

# Table of Contents – Part I

## Keynote Talks

|                                                                                                         |   |
|---------------------------------------------------------------------------------------------------------|---|
| Enabling Real Time Data Analysis .....                                                                  | 1 |
| <i>Divesh Srivastava</i>                                                                                |   |
| A New Paradigm of Thinking and Architecture for Real-Time<br>Information Processing at Fingertips ..... | 2 |
| <i>Sang Kyun Cha</i>                                                                                    |   |

## Query Processing and Optimization

|                                                                                |    |
|--------------------------------------------------------------------------------|----|
| Improving the Accuracy of Histograms for Geographic Data Objects....           | 3  |
| <i>Hai Thanh Mai, Jaeho Kim, and Myoung Ho Kim</i>                             |    |
| Improving Online Aggregation Performance for Skewed Data<br>Distribution ..... | 18 |
| <i>Yuxiang Wang, Junzhou Luo, Aibo Song, Jiahui Jin, and Fang Dong</i>         |    |
| A Relational-Based Approach for Aggregated Search in Graph<br>Databases .....  | 33 |
| <i>Thanh-Huy Le, Haytham Elghazel, and Mohand-Saïd Hacid</i>                   |    |

## Data Semantics and Interoperability

|                                                                                                         |    |
|---------------------------------------------------------------------------------------------------------|----|
| Discovery of Keys from SQL Tables .....                                                                 | 48 |
| <i>Van Bao Tran Le, Sebastian Link, and Mozghan Memari</i>                                              |    |
| A Framework for Realizing Artifact-Centric Business Processes in<br>Service-Oriented Architecture ..... | 63 |
| <i>Kan Ngamakeur, Sira Yongchareon, and Chengfei Liu</i>                                                |    |
| Appearance-Order-Based Schema Matching .....                                                            | 79 |
| <i>Guohui Ding, Han Dong, and Guoren Wang</i>                                                           |    |

## XML and Semi-structured Data I

|                                                                                        |     |
|----------------------------------------------------------------------------------------|-----|
| Fast Result Enumeration for Keyword Queries on XML Data .....                          | 95  |
| <i>Junfeng Zhou, Zhifeng Bao, Ziyang Chen, and Tok Wang Ling</i>                       |     |
| Stars on Steroids: Fast Evaluation of Multi-source Star Twig Queries<br>in RDBMS ..... | 110 |
| <i>Erwin Leonardi, Sourav S. Bhowmick, and Fengrong Li</i>                             |     |

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| Updating Typical XML Views .....                                | 126 |
| <i>Jixue Liu, Chengfei Liu, Theo Haerder, and Jeffery Xu Yu</i> |     |

## XML and Semi-structured Data II

|                                                                                                |     |
|------------------------------------------------------------------------------------------------|-----|
| Partitioned Indexes for Entity Search over RDF Knowledge Bases .....                           | 141 |
| <i>Fang Du, Yueguo Chen, and Xiaoyong Du</i>                                                   |     |
| SINBAD: Towards Structure-Independent Querying of Common<br>Neighbors in XML Databases .....   | 156 |
| <i>Ba Quan Truong, Sourav S. Bhowmick, and Curtis Dyreson</i>                                  |     |
| Top-Down SLCA Computation Based on List Partition .....                                        | 172 |
| <i>Junfeng Zhou, Zhifeng Bao, Ziyang Chen, Guoxiang Lan,<br/>Xudong Lin, and Tok Wang Ling</i> |     |
| Efficiently Identifying Contributors for XML Keyword Search .....                              | 185 |
| <i>Ya-Hui Chang, Po-Hsien Chien, and Yu-Kai Chang</i>                                          |     |

## Data Mining and Knowledge Discovery I

|                                                                                          |     |
|------------------------------------------------------------------------------------------|-----|
| Semi-supervised Clustering of Graph Objects: A Subgraph Mining<br>Approach .....         | 197 |
| <i>Xin Huang, Hong Cheng, Jiong Yang, Jeffery Xu Yu,<br/>Hongliang Fei, and Jun Huan</i> |     |
| Plink-LDA: Using Link as Prior Information in Topic Modeling .....                       | 213 |
| <i>Huan Xia, Juanzi Li, Jie Tang, and Marie-Francine Moens</i>                           |     |
| AnyOut: Anytime Outlier Detection on Streaming Data .....                                | 228 |
| <i>Ira Assent, Philipp Kranen, Corinna Baldauf, and Thomas Seidl</i>                     |     |

