The Ninth International Conference on Web-Age Information Management

WAIM 2008

Table of Contents

Welcome Message from the General Chairs ................................................................. xiii
Welcome Message from the Program Chairs .......................................................... xiv
Welcome Message from the Workshop Chairs ...................................................... xvi
Information about Zhangjiajie ................................................................................... xix
Organizing Committee ......................................................................................... xxiii
Program Committee .............................................................................................. xxv
Steering Committee ............................................................................................... xxvii
External Reviewers ............................................................................................... xxviii
About the WAIM 2008 Workshops ..................................................................... xxix
Keynote Addresses .................................................................................................. xxx
Tutorial Abstracts ..................................................................................................... xxxii
Industrial Report ..................................................................................................... xxxvii
Sponsors ................................................................................................................... xxxviii
Conference Program ............................................................................................... xxxix

The Ninth International Conference on Web-Age Information Management —
WAIM 2008

Research Session A1: Search

LG-ERM: An Entity-Level Ranking Mechanism for Deep Web Query ................................. 1
Yue Kou, Derong Shen, Ge Yu, and Tiezhen Nie

Weighting Links Using Lexical and Positional Analysis in Web Ranking .......................... 9
Yi Zhang, Yexin Wang, Lidong Bing, and Yan Zhang

Approximate Boolean + Ranking Query Answering Using Wavelets ............................. 17
Youngdae Kim and Seung-won Hwang
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-BITS : Social-Bookmarking Induced Topic Search</td>
<td>25</td>
</tr>
<tr>
<td>Tsubasa Takahashi and Hiroyuki Kitagawa</td>
<td></td>
</tr>
<tr>
<td>LCA-Based Keyword Search for Effectively Retrieving &quot;Information Unit&quot; from Web Pages</td>
<td>31</td>
</tr>
<tr>
<td>Xiaoming Song, Jianhua Feng, Guoliang Li, and Qin Hong</td>
<td></td>
</tr>
<tr>
<td><strong>Research Session B1: Ontology and Semantic Web</strong></td>
<td></td>
</tr>
<tr>
<td>Path Similarity Based Directory Ontology Matching</td>
<td>37</td>
</tr>
<tr>
<td>Qian Zhong, Juanzi Li, Jie Tang, Yi Li, and Lizhu Zhou</td>
<td></td>
</tr>
<tr>
<td>Using Ontology to Enhance Collaborative Recommendation Based on Community</td>
<td>45</td>
</tr>
<tr>
<td>Li Yu</td>
<td></td>
</tr>
<tr>
<td>Analyzing Distribution and Evolution of Research Interests by Term Extraction and Ontology Learning</td>
<td>50</td>
</tr>
<tr>
<td>Zhiqiang Gao, Wanying Zhu, Yuzhong Qu, and Zhisheng Huang</td>
<td></td>
</tr>
<tr>
<td>Approach to the Formal Representation of OWL-S Ontology Maintenance Requirements</td>
<td>56</td>
</tr>
<tr>
<td>Aihua Bao, Li Yao, Weiming Zhang, and Jinping Yuan</td>
<td></td>
</tr>
<tr>
<td>System II: A Native RDF Repository Based on the Hypergraph Representation for RDF Data Model</td>
<td>62</td>
</tr>
<tr>
<td>Gang Wu, Juanzi Li, Jianqiang Hu, and Kehong Wang</td>
<td></td>
</tr>
<tr>
<td><strong>Research Session C1: Query Processing</strong></td>
<td></td>
</tr>
<tr>
<td>Parallel Structural Join Algorithm on Shared-Memory Multi-Core Systems</td>
<td>70</td>
</tr>
<tr>
<td>Le Liu, Jianhua Feng, Guoliang Li, Qian Qian, and Jianhui Li</td>
<td></td>
</tr>
<tr>
<td>In-Network Execution of External Join for Sensor Networks</td>
<td>78</td>
</tr>
<tr>
<td>Yongxuan Lai, Yilong Chen, and Hong Chen</td>
<td></td>
</tr>
<tr>
<td>LOB: Bucket Based Index for Range Queries</td>
<td>86</td>
</tr>
<tr>
<td>Jieping Wang and Xiaoyong Du</td>
<td></td>
</tr>
<tr>
<td>A New Dynamic Hash Index for Flash-Based Storage</td>
<td>93</td>
</tr>
<tr>
<td>Xiang Li, Zhou Da, and Xiaofeng Meng</td>
<td></td>
</tr>
<tr>
<td>Efficient k-Closest-Pair Range-Queries in Spatial Databases</td>
<td>99</td>
</tr>
<tr>
<td>Shaojie Qiao, Changjie Tang, Jing Peng, Hongjun Li, and Shengqiao Ni</td>
<td></td>
</tr>
<tr>
<td><strong>Research Session A2: XML and Index Structures</strong></td>
<td></td>
</tr>
<tr>
<td>Exploiting Path Information for Syntax-Based XML Subtree Matching in RDBs</td>
<td>105</td>
</tr>
<tr>
<td>Wensxin Liang and Haruo Yokota</td>
<td></td>
</tr>
<tr>
<td>SIMP: Efficient XML Structural Index for Multiple Query Processing</td>
<td>113</td>
</tr>
<tr>
<td>Bo Zhang, Zhihua Geng, and Aoying Zhou</td>
<td></td>
</tr>
<tr>
<td>An Efficient Structural Index for Branching Path Queries</td>
<td>119</td>
</tr>
<tr>
<td>Yingjie Fan, Shuyun Wang, Chenghong Zhang, Haibing Ma, and Yunfa Hu</td>
<td></td>
</tr>
</tbody>
</table>
Effective Indices for Efficient Approximate String Search and Similarity Join .................................................................127
  Xuhui Liu, Guoliang Li, Jianhua Feng, and Lizhu Zhou
Efficient Processing of Complex Twig Pattern Matching ................................................................................................135
  Jingqing Zhu, Wei Wang, and Xiaofeng Meng

