

2012 IEEE 28th International Conference on Data Engineering (ICDE 2012)

**Arlington, Virginia, USA
1 – 5 April 2012**

Pages 1-713



**IEEE Catalog Number: CFP12026-PRT
ISBN: 978-1-4673-0042-1**

2012 IEEE 28th International Conference on Data Engineering

ICDE 2012

Table of Contents

| | |
|--|--------|
| Message from the Program Chairs and the General | |
| Chair | xvii |
| Conference Organization | xix |
| Awards | xxviii |
| Sponsors | xxx |

Keynotes

| | |
|---|---|
| Viewing the Web as a Distributed Knowledge Base | 1 |
| <i>Serge Abiteboul, Émilien Antoine, and Julia Stoyanovich</i> | |
| How Different is Big Data? | 5 |
| <i>Surajit Chaudhuri</i> | |
| Accountability and Trust in Cooperative Information Systems | 6 |
| <i>Peter Druschel</i> | |

Panel Paper

| | |
|--|---|
| The Future of Scientific Data Bases | 7 |
| <i>Michael Stonebraker, Anastasia Ailamaki, Jeremy Kepner, and Alex Szalay</i> | |

Session 1: Privacy

| | |
|---|----|
| Privacy in Social Networks: How Risky is Your Social Graph? | 9 |
| <i>Cuneyt Akcora, Barbara Carminati, and Elena Ferrari</i> | |
| Differentially Private Spatial Decompositions | 20 |
| <i>Graham Cormode, Cecilia Procopiuc, Divesh Srivastava, Entong Shen, and Ting Yu</i> | |
| Differentially Private Histogram Publication | 32 |
| <i>Jia Xu, Zhenjie Zhang, Xiaokui Xiao, Yin Yang, and Ge Yu</i> | |
| Privacy-Preserving and Content-Protecting Location Based Queries | 44 |
| <i>Russell Paulet, Md. Golam Koasar, Xun Yi, and Elisa Bertino</i> | |

Session 2: Web 2.0 Applications

| | |
|--|----|
| GeoFeed: A Location Aware News Feed System | 54 |
| <i>Jie Bao, Mohamed F. Mokbel, and Chi-Yin Chow</i> | |
| Entity Search Strategies for Mashup Applications | 66 |
| <i>Stefan Endrullis, Andreas Thor, and Erhard Rahm</i> | |
| CI-Rank: Ranking Keyword Search Results Based on Collective Importance | 78 |
| <i>Xiaohui Yu and Huxia Shi</i> | |
| Temporal Analytics on Big Data for Web Advertising | 90 |
| <i>Badrish Chandramouli, Jonathan Goldstein, and Songyun Duan</i> | |

Session 3: Storage Management

| | |
|--|-----|
| Lookup Tables: Fine-Grained Partitioning for Distributed Databases | 102 |
| <i>Aubrey L. Tatarowicz, Carlo Curino, Evan P.C. Jones, and Sam Madden</i> | |
| Temporal Support for Persistent Stored Modules | 114 |
| <i>Richard T. Snodgrass, Dengfeng Gao, Rui Zhang, and Stephen W. Thomas</i> | |
| Energy Efficient Storage Management Cooperated with Large Data Intensive Applications | 126 |
| <i>Norifumi Nishikawa, Miyuki Nakano, and Masaru Kitsuregawa</i> | |
| ISOBAR Preconditioner for Effective and High-throughput Lossless Data Compression | 138 |
| <i>Eric R. Schendel, Ye Jin, Neil Shah, Jackie Chen, C.S. Chang, Seung-Hoe Ku, Stephane Ethier, Scott Klasky, Robert Latham, Robert Ross, and Nagiza F. Samatova</i> | |