## Data Mining and Knowledge Discovery II

|                                                                                    |     |
|------------------------------------------------------------------------------------|-----|
| Ensemble Based Positive Unlabeled Learning for Time Series<br>Classification ..... | 243 |
| <i>Minh Nhut Nguyen, Xiao-Li Li, and See-Kiong Ng</i>                              |     |
| Efficient Mining Regularly Frequent Patterns in Transactional<br>Databases .....   | 258 |
| <i>Md. Mamunur Rashid, Md. Rezaul Karim, Byeong-Soo Jeong, and<br/>Ho-Jin Choi</i> |     |
| Fast Tree-Based Mining of Frequent Itemsets from Uncertain Data .....              | 272 |
| <i>Carson Kai-Sang Leung and Syed K. Tanbeer</i>                                   |     |

## Data Mining and Knowledge Discovery III

|                                                                                      |     |
|--------------------------------------------------------------------------------------|-----|
| On the Decidability and Complexity of Identity Knowledge Representation .....        | 288 |
| <i>Klaus-Dieter Schewe and Qing Wang</i>                                             |     |
| Privacy Preserving Mining Maximal Frequent Patterns in Transactional Databases ..... | 303 |
| <i>Md. Rezaul Karim, Md. Mamunur Rashid, Byeong-Soo Jeong, and Ho-Jin Choi</i>       |     |
| Data Privacy against Composition Attack .....                                        | 320 |
| <i>Muzammil M. Baig, Jiuyong Li, Jixue Liu, Xiaofeng Ding, and Hua Wang</i>          |     |

## Privacy and Anonymity

|                                                                                              |     |
|----------------------------------------------------------------------------------------------|-----|
| Protecting Sensitive Relationships against Inference Attacks in Social Networks .....        | 335 |
| <i>Xiangyu Liu and Xiaochun Yang</i>                                                         |     |
| You Can Walk Alone: Trajectory Privacy-Preserving through Significant Stays Protection ..... | 351 |
| <i>Zheng Huo, Xiaofeng Meng, Haibo Hu, and Yi Huang</i>                                      |     |
| Semi-Edge Anonymity: Graph Publication when the Protection Algorithm Is Available .....      | 367 |
| <i>Mingxuan Yuan and Lei Chen</i>                                                            |     |

## Data Management in the Web

|                                                                                                                 |     |
|-----------------------------------------------------------------------------------------------------------------|-----|
| On-the-Fly Generation of Facets as Navigation Signs for Web Objects .....                                       | 382 |
| <i>Yu Kawano, Hiroaki Ohshima, and Katsumi Tanaka</i>                                                           |     |
| Searching for Quality Microblog Posts: Filtering and Ranking Based on Content Analysis and Implicit Links ..... | 397 |
| <i>Jan Vosecky, Kenneth Wai-Ting Leung, and Wilfred Ng</i>                                                      |     |
| HotDigg: Finding Recent Hot Topics from Digg .....                                                              | 414 |
| <i>Younghoon Kim and Kyuseok Shim</i>                                                                           |     |
| Assessing Web Article Quality by Harnessing Collective Intelligence ....                                        | 428 |
| <i>Jingyu Han, Xueping Chen, Kejia Chen, and Dawei Jiang</i>                                                    |     |

## Graphs and Data Mining Applications

|                                                                                       |     |
|---------------------------------------------------------------------------------------|-----|
| Context Sensitive Tag Expansion with Information Inference .....                      | 440 |
| <i>Hongyun Cai, Zi Huang, Jie Shao, and Xue Li</i>                                    |     |
| Efficient Subgraph Similarity All-Matching .....                                      | 455 |
| <i>Gaoping Zhu, Ke Zhu, Wenjie Zhang, Xuemin Lin, and Chuan Xiao</i>                  |     |
| Efficient Algorithm for Mining Correlated Protein-DNA Binding<br>Cores .....          | 470 |
| <i>Po-Yuen Wong, Tak-Ming Chan, Man-Hon Wong, and<br/>Kwong-Sak Leung</i>             |     |
| A Novel Approach for Finding Alternative Clusterings Using Feature<br>Selection ..... | 482 |
| <i>Vinh Thanh Tao and JongHyeok Lee</i>                                               |     |

## Temporal and Spatial Data I

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| General Spatial Skyline Operator .....                              | 494 |
| <i>Qianlu Lin, Ying Zhang, Wenjie Zhang, and Aiping Li</i>          |     |
| Top- $k$ Similarity Join over Multi-valued Objects .....            | 509 |
| <i>Wenjie Zhang, Jing Xu, Xin Liang, Ying Zhang, and Xuemin Lin</i> |     |
| Indexing Network Voronoi Diagrams .....                             | 526 |
| <i>Ugur Demiryurek and Cyrus Shahabi</i>                            |     |

