

2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)

**Long Beach, California, USA
1 – 6 March 2010**

Pages 1-668



**IEEE Catalog Number: CFP10026-PRT
ISBN: 978-1-4244-5445-7**

Keynotes

- 1 **Large Scale Data Warehousing: Trends and Observations**
*Richard Winter*¹, *Pekka Kostamaa*²
¹*Winter Corporation, USA*; ²*Teradata, USA*
- 2 **DBMS: Lessons from the First 50 Years, Speculations for the Next 50**
Jeffrey F. Naughton, *University of Wisconsin-Madison, USA*
- 3 **How New is the Cloud?**
Donald Kossmann, *ETH Zürich, Switzerland*

■ indicates Short Paper

Research Session 1: KNN Queries

- 4 **K Nearest Neighbor Queries and KNN-Joins in Large Relational Databases (Almost) for Free**
Bin Yao, Feifei Li, Piyush Kumar, Florida State University, USA
- 16 **Quantile-Based KNN Over Multi-Valued Objects**
Wenjie Zhang, Xuemin Lin, Muhammad Aamir Cheema, Ying Zhang, Wei Wang, University of New South Wales, Australia
- 28 **Efficient Rank Based KNN Query Processing Over Uncertain Data**
Ying Zhang, Xuemin Lin, Gaoping Zhu, Wenjie Zhang, Qianlu Lin, University of New South Wales, Australia
-

Research Session 2: Distributed Data

- 40 **Reliable Storage and Querying for Collaborative Data Sharing Systems**
Nicholas E. Taylor, Zachary G. Ives, University of Pennsylvania, USA
- 52 **Strongly Consistent Replication for a Bargain**
Konstantinos Krikellas¹, Sameh Elnikety², Zografoula Vagena³, Orion Hodson²
¹University of Edinburgh, UK; ²Microsoft, UK; ³Concentra Consulting Ltd., UK
- 64 **Detecting Inconsistencies in Distributed Data**
Wenfei Fan, Floris Geerts, Shuai Ma, Heiko Müller, University of Edinburgh, UK

■ indicates Short Paper

Research Session 3: Stream Mining

- 76 **Optimal Load Shedding with Aggregates and Mining Queries**
Barzan Mozafari, Carlo Zaniolo, University of California at Los Angeles, USA
- 89 **Scheduling for Fast Response Multi-Pattern Matching Over Streaming Events**
Ying Yan¹, Jin Zhang¹, Ming-Chien Shan²
¹SAP, China; ²SAP, USA
- 101 **Discovery of Cross-Similarity in Data Streams ■**
Machiko Toyoda, Yasushi Sakurai, NTT, Japan
- 105 **Mining Distribution Change in Stock Order Streams ■**
*Xiaoyan Liu¹, Xindong Wu², Huaiqing Wang³, Rui Zhang¹, James Bailey¹,
Kotagiri Ramamohanarao¹*
*¹University of Melbourne, Australia; ²Hefei University of Technology, China; ³City University
of Hong Kong, China*
-

Research Session 4: Location Based Services

- 109 **TrajStore: An Adaptive Storage System for Very Large Trajectory Data Sets**
Philippe Cudre-Mauroux, Eugene Wu, Samuel R. Madden, MIT, USA
- 121 **C3: Concurrency Control on Continuous Queries Over Moving Objects**
Jing Dai, Chang-Tien Lu, Virginia Tech, USA
- 133 **Policy-Aware Sender Anonymity in Location Based Services**
Alin Deutsch¹, Richard Hull², Avinash Vyas³, Kevin Keliang Zhao¹
¹University of California at San Diego, USA; ²IBM, USA; ³Bell Labs Research, USA

■ indicates Short Paper

Research Session 5: Probabilistic Databases

- 145 **Approximate Confidence Computation in Probabilistic Databases**
Dan Olteanu¹, Jiewen Huang¹, Christoph Koch²
¹University of Oxford, UK; ²Cornell University, USA
- 157 **PIP: A Database System for Great and Small Expectations**
Oliver Kennedy, Christoph Koch, Cornell University, USA
- 169 **Generator-Recognizer Networks: A Unified Approach to Probabilistic Databases ■**
Ruiwen Chen, Yongyi Mao, Iluju Kiringa, University of Ottawa, Canada
- 173 **Probabilistic Declarative Information Extraction ■**
Daisy Zhe Wang¹, Eirinaios Michelakis¹, Michael J. Franklin¹, Minos Garofalakis², Joseph M. Hellerstein¹
¹University of California at Berkeley, USA; ²Technical University of Crete, Greece
-

