Lecture Notes in Computer Science 7418

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

Alfred Kobsa
University of California, Irvine, CA, 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
TU Dortmund University, Germany

Madhu Sudan
Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos
University of California, Los Angeles, CA, USA

Doug Tygar
University of California, Berkeley, CA, USA

Gerhard Weikum
Max Planck Institute for Informatics, Saarbruecken, Germany

Hong Gao Lipyeow Lim Wei Wang
Chuan Li Lei Chen (Eds.)

Web-Age Information Management

13th International Conference, WAIM 2012
Harbin, China, August 18-20, 2012
Proceedings

Springer

Volume Editors

Hong Gao
Harbin Institute of Technology
Harbin 150001, Heilongjiang, China
E-mail: honggao@hit.edu.cn

Lipyeow Lim
University of Hawaii
Honolulu 96822, HI, USA
E-mail: lipyeow@hawaii.edu

Wei Wang
Fudan University
Shanghai 200433, China
E-mail: weiwang1@fudan.edu.cn

Chuan Li
Sichuan University
Chengdu 610064, Sichuan, China
E-mail: lcharles@scu.edu.cn

Lei Chen
Hong Kong University of Science and Technology
Kowloon, Hong Kong, China
E-mail: leichen@cse.ust.hk

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-32280-8 e-ISBN 978-3-642-32281-5
DOI 10.1007/978-3-642-32281-5
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012943037

CR Subject Classification (1998): H.2.4, H.2.7-8, H.3.3-5, F.2.2, H.2, H.4, C.2, H.5, G.2.2, I.5.3

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

© Springer-Verlag Berlin Heidelberg 2012
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.
The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

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)

Preface

This volume contains the proceedings of the 13th International Conference on Web-Age Information Management (WAIM), held August 18–20, 2012, in Harbin, China. WAIM is a leading international conference on research, development, and applications of Web technologies, database systems and software engineering. WAIM is based in the Asia-Pacific region, and previous WAIM conferences were held in Shanghai (2000), Xi'an (2001), Beijing (2002), Chengdu (2003), Dalian (2004), Hangzhou (2005), Hong Kong (2006), Huangshan (2007), Zhangjiajie (2008), Suzhou (2009), Jiuzhaigou (2010), and Wuhan (2011). As the 13th event in the increasingly popular series, WAIM 2012, which was organized and supported by Harbin Institute of Technology, attracted outstanding researchers from all over the world to Harbin, China. In particular, this year WAIM and Microsoft Research Asia jointly sponsored a database summer school, which was collocated with WAIM.

This high-quality program would not have been possible without the authors who chose WAIM as a venue for their publications. Out of 178 submissions from various countries and regions, we selected 32 full papers and 10 short papers for publication. The acceptance rate for regular paper was 18% while the overall acceptance rate is 23.6% including short papers. The contributed papers address a wide range of topics such as spatial databases, query processing, XML and Web data, graph and uncertain data, distributed computing, information extraction and integration, data warehousing and data mining, similarity search, wireless sensor networks, social networks, data security, etc. We are grateful to our distinguished keynote speakers Wenfei Fan, Jiawei Han and Rakesh Agrawal for contributing to the event's success.

A conference like WAIM can only succeed as a team effort. We want to thank the Program Committee members and the reviewers for their invaluable efforts. Special thanks to the local Organizing Committee headed by Jizhou Luo. Many thanks also go to our Workshop Co-chairs (Xiaochun Yang and Hongzhi Wang), Publication Co-chairs (Lei Chen and Chuan Li), Publicity Co-chairs (Weiyi Meng and Guohui Li), Industrial chairs (Mukesh Mohania and Xiaoxin Wu), Panel Co-chairs (Jian Pei and Zhaonian Zou), and Finance Co-chairs (Howard Leung and Shengfei Shi). Last but not least, we wish to express our gratitude for the hard work of our webmaster (Hai Cao), and for our sponsors who generously supported the smooth running of our conference.

We hope that you enjoy reading the proceedings of WAIM 2012.