## Temporal and Spatial Data II

|                                                                                                              |     |
|--------------------------------------------------------------------------------------------------------------|-----|
| On Efficient Reverse $k$ -Skyband Query Processing .....                                                     | 544 |
| <i>Qing Liu, Yunjun Gao, Gang Chen, Qing Li, and Tao Jiang</i>                                               |     |
| <i>Co-spatial Searcher</i> : Efficient Tag-Based Collaborative Spatial Search<br>on Geo-social Network ..... | 560 |
| <i>Jinzeng Zhang, Xiaofeng Meng, Xuan Zhou, and Dongqi Liu</i>                                               |     |
| Traffic Aware Route Planning in Dynamic Road Networks .....                                                  | 576 |
| <i>Jiajie Xu, Limin Guo, Zhiming Ding, Xiling Sun, and Chengfei Liu</i>                                      |     |
| Author Index .....                                                                                           | 593 |

# Table of Contents – Part II

## Top-k and Skyline Query Processing

|                                                                                        |    |
|----------------------------------------------------------------------------------------|----|
| Top-k Best Probability Queries on Probabilistic Data . . . . .                         | 1  |
| <i>Trieu Minh Nhut Le and Jinli Cao</i>                                                |    |
| Probabilistic Reverse Skyline Query Processing over Uncertain Data<br>Stream . . . . . | 17 |
| <i>Mei Bai, Junchang Xin, and Guoren Wang</i>                                          |    |
| Malleability-Aware Skyline Computation on Linked Open Data . . . . .                   | 33 |
| <i>Christoph Lofi, Ulrich Güntzer, and Wolf-Tilo Balke</i>                             |    |

## Information Retrieval and Recommendation

|                                                                                             |    |
|---------------------------------------------------------------------------------------------|----|
| Effective Next-Items Recommendation via Personalized Sequential<br>Pattern Mining . . . . . | 48 |
| <i>Ghim-Eng Yap, Xiao-Li Li, and Philip S. Yu</i>                                           |    |
| Scalable Top- <i>k</i> Keyword Search in Relational Databases . . . . .                     | 65 |
| <i>Yanwei Xu, Jihong Guan, and Yoshiharu Ishikawa</i>                                       |    |
| Composition and Efficient Evaluation of Context-Aware Preference<br>Queries . . . . .       | 81 |
| <i>Patrick Roocks, Markus Endres, Stefan Mandl, and Werner Kießling</i>                     |    |

## Indexing and Search Systems

|                                                                                       |     |
|---------------------------------------------------------------------------------------|-----|
| An Automaton-Based Index Scheme for On-Demand XML Data<br>Broadcast . . . . .         | 96  |
| <i>Weiwei Sun, Peng Liu, Jingjing Wu, Yongrui Qin, and Baihua Zheng</i>               |     |
| Colored Range Searching on Internal Memory . . . . .                                  | 111 |
| <i>Haritha Bellam, Saladi Rahul, and Krishnan Rajan</i>                               |     |
| Circle of Friend Query in Geo-Social Networks . . . . .                               | 126 |
| <i>Weimo Liu, Weiwei Sun, Chunan Chen, Yan Huang,<br/>Yinan Jing, and Kunjie Chen</i> |     |
| A Power Saving Storage Method That Considers Individual Disk<br>Rotation . . . . .    | 138 |
| <i>Satoshi Hikida, Hieu Hanh Le, and Haruo Yokota</i>                                 |     |

## Cloud Computing and Scalability

|                                                                                           |     |
|-------------------------------------------------------------------------------------------|-----|
| ComMapReduce: An Improvement of MapReduce with Lightweight Communication Mechanisms ..... | 150 |
| <i>Linlin Ding, Junchang Xin, Guoren Wang, and Shan Huang</i>                             |     |
| Halt or Continue: Estimating Progress of Queries in the Cloud .....                       | 169 |
| <i>Yingjie Shi, Xiaofeng Meng, and Bingbing Liu</i>                                       |     |
| Towards a Scalable, Performance-Oriented OLAP Storage Engine .....                        | 185 |
| <i>Todd Eavis and Ahmad Taleb</i>                                                         |     |

## Industrial Papers I: Memory-Based Query Processing

|                                                                                                        |     |
|--------------------------------------------------------------------------------------------------------|-----|
| Highly Scalable Speech Processing on Data Stream Management System .....                               | 203 |
| <i>Shunsuke Nishii and Toyotaro Suzumura</i>                                                           |     |
| EVIS: A Fast and Scalable Episode Matching Engine for Massively Parallel Data Streams .....            | 213 |
| <i>Shinichiro Tago, Tatsuya Asai, Takashi Katoh, Hiroaki Morikawa, and Hiroya Inakoshi</i>             |     |
| Real-Time Analysis of ECG Data Using Mobile Data Stream Management System .....                        | 224 |
| <i>Seokjin Hong, Rana Prasad Sahu, M.R. Srikanth, Supriya Mandal, Kyoung-Gu Woo, and Il-Pyung Park</i> |     |
| A Main Memory Based Spatial DBMS: Kairos .....                                                         | 234 |
| <i>Hyeok Han and Seong-il Jin</i>                                                                      |     |

