

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

Yoshiharu Ishikawa Jianzhong Li
Wei Wang Rui Zhang Wenjie Zhang (Eds.)

Web Technologies and Applications

15th Asia-Pacific Web Conference, APWeb 2013
Sydney, Australia, April 4-6, 2013
Proceedings



Springer

Volume Editors

Yoshiharu Ishikawa
Nagoya University
Graduate School of Information Science
Nagoya 464-8601, Japan
E-mail: ishikawa@is.nagoya-u.ac.jp

Jianzhong Li
Harbin Institute of Technology
Department of Computer Science and Technology
Harbin 150006, China
E-mail: ljz@mail.banner.com

Wei Wang
Wenjie Zhang
University of New South Wales
School of Computer Science and Engineering
Sydney, NSW 2052, Australia
E-mail: {weiw, zhangw}@cse.unsw.edu.au

Rui Zhang
University of Melbourne
Department of Computing and Information Systems
Melbourne, VIC 3052, Australia
E-mail: rui@csse.unimelb.edu.au

ISSN 0302-9743
ISBN 978-3-642-37400-5
DOI 10.1007/978-3-642-37401-2
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349
e-ISBN 978-3-642-37401-2

Library of Congress Control Number: 2013934117

CR Subject Classification (1998): H.2.8, H.2, H.3, H.5, H.4, J.1, K.4, I.2

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

© Springer-Verlag Berlin Heidelberg 2013

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)

Message from the General Chairs

Welcome to APWeb 2013, the 15th Edition of the Asia Pacific Web Conference. APWeb is a leading international conference on research, development, and applications of Web technologies, database systems, information management, and software engineering, with a focus on the Asia-Pacific region. Previous APWeb conferences were held in Kunming (2012), Beijing (2011), Busan (2010), Suzhou (2009), Shenyang (2008), Huangshan (2007), Harbin (2006), Shanghai (2005), Hangzhou (2004), Xi'an (2003), Changsha (2001), Xi'an (2000), Hong Kong (1999), and Beijing (1998).

The APWeb 2013 conference was, for the first time, held in Sydney, Australia — a city blessed with a temperate climate, a beautiful harbor, and natural attractions surrounding it. These proceedings collect the technical papers selected for presentation at the conference, during April 4–6, 2013.

The APWeb 2013 program featured a main conference, a special track, and four satellite workshops. The main conference had three keynotes by eminent researchers H.V. Jagadish from the University of Michigan, USA, Mark Sanderson from RMIT University, Australia, and Dan Suci from the University of Washington, USA. Three tutorials were offered by Haixun Wang, Microsoft Research Asia, China, Yuqing Wu, Indiana University, USA, George Fletcher, Eindhoven University of Technology, The Netherlands, and Lei Chen, Hong Kong University of Science and Technology, Hong Kong, China. The conference received 165 paper submissions from North America, South America, Europe, Asia, and Oceania. Each submitted paper underwent a rigorous review by at least three independent referees, with detailed review reports. Finally, 39 full research papers and 22 short research papers were accepted, from Australia, Bangladesh, Canada, China, India, Ireland, Italy, Japan, New Zealand, Saudi Arabia, Sweden, Norway, UK, and USA. The special track of “Distributed Processing of Graph, XML and RDF Data: Theory and Practice” was organized by Alfredo Cuzzocrea. The conference had four workshops

- The Second International Workshop on Data Management for Emerging Network Infrastructure (DaMEN 2013)
- International Workshop on Location-Based Data Management (LBDM 2013)
- International Workshop on Management of Spatial Temporal Data (MSTD 2013)
- International Workshop on Social Media Analytics and Recommendation Technologies (SMART 2013)

We were extremely excited with our strong Program Committee, comprising outstanding researchers in the APWeb research areas. We would like to extend our sincere gratitude to the Program Committee members and external reviewers. Last but not least, we would like to thank the sponsors, for their strong support

of this conference, making it a big success. Special thanks go to the Chinese University of Hong Kong, the University of New South Wales, Macquarie University, and the University of Sydney.