August 2012

Hong Gao
Lipyeow Lim
Wei Wang

Conference Organization

General Co-chairs

Jianzhong Li	Harbin Institute of Technology, China
Qing Li	City University of Hong Kong

Program Committee Co-chairs

Hong Gao	Harbin Institute of Technology, China
Lipyeow Lim	University of Hawaii, USA
Wei Wang	Fudan University, China

Workshop Co-chairs

Xiaochun Yang	Northeast University, China
Hongzhi Wang	Harbin Institute of Technology, China

Panel Co-chairs

Jian Pei	Simon Fraser University, USA
Zhaonian Zou	Harbin Institute of Technology, China

Industry Co-chairs

Mukesh Mohania	IBM, India
Xiaoxin Wu	Huawei, China

Publication Co-chairs

Lei Chen	Hong Kong University of Science and Technology, Hong Kong
Chuan Li	Sichuan University, China

Publicity Co-chairs

Weiyi Meng	Binghamton University, USA
Guohui Li	Huazhong University of Science and Technology, China

Local Organization Chair

Jizhou Luo	Harbin Institute of Technology, China

Finance Co-chairs

Howard Leung	City University of Hong Kong
Shengfei Shi	Harbin Institute of Technology, China

Steering Committee Liaison

Xiaofeng Meng	Renmin University, China

CCF DBS Liaison

Zhiyong Peng	Wuhan University, China

Microsoft Summer School Liaison

Haixun Wang	Microsoft Research Asia, China

Program Committee

Alfredo Cuzzocrea	University of Calabria, Italy
Aoying Zhou	East China Normal University, China
Bin Cui	Peking University, China
Carson K. Leung	University of Manitoba, Canada
Chengkai Li	University of Texas at Arlington, USA
Chuan Li	Sichuan University, China
Donghui Zhang	Microsoft Jim Gray Systems Lab, USA
Feifei Li	Florida State University, USA
Ge Yu	Northeast University, China
Giovanna Guerrini	Università di Genova, Italy
Heng Tao Shen	University of Queensland, Australia
Hiroaki Ohshima	Kyoto University, Japan
Hong Chen	Chinese University of Hong Kong, Hong Kong
Hongzhi Wang	Harbin Institute of Technology, China
Hong Gao	Harbin Institute of Technology, China
Hwanjo Yu	Pohang University of Science and Technology, Korea
Jeffrey Yu	Chinese University of Hong Kong
Jianyong Wang	Tsinghua University, China
Jianzhong Li	Harbin Institute of Technology, China
JianliangXu	Hong Kong Baptist University, Hong Kong
Jimmy Huang	York University, Canada
Jizhou Luo	Harbin Institute of Technology, China

Jun Gao	Peking University, China
Johann Gamper	Free University of Bozen-Bolzano, Italy
Lei Duan	Sichuan University, China
Lei Chen	Hong Kong University of Science and Technology, Hong Kong
Lei Zou	Peking University, China
Ling Feng	Tsinghua University, China
Ning Jing	National University of Defense Technology, China
Ning Ruan	Kent State University, USA
Ning Yang	Sichuan University, China
Peng Wang	Fudan University, China
Raymond Ng	University of British Columbia, Canada
Shuai Ma	University of Edinburgh, UK
Shuigeng Zhou	Fudan University, China
Shuming Shi	Microsoft Research Asia, China
Tao Li	Florida International University, USA
Toshiyuki Amagasa	University of Tsukuba, Japan
Wei Wang	University of New South Wales, Australia
Weiyi Meng	State University of New York at Binghamton, USA
Xiangliang Zhang	King Abdullah University of Science and Technology, Saudi Arabia
Xiaofeng Meng	Renmin University of China
Xin Dong	AT&T Research, USA
Xingquan Zhu	University of Technology, Sydney
Xintao Wu	University of North Carolina at Charlotte, USA
Xuemin Lin	University of New South Wales, Australia
Xiaofang Zhou	University of Queensland, Australia
Xiaochun Yang	Northeastern University, China
Yanghua Xiao	Fudan University, China
Yan Jia	National University of Defense Technology, China
Yang-Sae Moon	Kangwon National University, Korea
YaokaiFeng	Kyushu University, Japan
Yi Cai	City University of Hong Kong
Ke Yi	Hong Kong University of Science and Technology, Hong Kong
Yoshiharu Ishikawa	Nagoya University, Japan
Yunjun Gao	Zhejianing University, China
Yuqing Wu	Indiana University at Bloomington, USA
Zhanhuai Li	Northwestern Polytechnical University, China
Zhaonian Zou	Harbin Institute of Technology, China
Zhiyong Peng	Wuhan University, China
Zongmin Ma	Northeastern University, China