Research Session 6: Spatial Indexing

- 177 **PARINET: A Tunable Access Method for In-Network Trajectories**
Julian Sandu Popa¹, Karine Zeitouni¹, Vincent Oria², Dominique Barth¹, Sandrine Vial¹
¹PRiSM, France; ²New Jersey Institute of Technology, USA
- 189 **Multi-Guarded Safe Zone: An Effective Technique to Monitor Moving Circular Range Queries**
Muhammad Aamir Cheema¹, Ljiljana Brankovic², Xuemin Lin¹, Wenjie Zhang¹, Wei Wang¹
¹University of New South Wales, Australia; ²University of Newcastle, Australia
- 201 **Geotagging with Local Lexicons to Build Indexes for Textually-Specified Spatial Data**
Michael D. Lieberman, Hanan Samet, Jagan Sankaranarayanan, University of Maryland at College Park, USA

■ indicates Short Paper

Research Session 7: Privacy Techniques

- 213 **On Optimal Anonymization for l^+ -Diversity**
Junqiang Liu, Ke Wang, Simon Fraser University, Canada
- 225 **Differential Privacy via Wavelet Transforms**
Xiaokui Xiao¹, Guozhang Wang², Johannes Gehrke²
¹Nanyang Technological University, Singapore; ²Cornell University, USA
- 237 **Efficient Verification of Shortest Path Search via Authenticated Hints**
Man Lung Yiu¹, Yimin Lin², Kyriakos Mouratidis²
¹Hong Kong Polytechnic University, China; ²Singapore Management University, Singapore
-

Research Session 8: Skyline Queries

- 249 **Evaluating Skylines in the Presence of Equijoins**
Wen Jin¹, Michael D. Morse¹, Jignesh M. Patel², Martin Ester³, Zengjian Hu³
¹University of Michigan, USA; ²University of Wisconsin-Madison, USA; ³Simon Fraser University, Canada
- 261 **Route Skyline Queries: A Multi-Preference Path Planning Approach**
Hans-Peter Kriegel, Matthias Renz, Matthias Schubert, LMU München, Germany
- 273 **Probabilistic Contextual Skylines**
Dimitris Sacharidis, Anastasios Arvanitis, Timos Sellis, Athena RC, Greece

■ indicates Short Paper

Research Session 9: Information Integration

- 285 **Schema Covering: A Step Towards Enabling Reuse in Information Integration**
Barna Saha¹, Ioana Stanoi², Kenneth L. Clarkson²
¹University of Maryland at College Park, USA; ²IBM, USA
- 297 **Managing Uncertainty of XML Schema Matching**
Reynold Cheng, Jian Gong, David W. Cheung, University of Hong Kong, China
- 309 **Propagating Updates Through XML Views Using Lineage Tracing**
Leonidas Fegaras, University of Texas at Arlington, USA
-

Research Session 10: Query Interfaces

- 321 **USHER: Improving Data Quality with Dynamic Forms** (*Best Student Paper Award*)
Kuang Chen¹, Harr Chen², Neil Conway¹, Joseph M. Hellerstein¹, Tapan S. Parikh¹
¹University of California at Berkeley, USA; ²MIT, USA
- 333 **Explaining Structured Queries in Natural Language**
Georgia Koutrika¹, Alkis Simitsis², Yannis E. Ioannidis³
¹Stanford University, USA; ²HP, USA; ³University of Athens, Greece
- 345 **ScoreFinder: A Method for Collaborative Quality Inference on User-Generated Content** ■
Yang Liao, Aaron Harwood, Kotagiri Ramamohanarao, University of Melbourne, Australia
- 349 **IQ^P: Incremental Query Construction, a Probabilistic Approach** ■
Elena Demidova¹, Xuan Zhou², Wolfgang Nejdl¹
¹L3S Research Center, Germany; ²CSIRO, Australia

■ indicates Short Paper

Research Session 11: Top-K Queries

- 353 **TASM: Top-k Approximate Subtree Matching** (*Best Paper Award*)
*Nikolaus Augsten*¹, *Denilson Barbosa*², *Michael Böhlen*¹, *Themis Palpanas*³
¹Free University of Bozen-Bolzano, Italy; ²University of Alberta, Canada; ³University of Trento, Italy
- 365 **Reverse Top-k Queries**
*Akrivi Vlachou*¹, *Christos Doulkeridis*¹, *Yannis Kotidis*², *Kjetil Nørnvåg*¹
¹NTNU, Norway; ²AUEB, Greece
- 377 **Top-K Aggregation Queries Over Large Networks** ■
*Xifeng Yan*¹, *Bin He*², *Feida Zhu*³, *Jiawei Han*⁴
¹University of California at Santa Barbara, USA; ²IBM, USA; ³Singapore Management University, Singapore; ⁴University of Illinois at Urbana-Champaign, USA
- 381 **TopCells: Keyword-Based Search of Top-k Aggregated Documents in Text Cube** ■
Bolin Ding, *Bo Zhao*, *Cindy Xide Lin*, *Jiawei Han*, *Chengxiang Zhai*, University of Illinois at Urbana-Champaign, USA

