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 12th 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  
Guohui Li  
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
Jianyong 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  
Shinsuke Nakajima  
Shuai Ma  
Shuigeng Zhou  
Shuming Shi  
Tao Li  
Tengjiao Wang  
Ting Yu  
Toshiyuki Amagasa  
Wei Wang  
Weiyi Meng  
Weizhu Chen  
Xiaochun Yang  
Xiaoyong Du  
Xin (Luna)  
Xingquan Hill Zhu  
Xu Yu  
Xuanjing Huang  
Xueming Lin  
Yan Jia  
Yanghua Xiao  
Yaokai Feng  
Yi Cai  
Yi Ke  
Yingzhi Jin  
Yoshiharu Ishikawa  
Yuchen Fu  
Yunjun Gao  
Yuqing Wu  
Zhanhui Li  
Zhongfei Zhang  
Zhongyuan Wang

Pohang University of Science and Technology, Korea  
Kyoto Sangyo University, Japan  
University of Edinburgh, UK  
Fudan University, China  
Microsoft Research Asia  
Florida International University, USA  
Peking University, China  
North Carolina State University, USA  
University of Tsukuba, Japan  
University of New South Wales, Australia  
State University of New York at Binghamton, USA  
Microsoft Research Asia  
Northeastern University, China  
Renmin University of China, China  
Renmin University of China, China  
Dong AT&T Research, USA  
University of Technology, Sydney  
University of North Carolina at Charlotte, USA  
Teradata, USA  
Fudan University, China  
University of New South Wales, Australia  
National University of Defense Technology, China  
Fudan University, China  
Kyushu University, Japan  
City University of Hong Kong, Hong Kong  
Hong Kong University of Science and Technology, Hong Kong  
University of Tokyo, Japan  
Nagoya University, Japan  
Soochow University, China  
Zhejiang University, China  
Indiana University at Bloomington, USA  
Northwestern Polytechnical University, China  
State University of New York at Binghamton, USA  
Microsoft Research Asia
# Table of Contents

## Keynotes

Analytics for Info-plosion Including Information Diffusion Studies for the 3.11 Disaster ................................. 1  
*Masaru Kitsuregawa and Masashi Toyoda*

Using the Web for Collaborative Language Learning and Teaching .... 2  
*Werner Winiwarter*

Data-Driven Modeling and Analysis of Online Social Networks......... 3  
*Divyakant Agrawal, Bassam Bamieh, Ceren Budak, Amr El Abbadi, Andrew Flanagin, and Stacy Patterson*

## Session 1A: Query Processing

Efficient Filter Algorithms for Reverse k-Nearest Neighbor Query ...... 18  
*Shengsheng Wang, Qiannan Lv, Dayou Liu, and Fangming Gu*

Keyword Query Cleaning with Query Logs ............................ 31  
*Lei Gao, Xiaohui Yu, and Yang Liu*

A Self-adaptive Cross-Domain Query Approach on the Deep Web ...... 43  
*Yingjun Li, Derong Shen, Tiezheng Nie, Ge Yu, Jing Shan, and Kou Yue*

## Session 1B: Uncertain Data

SPARQL Query Answering with RDFS Reasoning on Correlated Probabilistic Data ................................................. 56  
*Chi-Cheong Szeto, Edward Hung, and Yu Deng*

Probabilistic Threshold Join over Distributed Uncertain Data .......... 68  
*Lei Deng, Fei Wang, and Benxiong Huang*

Bayesian Classifiers for Positive Unlabeled Learning .................. 81  
*Jiazhen He, Yang Zhang, Xue Li, and Yong Wang*

## Session 1C: Social Media (1)

Measuring Social Tag Confidence: Is It a Good or Bad Tag? .......... 94  
*Xiwu Gu, Xianbing Wang, Ruixuan Li, Kunmei Wen, Yufei Yang, and Weijun Xiao*
# Table of Contents

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

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

## Session 2A: Semantics

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

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

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

## Session 2B: Data Mining (1)

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

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

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

## Session 2C: Social Media (2)

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

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

DRScribe: An Improved Topic-Based Publish-Subscribe System with Dynamic Routing.................................................. 226  
Guohui Li and Sheng Gao
Session 3A: Cloud Data

An Efficient Quad-Tree Based Index Structure for Cloud Data Management .................................................. 238
  Linlin Ding, Baiyou Qiao, Guoren Wang, and Chen Chen

Efficient Duplicate Detection on Cloud Using a New Signature Scheme .................................................. 251
  Chuitian Rong, Wei Lu, Xiaoyong Du, and Xiao Zhang

A Secure and Efficient Role-Based Access Policy towards Cryptographic Cloud Storage ........................................ 264
  Cheng Hong, Zhiquan lv, Min Zhang, and Dengguo Feng

Session 3B: Multimedia Data

Tagging Image by Exploring Weighted Correlation between Visual Features and Tags ........................................ 277
  Xiaoming Zhang, Zhoujun Li, and Yun Long

Credibility-Oriented Ranking of Multimedia News Based on a Material-Opinion Model ........................................ 290
  Ling Xu, Qiang Ma, and Masatoshi Yoshikawa

Actions in Still Web Images: Visualization, Detection and Retrieval ................................................................. 302
  Piji Li, Jun Ma, and Shuai Gao