Finally, we wish to thank the APWeb Steering Committee, led by Xuemin Lin, for offering us the opportunity to organize APWeb 2013 in Sydney. We also wish to thank the host organization, the University of New South Wales, and Local Arrangements Committee and volunteers for their assistance in organizing this conference.

February 2013

Vijay Varadharajan
Jeffrey Xu Yu

Conference Organization

Conference Co-chairs

Vijay Varadharajan	Macquarie University, Australia
Jeffrey Xu Yu	Chinese University of Hong Kong, China

Program Committee Co-chairs

Yoshiharu Ishikawa	Nagoya University, Japan
Jianzhong Li	Harbin Institute of Technology, China
Wei Wang	University of New South Wales, Australia

Local Organization Co-chairs

Muhammad Aamir Cheema	University of New South Wales, Australia
Ying Zhang	University of New South Wales, Australia

Workshop Co-chairs

James Bailey	University of Melbourne, Australia
Xiaochun Yang	Northeastern University, China

Tutorial/Panel Co-chairs

Sanjay Chawla	University of Sydney, Australia
Xiaofeng Meng	Renmin University of China, China

Industrial Co-chairs

Marek Kowalkiewicz	SAP Research in Brisbane, Australia
Mukesh Mohania	IBM Research, India

Publication Co-chairs

Rui Zhang	University of Melbourne, Australia
Wenjie Zhang	University of New South Wales, Australia

Publicity Co-chairs

Alfredo Cuzzocrea
Jiaheng Lu

University of Calabria, Italy
Renmin University of China, China

Demo Co-chairs

Wook-Shin Han
Helen Huang

Kyungpook National University, Korea
University of Queensland, Australia

APWeb Steering Committee Liaison

Xuemin Lin

University of New South Wales, Australia

WISE Society Liaison

Yanchun Zhang

Victoria University, Australia

WAIM Steering Committee Liaison

Qing Li

City University of Hong Kong, China

Webmasters

Yu Zheng
Chen Chen

East China Normal University, China
University of New South Wales, Australia

Program Committee

Toshiyuki Amagasa
Djamal Benslimane
Jae-Woo Chang
Haiming Chen
Jinchuan Chen
David Cheung
Bin Cui
Alfredo Cuzzocrea
Ting Deng
Jianlin Feng
Yaokai Feng

University of Tsukuba
University of Lyon
Chonbuk National University
Chinese Academy of Sciences
Renmin University of China
The University of Hong Kong
Beijing University
ICAR-CNR & University of Calabria
Beihang University
Sun Yat-Sen University
Kyushu University

Sergio Flesca	University of Calabria
Hong Gao	Harbin Institute of Technology
Yunjun Gao	Zhejiang University
Stephane Grumbach	INRIA
Giovanna Guerrini	University of Genoa
Mohand-Said Hacid	University of Lyon 1
Qi He	IBM
Jun Hong	Queen's University Belfast
Michael Houle	National Institute of Informatics
Bin Hu	Lanzhou University
Zi Huang	University of Queensland
Jeong-Hyon Hwang	State University of New York at Albany
Seung-won Hwang	POSTECH
Mizuho Iwaihara	Waseda University
Adam Jatowt	Kyoto University
Cheqing Jin	East China Normal University
Anastasios Kementsietsidis	IBM T.J. Watson Research Center
Jin-Ho Kim	Kangwon National University
Markus Kirchberg	Institute for Infocomm Research
Manolis Koubarakis	University of Athens
Byung Lee	Vermont University
Chiang Lee	National Cheng Kung University
Jae-Gil Lee	KAIST
SangKeun Lee	Korea University
Carson Leung	University of Manitoba
Jianxin Li	Swinburne University
Xue Li	Queensland University
Yingshu Li	Georgia State University
Yinsheng Li	Fudan University
Zhanhuai Li	Northwestern Polytechnical University
Xiang Lian	University of Texas - Pan American
Guimei Liu	National University of Singapore
Mengchi Liu	Carleton University
Chengfei Liu	Swinburne University of Technology
Bo Luo	University of Kansas
Jiangang Ma	University of Adelaide
Qiang Ma	Kyoto University
Shuai Ma	Beihang University
Zakaria Maamar	Zayed University
Sanjay Madria	University of Missouri-Rolla
Weiyi Meng	State University of New York at Binghamton
Yang-Sae Moon	Kangwon National University
Michael Mrissa	University of Lyon
Akiyo Nadamoto	Konan University
Shinsuke Nakajima	Kyoto Sangyo University