■ indicates Short Paper

Research Session 12: Workflow and Workload Management

- 385 **Optimizing ETL Workflows for Fault-Tolerance**
Alkis Simitsis, Kevin Wilkinson, Umeshwar Dayal, Malu Castellanos, HP, USA
- 397 **Q-Cop: Avoiding Bad Query Mixes to Minimize Client Timeouts Under Heavy Loads**
Sean Tozer, Tim Brecht, Ashraf Aboulnaga, University of Waterloo, Canada
- 409 **Admission Control Mechanisms for Continuous Queries in the Cloud ■**
*Lory Al Moakar¹, Panos K. Chrysanthis¹, Christine Chung², Shenoda Guirguis¹,
Alexandros Labrinidis¹, Panayiotis Neophytou¹, Kirk Pruhs¹*
¹University of Pittsburgh, USA; ²Connecticut College, USA
- 413 **Interaction-Aware Prediction of Business Intelligence Workload Completion Times ■**
Mumtaz Ahmad¹, Songyun Duan², Ashraf Aboulnaga¹, Shivnath Babu²
¹University of Waterloo, Canada; ²Duke University, USA

■ indicates Short Paper

Research Session 13: Indexing and Hashing

- 417 **Fast In-Memory XPath Search Using Compressed Indexes**
Diego Arroyuelo¹, Francisco Claude², Sebastian Maneth³, Veli Mäkinen⁴, Gonzalo Navarro⁵, Kim Nguyễn³, Jouni Sirén⁴, Niko Välimäki⁴
¹Yahoo!, Chile; ²University of Waterloo, Canada; ³NICTA, Australia; ⁴University of Helsinki, Finland; ⁵University of Chile, Chile
- 429 **Hashing Tree-Structured Data: Methods and Applications**
Shirish Tatikonda, Srinivasan Parthasarathy, Ohio State University, USA
- 441 **Estimating the Compression Fraction of an Index Using Sampling** ■
Stratos Idreos¹, Raghav Kaushik², Vivek Narasayya², Ravishankar Ramamurthy²
¹CWI, The Netherlands; ²Microsoft, USA
- 445 **The Hybrid-Layer Index: A Synergic Approach to Answering Top-k Queries in Arbitrary Subspaces** ■
Jun-Seok Heo¹, Junghoo Cho², Kyu-Young Whang¹
¹KAIST, Korea; ²University of California at Los Angeles, USA

■ indicates Short Paper

Research Session 14: Scientific Data Mining

- 449 **The Model-Summary Problem and a Solution for Trees**
*Biswanath Panda*¹, *Mirek Riedewald*², *Daniel Fink*³
¹Google, USA; ²Northeastern University, USA; ³Cornell University, USA
- 461 **Efficient and Accurate Discovery of Patterns in Sequence Datasets**
*Avrilia Floratou*¹, *Sandeep Tata*², *Jignesh M. Patel*¹
¹University of Wisconsin-Madison, USA; ²IBM, USA
- 473 **Mining Mutation Chains in Biological Sequences**
*Chang Sheng*¹, *Wynne Hsu*¹, *Mong Li Lee*¹, *Joo Chuan Tong*², *See-Kiong Ng*²
¹National University of Singapore, Singapore; ²Institute of Infocomm Research, Singapore
-

Research Session 15: Database Performance and Reliability

- 485 **Exploring Power-Performance Tradeoffs in Database Systems**
*Zichen Xu*¹, *Yi-Cheng Tu*¹, *Xiaorui Wang*²
¹University of South Florida, USA; ²University of Tennessee, USA
- 497 **Workload Driven Index Defragmentation**
Vivek Narasayya, *Manoj Syamala*, Microsoft, USA
- 509 **Impact of Disk Corruption on Open-Source DBMS**
Sriram Subramanian, *Yupu Zhang*, *Rajiv Vaidyanathan*, *Haryadi S. Gunawi*,
Andrea C. Arpaci-Dusseau, *Remzi H. Arpaci-Dusseau*, *Jeffrey F. Naughton*, University of
Wisconsin-Madison, USA

■ indicates Short Paper

Research Session 16: Spatial Databases

- 521 **Locating Mapped Resources in Web 2.0**
Dongxiang Zhang, Beng Chin Ooi, Anthony K.H. Tung, National University of Singapore, Singapore
- 533 **Preference Queries in Large Multi-Cost Transportation Networks**
Kyriakos Mouratidis¹, Yimin Lin¹, Man Lung Yiu²
¹Singapore Management University, Singapore; ²Hong Kong Polytechnic University, China
- 545 **Approximate String Search in Spatial Databases**
Bin Yao¹, Feifei Li¹, Marios Hadjieleftheriou², Kun Hou¹
¹Florida State University, USA; ²AT&T Labs Research, USA
-

Research Session 17: Sensor Networks

- 557 **Global Iceberg Detection Over Distributed Data Streams**
Haiquan Zhao¹, Ashwin Lall¹, Mitsunori Ogihara², Jun Xu¹
¹Georgia Institute of Technology, USA; ²University of Miami, USA
- 569 **Non-Dyadic Haar Wavelets for Streaming and Sensor Data**
Chetan Gupta, Choudur Lakshminarayan, Song Wang, Abhay Mehta, HP, USA
- 581 **Ratio Threshold Queries Over Distributed Data Sources ■**
Rajeev Gupta¹, Krithi Ramamritham², Mukesh Mohania¹
¹IBM, India; ²IIT Bombay, India
- 585 **Probabilistic Top-k Query Processing in Distributed Sensor Networks ■**
Mao Ye¹, Xingjie Liu¹, Wang-Chien Lee¹, Dik Lun Lee²
¹Pennsylvania State University, USA; ²Hong Kong University of Science & Technology, China