## Industrial Papers II: Semantic and Decision Support Systems

|                                                                                              |     |
|----------------------------------------------------------------------------------------------|-----|
| Study on the International Standardization for the Semantic Metadata Mapping Procedure ..... | 243 |
| <i>Sungjoon Lim, Taesul Seo, Changhan Lee, and Soungsoo Shin</i>                             |     |
| Semantics and Usage Statistics for Multi-dimensional Query Expansion .....                   | 250 |
| <i>Raphaël Thollot, Nicolas Kuchmann-Beauger, and Marie-Aude Aufaure</i>                     |     |
| Hierarchy-Based Update Propagation in Decision Support Systems .....                         | 261 |
| <i>Haitang Feng, Nicolas Lumineau, Mohand-Saïd Hacid, and Richard Doms</i>                   |     |
| An Experiment with Asymmetric Algorithm: CPU Vs. GPU .....                                   | 272 |
| <i>Sujatha R. Upadhyaya and David Toth</i>                                                   |     |

## Demo Papers I: Social Data

|                                                                                 |     |
|---------------------------------------------------------------------------------|-----|
| Tag Association Based Graphical Password Using Image Feature Matching .....     | 282 |
| <i>Kyoji Kawagoe, Shinichi Sakaguchi, Yuki Sakon, and Hung-Hsuan Huang</i>      |     |
| ACARP: Author-Centric Analysis of Research Papers .....                         | 287 |
| <i>Weiming Zhang, Xueqing Gong, Weining Qian, and Aoying Zhou</i>               |     |
| iParticipate: Automatic Tweet Generation from Local Government Data .....       | 295 |
| <i>Christoph Lofi and Ralf Krestel</i>                                          |     |
| gRecs: A Group Recommendation System Based on User Clustering ....              | 299 |
| <i>Irene Ntoutsis, Kostas Stefanidis, Kjetil Norvag, and Hans-Peter Kriegel</i> |     |

## Demo Papers II: Data Mining

|                                                                                                                                 |     |
|---------------------------------------------------------------------------------------------------------------------------------|-----|
| PEACOD: A Platform for Evaluation and Comparison of Database Partitioning Schemes .....                                         | 304 |
| <i>Xiaoyan Guo, Jidong Chen, Yu Cao, and Mengdong Yang</i>                                                                      |     |
| Stream Data Mining Using the MOA Framework .....                                                                                | 309 |
| <i>Philipp Kranen, Hardy Kremer, Timm Jansen, Thomas Seidl, Albert Bifet, Geoff Holmes, Bernhard Pfahringer, and Jesse Read</i> |     |
| Shot Classification Using Domain Specific Features for Movie Management .....                                                   | 314 |
| <i>Muhammad Abul Hasan, Min Xu, Xiangjian He, and Ling Chen</i>                                                                 |     |
| PA-Miner: Process Analysis Using Retrieval, Modeling, and Prediction .....                                                      | 319 |
| <i>Anca Maria Ivanescu, Philipp Kranen, Manfred Smieschek, Philip Driessen, and Thomas Seidl</i>                                |     |

## Panel

|                                                                                                                              |     |
|------------------------------------------------------------------------------------------------------------------------------|-----|
| Data Management Challenges and Opportunities in Cloud Computing (Abstract) .....                                             | 323 |
| <i>Kyuseok Shim, Sang Kyun Cha, Lei Chen, Wook-Shin Han, Divesh Srivastava, Katsumi Tanaka, Hwanjo Yu, and Xiaofang Zhou</i> |     |

Tutorials

Detecting Clones, Copying and Reuse on the Web (DASFAA 2012 Tutorial) ..... 324  
*Xin Luna Dong and Divesh Srivastava*

Query Processing over Uncertain and Probabilistic Databases ..... 326  
*Lei Chen and Xiang Lian*

Tutorial: Data Stream Mining and Its Applications (Abstract) ..... 328  
*Latifur Khan and Wei Fan*

Storing, Querying, Summarizing, and Comparing Molecular Networks: The State-of-the-Art ..... 330  
*Sourav S. Bhowmick and Boon-Siew Seah*

Author Index ..... 333