

Table of Contents

- xv **Conference Organizers**
- xvii **Core Database Technology Program Committee**
- xix **Infrastructure for Information Systems Program Committee**
- xx **Industrial and Applications Program Committee**
- xx **Demonstrations Program Committee**
- xxi **Additional Reviewers**
- xxiv **VLDB Endowment Board of Trustees**

Keynote

- 1 **Why Search Engines are Used Increasingly to Offload Queries from Databases**
Bjørn Olstad (CTO, FAST Search & Transfer, Norway)

Ten-Year Best Paper Award

- 2 **Database-Inspired Search**
David Konopnicki (IBM Haifa Research Laboratory, Israel), Oded Shmueli (Technion, Haifa, Israel)

Research Sessions

Streams and Stream-based Processing

- 13 **Sketching Streams Through the Net: Distributed Approximate Query Tracking**
Graham Cormode (Bell Labs, Lucent Technologies, USA), Minos Garofalakis (Intel Research, Berkeley, USA)
- 25 **Summarizing and Mining Inverse Distributions on Data Streams via Dynamic Inverse Sampling**
Graham Cormode (Bell Labs, Lucent Technologies, USA), S. Muthukrishnan, Irina Rozenbaum (Rutgers University, USA)
- 37 **Adaptive Stream Filters for Entity-based Queries with Non-Value Tolerance**
Reynold Cheng (The Hong Kong Polytechnic University, China), Ben Kao (University of Hong Kong, China), Sunil Prabhakar (Purdue University, USA), Alan Kwan (University of Hong Kong, China), Yicheng Tu (Purdue University, USA)

Database Architectures for New Hardware

- 49 **Improving Database Performance on Simultaneous Multithreading Processors**
Jingren Zhou (Microsoft Research, USA), John Cieslewicz, Kenneth Ross, Mihir Shah (Columbia University, USA)

- 61 **Parallel Querying with Non-Dedicated Computers**
Vijayshankar Raman, Wei Han, Inderpal Narang (IBM Almaden Research Center, USA)
- 73 **Optimistic Intra-Transaction Parallelism on Chip Multiprocessors**
Christopher Colohan, Anastassia Ailamaki (Carnegie Mellon University, USA), J. Gregory Steffan (University of Toronto, Canada), Todd Mowry (Carnegie Mellon University and Intel Research Corporation, USA)

Integration and Mapping #1

- 85 **Information Preserving XML Schema Embedding**
Philip Bohannon (Bell Labs, Lucent Technologies, USA), Wenfei Fan (University of Edinburgh & Bell Laboratories, UK), Michael Flaster, P.P.S. Narayan (Bell Labs, Lucent Technologies, USA)
- 97 **Light-weight Domain-based Form Assistant: Querying Web Databases On the Fly**
Zhen Zhang, Bin He, Kevin Chen-Chuan Chang (University of Illinois at Urbana-Champaign, USA)
- 109 **Designing Information-Preserving Mapping Schemes for XML**
Denilson Barbosa (University of Calgary, Canada), Juliana Freire (University of Utah, USA), Alberto O. Mendelzon (University of Toronto, Canada)

XML Query Processing #1

- 121 **Rewriting XPath Queries Using Materialized Views**
Wanhong Xu, Z. Meral Özsoyoglu (Case Western Reserve University, USA)
- 133 **Benefits of Path Summaries in an XML Query Optimizer Supporting Multiple Access Methods**
Attila Barta, Mariano P. Consens, Alberto O. Mendelzon (University of Toronto, Canada)
- 145 **Efficient Processing of XML Path Queries Using the Disk-based F&B Index**
Wei Wang (The University of New South Wales, Australia), Hongzhi Wang (The University of New South Wales, Australia & Harbin Institute of Technology, China), Hongjun Lu (Hongkong University of Science & Technology, China), Haifeng Jiang (IBM, USA), Xuemin Lin (The University of New South Wales, Australia), Jianzhong Li (Harbin Institute of Technology, China)

Streams

- 157 **Customizable Parallel Execution of Scientific Stream Queries**
Milena Ivanova, Tore Risch (Uppsala University, Sweden)

Table of Contents

169 **Using Association Rules for Fraud Detection in Web Advertising Networks**

Ahmed Metwally, Divyakant Agrawal, Amr El Abbadi (University of California, Santa Barbara, USA)