■ indicates Short Paper

Research Session 18: Query Optimization

- 589 **Polynomial Heuristics for Query Optimization**
Nicolas Bruno, César Galindo-Legaria, Milind Joshi, Microsoft, USA
- 601 **Optimized Query Evaluation Using Cooperative Sorts**
Yu Cao, Ramadhana Bramandia, Chee-Yong Chan, Kian-Lee Tan, National University of Singapore, Singapore
- 613 **Generating Code for Holistic Query Evaluation**
Konstantinos Krikellas, Stratis D. Viglas, Marcelo Cintra, University of Edinburgh, UK
-

Research Session 19: Graph Mining

- 625 **Finding Clusters in Subspaces of Very Large, Multi-Dimensional Datasets**
Robson L.F. Cordeiro¹, Agma J.M. Traina¹, Christos Faloutsos², Caetano Traina Jr.¹
¹Universidade de São Paulo, Brazil; ²Carnegie Mellon University, USA
- 637 **Similarity Search on Supergraph Containment**
Haichuan Shang¹, Ke Zhu¹, Xuemin Lin¹, Ying Zhang¹, Ryutaro Ichise²
¹University of New South Wales, Australia; ²National Institute of Informatics, Japan
- 649 **Finding Top-k Maximal Cliques in an Uncertain Graph ■**
Zhaonian Zou, Jianzhong Li, Hong Gao, Shuo Zhang, Harbin Institute of Technology, China
- 653 **Progressive Clustering of Networks Using Structure-Connected Order of Traversal ■**
Dustin Bortner, Jiawei Han, University of Illinois at Urbana-Champaign, USA

■ indicates Short Paper

Research Session 20: Parallel Processing

- 657 **Osprey: Implementing MapReduce-Style Fault Tolerance in a Shared-Nothing Distributed Database**
Christopher Yang, Christine Yen, Ceryen Tan, Samuel R. Madden, MIT, USA
- 669 **FPGA Acceleration for the Frequent Item Problem**
Jens Teubner, Rene Mueller, Gustavo Alonso, ETH Zürich, Switzerland
- 681 **Estimating the Progress of MapReduce Pipelines ■**
Kristi Morton, Abram Friesen, Magdalena Balazinska, Dan Grossman, University of Washington, USA
- 685 **Scalable Distributed-Memory External Sorting ■**
Mirko Rahn, Peter Sanders, Johannes Singler, Karlsruhe Institute of Technology, Germany

■ indicates Short Paper

Research Session 21: Keyword Search

- 689 **Supporting Top-K Keyword Search in XML Databases**
Liang Jeff Chen, Yannis Papakonstantinou, University of California at San Diego, USA
- 701 **Personalized Web Search with Location Preferences**
Kenneth Wai-Ting Leung¹, Dik Lun Lee¹, Wang-Chien Lee²
¹Hong Kong University of Science & Technology, China; ²Pennsylvania State University, USA
- 713 **Fuzzy Matching of Web Queries to Structured Data ■**
Tao Cheng¹, Hady W. Lauw², Stelios Paparizos²
¹University of Illinois at Urbana-Champaign, USA; ²Microsoft, USA
- 717 **Toward Industrial-Strength Keyword Search Systems Over Relational Data ■**
Akanksha Baid, Ian Rae, AnHai Doan, Jeffrey F. Naughton, University of Wisconsin-Madison, USA
-

Research Session 22: Query Processing

- 721 **Efficient Processing of Substring Match Queries with Inverted q-Gram Indexes**
Younghoon Kim¹, Kyoung-Gu Woo², Hyoungmin Park¹, Kyuseok Shim¹
¹Seoul National University, Korea; ²Samsung Electronics, Korea
- 733 **Progressive Result Generation for Multi-Criteria Decision Support Queries**
Venkatesh Raghavan, Elke A. Rundensteiner, Worcester Polytechnic Institute, USA
- 745 **Nb-GCLOCK: A Non-Blocking Buffer Management Based on the Generalized CLOCK**
Makoto Yui¹, Jun Miyazaki², Shunsuke Uemura³, Hayato Yamana¹
¹Waseda University, Japan; ²NAIST, Japan; ³Nara Sangyo University, Japan