Miyuki Nakano	University of Tokyo
Werner Nutt	Free University of Bozen-Bolzano
Satoshi Oyama	Hokkaido University
Helen Paik	University of New South Wales
Chaoyi Pang	CSIRO
Apostolos Papadopoulos	Aristotle University
Eric Pardede	Latrobe University
Sanghyun Park	Yonsei University
Zhiyong Peng	Wuhan University
Tieyun Qian	Wuhan University
Weining Qian	East China Normal University
Joao Rocha-Junior	Univ. Estadual de Feira de Santana
KeunHo Ryu	Chungbuk National University
Markus Schneider	University of Florida
Marc Scholl	Universität Konstanz
Aviv Segev	KAIST
Bin Shao	Microsoft Research Asia
Derong Shen	Northeastern University
Heng Shen	Queensland University
Jialie Shen	Singapore Management University
Timothy Shih	National Taipei University of Education
Lidan Shou	Zhejiang University
Shaoxu Song	Tsinghua University
Konstantinos Stefanidis	Norwegian University of Science and Technology
Kazutoshi Sumiya	University of Hyogo
Aixin Sun	Nanyang Technological University
Claudia Szabo	University of Adelaide
Changjie Tang	Sichuan University
Nan Tang	University of Edinburgh
David Taniar	Monash University
Alex Thomo	University of Victoria
Chaokun Wang	Tsinghua University
Daling Wang	Northeastern University
Fan Wang	Microsoft
Guoren Wang	Northeastern University
Hongzhi Wang	Harbin Institute of Technology
Hua Wang	University of Southern Queensland
Jianyong Wang	Tsinghua University
X. Wang	Fudan University
Xiaoling Wang	East China Normal University
Jef Wijzen	University of Mons-Hainaut
Jianliang Xu	Hong Kong Baptist University
Xiaochun Yang	Northeastern University
Jian Yin	Sun Yat-Sen University
Haruo Yokota	Tokyo Institute of Technology

Jian Yu	Swinburne University of Technology
Ming Zhang	Beijing University
Xiao Zhang	Renmin University of China
Baihua Zheng	Singapore Management University
Rui Zhou	Swinburne University
Shuigeng Zhou	Fudan University
Xiaofang Zhou	University of Queensland
Xuan Zhou	Renmin University
Zhaonian Zou	Harbin Institute of Technology

External Reviewers

Xuefei Li	Mahmoud Barhamgi
Hongyun Cai	Xian Li
Jingkuan Song	Yu Jiang
Yang Yang	Saurav Acharya
Xiaofeng Zhu	Syed K. Tanbeer
Scott Bourne	Hongda Ren
Yasser Salem	Wei Shen
Shi Feng	Zhenhua Song
Jianwei Zhang	Jianhua Yin
Kenta Oku	Liu Chen
Sukhwan Jung	Wei Song

An Overview of Probabilistic Databases

Dan Suciu

University of Washington
suciu@cs.washington.edu

<http://homes.cs.washington.edu/~suciu/>

A major challenge in modern data management is how to cope with uncertainty in the data. Uncertainty may exist because the data was extracted automatically from text, or was derived from the physical world such as RFID data, or was obtained by integrating several data sets using fuzzy matches, or may be the result of complex stochastic models. In a *probabilistic database* uncertainty is modeled using probabilities, and data management techniques are extended to cope with probabilistic data.

