

*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*

Haixun Wang Shijun Li Satoshi Oyama  
Xiaohua Hu Tieyun Qian (Eds.)

# Web-Age Information Management

12th International Conference, WAIM 2011  
Wuhan, China, September 14-16, 2011  
Proceedings

## Volume Editors

Haixun Wang

Microsoft Research Asia, Beijing, 100190, China

E-mail: haixunw@microsoft.com

Shijun Li

Wuhan University, Hubei 430072, China

E-mail: shjli@whu.edu.cn

Satoshi Oyama

Hokkaido University, Sapporo, Hokkaido 060-0814, Japan

E-mail: oyama@ist.hokudai.ac.jp

Xiaohua Hu

Drexel University, Philadelphia, PA 19104, USA

E-mail: xh29@drexel.edu

Tieyun Qian

Wuhan University, Hubei, 430072, China

E-mail: qty@whu.edu.cn

ISSN 0302-9743

e-ISSN 1611-3349

ISBN 978-3-642-23534-4

e-ISBN 978-3-642-23535-1

DOI 10.1007/978-3-642-23535-1

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011934878

CR Subject Classification (1998): H.3, H.4, I.2, C.2, H.2, H.5

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

© Springer-Verlag Berlin Heidelberg 2011

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

WAIM has been a leading international conference on research, development, and applications of Web technologies, database systems, information management and software engineering. WAIM is focused 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), and Jiuzhaigou (2010). As the 12<sup>th</sup> event in the increasingly popular series, WAIM 2011 attracted outstanding researchers from all over the world to Wuhan, 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 181 submitted papers from various countries and regions, we selected 53 full papers for publication. The acceptance rate for regular full papers is 29%. The contributed papers addressed a wide scope of issues in the fields of Web-age information management and advanced applications, including Web data mining, knowledge discovery from streaming data, query processing, multidimensional data analysis, data management support to advanced applications, etc.

A conference like this can only succeed as a team effort. We want to thank the Program Committee Chairs, Program Committee members and the reviewers for their invaluable efforts. Special thanks to the local Organizing Committee headed by Liang Hong and Ming Zhong. Many thanks also go to the Workshop Co-chairs (Chengfei Liu and Liwei Wang), Finance Chairs (Howard Leung and Xuhui Li), and Publicity Co-chairs (Weiyi Meng, Lei Chen and Guohui Li). We also thank the conference sponsors for their generous support.

We hope that you enjoy reading the proceedings of WAIM 2011.

September 2011

Haixun Wang  
Shijun Li  
Satoshi Oyama

# Conference Organization

## Honorary General Co-chairs

Yanxiang He  
Changjie Tang

Wuhan University, China  
Sichuan University, China

## General Co-chairs

Katsumi Tanaka  
Zhiyong Peng

Kyoto University, Japan  
Wuhan University, China

## Program Committee Co-chairs

Haixun Wang  
Shijun Li  
Satoshi Oyama

Microsoft Research Asia  
Wuhan University, China  
Hokkaido University, Japan

## Workshop Co-chairs

Chengfei Liu  
Liwei Wang

Swinburne University of Technology, Australia  
Wuhan University, China

## Industry Co-chairs

Mukesh Mohania  
Xiaoxin Wu

IBM, India  
Huawei, China

## Publication Co-chairs

Xiaohua Hu  
Tieyun Qian

Drexel University, USA  
Wuhan University, China

## Publicity Co-chairs

Weiyi Meng  
Lei Chen

Binghamton University, USA  
Hong Kong University of Science and  
Technology, Hong Kong  
Huazhong University of Science and  
Technology, China

## Local Organization Co-chairs

Liang Hong	Wuhan University, China
Ming Zhong	Wuhan University, China

## Finance Co-chairs

Howard Leung	City University of Hong Kong
Xuhui Li	Wuhan University, China

## Steering Committee Liaison

Qing Li	City University of Hong Kong
---------	------------------------------

## CCF DBS Liaison

Xiaofeng Meng	Renmin University, China
---------------	--------------------------