■ indicates Short Paper

Research Session 23: Web and Collaborative Applications

- 757 **Effective Automated Object Matching**
Diego Zardetto¹, Monica Scannapieco¹, Tiziana Catarci²
¹Istituto Nazionale di Statistica, Italy; ²Università di Roma "La Sapienza", Italy
- 769 **Efficient Identification of Coupled Entities in Document Collections ■**
Nikos Sarkas¹, Albert Angel¹, Nick Koudas¹, Divesh Srivastava²
¹University of Toronto, Canada; ²AT&T Labs Research, USA
- 773 **On Supporting Effective Web Extraction ■**
Wook-Shin Han¹, Wooseong Kwak¹, Hwanjo Yu²
¹Kyungpook National University, Korea; ²POSTECH, Korea
- 776 **A Partial Persistent Data Structure to Support Consistency in Real-Time Collaborative Editing ■**
Qinyi Wu¹, Calton Pu¹, João Eduardo Ferreiar²
¹Georgia Institute of Technology, USA; ²Universidade de São Paulo, Brazil
- 780 **Detecting Bursty Events in Collaborative Tagging Systems ■**
Junjie Yao, Bin Cui, Yuxin Huang, Yanhong Zhou, Peking University, China

■ indicates Short Paper

Research Session 24: Scientific Databases

- 784 **Credibility-Enhanced Curated Database: Improving the Value of Curated Databases**
Qun Ni, Elisa Bertino, Purdue University, USA
- 796 **UV-Diagram: A Voronoi Diagram for Uncertain Data**
Reynold Cheng¹, Xike Xie¹, Man Lung Yiu², Jinchuan Chen³, Liwen Sun¹
¹University of Hong Kong, China; ²Hong Kong Polytechnic University, China; ³Renmin University of China, China
- 808 **Supporting Real-World Activities in Database Management Systems ■**
Mohamed Y. Eltabakh, Walid G. Aref, Ahmed K. Elmagarmid, Yasin N. Silva, Mourad Ouzzani, Purdue University, USA
- 812 **XML-Based Computation for Scientific Workflows ■**
Daniel Zinn¹, Shawn Bowers², Bertram Ludäscher¹
¹University of California at Davis, USA; ²Gonzaga University, USA

■ indicates Short Paper

Research Session 25: Tree Queries and Semi-Structured Databases

- 816 **ViewJoin: Efficient View-Based Evaluation of Tree Pattern Queries**
Ding Chen, Chee-Yong Chan, National University of Singapore, Singapore
- 828 **FlexPref: A Framework for Extensible Preference Evaluation in Database Systems**
Justin J. Levandoski, Mohamed F. Mokbel, Mohamed E. Khalefa, University of Minnesota, USA
- 840 **Optimal Tree Node Ordering for Child/Descendant Navigations ■**
Atsuyuki Morishima¹, Keishi Tajima², Masateru Tadaishi¹
¹University of Tsukuba, Japan; ²Kyoto University, Japan
- 844 **XMorph: A Shape-Polymorphic, Domain-Specific XML Data Transformation Language ■**
Curtis Dyreson¹, Sourav Bhowmick², Aswani Rao Jannu¹, Kirankanth Mallampalli¹, Shuohao Zhang³
¹Utah State University, USA; ²Nanyang Technological University, Singapore; ³Marvel, USA

■ indicates Short Paper

Research Session 26: Query Ranking and Database Testing

- 848 **Surrogate Ranking for Very Expensive Similarity Queries**
Fei Xu¹, Ravi Jampani¹, Mingxi Wu², Chris Jermaine¹, Tamer Kahveci¹
¹University of Florida, USA; ²Oracle, USA
- 860 **Semantic Ranking and Result Visualization for Life Sciences Publications**
Julia Stoyanovich, William Mee, Kenneth A. Ross, Columbia University, USA
- 872 **Ranked Queries Over Sources with Boolean Query Interfaces without Ranking Support** ■
Vagelis Hristidis¹, Yuheng Hu¹, Panagiotis G. Ipeirotis²
¹Florida International University, USA; ²New York University, USA
- 876 **X-Data: Generating Test Data for Killing SQL Mutants** ■
Bhanu Pratap Gupta, Devang Vira, S. Sudarshan, IIT Bombay, India

■ indicates Short Paper

Research Session 27: Social Networks and Similarity Queries

- 880 **Discovery-Driven Graph Summarization**
Ning Zhang¹, Yuanyuan Tian², Jignesh M. Patel¹
¹University of Wisconsin-Madison, USA; ²IBM, USA
- 892 **The Similarity Join Database Operator**
Yasin N. Silva¹, Walid G. Aref¹, Mohamed H. Ali²
¹Purdue University, USA; ²Microsoft, USA
- 904 **Anonymizing Weighted Social Network Graphs ■**
Sudipto Das, Ömer Eğecioğlu, Amr El Abbadi, University of California at Santa Barbara, USA
- 908 **Efficient Similarity Matching of Time Series Cliques with Natural Relations ■**
Zhe Zhao¹, Bin Cui¹, Wee Hyong Tok², Jiakui Zhao³
¹Peking University, China; ²Microsoft, China; ³China Electric Power Research Institute, China