The main challenge is query evaluation. For each answer to the query, its degree of uncertainty is the probability that its lineage formula is true. Thus, query evaluation reduces to the problem of computing the probability of a Boolean formula. This problem generalizes model counting, which has been extensively studied in the AI and model checking literature. Today's state of the art methods for computing the exact probability are extensions of Davis Putnam's (DP) procedure [3, 2, 1, 4]. In probabilistic databases we can take a new approach, because here we can fix the query, and consider only the database as variable input (called *data complexity* [7]). An interesting dichotomy theorem holds: for every query, either its complexity is in PTIME or is $\#P$ -hard. A new probabilistic inference algorithm was needed in order to compute *all* PTIME queries, which uses the inclusion/exclusion principle [6]. This technique is missing from today's extensions of DP, yet necessary: without it one can show that probabilistic inference for certain simple PTIME queries requires exponential time [5].

References

1. Bacchus, F., Dalmao, S., Pitassi, T.: Algorithms and complexity results for $\#sat$ and bayesian inference. In: FOCS, pp. 340–351 (2003)
2. Birnbaum, E., Lozinskii, E.L.: The good old davis-putnam procedure helps counting models. *J. Artif. Int. Res.* 10(1), 457–477 (1999)
3. Davis, M., Logemann, G., Loveland, D.: A machine program for theorem-proving. *Commun. ACM* 5(7), 394–397 (1962)
4. Gomes, C.P., Sabharwal, A., Selman, B.: Model counting. In: *Handbook of Satisfiability*, pp. 633–654 (2009)
5. Jha, A.K., Suciu, D.: Knowledge compilation meets database theory: compiling queries to decision diagrams. In: ICDT, pp. 162–173 (2011)
6. Suciu, D., Olteanu, D., Ré, C., Koch, C.: *Probabilistic Databases*. In: *Synthesis Lectures on Data Management*. Morgan & Claypool Publishers (2011)
7. Vardi, M.Y.: The complexity of relational query languages (extended abstract). In: STOC, pp. 137–146 (1982)

Challenges with Big Data on the Web

H.V. Jagadish*

University of Michigan
jag@umich.edu

The promise of data-driven decision-making is now being recognized broadly, and there is growing enthusiasm for the notion of “Big Data.” This is true of Big Data in the enterprise, but this is even more true of Big Data on the web. While the promise of Big Data is real – for example, it is estimated that Google alone contributed 54 billion dollars to the US economy in 2009 – there is currently a wide gap between its potential and its realization.

Heterogeneity, scale, timeliness, complexity, and privacy problems with Big Data impede progress at all phases of the pipeline that can create value from data. The problems start right away during data acquisition, when the data tsunami requires us to make decisions, currently in an ad hoc manner, about what data to keep and what to discard, and how to store what we keep reliably with the right metadata. Much data today is not natively in structured format; for example, tweets and blogs are weakly structured pieces of text, while images and video are structured for storage and display, but not for semantic content and search: transforming such content into a structured format for later analysis is a major challenge. The value of data explodes when it can be linked with other data, thus data integration is a major creator of value. Since most data is directly generated in digital format today, we have the opportunity and the challenge both to influence the creation to facilitate later linkage and to automatically link previously created data. Data analysis, organization, retrieval, and modeling are other foundational challenges. Finally, presentation of the results and its interpretation by non-technical domain experts is crucial to extracting actionable knowledge.

A recent white paper[CCC12] mapped out the many challenges in this space. In this talk, drawing upon this white paper, I will present these challenges, particularly as they relate to the web. I will draw upon examples from database usability to show how size and complexity of Big Data can create difficulties for a user, and mention some directions of work in this regard. In particular, I will highlight how Big Data issues arise in surprising contexts, such as in browsing[SIGMOD12].

References

- [CCC12] Jagadish, H.V., et al: Challenges and Opportunities with Big Data, <http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf>
[SIGMOD12] Singh, M., Nandi, A., Jagadish, H.V.: Skimmer: rapid scrolling of relational query results. In: SIGMOD Conference, pp. 181–192 (2012)

* Supported in part by NSF under grant IIS-1017296.

Twenty Years of Web Search – Where to Next?