Session 4: Data Streams Processing

| | |
|---|-----|
| Physically Independent Stream Merging | 150 |
| <i>Badrish Chandramouli, David Maier, and Jonathan Goldstein</i> | |
| A General Method for Estimating Correlated Aggregates over a Data Stream | 162 |
| <i>Srikanta Tirthapura and David P. Woodruff</i> | |
| Accuracy-Aware Uncertain Stream Databases | 174 |
| <i>Tingjian Ge and Fujun Liu</i> | |
| On Discovery of Traveling Companions from Streaming Trajectories | 186 |
| <i>Lu-An Tang, Yu Zheng, Jing Yuan, Jiawei Han, Alice Leung, Chih-Chieh Hung, and Wen-Chih Peng</i> | |

Session 5: Graphs

| | |
|--|-----|
| Iterative Graph Feature Mining for Graph Indexing | 198 |
| <i>Dayu Yuan, Prasenjit Mitra, Huiwen Yu, and C. Lee Giles</i> | |
| An Efficient Graph Indexing Method | 210 |
| <i>Xiaoli Wang, Xiaofeng Ding, Anthony K.H. Tung, Shanshan Ying, and Hai Jin</i> | |
| PRAGUE: Towards Blending Practical Visual Subgraph Query Formulation and Query Processing | 222 |
| <i>Changjiu Jin, Sourav S. Bhowmick, Byron Choi, and Shuigeng Zhou</i> | |
| Ego-centric Graph Pattern Census | 234 |
| <i>Walaa Eldin Moustafa, Amol Deshpande, and Lise Getoor</i> | |

Session 6: Uncertain and Probabilistic Databases

| | |
|--|-----|
| Searching Uncertain Data Represented by Non-axis Parallel Gaussian Mixture Models | 246 |
| <i>Katrin Haegler, Frank Fiedler, and Christian Böhm</i> | |
| Aggregate Query Answering on Possibilistic Data with Cardinality Constraints | 258 |
| <i>Graham Cormode, Divesh Srivastava, Entong Shen, and Ting Yu</i> | |
| Discovering Threshold-based Frequent Closed Itemsets over Probabilistic Data | 270 |
| <i>Yongxin Tong, Lei Chen, and Bolin Ding</i> | |
| Ranking Query Answers in Probabilistic Databases: Complexity and Efficient Algorithms | 282 |
| <i>Dan Olteanu and Hongkai Wen</i> | |

Session 7: Data Integration and Extraction

| | |
|---|-----|
| Joint Entity Resolution | 294 |
| <i>Steven Euijong Whang and Hector Garcia-Molina</i> | |
| A Self-Configuring Schema Matching System | 306 |
| <i>Eric Peukert, Julian Eberius, and Erhard Rahm</i> | |
| Incremental Detection of Inconsistencies in Distributed Data | 318 |
| <i>Wenfei Fan, Jianzhong Li, Nan Tang, and Wenyuan Yu</i> | |
| Recomputing Materialized Instances after Changes to Mappings and Data | 330 |
| <i>Todd J. Green and Zachary G. Ives</i> | |

Session 8: Spatio-Temporal Data Management

| | |
|--|-----|
| SWST: A Disk Based Index for Sliding Window Spatio-Temporal Data | 342 |
| <i>Manish Singh, Qiang Zhu, and H.V. Jagadish</i> | |
| Querying Uncertain Spatio-Temporal Data | 354 |
| <i>Tobias Emrich, Hans-Peter Kriegel, Nikos Mamoulis, Matthias Renz, and Andreas Züfle</i> | |

| | |
|---|-----|
| The Min-dist Location Selection Query | 366 |
| <i>Jianzhong Qi, Rui Zhang, Lars Kulik, Dan Lin, and Yuan Xue</i> | |

| | |
|--|-----|
| Bi-level Locality Sensitive Hashing for k-Nearest Neighbor Computation | 378 |
| <i>Jia Pan and Dinesh Manocha</i> | |