## Program Co-chairs

Haixun Wang	Microsoft Research Asia
Shijun Li	Wuhan University, China
Satoshi Oyama	Hokkaido University, Japan

## Area Chairs

Bin Cui	Peking University, China
Danushka Bollegala	University of Tokyo, Japan
Jianyong Wang	Tsinghua University, China
Kenny Zhu	Shanghai Jiao Tong University, China
Seung-won Hwang	Pohang University of Science and Technology, Korea
Wei Wang	University of New South Wales, Australia
Xiaochun Yang	Northeastern University, China
Alfredo Cuzzocrea	University of Calabria, Italy

## Program Committee

Anthony Tung	National University of Singapore, Singapore
Aoying Zhou	East China Normal University, China
Baihua Zheng	Singapore Management University, Singapore
Bin Cui	Peking University, China
Bingsheng He	Chinese University of Hong Kong, Hong Kong
Chengkai Li	University of Texas at Arlington, USA

Danushka Bollegala	The University of Tokyo, Japan
David Cheung	The University of Hong Kong
Donghui Zhang	Microsoft Jim Gray Systems Lab, USA
Fei Xu	Microsoft
Feifei Li	Florida State University, USA
Ge Yu	Northeastern University, China
Guoren Wang	Northeastern University, China
Guozhu Dong	Wright State University, USA
Heng Tao Shen	University of Queensland, Australia
Hiroaki Ohshima	Kyoto University, Japan
Hong Chen	Chinese University of Hong Kong, Hong Kong
Hongzhi Wang	Harbin Industry University, China
Hua Wang	University of Southern Queensland, Australia
Huan Liu	Arizona State University, USA
Hwanjo Yu	Pohang University of Science and Technology, Korea
Jaewoo Kang	Korea University
Jeffrey Yu	Chinese University of Hong Kong
Jianliang Xu	Hong Kong Baptist University, Hong Kong
Jianyang Wang	Tsinghua University, China
Jie Tang	Tsinghua University, China
Jimmy Huang	York University, Canada
Jun Gao	Peking University, China
Ke Wang	Simon Fraser University, Canada
Kenny Zhu	Shanghai Jiao Tong University, China
Lei Chen	Hong Kong University of Science and Technology, Hong Kong
Lei Duan	Sichuan University, China
Lin Li	Wuhan University of Technology, China
Lei Zou	Peking University, China
Lipeow Lim	University of Hawaii at Manoa, USA
Min Wang	HP Lab China
Nick Koudas	University of Toronto, Canada
Ning Jing	National University of Defense Technology, China
Peiquan Jin	University of Science and Technology China, China
Peng Wang	Fudan University, China
Philip Yu	University of Illinois at Chicago, USA
Qiang Ma	Kyoto University, Japan
Qiang Zhu	University of Michigan at Dearborn, USA
Raymond Ng	University of British Columbia, Canada
Ruili Wang	Massey University, New Zealand
Ruoming Jin	Kent State University, USA

Seung-won Hwang	Pohang University of Science and Technology, Korea
Shinsuke Nakajima	Kyoto Sangyo University, Japan
Shuai Ma	University of Edinburgh, UK
Shuigeng Zhou	Fudan University, China
Shuming Shi	Microsoft Research Asia
Tao Li	Florida International University, USA
Tengjiao Wang	Peking University, China
Ting Yu	North Carolina State 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
Weizhu Chen	Microsoft Research Asia
Xiaochun Yang	Northeastern University, China
Xiaofeng Meng	Renmin University of China, China
Xiaoyong Du	Renmin University of China, China
Xin (Luna)	Dong AT&T Research, USA
Xingquan Hill Zhu	University of Technology, Sydney
Xintao Wu	University of North Carolina at Charlotte, USA
Xu Yu	Teradata, USA
Xuanjing Huang	Fudan University, China
Xuemin Lin	University of New South Wales, Australia
Yan Jia	National University of Defense Technology, China
Yanghua Xiao	Fudan University, China
Yaokai Feng	Kyushu University, Japan
Yi Cai	City University of Hong Kong, Hong Kong
Yi Ke	Hong Kong University of Science and Technology, Hong Kong
Yingzi Jin	University of Tokyo, Japan
Yoshiharu Ishikawa	Nagoya University, Japan
Yuchen Fu	Soochow University, China
Yunjun Gao	Zhejiang University, China
Yuqing Wu	Indiana University at Bloomington, USA
Zhanhuai Li	Northwestern Polytechnical University, China
Zhongfei Zhang	State University of New York at Binghamton, USA
Zhongyuan Wang	Microsoft Research Asia