Mark Sanderson

School of Computer Science and Information Technology
RMIT University
GPO Box 2476, Melbourne 3001
Victoria, Australia

Abstract. This year, (2013) marks the 20th anniversary of the first public web search engine JumpStation launched in late 1993. For those who were around in those early days, it was becoming clear that an information provision and an information access revolution was on its way; though very few, if any would have predicted the state of the information society we have today. It is perhaps worth reflecting on what has been achieved in the field of information retrieval since these systems were first created, and consider what remains to be accomplished. It is perhaps easy to see the success of systems like Google and ask what else is there to achieve? However, in some ways, Google has it easy. In this talk, I will explain why Web search can be viewed as a relatively easy task and why other forms of search are much harder to perform accurately.

Search engines require a great deal of tuning, currently achieved empirically. The tuning carried out depends greatly on the types of queries submitted to a search engine and the types of document collections the queries will search over. It should be possible to study the population of queries and documents and predictively configure a search engine. However, there is little understanding in either the research or practitioner communities on how query and collection properties map to search engine configurations. I will present the some of the early work we have conducted at RMIT to start charting the problems in this particular space.

Another crucial challenge for search engine companies is how to ensure that users are delivered the best quality content. There is a growth in systems that recommend content based not only on queries, but also on user context. The problem is that the quality of these systems is highly variable; one way of tackling this problem is gathering context from a wider range of places. I will present some of the possible new approaches to providing that context to search engines. Here diverse social media, and advances in location technologies will be emphasized.

Finally, I will describe what I see as one of the more important challenges that face the whole of the information community, namely the penetration of computer systems to virtually every person on the planet and the challenges that such an expansion presents.

Table of Contents

Tutorials

Understanding Short Texts	1
<i>Hairun Wang</i>	
Managing the Wisdom of Crowds on Social Media Services	2
<i>Lei Chen</i>	
Search on Graphs: Theory Meets Engineering	3
<i>Yuqing Wu and George H.L. Fletcher</i>	

Distributed Processing:

A Simple XSLT Processor for Distributed XML	7
<i>Hiroki Mizumoto and Nobutaka Suzuki</i>	
Ontology Usage Network Analysis Framework	19
<i>Jamshaid Ashraf and Omar Khadeer Hussain</i>	
Energy Efficiency in W-Grid Data-Centric Sensor Networks via Workload Balancing	31
<i>Alfredo Cuzzocrea, Gianluca Moro, and Claudio Sartori</i>	
Update Semantics for Interoperability among XML, RDF and RDB: A Case Study of Semantic Presence in CISCO's Unified Presence Systems	43
<i>Muhammad Intizar Ali, Nuno Lopes, Owen Friel, and Alessandra Mileo</i>	

Graphs

GPU-Accelerated Bidirected De Bruijn Graph Construction for Genome Assembly	51
<i>Mian Lu, Qiong Luo, Bingqiang Wang, Junkai Wu, and Jiuxin Zhao</i>	
K Hops Frequent Subgraphs Mining for Large Attribute Graph	63
<i>Haiwei Zhang, Simeng Jin, Xiangyu Hu, Ying Zhang, Yanlong Wen, and Xiaojie Yuan</i>	
Privacy Preserving Graph Publication in a Distributed Environment	75
<i>Mingxuan Yuan, Lei Chen, Philip S. Yu, and Hong Mei</i>	