■ indicates Short Paper

Research Session 28: Stream Processing

- 912 **Continuous Query Evaluation Over Distributed Sensor Networks**
Oana Jurca, Sebastian Michel, Alexandre Herrmann, Karl Aberer, EPFL, Switzerland
- 924 **Space-Efficient Online Approximation of Time Series Data: Streams, Amnesia, and Out-of-Order**
Sorabh Gandhi, Luca Foschini, Subhash Suri, University of California at Santa Barbara, USA
- 936 **Approximation Trade-Offs in Markovian Stream Processing: An Empirical Study ■**
Julie Letchner¹, Christopher Ré², Magdalena Balazinska¹, Matthai Philipose³
¹University of Washington, USA; ²University of Wisconsin-Madison, USA; ³Intel, USA
- 940 **FENCE: Continuous Access Control Enforcement in Dynamic Data Stream Environments ■**
Rimma V. Nehme¹, Hyo-Sang Lim², Elisa Bertino²
¹Microsoft, USA; ²Purdue University, USA

■ indicates Short Paper

Research Session 29: Publishing Privacy

- 944 **A Privacy-Preserving Approach to Policy-Based Content Dissemination**
Ning Shang, Mohamed Nabeel, Federica Paci, Elisa Bertino, Purdue University, USA
- 956 **Global Privacy Guarantee in Serial Data Publishing** ■
Raymond Chi-Wing Wong¹, Ada Wai-Chee Fu², Jia Liu², Ke Wang³, Yabo Xu⁴
¹Hong Kong University of Science & Technology, China; ²Chinese University of Hong Kong, China; ³Simon Fraser University, Canada; ⁴Sun Yat-sen University, China
- 960 **XColor: Protecting General Proximity Privacy** ■
Ting Wang, Ling Liu, Georgia Institute of Technology, USA
- 964 **Correlation Hiding by Independence Masking** ■
Yufei Tao¹, Jian Pei², Jiexing Li¹, Xiaokui Xiao³, Ke Yi⁴, Zhengzheng Xing²
¹Chinese University of Hong Kong, China; ²Simon Fraser University, Canada; ³Nanyang Technological University, Singapore; ⁴Hong Kong University of Science & Technology, China

■ indicates Short Paper

Research Session 30: Data Clouds

- 968 **Monitoring Continuous State Violation in Datacenters: Exploring the Time Dimension**
Shicong Meng, Ting Wang, Ling Liu, Georgia Institute of Technology, USA
- 980 **Cost-Efficient and Differentiated Data Availability Guarantees in Data Clouds ■**
Nicolas Bonvin, Thanasis G. Papaioannou, Karl Aberer, EPFL, Switzerland
- 984 **Intensional Associations in Dataspaces ■**
Marcos Antonio Vaz Salles¹, Jens Dittrich², Lukas Blunschi³
¹Cornell University, USA; ²Saarland University, Germany; ³ETH Zürich, Switzerland
- 988 **A Tuple Space for Social Networking on Mobile Phones ■**
Emre Sarigöl, Oriana Riva, Gustavo Alonso, ETH Zürich, Switzerland
- 992 **Overlapping Community Search for Social Networks ■**
Arnau Padrol-Sureda, Guillem Perarnau-Llobet, Julian Pfeifle, Victor Muntés-Muleor,
Universitat Politècnica de Catalunya, Spain

■ indicates Short Paper

Industry Session 1: Data Warehousing

- 996 **Hive — A Petabyte Scale Data Warehouse Using Hadoop**
Ashish Thusoo, Joydeep Sen Sarma, Namit Jain, Zheng Shao, Prasad Chakka, Ning Zhang, Suresh Antony, Hao Liu, Raghotham Murthy, Facebook, USA
- 1006 **Tuning Servers, Storage and Database for Energy Efficient Data Warehouses**
Meikel Poess¹, Raghunath Othayoth Nambiar²
¹Oracle, USA; ²HP, USA
- 1018 **A New Algorithm for Small-Large Table Outer Joins in Parallel DBMS**
Yu Xu, Pekka Kostamaa, Teradata, USA
-

Industry Session 2: Data, Data, and More Data

- 1025 **Data Cleansing as a Transient Service**
Tanveer A. Faruque, Hima Prasad K., L. Venkata Subramaniam, Mukesh Mohania, Girish Venkatachaliah, Shrinivas Kulkarni, Pramit Basu, IBM, India
- 1037 **XBRL Repository — An Industrial approach of Management of XBRL Documents**
Zhen Hua Liu, Thomas Baby, Sriram Krishnamurthy, Ying Lu, Qin Yu, Anguel Novoselsky, Vikas Arora, Oracle, USA
- 1048 **Visualizing Large-Scale RDF Data Using Subsets, Summaries, and Sampling in Oracle**
Seema Sundara, Medha Atre, Vladimir Kolovski, Souripriya Das, Zhe Wu, Eugene Inseok Chong, Jagannathan Srinivasan, Oracle, USA