Research Session B2: Data Stream

Exploiting Application Semantics in Monitoring Real-Time Data Streams .................................................................141
  Hongya Wang, LihChyun Shu, Zhidong Qin, Xiaoqiang Liu, Jing Cong, and Hui Song
Approximate Validity of XML Streaming Data ..................................................................................................................149
  Huang Cheng, Li Jun, and Michel de Rougemont
Cost-Efficient Processing of Continuous Extreme Queries over Distributed Data Streams ..................................................157
  Li Tian, Peng Zou, Li Zhang, and Aiping Li
Advanced Star Coordinates ....................................................................................................................................................165
  Yang Sun, Jiuyang Tang, Daquan Tang, and Weidong Xiao

Research Session C2: Web Services and Service Applications

QoS-Aware Scheduling of Web Services ........................................................................................................................................171
  An Liu, Qing Li, Liusheng Huang, Mingjun Xiao, and Hai Liu
Modeling and Reasoning about Semantic Web Services Contract Using Description Logic ..................................................................................................................179
  Hai Liu, Qing Li, Naijie Gu, and An Liu
Modeling and Service Capability Evaluation for RFID Complex Event Processing ..................................................................................................................................................187
  Yu Gu, Yanfei Lv, Ge Yu, and Jia Xu
Research on Web-Based Multi-Agent System for Aeroengine Fault Diagnosis .................................................................195
  Jianrong Wang, Tianbiao Yu, Wanshan Wang, and Ge Yu