Correlation Mining in Graph Databases with a New Measure	88
<i>Md. Samiullah, Chowdhury Farhan Ahmed, Manziba Akanda Nishi, Anna Fariha, S M Abdullah, and Md. Rafiqul Islam</i>	
Improved Parallel Processing of Massive De Bruijn Graph for Genome Assembly	96
<i>Li Zeng, Jiefeng Cheng, Jintao Meng, Bingqiang Wang, and Shengzhong Feng</i>	
B3Clustering: Identifying Protein Complexes from Protein-Protein Interaction Network	108
<i>Eunjung Chin and Jia Zhu</i>	
Detecting Event Rumors on Sina Weibo Automatically	120
<i>Shengyun Sun, Hongyan Liu, Jun He, and Xiaoyong Du</i>	
Uncertain Subgraph Query Processing over Uncertain Graphs	132
<i>Wenjing Ruan, Chaokun Wang, Lu Han, Zhuo Peng, and Yiyuan Bai</i>	
Web Search and Web Mining	
Improving Keyphrase Extraction from Web News by Exploiting Comments Information	140
<i>Zhunchen Luo, Jintao Tang, and Ting Wang</i>	
A Two-Layer Multi-dimensional Trustworthiness Metric for Web Service Composition	151
<i>Han Jiao, Jixue Liu, Jiuyong Li, and Chengfei Liu</i>	
An Integrated Approach for Large-Scale Relation Extraction from the Web	163
<i>Naimdjon Takhirov, Fabien Duchateau, Trond Aalberg, and Ingeborg Sølberg</i>	
Multi-QoS Effective Prediction in Web Service Selection	176
<i>Zhongjun Liang, Hua Zou, Jing Guo, Fangchun Yang, and Rongheng Lin</i>	
Accelerating Topic Model Training on a Single Machine	184
<i>Mian Lu, Ge Bai, Qiong Luo, Jie Tang, and Jiuxin Zhao</i>	
Collusion Detection in Online Rating Systems	196
<i>Mohammad Allahbakhsh, Aleksandar Ignjatovic, Boualem Benatallah, Seyed-Mehdi-Reza Beheshti, Elisa Bertino, and Norman Foo</i>	
A Recommender System Model Combining Trust with Topic Maps	208
<i>Zukun Yu, William Wei Song, Xiaolin Zheng, and Deren Chen</i>	

A Novel Approach to Large-Scale Services Composition	220
<i>Hongbing Wang and Xiaojun Wang</i>	

XML, RDF Data and Query Processing

The Consistency and Absolute Consistency Problems of XML Schema Mappings between Restricted DTDs	228
<i>Hayato Kuwada, Kenji Hashimoto, Yasunori Ishihara, and Toru Fujiwara</i>	
Linking Entities in Unstructured Texts with RDF Knowledge Bases	240
<i>Fang Du, Yueguo Chen, and Xiaoyong Du</i>	
An Approach to Retrieving Similar Source Codes by Control Structure and Method Identifiers	252
<i>Yoshihisa Udagawa</i>	
Complementary Information for Wikipedia by Comparing Multilingual Articles	260
<i>Yuya Fujiwara, Yu Suzuki, Yukio Konishi, and Akiyo Nadamoto</i>	

Social Networks

Identification of Sybil Communities Generating Context-Aware Spam on Online Social Networks	268
<i>Faraz Ahmed and Muhammad Abulaish</i>	
Location-Based Emerging Event Detection in Social Networks	280
<i>Sayan Unankard, Xue Li, and Mohamed A. Sharaf</i>	
Measuring Strength of Ties in Social Network	292
<i>Dakui Sheng, Tao Sun, Sheng Wang, Ziqi Wang, and Ming Zhang</i>	
Finding Diverse Friends in Social Networks	301
<i>Syed Khairuzzaman Tanbeer and Carson Kai-Sang Leung</i>	
Social Network User Influence Dynamics Prediction	310
<i>Jingzuan Li, Wei Peng, Tao Li, and Tong Sun</i>	
Credibility-Based Twitter Social Network Analysis	323
<i>Jebrin Al-Sharawneh, Suku Sinnappan, and Mary-Anne Williams</i>	
Design and Evaluation of Access Control Model Based on Classification of Users' Network Behaviors	332
<i>Peipeng Liu, Jinqiao Shi, Fei Xu, Lihong Wang, and Li Guo</i>	
Two Phase Extraction Method for Extracting Real Life Tweets Using LDA	340
<i>Shuheii Yamamoto and Tetsuji Satoh</i>	