■ indicates Short Paper

Industry Session 3: Query Optimization

- 1060 **Incorporating Partitioning and Parallel Plans into the SCOPE Optimizer**
Jingren Zhou, Per-Ake Larson, Ronnie Chaiken, Microsoft, USA
- 1072 **Rule Profiling for Query Optimizers and Their Implications**
Surajit Chaudhuri, Leo Giakoumakis, Vivek Narasayya, Ravishankar Ramamurthy, Microsoft, USA
- 1081 **Data Desensitization of Customer Data for Use in Optimizer Performance Experiments**
Malu Castellanos¹, Bin Zhang¹, Ivo Jimenez¹, Perla Ruiz², Miguel Durazo², Umeshwar Dayal¹, Lily Jow¹
¹HP, USA; ²University of Sonora, Mexico

■ indicates Short Paper

Demo Session 1: Events, Streams, Services, Mashups and Search

- 1093 **A Demonstration of the MaxStream Federated Stream Processing System**
*Irina Botan¹, Younggoo Cho², Roozbeh Derakhshan¹, Nihal Dindar¹, Ankush Gupta¹,
Laura M. Haas³, Kihong Kim², Chulwon Lee², Girish Mundada⁴, Ming-Chien Shan⁴,
Nesime Tatbul¹, Ying Yan⁵, Beomjin Yun², Jin Zhang⁵*
¹ETH Zürich, Switzerland; ²SAP, Korea; ³IBM, USA; ⁴SAP, USA; ⁵SAP, China
- 1097 **E-Cube: Multi-Dimensional Event Sequence Processing Using Concept and Pattern Hierarchies**
*Mo Liu¹, Elke A. Rundensteiner¹, Kara Greenfield¹, Chetan Gupta², Song Wang²,
Ismail Ari³, Abhay Mehta²*
¹Worcester Polytechnic Institute, USA; ²HP, USA; ³Ozyegin University, Turkey
- 1101 **TargetSearch: A Ranking Friendly XML Keyword Search Engine**
Ziyang Liu, Yichuan Cai, Yi Chen, Arizona State University, USA
- 1105 **Efficient Fuzzy Type-Ahead Search in TASTIER**
Guoliang Li¹, Shengyue Ji², Chen Li², Jiannan Wang¹, Jianhua Feng¹
¹Tsinghua University, China; ²University of California at Irvine, USA
- 1109 **MASS: A Multi-Facet Domain-Specific Influential Blogger Mining System**
Yichuan Cai, Yi Chen, Arizona State University, USA
- 1113 **Product EntityCube: A Recommendation and Navigation System for Product Search**
Jongwuk Lee¹, Seung-won Hwang¹, Zaiqing Nie², Ji-Rong Wen²
¹POSTECH, Korea; ²Microsoft, China
- 1117 **Navigating Through Mashed-Up Applications with COMPASS**
Daniel Deutch, Ohad Greenshpan, Tova Milo, Tel-Aviv University, Israel

... continues next page ...

■ indicates Short Paper

- 1121 **GenerIE: Information Extraction Using Database Queries**
Luis Tari¹, Phan Huy Tu¹, Jörg Hakenberg¹, Yi Chen¹, Tran Cao Son², Graciela Gonzalez¹, Chitta Baral¹
¹Arizona State University, USA; ²New Mexico State University, USA
- 1125 **Power-Aware Data Analysis in Sensor Networks**
Daniel Klan¹, Katja Hose¹, Marcel Karnstedt², Kai-Uwe Sattler¹
¹Ilmenau University of Technology, Germany; ²NUI Galway, Ireland
- 1129 **A View-Based Monitoring for Privacy-Aware Web Services**
Hassina Meziane¹, Salima Benbernou¹, Aouda K. Zerdali¹, Mohand-Said Hacid², Mike Papazoglou³
¹Université Paris Descartes, France; ²Université de Lyon, France; ³Tilburg University, The Netherlands
- 1133 **Viewing a World of Annotations Through AnnoVIP**
Konstantinos Karanasos, Spyros Zoupanos, INRIA, France
- 1137 **MashRank: Towards Uncertainty-Aware and Rank-Aware Mashups**
Mohamed A. Soliman, Mina Saleeb, Ihab F. Ilyas, University of Waterloo, Canada
- 1141 **T-Warehouse: Visual OLAP Analysis on Trajectory Data**
Luca Leonardi¹, Gerasimos Marketos², Elias Frentzos², Nikos Giatrakos², Salvatore Orlando¹, Nikos Pelekis², Alessandra Raffaetà¹, Alessandro Roncato¹, Claudio Silvestri¹, Yannis Theodoridis²
¹Università Ca' Foscari Venezia, Italy; ²University of Piraeus, Greece
- 1145 **WikiAnalytics: Ad-Hoc Querying of Highly Heterogeneous Structured Data**
Andrey Balmin¹, Emiran Curtmola²
¹IBM, USA; ²University of California at San Diego, USA
- 1149 **SMARTINT: A System for Answering Queries Over Web Databases Using Attribute Dependencies**
Ravi Gummadi, Anupam Khulbe, Aravind Kalavagattu, Sanil Salvi, Subbarao Kambhampati, Arizona State University, USA