Research Session A3: Classification

An Algorithm for Classifying Articles and Patent Documents Using Link Structure ........................................................................................................................................203
  Kishore Varma Indukuri, Pranav Mirajkar, and Ashish Sureka
Automatic Web Query Classification Using Large Unlabeled Web Pages ..................................................................................................................211
  Jingbo Yu and Na Ye
Self-Switching Classification Framework for Titled Documents ..................................................................................................................216
  Hang Guo, Lizhu Zhou, and Ling Feng
Subject-Oriented Classification Based on Scale Probing in the Deep Web ...............................................................................224
  Tiezheng Nie, Derong Shen, Ge Yu, and Yue Kou
Research Session B3: KNN and Top-N Queries
Towards Efficient and Flexible KNN Query Processing in Real-Life Road Networks ................................................................. 230
  Yang Lu, Bin Cui, Jiakui Zhao, Hua Lu, and Jialie Shen
Continuous Reverse Nearest Neighbor Queries on Moving Objects in Road Networks ................................................................. 238
  Huan-Liang Sun, Chao Jiang, Jun-Ling Liu, and Limei Sun
Evaluation of a Stream of Top-N Selection Queries in Relational Databases ................................................................. 246
  Liang Zhu, Chunnian Liu, Yanchao Feng, and Shenda Ji
A Cost-Efficient Method for Continuous Top-k Processing over Data Stream ................................................................. 254
  Li Zhang, Li Tian, Peng Zou, and Aiping Li

Research Session C3: Information Retrieval
A Query Dependent Approach to Learning to Rank for Information Retrieval ................................................................. 262
  Weijian Ni, Yalou Huang, and Maoqiang Xie
PerRank: Personalized Rank Retrieval with Categorical and Numerical Attributes ................................................................. 270
  Sangkyum Kim, Jaebum Kim, Younhee Ko, Seung-won Hwang, and Jiawei Han
Categorizing Queries by Topic Directory ................................................................. 278
  Miao He, Michal Cutler, and Kelvin Wu
A New Performance Benchmark for Content-Based 3D Model Retrieval ................................................................. 285
  Jinjie Lin, Yubin Yang, Tong Lu, Jiabin Ruan, and Wei Wei

Research Session A4: Social Networks and Community
Expertise Search in a Time-Varying Social Network ................................................................. 293
  Yize Li and Jie Tang
REC: A Novel Model to Rank Experts in Communities ................................................................. 301
  Chen Lin, Haofeng Zhou, Zhenhua Huang, and Wei Wang
Recommendation over a Heterogeneous Social Network ................................................................. 309
  Jing Zhang, Jie Tang, Bangyong Liang, Zi Yang, Sijie Wang, Jingjing Zuo, and Juanzi Li
Co-Clustering Tags and Social Data Sources ................................................................. 317
  Eirini Giannakidou, Vassiliki Koutsonikola, Athena Vakali, and Yiannis Kompatsiaris

Research Session B4: Privacy and Security
Set-Expression Based Method for Effective Privacy Preservation ................................................................. 325
  Yu Liu, Dapeng Lv, Yang Ye, Jianhua Feng, and Qin Hong
A New Approach for Detecting Anonymity of Patterns ................................................................. 333
  Zhihui Wang, Wei Wang, and Baile Shi
A Personalized (?,k)-Anonymity Model ................................................................. 341
  Xiaojun Ye, Yawei Zhang, and Ming Liu
Scientific Workflow Provenance Querying with Security Views ................................................................. 349
   Artem Chebotko, Seunghan Chang, Shiyong Lu, Farshad Fotouhi, and Ping Yang

Research Session C4: Data Mining

Name Disambiguation Using Atomic Clusters ........................................................................................ 357
   Feng Wang, Juanzi Li, Jie Tang, Jing Zhang, and Kehong Wang
Fuzzy K-Means with Variable Weighting in High Dimensional Data Analysis ........................................ 365
   Qiang Wang, Yunming Ye, and Joshua Zhexue Huang
Outlier Detection for Transaction Databases Using Association Rules .................................................. 373
   Kazuyo Narita and Hiroyuki Kitagawa
Scalable Community Discovery of Large Networks ................................................................................. 381
   Zhemin Zhu, Chen Wang, Li Ma, Yue Pan, and Zhiming Ding