Session 9: Query Processing

| | |
|--|-----|
| Learning-based Query Performance Modeling and Prediction | 390 |
| <i>Mert Akdere, Ugur Çetintemel, Matteo Riondato, Eli Upfal, and Stanley B. Zdonik</i> | |

| | |
|--|-----|
| Parametric Plan Caching Using Density-Based Clustering | 402 |
| <i>Günes Aluç, David E. DeHaan, and Ivan T. Bowman</i> | |

| | |
|---|-----|
| Effective and Robust Pruning for Top-Down Join Enumeration Algorithms | 414 |
| <i>Pit Fender, Guido Moerkotte, Thomas Neumann, and Viktor Leis</i> | |

| | |
|---|-----|
| Towards Preference-aware Relational Databases | 426 |
| <i>Anastasios Arvanitis and Georgia Koutrika</i> | |

Session 10: Location Aware Data Processing

| | |
|---|-----|
| A Foundation for Efficient Indoor Distance-Aware Query Processing | 438 |
| <i>Hua Lu, Xin Cao, and Christian S. Jensen</i> | |

| | |
|--|-----|
| LARS: A Location-Aware Recommender System | 450 |
| <i>Justin J. Levandoski, Mohamed Sarwat, Ahmed Eldawy, and Mohamed F. Mokbel</i> | |

| | |
|---|-----|
| Approximate Shortest Distance Computing: A Query-Dependent Local Landmark Scheme | 462 |
| <i>Miao Qiao, Hong Cheng, Lijun Chang, and Jeffrey Xu Yu</i> | |

| | |
|---|-----|
| DESKS: Direction-Aware Spatial Keyword Search | 474 |
| <i>Guoliang Li, Jianhua Feng, and Jing Xu</i> | |

Session 11: Map-Reduce Based Data Processing

| | |
|---|-----|
| Extending Map-Reduce for Efficient Predicate-Based Sampling | 486 |
| <i>Raman Grover and Michael J. Carey</i> | |

| | |
|--|-----|
| Fuzzy Joins Using MapReduce | 498 |
| <i>Foto N. Afrati, Anish Das Sarma, David Menestrina, Aditya Parameswaran, and Jeffrey D. Ullman</i> | |

| | |
|---|-----|
| Parallel Top-K Similarity Join Algorithms Using MapReduce | 510 |
| <i>Younghoon Kim and Kyuseok Shim</i> | |

| | |
|--|-----|
| Load Balancing in MapReduce Based on Scalable Cardinality Estimates | 522 |
| <i>Benjamin Gufler, Nikolaus Augsten, Angelika Reiser, and Alfons Kemper</i> | |

Session 12: Social Media

| | |
|---|-----|
| Community Detection with Edge Content in Social Media Networks | 534 |
| <i>Guo-Jun Qi, Charu C. Aggarwal, and Thomas Huang</i> | |
| Cross Domain Search by Exploiting Wikipedia | 546 |
| <i>Chen Liu, Sai Wu, Shouxu Jiang, and Anthony K.H. Tung</i> | |
| Provenance-based Indexing Support in Micro-blog Platforms | 558 |
| <i>Junjie Yao, Bin Cui, Zijun Xue, and Qingyun Liu</i> | |
| Learning Stochastic Models of Information Flow | 570 |
| <i>Luke Dickens, Ian Molloy, Jorge Lobo, Pau-Chen Cheng, and Alessandra Russo</i> | |

Session 13: P2P and Distributed Processing

| | |
|---|-----|
| BestPeer++: A Peer-to-Peer Based Large-Scale Data Processing Platform | 582 |
| <i>Gang Chen, Tianlei Hu, Dawei Jiang, Peng Lu, Kian-Lee Tan, Hoang Tam Vo, and Sai Wu</i> | |
| Effective Data Density Estimation in Ring-Based P2P Networks | 594 |
| <i>Minqi Zhou, Heng Tao Shen, Xiaofang Zhou, Weining Qian, and Aoying Zhou</i> | |
| Processing of Rank Joins in Highly Distributed Systems | 606 |
| <i>Christos Doulkeridis, Akrivi Vlachou, Kjetil Nørvåg, Yannis Kotidis, and Neoklis Polyzotis</i> | |
| Load Balancing for MapReduce-based Entity Resolution | 618 |
| <i>Lars Kolb, Andreas Thor, and Erhard Rahm</i> | |