■ indicates Short Paper

Demo Session 2: Scalability, Design, Optimization and Miscellaneous

- 1153 **Mini-Me: A Min-Repro System for Database Software**
Nicolas Bruno, Rimma V. Nehme, Microsoft, USA
- 1157 **I/O-Efficient Statistical Computing with RIOT**
Yi Zhang, Weiping Zhang, Jun Yang, Duke University, USA
- 1161 **Interactive Physical Design Tuning**
Nicolas Bruno, Surajit Chaudhuri, Microsoft, USA
- 1165 **Visualizing Cost-Based XQuery Optimization**
Andreas M. Weiner, Theo Härder, Renato Oliveira da Silva, University of Kaiserslautern, Germany
- 1169 **XML Reasoning Made Practical**
Pierre Genevès¹, Nabil Layaïda²
¹CNRS, France; ²INRIA, France
- 1173 **TransScale: Scalability Transformations for Declarative Applications**
Alexander Böhm, Erich Marth, Carl-Christian Kanne, University of Mannheim, Germany
- 1177 **Reverse Engineering Models from Databases to Bootstrap Application Development**
Ankit Malpani¹, Philip A. Bernstein², Sergey Melnik³, James F. Terwilliger²
¹IIT Madras, India; ²Microsoft, USA; ³Google, USA
- 1181 **HECATAEUS: Regulating Schema Evolution**
George Papastefanatos¹, Panos Vassiliadis², Alkis Simitsis³, Yannis Vassiliou¹
¹National Technical University of Athens, Greece; ²University of Ioannina, Greece; ³HP, USA

■ indicates Short Paper

- 1185 **ROX: The Robustness of a Run-Time XQuery Optimizer Against Correlated Data**
Riham Abdel Kader¹, Peter A. Boncz², Stefan Manegold², Maurice van Keulen¹
¹University of Twente, The Netherlands; ²CWI, The Netherlands
- 1189 **Symphony: A Platform for Search-Driven Applications**
John C. Shafer, Rakesh Agrawal, Hady W. Lauw, Microsoft, USA
- 1193 **ProbClean: A Probabilistic Duplicate Detection System**
George Beskales, Mohamed A. Soliman, Ihab F. Ilyas, Shai Ben-David, Yubin Kim, University of Waterloo, Canada
- 1197 **TransDec: A Spatiotemporal Query Processing Framework for Transportation Systems**
Ugur Demiryurek, Farnoush Banaei-Kashani, Cyrus Shahabi, University of Southern California, USA
- 1201 **Provenance Browser: Displaying and Querying Scientific Workflow Provenance Graphs**
Manish Kumar Anand¹, Shawn Bowers², Bertram Ludäscher¹
¹University of California at Davis, USA; ²Gonzaga University, USA
- 1205 **Inconsistency Resolution in Online Databases**
Yannis Katsis¹, Alin Deutsch¹, Yannis Papakonstantinou¹, Vasilis Vassalos²
¹University of California at San Diego, USA; ²Athens University of Economics & Business, Greece

■ indicates Short Paper

Panels

- 1209 **Cloudy Skies for Data Management**
*David Campbell*¹, *Brian Cooper*², *Dean Jacobs*³, *Ashok Joshi*⁴, *Volker Markl*⁵,
*Srinivas Narayanan*⁶
¹Microsoft, USA; ²Yahoo!, USA; ³SAP, Germany; ⁴Oracle, USA; ⁵Technische Universität
Berlin, Germany; ⁶Facebook, USA
- 1210 **Database Architecture (R)evolution: New Hardware vs. New Software**
*Stavros Harizopoulos*¹, *Tassos Argyros*², *Peter A. Boncz*³, *Dan Dietterich*⁴,
*Samuel R. Madden*⁵, *Florian M. Waas*⁶
¹HP, USA; ²Aster Data, USA; ³CWI, The Netherlands; ⁴Netezza, USA; ⁵MIT, USA;
⁶Greenplum, USA

■ indicates Short Paper

Seminars

- 1211 **Anonymized Data: Generation, Models, Usage**
Graham Cormode, Divesh Srivastava, AT&T Labs Research, USA
- 1213 **Privacy in Data Publishing**
Johannes Gehrke¹, Daniel Kifer², Ashwin Machanavajjhala³
¹Cornell University, USA; ²Pennsylvania State University, USA; ³Yahoo!, USA
- 1214 **Representation, Composition and Application of Preferences in Databases**
Georgia Koutrika¹, Evaggelia Pitoura², Kostas Stefanidis²
¹Stanford University, USA; ²University of Ioannina, Greece
- 1216 **Database as a Service (DBaaS)**
Wolfgang Lehner¹, Kai-Uwe Sattler²
¹Dresden University of Technology, Germany; ²Ilmenau University of Technology, Germany
- 1218 **Techniques for Efficiently Searching in Spatial, Temporal, Spatio-Temporal, and Multimedia Databases**
Hans-Peter Kriegel, Peer Kröger, Matthias Renz, LMU München, Germany
- indicates Short Paper