Table of Contents

Keynotes

Data Quality: Theory and Practice 1
Wenfei Fan

Construction of Web-Based, Service-Oriented Information Networks: A Data Mining Perspective (Abstract) 17
Jiawei Han

Electronic Textbooks and Data Mining 20
Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, and Krishnaram Kenthapadi

Session 1: Wireless Sensor Networks

Topology-Based Data Compression in Wireless Sensor Networks 22
Shangfeng Mo, Hong Chen, and Yinglong Li

A Residual Energy-Based Fairness Scheduling MAC Protocol for Wireless Sensor Networks 35
Long Tan

Topology-Aided Geographic Routing Protocol for Wireless Sensor Networks 47
Guilin Li, Longjiang Guo, Jian Zhang, and Minghong Liao

POLARIS: A Fingerprint-Based Localization System over Wireless Networks 58
Nan Zhang and Jianhua Feng

Session 2: Data Warehouse and Data Mining

A High-Performance Algorithm for Frequent Itemset Mining 71
Jun-Feng Qu and Mengchi Liu

Mining Link Patterns in Linked Data 83
Xiang Zhang, Cuifang Zhao, Peng Wang, and Fengbo Zhou

Detecting Positive Opinion Leader Group from Forum 95
Kaisong Song, Daling Wang, Shi Feng, Dong Wang, and Ge Yu

D'MART: A Tool for Building and Populating Data Warehouse Model from Existing Reports and Tables 102
Sumit Negi, Manish A. Bhide, Vishal S. Batra, Mukesh K. Mohania, and Sunil Bajpai

Session 3: Query Proceeding (1)

Continuous Skyline Queries with Integrity Assurance in Outsourced Spatial Databases 114
Xin Lin, Jianliang Xu, and Junzhong Gu

Assessing Quality Values of Wikipedia Articles Using Implicit Positive and Negative Ratings 127
Yu Suzuki

Form-Based Instant Search and Query Autocompletion on Relational Data 139
Hao Wu and Lizhu Zhou

Session 4: Query Proceeding (2)

Range Query Estimation for Dirty Data Management System 152
Yan Zhang, Long Yang, and Hongzhi Wang

Top-k Most Incremental Location Selection with Capacity Constraint . . . 165
Yu Sun, Jin Huang, Yueguo Chen, Xiaoyong Du, and Rui Zhang

An Approach of Text-Based and Image-Based Multi-modal Search for Online Shopping 172
Renfei Li, Daling Wang, Yifei Zhang, Shi Feng, and Ge Yu

Categorizing Search Results Using WordNet and Wikipedia 185
Reza Taghizadeh Hemayati, Weiyi Meng, and Clement Yu

Session 5: Spatial Database

Optimal Sequenced Route Query Algorithm Using Visited POI Graph 198
Htoo Htoo, Yutaka Ohsawa, Noboru Sonehara, and Masao Sakauchi

A Packaging Approach for Massive Amounts of Small Geospatial Files with HDFS 210
Jifeng Cui, Yong Zhang, Chao Li, and ChunXiao Xing

Adaptive Update Workload Reduction for Moving Objects in Road Networks 216
Miao Li, Yu Gu, Jia Xu, and Ge Yu

Session 6: Similarity Search and Queries

An Adaptive Distributed Index for Similarity Queries in Metric Spaces 222
Mingdong Zhu, Derong Shen, Yue Kou, Tiezheng Nie, and Ge Yu