Session 3C: User Models

Social Analytics for Personalization in Work Environments ................................................................. 314
  Qihua Wang and Hongxia Jin

Finding Appropriate Experts for Collaboration ................................................................. 327
  Zhenjiang Zhan, Lichun Yang, Shenghua Bao, Dingyi Han, Zhong Su, and Yong Yu

Location Privacy Protection in the Presence of Users’ Preferences ................................................................. 340
  Weiwei Ni, Jinwang Zheng, Zhihong Chong, Shan Lu, and Lei Hu

Session 4A: Data Management

Layered Graph Data Model for Data Management of DataSpace Support Platform .................................................. 353
  Dan Yang, Derong Shen, Tiezheng Nie, Ge Yu, and Yue Kou

A Study of RDB-Based RDF Data Management Techniques ................................................................. 366
  Vahid Jalali, Mo Zhou, and Yuqing Wu
<table>
<thead>
<tr>
<th>Session 4B: Graph Data</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answering Subgraph Queries over Large Graphs</td>
<td>390</td>
</tr>
<tr>
<td>Weiguo Zheng, Lei Zou, and Dongyan Zhao</td>
<td></td>
</tr>
<tr>
<td>Finding Relevant Papers Based on Citation Relations</td>
<td>403</td>
</tr>
<tr>
<td>Yicong Liang, Qing Li, and Tieyun Qian</td>
<td></td>
</tr>
<tr>
<td>ASAP: Towards Accurate, Stable and Accelerative Penetrating-Rank Estimation on Large Graphs</td>
<td>415</td>
</tr>
<tr>
<td>Xuefei Li, Weiren Yu, Bo Yang, and Jiajin Le</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 4C: Name Disambiguation</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficient Name Disambiguation in Digital Libraries</td>
<td>430</td>
</tr>
<tr>
<td>Jia Zhu, Gabriel Fung, and Liwei Wang</td>
<td></td>
</tr>
<tr>
<td>A Classification Framework for Disambiguating Web People Search Result Using Feedback</td>
<td>442</td>
</tr>
<tr>
<td>Ou Jin, Shenghua Bao, Zhong Su, and Yong Yu</td>
<td></td>
</tr>
<tr>
<td>Incorporating User Feedback into Name Disambiguation of Scientific Cooperation Network</td>
<td>454</td>
</tr>
<tr>
<td>Yuhua Li, Aiming Wen, Quan Lin, Ruixuan Li, and Zhengding Lu</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 5A: Performance</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-core vs. I/O Wall: The Approaches to Conquer and Cooperate</td>
<td>467</td>
</tr>
<tr>
<td>Yansong Zhang, Min Jiao, Zhanwei Wang, Shan Wang, and Xuan Zhou</td>
<td></td>
</tr>
<tr>
<td>W-Order Scan: Minimizing Cache Pollution by Application Software Level Cache Management for MMDB</td>
<td>480</td>
</tr>
<tr>
<td>Yansong Zhang, Min Jiao, Zhanwei Wang, Shan Wang, and Xuan Zhou</td>
<td></td>
</tr>
<tr>
<td>Index Structure for Cross-Class Query in Object Deputy Database</td>
<td>493</td>
</tr>
<tr>
<td>Yuwei Peng, Hefei Ge, Mao Ding, Zeqian Huang, and Chuanjian Wang</td>
<td></td>
</tr>
</tbody>
</table>
Session 5B: Data Mining (2)

Ensemble Pruning via Base-Classifier Replacement ............................. 505
   Huaping Guo and Ming Fan

Efficient Approximate Similarity Search Using Random Projection Learning ........................................ 517
   Peisen Yuan, Chaofeng Sha, Xiaoling Wang, Bin Yang, and Aoying Zhou

Informed Prediction with Incremental Core-Based Friend Cycle Discovering ........................................ 530
   Yue Wang, Weijing Huang, Wei Chen, Tengjiao Wang, and Dongqing Yang

Session 5C: Temporal Data

Early Prediction of Temporal Sequences Based on Information Transfer ....................................................... 542
   Ning Yang, Jian Peng, Yu Chen, and Changjie Tang

Mining Event Temporal Boundaries from News Corpora through Evolution Phase Discovery ........................ 554
   Liang Kong, Rui Yan, Han Jiang, Yan Zhang, Yan Gao, and Li Fu

Event Detection over Live and Archived Streams .................................... 566
   Shanglian Peng, Zhanhuai Li, Qiang Li, Qun Chen, Wei Pan, Hailong Liu, and Yanming Nie

Session 6A: XML

Renda-RX: A Benchmark for Evaluating XML-Relational Database System ...................................................... 578
   Xiao Zhang, Kuicheng Liu, Li Zou, Xiaoyong Du, and Shan Wang

Energy-Conserving Fragment Methods for Skewed XML Data Access in Push-Based Broadcast .......................... 590
   Jingjing Wu, Peng Liu, Lu Gan, Yongrui Qin, and Weiwei Sun

Session 6B: Spatial Data

Evaluating Probabilistic Spatial-Range Closest Pairs Queries over Uncertain Objects ........................................ 602
   Mo Chen, Zixi Jia, Yu Gu, Ge Yu, and Chuanwen Li

Synthesizing Routes for Low Sampling Trajectories with Absorbing Markov Chains ...................................... 614
   Chengxuan Liao, Jiaheng Lu, and Hong Chen
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