Session 14: XML and RDF Data Management

| | |
|--|-----|
| Mapping XML to a Wide Sparse Table | 630 |
| <i>Liang Jeff Chen, Philip A. Bernstein, Peter Carlin, Dimitrije Filipovic, Michael Rys, Nikita Shamgunov, James F. Terwilliger, Milos Todic, Sasa Tomasevic, and Dragan Tomic</i> | |
| Querying XML Data: As You Shape It | 642 |
| <i>Curtis E. Dyreson and Sourav S. Bhowmick</i> | |
| Branch Code: A Labeling Scheme for Efficient Query Answering on Trees | 654 |
| <i>Yanghua Xiao, Ji Hong, Wanyun Cui, Zhenying He, Wei Wang, and Guodong Feng</i> | |
| Scalable Multi-query Optimization for SPARQL | 666 |
| <i>Wangchao Le, Anastasios Kementsietsidis, Songyun Duan, and Feifei Li</i> | |

Session 15: Performance

| | |
|--|-----|
| GSLPI: A Cost-Based Query Progress Indicator | 678 |
| <i>Jiexing Li, Rimma V. Nehme, and Jeffrey Naughton</i> | |
| Micro-Specialization in DBMSes | 690 |
| <i>Rui Zhang, Richard T. Snodgrass, and Saumya Debray</i> | |
| Towards Multi-tenant Performance SLOs | 702 |
| <i>Willis Lang, Srinath Shankar, Jignesh M. Patel, and Ajay Kalhan</i> | |

| | |
|---|-----|
| Multi-version Concurrency via Timestamp Range Conflict Management | 714 |
| <i>David Lomet, Alan Fekete, Rui Wang, and Peter Ward</i> | |

Session 16: Data Extraction and Quality

| | |
|---|-----|
| Automatic Extraction of Structured Web Data with Domain Knowledge | 726 |
| <i>Nora Derouiche, Bogdan Cautis, and Talel Abdessalem</i> | |
| Discovering Conservation Rules | 738 |
| <i>Lukasz Golab, Howard Karloff, Flip Korn, Barna Saha, and Divesh Srivastava</i> | |
| Answering Why-not Questions on Top-k Queries | 750 |
| <i>Zhian He and Eric Lo</i> | |
| An Efficient Trie-based Method for Approximate Entity Extraction with Edit-Distance Constraints | 762 |
| <i>Dong Deng, Guoliang Li, and Jianhua Feng</i> | |

Session 17: Top-K Processing

| | |
|--|-----|
| On Top-k Structural Similarity Search | 774 |
| <i>Pei Lee, Laks V.S. Lakshmanan, and Jeffrey Xu Yu</i> | |
| Relevance Matters: Capitalizing on Less (Top-k Matching in Publish/Subscribe) | 786 |
| <i>Mohammad Sadoghi and Hans-Arno Jacobsen</i> | |
| Efficiently Monitoring Top-k Pairs over Sliding Windows | 798 |
| <i>Zhitao Shen, Muhammad Aamir Cheema, Xuemin Lin, Wenjie Zhang, and Haixun Wang</i> | |
| Processing and Notifying Range Top-k Subscriptions | 810 |
| <i>Albert Yu, Pankaj K. Agarwal, and Jun Yang</i> | |

Session 18: Similarity

| | |
|---|-----|
| Efficient Exact Similarity Searches Using Multiple Token Orderings | 822 |
| <i>Jongik Kim and Hongrae Lee</i> | |
| Efficient Graph Similarity Joins with Edit Distance Constraints | 834 |
| <i>Xiang Zhao, Chuan Xiao, Xuemin Lin, and Wei Wang</i> | |
| Parameter-Free Determination of Distance Thresholds for Metric Distance Constraints | 846 |
| <i>Shaoxu Song, Lei Chen, and Hong Cheng</i> | |
| Random Error Reduction in Similarity Search on Time Series: A Statistical Approach | 858 |
| <i>Wush Chi-Hsuan Wu, Mi-Yen Yeh, and Jian Pei</i> | |