Finding Relevant Tweets 228
Deepak P. and Sutanu Chakraborti

Fgram-Tree: An Index Structure Based on Feature Grams for String Approximate Search 241
Xing Tong and Hongzhi Wang

Session 7: XML and Web Data

Efficient Processing of Updates in Dynamic Graph-Structured XML Data 254
Lizhen Fu and Xiaofeng Meng

Extracting Focused Time for Web Pages 266
Sheng Lin, Peiquan Jin, Xujian Zhao, Jie Zhao, and Lihua Yue

Top-Down SLCA Computation Based on Hash Search 272
Junfeng Zhou, Guoxiang Lan, Ziyang Chen, Xian Tang, and Jingfeng Guo

Session 8: Graph and Uncertain Data

Top-K Graph Pattern Matching: A Twig Query Approach 284
Xianggang Zeng, Jiefeng Cheng, Jeffrey Xu Yu, and Shengzhong Feng

Dynamic Graph Shortest Path Algorithm 296
Xueli Liu and Hongzhi Wang

A Framework for High-Quality Clustering Uncertain Data Stream over Sliding Windows 308
Keyan Cao, Guoren Wang, Donghong Han, Yue Ma, and Xianzhe Ma

Bayesian Network Structure Learning from Attribute Uncertain Data . . . 314
Wenting Song, Jeffrey Xu Yu, Hong Cheng, Hongyan Liu, Jun He, and Xiaoyong Du

Session 9: Distributed Computing

Bandwidth-Aware Medical Image Retrieval in Mobile Cloud Computing Network 322
Yi Zhuang, Nan Jiang, Zhiang Wu, Dickson Chiu, Guochang Jiang, and Hua Hu

Efficient Algorithms for Constrained Subspace Skyline Query in Structured Peer-to-Peer Systems 334
Khaled M. Banafaa and Ruixuan Li

Processing All k-Nearest Neighbor Queries in Hadoop 346
Takuya Yokoyama, Yoshiharu Ishikawa, and Yu Suzuki

Session 10: Data Security and Management

RGH: An Efficient RSU-Aided Group-Based Hierarchical Privacy Enhancement Protocol for VANETs 352
Tao Yang, Lingbo Kong, Liangwen Yu, Jianbin Hu, and Zhong Chen

Locating Encrypted Data Precisely without Leaking Their Distribution 363
Liqing Huang and Yi Tang

LB-Logging: A Highly Efficient Recovery Technique for Flash-Based Database 375
Zeping Lu, Xiaoying Qi, Wei Cao, and Xiaofeng Meng

Session 11: I nformation Extraction and Integration

An Under-Sampling Approach to Imbalanced Automatic Keyphrase Extraction 387
Weijian Ni, Tong Liu, and Qingtian Zeng

A Tourist Itinerary Planning Approach Based on Ant Colony Algorithm 399
Lei Yang, Richong Zhang, Hailong Sun, Xiaohui Guo, and Jinpeng Huai

A Transparent Approach for Database Schema Evolution Using View Mechanism 405
Jianxin Xue, Derong Shen, Tiezheng Nie, Yue Kou, and Ge Yu

WYSIWYE: An Algebra for Expressing Spatial and Textual Rules for Information Extraction 419
Vijil Chenthamarakshan, Ramakrishna Varadarajan, Prasad M. Deshpande, Raghuram Krishnapuram, and Knut Stolze

Session 12: Social Networks and Modern Web Services

A Scalable Algorithm for Detecting Community Outliers in Social Networks 434
Tengfei Ji, Jun Gao, and Dongqing Yang

An Efficient Index for Top-k Keyword Search on Social Networks 446
Xudong Du

Engineering Pathway for User Personal Knowledge Recommendation ... 459
Yunlu Zhang, Guofu Zhou, Jingxing Zhang, Ming Xie, Wei Yu, and Shijun Li

Pick-Up Tree Based Route Recommendation from Taxi Trajectories 471
Haoran Hu, Zhiang Wu, Bo Mao, Yi Zhuang, Jie Cao, and Jingui Pan

Author Index 485