# Table of Contents

## Keynotes

Analytics for Info-plosion Including Information Diffusion Studies for the 3.11 Disaster . . . . .	1
<i>Masaru Kitsuregawa and Masashi Toyoda</i>	
Using the Web for Collaborative Language Learning and Teaching . . . . .	2
<i>Werner Winiwarter</i>	
Data-Driven Modeling and Analysis of Online Social Networks . . . . .	3
<i>Divyakant Agrawal, Bassam Bamieh, Ceren Budak, Amr El Abbadi, Andrew Flanagin, and Stacy Patterson</i>	

## Session 1A: Query Processing

Efficient Filter Algorithms for Reverse k-Nearest Neighbor Query . . . . .	18
<i>Shengsheng Wang, Qiannan Lv, Dayou Liu, and Fangming Gu</i>	
Keyword Query Cleaning with Query Logs . . . . .	31
<i>Lei Gao, Xiaohui Yu, and Yang Liu</i>	
A Self-adaptive Cross-Domain Query Approach on the Deep Web . . . . .	43
<i>Yingjun Li, Derong Shen, Tiezheng Nie, Ge Yu, Jing Shan, and Kou Yue</i>	

## Session 1B: Uncertain Data

SPARQL Query Answering with RDFS Reasoning on Correlated Probabilistic Data . . . . .	56
<i>Chi-Cheong Szeto, Edward Hung, and Yu Deng</i>	
Probabilistic Threshold Join over Distributed Uncertain Data . . . . .	68
<i>Lei Deng, Fei Wang, and Benxiong Huang</i>	
Bayesian Classifiers for Positive Unlabeled Learning . . . . .	81
<i>Jiazhen He, Yang Zhang, Xue Li, and Yong Wang</i>	

## Session 1C: Social Media (1)

Measuring Social Tag Confidence: Is It a Good or Bad Tag? . . . . .	94
<i>Xiwu Gu, Xianbing Wang, Ruixuan Li, Kunmei Wen, Yufei Yang, and Weijun Xiao</i>	

A New Vector Space Model Exploiting Semantic Correlations of Social Annotations for Web Page Clustering . . . . .	106
<i>Xiwu Gu, Xianbing Wang, Ruixuan Li, Kunmei Wen, Yufei Yang, and Weijun Xiao</i>	

A Generalization Based Approach for Anonymizing Weighted Social Network Graphs . . . . .	118
<i>Xiangyu Liu and Xiaochun Yang</i>	

**Session 2A: Semantics**

Incremental Reasoning over Multiple Ontologies . . . . .	131
<i>Jing Lu, Xingzhi Sun, Linhao Xu, and Haofen Wang</i>	

General-Purpose Ontology Enrichment from the WWW . . . . .	144
<i>Mohammed Maree, Mohammed Belkhatir, and Saadat M. Alhashmi</i>	

QuerySem: Deriving Query Semantics Based on Multiple Ontologies . . .	157
<i>Mohammed Maree, Saadat M. Alhashmi, and Mohammed Belkhatir</i>	

**Session 2B: Data Mining (1)**

Getting Critical Categories of a Data Set . . . . .	169
<i>Cheqing Jin, Yizhen Zhang, and Aoying Zhou</i>	