Session 19: Text and Strings

| | |
|---|-----|
| Optimizing Statistical Information Extraction Programs over Evolving Text | 870 |
| <i>Fei Chen, Xixuan Feng, Christopher Ré, and Min Wang</i> | |

| | |
|--|-----|
| Approximate String Membership Checking: A Multiple Filter, Optimization-Based Approach | 882 |
| <i>Chong Sun, Jeffrey F. Naughton, and Siddharth Barman</i> | |
| On Text Clustering with Side Information | 894 |
| <i>Charu C. Aggarwal, Yuchen Zhao, and Philip S. Yu</i> | |
| Fast SLCA and ELCA Computation for XML Keyword Queries Based on Set Intersection | 905 |
| <i>Junfeng Zhou, Zhifeng Bao, Wei Wang, Tok Wang Ling, Ziyang Chen, Xudong Lin, and Jingfeng Guo</i> | |

Session 20: Query Processing II

| | |
|---|-----|
| Optimization of Massive Pattern Queries by Dynamic Configuration Morphing | 917 |
| <i>Nikolay Laptev and Carlo Zaniolo</i> | |
| Three-Level Processing of Multiple Aggregate Continuous Queries | 929 |
| <i>Shenoda Guirguis, Mohamed A. Sharaf, Panos K. Chrysanthis, and Alexandros Labrinidis</i> | |
| Accelerating Range Queries for Brain Simulations | 941 |
| <i>Farhan Tauheed, Laurynas Biveinis, Thomas Heinis, Felix Schürmann, Henry Markram, and Anastasia Ailamaki</i> | |
| Keyword Query Reformulation on Structured Data | 953 |
| <i>Junjie Yao, Bin Cui, Liansheng Hua, and Yuxin Huang</i> | |

Session 21: Data Mining

| | |
|---|------|
| Predicting Approximate Protein-DNA Binding Cores Using Association Rule Mining | 965 |
| <i>Po-Yuen Wong, Tak-Ming Chan, Man-Hon Wong, and Kwong-Sak Leung</i> | |
| Upgrading Uncompetitive Products Economically | 977 |
| <i>Hua Lu and Christian S. Jensen</i> | |
| Attribute-Based Subsequence Matching and Mining | 989 |
| <i>Yu Peng, Raymond Chi-Wing Wong, Liangliang Ye, and Philip S. Yu</i> | |
| Integrating Frequent Pattern Mining from Multiple Data Domains for Classification | 1001 |
| <i>Dhaval Patel, Wynne Hsu, and Mong Li Lee</i> | |

Session 22: Scientific Data, Analysis and Visualization

| | |
|---|------|
| Efficient Versioning for Scientific Array Databases | 1013 |
| <i>Adam Seering, Philippe Cudre-Mauroux, Samuel Madden, and Michael Stonebraker</i> | |
| Multidimensional Analysis of Atypical Events in Cyber-Physical Data | 1025 |
| <i>Lu-An Tang, Xiao Yu, Sangkyum Kim, Jiawei Han, Wen-Chih Peng, Yizhou Sun, Hector Gonzalez, and Sebastian Seith</i> | |
| HiCS: High Contrast Subspaces for Density-Based Outlier Ranking | 1037 |
| <i>Fabian Keller, Emmanuel Müller, and Klemens Böhm</i> | |

| | |
|---|------|
| Extracting Analyzing and Visualizing Triangle K-Core Motifs within Networks | 1049 |
| <i>Yang Zhang and Srinivasan Parthasarathy</i> | |

Session 23: Similarity Search and Detection

| | |
|--|------|
| Horizontal Reduction: Instance-Level Dimensionality Reduction for Similarity Search in Large Document Databases | 1061 |
| <i>Min Soo Kim, Kyu-Young Whang, and Yang-Sae Moon</i> | |
| Adaptive Windows for Duplicate Detection | 1073 |
| <i>Uwe Draisbach, Felix Naumann, Sascha Szott, and Oliver Wonneberg</i> | |
| Efficient Dual-Resolution Layer Indexing for Top-k Queries | 1084 |
| <i>Jongwuk Lee, Hyunsouk Cho, and Seung-won Hwang</i> | |
| Evaluating Probabilistic Queries over Uncertain Matching | 1096 |
| <i>Reynold Cheng, Jian Gong, David W. Cheung, and Jiefeng Cheng</i> | |