Probabilistic Queries

- A Probabilistic Model for Diversifying Recommendation Lists 348
Yutaka Kabutoya, Tomoharu Iwata, Hiroyuki Toda, and Hiroyuki Kitagawa
- A Probabilistic Data Replacement Strategy for Flash-Based Hybrid Storage System 360
Yanfei Lv, Xuexuan Chen, Guangyu Sun, and Bin Cui
- An Influence Strength Measurement via Time-Aware Probabilistic Generative Model for Microblogs 372
Zhaoyun Ding, Yan Jia, Bin Zhou, Jianfeng Zhang, Yi Han, and Chunfeng Yu
- A New Similarity Measure Based on Preference Sequences for Collaborative Filtering 384
Tianfeng Shang, Qing He, Fuzhen Zhuang, and Zhongzhi Shi

Multimedia and Visualization

- Visually Extracting Data Records from Query Result Pages 392
Neil Anderson and Jun Hong
- Leveraging Visual Features and Hierarchical Dependencies for Conference Information Extraction 404
Yue You, Guandong Xu, Jian Cao, Yanchun Zhang, and Guangyan Huang
- Aggregation-Based Probing for Large-Scale Duplicate Image Detection 417
Ziming Feng, Jia Chen, Xian Wu, and Yong Yu
- User Interest Based Complex Web Information Visualization 429
Shibli Saleheen and Wei Lai

Spatial-Temporal Databases

- FIMO: A Novel WiFi Localization Method 437
Yao Zhou, Leilei Jin, Cheqing Jin, and Aoying Zhou
- An Algorithm for Outlier Detection on Uncertain Data Stream 449
Keyan Cao, Donghong Han, Guoren Wang, Yachao Hu, and Ye Yuan
- Improved Spatial Keyword Search Based on IDF Approximation 461
Xiaoling Zhou, Yifei Lu, Yifang Sun, and Muhammad Aamir Cheema

Efficient Location-Dependent Skyline Retrieval with Peer-to-Peer Sharing	473
<i>Yingyuan Xiao, Yan Shen, Hongya Wang, and Xiaoye Wang</i>	

Data Mining and Knowledge Discovery

What Can We Get from Learning Resource Comments on Engineering Pathway	482
<i>Yunlu Zhang, Wei Yu, and Shijun Li</i>	
Tuned X-HYBRIDJOIN for Near-Real-Time Data Warehousing	494
<i>M. Asif Naeem</i>	
Exploiting Interaction Features in User Intent Understanding	506
<i>Vincenzo Deufemia, Massimiliano Giordano, Giuseppe Polese, and Luigi Marco Simonetti</i>	
Identifying Semantic-Related Search Tasks in Query Log	518
<i>Shuai Gong, Jinhua Xiong, Cheng Zhang, and Zhiyong Liu</i>	

Privacy and Security

Multi-verifier: A Novel Method for Fact Statement Verification	526
<i>Teng Wang, Qing Zhu, and Shan Wang</i>	
An Efficient Privacy-Preserving RFID Ownership Transfer Protocol	538
<i>Wei Xin, Zhi Guan, Tao Yang, Huiping Sun, and Zhong Chen</i>	
Fractal Based Anomaly Detection over Data Streams	550
<i>Xueqing Gong, Weining Qian, Shouke Qin, and Aoying Zhou</i>	
Preservation of Proximity Privacy in Publishing Categorical Sensitive Data	563
<i>Yujia Li, Xianmang He, Wei Wang, Huahui Chen, and Zhihui Wang</i>	

Performance

S2MART: Smart Sql to Map-Reduce Translators	571
<i>Narayan Gouraj, Prasanna Venkatesh Ravi, Mouniga V, and M.R. Sumalatha</i>	
MisDis: An Efficient Misbehavior Discovering Method Based on Accountability and State Machine in VANET	583
<i>Tao Yang, Wei Xin, Liangwen Yu, Yong Yang, Jianbin Hu, and Zhong Chen</i>	

A Scalable Approach for LRT Computation in GPGPU Environments	595
<i>Linsey Xiaolin Pang, Sanjay Chawla, Bernhard Scholz, and Georgina Wilcox</i>	
ASAWA: An Automatic Partition Key Selection Strategy	609
<i>Xiaoyan Wang, Jinchuan Chen, and Xiaoyong Du</i>	