Research Session: Invited Papers

PLEDS: A Personalized Entity Detection System Based on Web Log Mining Techniques .......................................................... 389
   Kathleen Tsoukalas, Bin Zhou, Jian Pei, and Davor Cubranic
Temperature-Aware Scheduling: When is System-Throttling Good Enough? .......................................... 397
   Deepak Rajan and Philip S. Yu
Managing Uncertain Data: Probabilistic Approaches ............................................................................... 405
   Wenjie Zhang, Xuemin Lin, Jian Pei, and Ying Zhang

Research Session B5: Correlation Analysis and Security

Finding Correlated Item Pairs through Efficient Pruning with a Given Threshold ..................................... 413
   Bo Wang, Liang Su, Aiping Li, and Peng Zou
Co-Occurring Patterns of Amino Acid Physicochemical Properties in Proteins ..................................... 421
   Gouchol Pok and Keun Ho Ryu
COCA: More Accurate Multidimensional Histograms out of More Accurate Correlations Detection ........... 429
   Wei Cao, Xiongpai Qin, and Shan Wang
Research on Malicious Transaction Processing Method of Database System ......................................... 435
   Chi Chen, Deng-guo Feng, Min Zhang, De-sheng Zhang, and He-qun Xian

Research Session C5: Distributed Systems and User Profile

Research on Communication-Efficient Method for Distributed Threshold Monitoring ......................................................... 441
   Li Tian, Peng Zou, Feng Wu, and AiPing Li
Reducing Data Decryption Cost by Broadcast Encryption and Account Assignment for Web Applications .................................................................................................................. 449
   Junpei Kawamoto, Qiang Ma, and Masatoshi Yoshikawa
Churn Impact on Replicated Data Duration in Structured P2P Networks ..................................................455
   Guangping Xu, Wenhui Ma, Gang Wang, Xiaoguang Liu, and Jing Liu

An Adaptive User Profile Based on Memory Model ..........................................................................................461
   Dingming Wu, Dongyan Zhao, and Xue Zhang

Demo Sessions

PEDS-VM: A Variation Management Prototype for Pattern Evolving Data Streams ..................................................469
   Xinbiao Zhou, Gaoshan Miao, Hongyan Li, Xin Wei, and Lu-An Tang

A Publicly Verifiable Dynamic Sharing Protocol for Data Secure Storage ..........................................................471
   Jia Yu, Fanyu Kong, Rong Hao, and Zhen Cheng

PASE: A Prototype for Ad-hoc Process-Aware Information System
Declaratively Constructing Environment ..............................................................................................................473
   Bin Zhou, Hongyan Li, Lei Wang, Huaqiang Zhang, and Meimei Li

Workshop on Internet Security

Workshop on Internet Security: Session B1

A Survey of Alert Fusion Techniques for Security Incident ..................................................................................475
   Tianning Zang, Xiaochun Yun, and Yongzheng Zhang

A Survey of Secure Routing in Ad Hoc Networks ..........................................................................................482
   Dongbin Wang, Mingzeng Hu, and Hui Zhi

An Admission Control Policy Based on Social Networks for P2P Systems ..................................................487
   Yuan Liang and Haixin Duan

Protecting Mobile Codes Using the Decentralized Label Model .........................................................................493
   Jian-Wei Ye, Bin-Xing Fang, Jin-Qiao Shi, and Zhi-Gang Wu

Workshop on Internet Security: Session B2

AMCAS: An Automatic Malicious Code Analysis System ..................................................................................501
   Jia Zhang, Yuntao Guan, Xiaoxin Jiang, Haixin Duan, and Jianping Wu

A Scalable Framework of Network Traffic Identification ..................................................................................508
   Zhang Shuzhuang, Fang Binxing, and Luo Hao

Counting Data Stream Based on Improved Counting Bloom Filter .................................................................512
   Zhijian Yuan, Jiajia Miao, Yan Jia, and Le Wang

Workshop on Internet Security: Session B3

Optimizing Traffic Classification Using Hybrid Feature Selection .....................................................................520
   Dai Lei, Yun Xiaochun, and Xiao Jun