Session 24: Sensors Network and Trajectory

| | |
|--|------|
| Detecting Outliers in Sensor Networks Using the Geometric Approach | 1108 |
| <i>Sabbas Burdakis and Antonios Deligiannakis</i> | |
| Efficient Threshold Monitoring for Distributed Probabilistic Data | 1120 |
| <i>Mingwang Tang, Feifei Li, Jeff M. Phillips, and Jeffrey Jestes</i> | |
| Incorporating Duration Information for Trajectory Classification | 1132 |
| <i>Dhaval Patel, Chang Sheng, Wynne Hsu, and Mong Li Lee</i> | |
| Reducing Uncertainty of Low-Sampling-Rate Trajectories | 1144 |
| <i>Kai Zheng, Yu Zheng, Xing Xie, and Xiaofang Zhou</i> | |

Session 25: Error Reduction and Data Security

| | |
|---|------|
| Efficient Similarity Search over Encrypted Data | 1156 |
| <i>Mehmet Kuzu, Mohammad Saiful Islam, and Murat Kantarcioglu</i> | |
| Obfuscating the Topical Intention in Enterprise Text Search | 1168 |
| <i>Hwee Hwa Pang, Xiaokui Xiao, and Jialie Shen</i> | |
| Correlation Support for Risk Evaluation in Databases | 1180 |
| <i>Katrin Eisenreich, Jochen Adamek, Philipp Rösch, Volker Markl, and Gregor Hackenbroich</i> | |
| A Game-Theoretic Approach for High-Assurance of Data Trustworthiness in Sensor Networks | 1192 |
| <i>Hyo-Sang Lim, Gabriel Ghinita, Elisa Bertino, and Murat Kantarcioglu</i> | |

Seminar 1

| | |
|--|------|
| Data Management Issues on the Semantic Web | 1204 |
| <i>Oktie Hassanzadeh, Anastasios Kementsietsidis, and Yannis Velegarakis</i> | |

Seminar 2

| | |
|---|------|
| Discovering Multiple Clustering Solutions: Grouping Objects in Different Views of the Data | 1207 |
| <i>Emmanuel Müller, Stephan Günnemann, Ines Färber, and Thomas Seidl</i> | |

Seminar 3

| | |
|--|------|
| Detecting Clones, Copying and Reuse on the Web | 1211 |
| <i>Xin Luna Dong and Divesh Srivastava</i> | |

Seminar 4

| | |
|--|------|
| Mining Knowledge from Data: An Information Network Analysis Approach | 1214 |
| <i>Jiawei Han, Yizhou Sun, Xifeng Yan, and Philip S. Yu</i> | |

Seminar 5

| | |
|--|------|
| Emerging Graph Queries in Linked Data | 1218 |
| <i>Arijit Khan, Yinghui Wu, and Xifeng Yan</i> | |

Seminar 6

| | |
|--|------|
| Boolean Matrix Decomposition Problem: Theory, Variations and Applications to Data Engineering | 1222 |
| <i>Jaideep Vaidya</i> | |

Demo Group 1

| | |
|--|------|
| SMIX Live—A Self-Managing Index Infrastructure for Dynamic Workloads | 1225 |
| <i>Thomas Kissinger, Hannes Voigt, and Wolfgang Lehner</i> | |
| Multi-query Stream Processing on FPGAs | 1229 |
| <i>Mohammad Sadoghi, Rija Javed, Naif Tarafdar, Harsh Singh, Rohan Palaniappan, and Hans-Arno Jacobsen</i> | |
| EUDEMON: A System for Online Video Frame Copy Detection by Earth Mover’s Distance | 1233 |
| <i>Jia Xu, Qiushi Bai, Yu Gu, Anthony K.H. Tung, Guoren Wang, Ge Yu, and Zhenjie Zhang</i> | |
| A Dataset Search Engine for the Research Document Corpus | 1237 |
| <i>Meiyu Lu, Srinivas Bangalore, Graham Cormode, Marios Hadjieleftheriou, and Divesh Srivastava</i> | |

