
Sai Wu
Dawei Jiang

International Workshop on Graph Data Management and Mining (GDMM 2012)

Workshop Co-chairs

Yu Gu	Northeastern University, China
Sai Wu	Zhejiang University, China
Dawei Jiang	National University of Singapore, Singapore

Program Committee

Guoliang Li	Tsinghua University, China
Dongxiang Zhang	National University of Singapore, Singapore
Ke Chen	Zhejiang University, China
Zhenjie Zhang	ADSC, UIUC, Singapore
Hongzhi Wang	Harbin Institute of Technology, China
Quanqing Xu	National ICT Australia (NICTA), Australia
Wei Wu	Institute for Infocomm Research, Singapore
Yuting Lin	Google, USA
Jeffrey Xu Yu	The Chinese University of Hong Kong, Hong Kong, SAR China
TieZheng Nie	Northeastern University, China
Yueguo Chen	Renmin University of China, China

Table of Contents

The First International Workshop on Graph Data Management and Mining (GDMM 2012)

The Second International Wireless Sensor Networks Workshop (IWSN 2012)

The First International Workshop on Massive Data Storage and Processing (MSDP 2012)

The Third International Workshop on Unstructured Data Management (USDM 2012)

The Forth International Workshop on XML Data Management (XMLDM 2012)