MFCluster: Mining Maximal Fault-Tolerant Constant Row Biclusters in Microarray Dataset . . . . .	181
<i>Miao Wang, Xuequn Shang, Miao Miao, Zhanhuai Li, and Wenbin Liu</i>	

Expansion Finding for Given Acronyms Using Conditional Random Fields . . . . .	191
<i>Jie Liu, Jimeng Chen, Tianbi Liu, and Yalou Huang</i>	

**Session 2C: Social Media (2)**

Leveraging Communication Information among Readers for RFID Data Cleaning . . . . .	201
<i>Tao Jiang, Yingyuan Xiao, Xiaoye Wang, and Yukun Li</i>	

Web Article Quality Assessment in Multi-dimensional Space . . . . .	214
<i>Jingyu Han, Xiong Fu, Kejia Chen, and Chuandong Wang</i>	

DRScribe: An Improved Topic-Based Publish-Subscribe System with Dynamic Routing . . . . .	226
<i>Guohui Li and Sheng Gao</i>	

## Session 3A: Cloud Data

An Efficient Quad-Tree Based Index Structure for Cloud Data Management .....	238
<i>Linlin Ding, Baiyou Qiao, Guoren Wang, and Chen Chen</i>	
Efficient Duplicate Detection on Cloud Using a New Signature Scheme .....	251
<i>Chuitian Rong, Wei Lu, Xiaoyong Du, and Xiao Zhang</i>	
A Secure and Efficient Role-Based Access Policy towards Cryptographic Cloud Storage .....	264
<i>Cheng Hong, Zhiquan lv, Min Zhang, and Dengguo Feng</i>	

## Session 3B: Multimedia Data

Tagging Image by Exploring Weighted Correlation between Visual Features and Tags .....	277
<i>Xiaoming Zhang, Zhoujun Li, and Yun Long</i>	
Credibility-Oriented Ranking of Multimedia News Based on a Material-Opinion Model .....	290
<i>Ling Xu, Qiang Ma, and Masatoshi Yoshikawa</i>	
Actions in Still Web Images: Visualization, Detection and Retrieval ....	302
<i>Piji Li, Jun Ma, and Shuai Gao</i>	

## Session 3C: User Models

Social Analytics for Personalization in Work Environments .....	314
<i>Qihua Wang and Hongxia Jin</i>	
Finding Appropriate Experts for Collaboration .....	327
<i>Zhenjiang Zhan, Lichun Yang, Shenghua Bao, Dingyi Han, Zhong Su, and Yong Yu</i>	
Location Privacy Protection in the Presence of Users' Preferences .....	340
<i>Weiwei Ni, Jinwang Zheng, Zhihong Chong, Shan Lu, and Lei Hu</i>	

## Session 4A: Data Management

Layered Graph Data Model for Data Management of DataSpace Support Platform .....	353
<i>Dan Yang, Derong Shen, Tiezheng Nie, Ge Yu, and Yue Kou</i>	
A Study of RDB-Based RDF Data Management Techniques .....	366
<i>Vahid Jalali, Mo Zhou, and Yuqing Wu</i>	

Requirement-Based Query and Update Scheduling in Real-Time Data Warehouses .....	379
<i>Fangling Leng, Yubin Bao, Ge Yu, Jingang Shi, and Xiaoyan Cai</i>	

## Session 4B: Graph Data

Answering Subgraph Queries over Large Graphs .....	390
<i>Weiguo Zheng, Lei Zou, and Dongyan Zhao</i>	
Finding Relevant Papers Based on Citation Relations.....	403
<i>Yicong Liang, Qing Li, and Tieyun Qian</i>	
ASAP: Towards <u>A</u> ccurate, <u>S</u> table and <u>A</u> ccelerative <u>P</u> enetrating-Rank Estimation on Large Graphs.....	415
<i>Xuefei Li, Weiren Yu, Bo Yang, and Jiajin Le</i>	