| | |
|--|------|
| AskFuzzy: Attractive Visual Fuzzy Query Builder | 1241 |
| <i>Keivan Kianmehr, Negar Koochakzadeh, and Reda Alhajj</i> | |
| F2DB: The Flash-Forward Database System | 1245 |
| <i>Ulrike Fischer, Frank Rosenthal, and Wolfgang Lehner</i> | |
| Provenance-Based Debugging and Drill-Down in Data-Oriented Workflows | 1249 |
| <i>Robert Ikeda, Junsang Cho, Charlie Fang, Semih Salihoglu, Satoshi Torikai, and Jennifer Widom</i> | |

Demo Group 2

| | |
|--|------|
| M3: Stream Processing on Main-Memory MapReduce | 1253 |
| <i>Ahmed M. Aly, Asmaa Sallam, Bala M. Gnanasekaran, Long-Van Nguyen-Dinh, Walid G. Aref, Mourad Ouzzani, and Arif Ghafoor</i> | |
| A Deep Embedding of Queries into Ruby | 1257 |
| <i>Torsten Grust and Manuel Mayr</i> | |
| Asking the Right Questions in Crowd Data Sourcing | 1261 |
| <i>Rubi Boim, Ohad Greenshpan, Tova Milo, Slava Novgorodov, Neoklis Polyzotis, and Wang-Chiew Tan</i> | |
| LotusX: A Position-Aware XML Graphical Search System with Auto-Completion | 1265 |
| <i>Chunbin Lin, Jiaheng Lu, Tok Wang Ling, and Bogdan Cautis</i> | |
| Efficient Top-k Keyword Search in Graphs with Polynomial Delay | 1269 |
| <i>Mehdi Kargar and Aijun An</i> | |
| TEDAS: A Twitter-based Event Detection and Analysis System | 1273 |
| <i>Rui Li, Kin Hou Lei, Ravi Khadiwala, and Kevin Chen-Chuan Chang</i> | |
| AutoDict: Automated Dictionary Discovery | 1277 |
| <i>Fei Chiang, Periklis Andritsos, Erkang Zhu, and Renée J. Miller</i> | |

Demo Group 3

| | |
|---|------|
| Trust and Share: Trusted Information Sharing in Online Social Networks | 1281 |
| <i>Barbara Carminati, Elena Ferrari, and Jacopo Girardi</i> | |
| Evaluation of Clusterings—Metrics and Visual Support | 1285 |
| <i>Elke Aichert, Sascha Goldhofer, Hans-Peter Kriegel, Erich Schubert, and Arthur Zimek</i> | |
| Horton: Online Query Execution Engine for Large Distributed Graphs | 1289 |
| <i>Mohamed Sarwat, Sameh Elnikety, Yuxiong He, and Gabriel Kliot</i> | |
| MXQuery with Hardware Acceleration | 1293 |
| <i>Peter M. Fischer and Jens Teubner</i> | |
| Data3—A Kinect Interface for OLAP Using Complex Event Processing | 1297 |
| <i>Steffen Hirte, Andreas Seifert, Stephan Baumann, Daniel Klan, and Kai-Uwe Sattler</i> | |
| Analyzing Query Optimization Process: Portraits of Join Enumeration Algorithms | 1301 |
| <i>Anisoara Nica, Ian Charlesworth, and Maysum Panju</i> | |

| | |
|--|------|
| DPCube: Releasing Differentially Private Data Cubes for Health Information | 1305 |
| <i>Yonghui Xiao, James Gardner, and Li Xiong</i> | |