Design and Implementation of Multi-Version Disk Backup Data Merging Algorithm ....................................526
   Guangjun Wu, Xiaochun Yun, and Shupeng Wang

Network Dynamic Risk Assessment Based on the Threat Stream Analysis ..................................................532
   Wencong Cheng, Xishan Xu, Yan Jia, and Peng Zou
Workshop on Internet Security: Session B4

Accelerating Multiple String Matching by Using Cache-Efficient Strategy .................................................. 539
  Jianlong Tan, Yanbing Liu, and Ping Liu
Methodology of Network Intrusion Detection System Penetration Testing .................................................. 546
  Hui Liu and Zhitang Li
Network Security Analysis Based on Security Status Space ............................................................................. 552
  Tao Zhang and Chong Wu
A Quasi Word-Based Compression Method of English Text Using Byte-Oriented Coding Scheme .................. 558
  Wei-ling Chang, Xiao-chun Yun, Bin-xing Fang, Shu-peng Wang, and Shu-hao Li

First International Workshop on Self-Adaptive Information Management

First International Workshop on Self-Adaptive Information Management: Session C1

Research on Web QoS Control Strategy Based on User Behaviour ................................................................. 564
  Xiaolei Guo, Zhiguang Shan, and Cong Wang
Component Based Context Model .................................................................................................................. 569
  Jian-Quan Ouyang, Bo Ding, Huai-Min Wang, and Dian-Xi Shi
A New Database Adaptive Buffer Replacement Algorithm Based on Call Graph Optimization ...................... 575
  Ya-Dan Deng, Ning Jing, and Wei Xiong

Workshop on Information System Security for the Web Age

Workshop on Information System Security for the Web Age: Session C2

Controlling Information Leakage of Fine-Grained Access Model in DBMSs ................................................... 583
  Hong Zhu, Jie Shi, Yuanzhen Wang, and Yucai Feng
DTAD: A Dynamic Taint Analysis Detector for Information Security ............................................................ 591
  Zhiwen Bai, Liming Wang, Jinglin Chen, Lin Xu, Jian Liu, and Xiyang Liu
A Kind of Formal Modelling for Network Security Situational Awareness Based on HMM ............................. 598
  Ying Liang, Huiqiang Wang, and Yonggang Pang
An Efficient Algorithm of Role Mapping across Security Domains in Data-Sharing Environments ................. 606
  Wang Xiaoqing, Gu Tianyang, Guo Yong, Zheng Yanxing, and Zong Jianjian

Workshop on Information System Security for the Web Age: Session C3

An Improved Method of Hybrid Worm Simulation .......................................................................................... 612
  Qi Liu, Jiwu Jing, and Yuewu Wang
A New-Style Domain Integrating Management of Windows and UNIX ......................................................... 619
  Shengli Liu, Wenbing Wang, and Yuefei Zhu
A Prediction Model of DoS Attack's Distribution Discrete Probability ........................................................................625
  Wentao Zhao, Jianping Yin, and Jun Long
SOA-Based Service Recovery Framework .................................................................................................................629
  Jie-Ying Chen, Yong-Jun Wang, and Yi Xiao

Workshop on Information System Security for the Web Age: Session C4

A Strategy Selection Model for Network Survivability Based on Fuzzy Matrix
Game ........................................................................................................................................................................636
  Guosheng Zhao, Huiqiang Wang, and Jian Wang
Measuring Network Vulnerability Based on Pathology ...............................................................................................640
  Yulong Wang, Fangchun Yang, and Qibo Sun
Data-Flow Based Analysis of Java Bytecode Vulnerability ........................................................................................647
  Gang Zhao, Hua Chen, and Dongxia Wang
ASG — Automated Signature Generation for Worm-Like P2P Traffic Patterns ..........................................................654
  Fengtao Xiao, HuaPing Hu, Xin Chen, and Bo Liu
Research of a Secure File System for Protection of Intellectual Property Right ......................................................661
  Jun Ma, Jiangchun Ren, Zhiying Wang, and Yaokai Zhu

Author Index ................................................................................................................................................................666