## Session 4C: Name Disambiguation

Efficient Name Disambiguation in Digital Libraries .....	430
<i>Jia Zhu, Gabriel Fung, and Liwei Wang</i>	
A Classification Framework for Disambiguating Web People Search Result Using Feedback.....	442
<i>Ou Jin, Shenghua Bao, Zhong Su, and Yong Yu</i>	
Incorporating User Feedback into Name Disambiguation of Scientific Cooperation Network.....	454
<i>Yuhua Li, Aiming Wen, Quan Lin, Ruixuan Li, and Zhengding Lu</i>	

## Session 5A: Performance

Multi-core vs. I/O Wall: The Approaches to Conquer and Cooperate ...	467
<i>Yansong Zhang, Min Jiao, Zhanwei Wang, Shan Wang, and Xuan Zhou</i>	
W-Order Scan: Minimizing Cache Pollution by Application Software Level Cache Management for MMDB .....	480
<i>Yansong Zhang, Min Jiao, Zhanwei Wang, Shan Wang, and Xuan Zhou</i>	
Index Structure for Cross-Class Query in Object Deputy Database .....	493
<i>Yuwei Peng, Hefei Ge, Mao Ding, Zeqian Huang, and Chuanjian Wang</i>	

## Session 5B: Data Mining (2)

Ensemble Pruning via Base-Classifer Replacement .....	505
<i>Huaping Guo and Ming Fan</i>	
Efficient Approximate Similarity Search Using Random Projection Learning .....	517
<i>Peisen Yuan, Chaofeng Sha, Xiaoling Wang, Bin Yang, and Aoying Zhou</i>	
Informed Prediction with Incremental Core-Based Friend Cycle Discovering .....	530
<i>Yue Wang, Weijing Huang, Wei Chen, Tengjiao Wang, and Dongqing Yang</i>	

## Session 5C: Temporal Data

Early Prediction of Temporal Sequences Based on Information Transfer .....	542
<i>Ning Yang, Jian Peng, Yu Chen, and Changjie Tang</i>	
Mining Event Temporal Boundaries from News Corpora through Evolution Phase Discovery .....	554
<i>Liang Kong, Rui Yan, Han Jiang, Yan Zhang, Yan Gao, and Li Fu</i>	
Event Detection over Live and Archived Streams .....	566
<i>Shanglian Peng, Zhanhuai Li, Qiang Li, Qun Chen, Wei Pan, Hailong Liu, and Yanming Nie</i>	

## Session 6A: XML

Renda-RX: A Benchmark for Evaluating XML-Relational Database System .....	578
<i>Xiao Zhang, Kuicheng Liu, Li Zou, Xiaoyong Du, and Shan Wang</i>	
Energy-Conserving Fragment Methods for Skewed XML Data Access in Push-Based Broadcast .....	590
<i>Jingjing Wu, Peng Liu, Lu Gan, Yongrui Qin, and Weiwei Sun</i>	

## Session 6B: Spatial Data

Evaluating Probabilistic Spatial-Range Closest Pairs Queries over Uncertain Objects .....	602
<i>Mo Chen, Zixi Jia, Yu Gu, Ge Yu, and Chuanwen Li</i>	
Synthesizing Routes for Low Sampling Trajectories with Absorbing Markov Chains .....	614
<i>Chengxuan Liao, Jiaheng Lu, and Hong Chen</i>	

Approximate Continuous  $K$ -Nearest Neighbor Queries for Uncertain  
Objects in Road Networks ..... 627  
*Guohui Li, Ping Fan, and Ling Yuan*

**Session 6C: Event Detection**

Parallel Detection of Temporal Events from Streaming Data ..... 639  
*Hao Wang, Ling Feng, and Wenwei Xue*

Towards Effective Event Detection, Tracking and Summarization on  
Microblog Data ..... 652  
*Rui Long, Haofen Wang, Yuqiang Chen, Ou Jin, and Yong Yu*

**Author Index** ..... 665