Demo Group 4

| | |
|---|------|
| NYAYA: A System Supporting the Uniform Management of Large Sets of Semantic Data | 1309 |
| <i>Roberto De Virgilio, Giorgio Orsi, Letizia Tanca, and Riccardo Torlone</i> | |
| R2DB: A System for Querying and Visualizing Weighted RDF Graphs | 1313 |
| <i>Songling Liu, Juan P. Cedeño, K. Selçuk Candan, Maria Luisa Sapino, Shengyu Huang, and Xinsheng Li</i> | |
| Project Daytona: Data Analytics as a Cloud Service | 1317 |
| <i>Roger S. Barga, Jaliya Ekanayake, and Wei Lu</i> | |
| Interactive User Feedback in Ontology Matching Using Signature Vectors | 1321 |
| <i>Isabel F. Cruz, Cosmin Stroe, and Matteo Palmonari</i> | |
| DObjects+: Enabling Privacy-Preserving Data Federation Services | 1325 |
| <i>Pawel Jurczyk, Li Xiong, and Slawomir Goryczka</i> | |
| DRAGOON: An Information Accountability System for High-Performance Databases | 1329 |
| <i>Kyriacos E. Pavlou and Richard T. Snodgrass</i> | |
| Intuitive Interaction with Encrypted Query Execution in DataStorm | 1333 |
| <i>Ken Smith, Ameet Kini, William Wang, Chris Wolf, M. David Allen, and Andrew Sillers</i> | |

Industrial Session 1: Support for Large Scale Data Analytics

| | |
|--|------|
| Exploiting Common Subexpressions for Cloud Query Processing | 1337 |
| <i>Yasin N. Silva, Paul-Ake Larson, and Jingren Zhou</i> | |
| Vectorwise: A Vectorized Analytical DBMS | 1349 |
| <i>Marcin Zukowski, Mark van de Wiel, and Peter Boncz</i> | |
| Scalable and Numerically Stable Descriptive Statistics in SystemML | 1351 |
| <i>Yuanyuan Tian, Shirish Tatikonda, and Berthold Reinwald</i> | |

Industrial Session 2: Evolving Platforms for New Applications

| | |
|---|------|
| Earlybird: Real-Time Search at Twitter | 1360 |
| <i>Michael Busch, Krishna Gade, Brian Larson, Patrick Lok, Samuel Luckenbill, and Jimmy Lin</i> | |

| | |
|--|------|
| Data Infrastructure at LinkedIn | 1370 |
| <i>Aditya Auradkar, Chavdar Botev, Shirshanka Das, Dave De Maagd, Alex Feinberg, Phanindra Ganti, Lei Gao, Bhaskar Ghosh, Kishore Gopalakrishna, Brendan Harris, Joel Koshy, Kevin Krawez, Jay Kreps, Shi Lu, Sunil Nagaraj, Neha Narkhede, Sasha Pachev, Igor Perisic, Lin Qiao, Tom Quiggle, Jun Rao, Bob Schulman, Abraham Sebastian, Oliver Seeliger, Adam Silberstein, BBoris Shkolnik, Chinmay Soman, Roshan Sumbaly, Kapil Surlaker, Sajid Topiwala, Cuong Tran, Balaji Varadarajan, Jemiah Westerman, Zach White, David Zhang, and Jason Zhang</i> | |
| The Credit Suisse Meta-data Warehouse | 1382 |
| <i>Claudio Jossen, Lukas Blunschi, Magdalini Mori, Donald Kossmann, and Kurt Stockinger</i> | |

Industrial Session 3: Indexing, Updates and Processing

| | |
|---|------|
| Efficient Support of XQuery Update Facility in XML Enabled RDBMS | 1394 |
| <i>Zhen Hua Liu, Hui J. Chang, and Balasubramanyam Sthanikam</i> | |
| Making Unstructured Data SPARQL Using Semantic Indexing in Oracle Database | 1405 |
| <i>Souripriya Das, Seema Sundara, Matthew Perry, Jagannathan Srinivasan, Jayanta Banerjee, and Aravind Yalamanchi</i> | |
| A Meta-language for MDX Queries in eLog Business Solution | 1417 |
| <i>Sonia Bergamaschi, Matteo Interlandidi, Mario Longo, Laura Po, and Maurizio Vincini</i> | |

Author Index