Query Processing and Optimization

An Active Service Reselection Triggering Mechanism	621
<i>Ying Yin, Tiancheng Zhang, Bin Zhang, Gang Sheng, Yuhai Zhao, and Ming Li</i>	
Linked Data Informativeness	629
<i>Rouzbeh Meymandpour and Joseph G. Davis</i>	
Harnessing the Wisdom of Crowds for Corpus Annotation through CAPTCHA	638
<i>Yini Cao and Xuan Zhou</i>	
A Framework for OLAP in Column-Store Database: One-Pass Join and Pushing the Materialization to the End	646
<i>Yuean Zhu, Yansong Zhang, Xuan Zhou, and Shan Wang</i>	
A Self-healing Framework for QoS-Aware Web Service Composition via Case-Based Reasoning	654
<i>Guoqiang Li, Lejian Liao, Dandan Song, Jingang Wang, Fuzhen Sun, and Guangcheng Liang</i>	

The Second International Workshop on Data Management for Emerging Network Infrastructure

Workload-Aware Cache for Social Media Data	662
<i>Jinxian Wei, Fan Xia, Chaofeng Sha, Chen Xu, Xiaofeng He, and Aoying Zhou</i>	
Shortening the Tour-Length of a Mobile Data Collector in the WSN by the Method of Linear Shortcut	674
<i>Md. Shaifur Rahman and Mahmuda Naznin</i>	
Towards Fault-Tolerant Chord P2P System: Analysis of Some Replication Strategies	686
<i>Rafał Kapelko</i>	
A MapReduce-Based Method for Learning Bayesian Network from Massive Data	697
<i>Qiyu Fang, Kun Yue, Xiaodong Fu, Hong Wu, and Weiyi Liu</i>	

Practical Duplicate Bug Reports Detection in a Large Web-Based Development Community	709
<i>Liang Feng, Leyi Song, Chaofeng Sha, and Xueqing Gong</i>	

Selecting a Diversified Set of Reviews	721
<i>Wenzhe Yu, Rong Zhang, Xiaofeng He, and Chaofeng Sha</i>	

International Workshop on Social Media Analytics and Recommendation Technologies

Detecting Community Structures in Microblogs from Behavioral Interactions	734
<i>Ping Zhang, Kun Yue, Jin Li, Xiaodong Fu, and Weiyi Liu</i>	

Towards a Novel and Timely Search and Discovery System Using the Real-Time Social Web.	746
<i>Owen Phelan, Kevin McCarthy, and Barry Smyth</i>	

GWMF: Gradient Weighted Matrix Factorisation for Recommender Systems	758
<i>Nipa Chowdhury and Xiongcai Cai</i>	

Collaborative Ranking with Ranking-Based Neighborhood	770
<i>Chaosheng Fan and Zuoquan Lin</i>	

International Workshop on Management of Spatial Temporal Data

Probabilistic Top- k Dominating Query over Sliding Windows	782
<i>Xing Feng, Xiang Zhao, Yunjun Gao, and Ying Zhang</i>	

Distributed Range Querying Moving Objects in Network-Centric Warfare	794
<i>Bin Ge, Chong Zhang, Da-quan Tang, and Wei-dong Xiao</i>	

An Efficient Approach on Answering Top- k Queries with Grid Dominant Graph Index	804
<i>Aiping Li, Jinghu Xu, Liang Gan, Bin Zhou, and Yan Jia</i>	

A Survey on Clustering Techniques for Situation Awareness	815
<i>Stefan Mitsch, Andreas Müller, Werner Retschitzegger, Andrea Salfinger, and Wieland Schwinger</i>	

Parallel k -Skyband Computation on Multicore Architecture	827
<i>Xing Feng, Yunjun Gao, Tao Jiang, Lu Chen, Xiaoye Miao, and Qing Liu</i>	
Moving Distance Simulation for Electric Vehicle Sharing Systems for Jeju City Area	838
<i>Junghoon Lee and Gyung-Leen Park</i>	
Author Index	843