181 **Parameter Free Bursty Events Detection in Text Streams**

Gabriel Pui Cheong Fung, Jeffrey Xu Yu (The Chinese University of Hong Kong, China), Philip S. Yu (IBM T. J. Watson Research Center, USA), Hongjun Lu (Hongkong University of Science & Technology, China)

XML Query Processing #2

193 **From Region Encoding To Extended Dewey: On Efficient Processing of XML Twig Pattern Matching**

Jiaheng Lu, Tok Wang Ling, Chee-Yong Chan, Ting Chen (National University of Singapore, Singapore)

205 **Tree-Pattern Queries on a Lightweight XML Processor**

Mirella Moro, Zografoula Vagena, Vassilis Tsotras (University of California, Riverside, USA)

217 **FiST: Scalable XML Document Filtering by Sequencing Twig Patterns**

Joonho Kwon (Seoul National University, Korea), Praveen Rao, Bongki Moon (University of Arizona, USA), Sukho Lee (Seoul National University, Korea)

Skyline Querying

229 **Maximal Vector Computation in Large Data Sets**

Parke Godfrey (York University, Canada), Ryan Shipley (The College of William and Mary, USA), Jarek Gryz (York University, Canada)

241 **Efficient Computation of the Skyline Cube**

Yidong Yuan, Xuemin Lin, Qing Liu, Wei Wang (The University of New South Wales, Australia), Jeffrey Xu Yu (The Chinese University of Hong Kong, China), Qing Zhang (The University of New South Wales, Australia)

253 **Catching the Best Views of Skyline: A Semantic Approach Based on Decisive Subspaces**

Jian Pei, Wen Jin, Martin Ester (Simon Fraser University, Canada), Yufei Tao (City University of Hong Kong, China)

XML Query Processing #3

265 **Efficient Evaluation of XQuery over Streaming Data**

Xiaogang Li, Gagan Agrawal (Ohio State University, USA)

277 **Semantic Query Optimization for XQuery over XML Streams**

Hong Su, Elke Rundensteiner, Murali Mani (Worcester Polytechnic Institute, USA)

289 **Statistical Learning Techniques for Costing XML Queries**

Ning Zhang (University of Waterloo, Canada), Peter J. Haas, Vanja Josifovski, Guy M. Lohman, Chun Zhang (IBM Almaden Research Center, USA)

Text Data Management

301 **Approximate Matching of Hierarchical Data Using pq-Grams**

Nikolaus Augsten, Michael Böhlen, Johann Gamper (Free University of Bozen-Bolzano, Italy)

313 **The TEXTURE Benchmark: Measuring Performance of Text Queries on a Relational DBMS**

Vuk Ercegovic, David J. DeWitt, Raghu Ramakrishnan (University of Wisconsin – Madison, USA)

325 **n-Gram/2L: A Space and Time Efficient Two-Level n-Gram Inverted Index Structure**

Min-Soo Kim, Kyu-Young Whang, Jae-Gil Lee, Min-Jae Lee (Korea Advanced Institute of Science & Technology (KAIST), Korea)

XML Query Processing #4

337 **Query Translation from XPath to SQL in the Presence of Recursive DTDs**

Wenfei Fan (University of Edinburgh & Bell Laboratories, UK), Jeffrey Xu Yu (The Chinese University of Hong Kong, China), Hongjun Lu (Hong Kong University of Science & Technology, China), Jianhua Lu (The Chinese University of Hong Kong, China), Rajeev Rastogi (Bell Labs, India)

349 **Pattern Tree Algebras: Sets or Sequences?**

Stelios Paparizos, Hosagrahar Visvesvaraya Jagadish (University of Michigan, USA)

361 **Structure and Content Scoring for XML**

Sihem Amer-Yahia (AT&T Labs – Research, USA), Nick Koudas (University of Toronto, Canada), Amélie Marian (Columbia University, USA), Divesh Srivastava (AT&T Labs – Research, USA), David Toman (University of Waterloo, Canada)

Query Optimization #1

373 **Consistently Estimating the Selectivity of Conjunctions of Predicates**

Volker Markl, Nimrod Megiddo (IBM Almaden Research Center, USA), Marcel Kutsch (IBM, Germany), Tam Minh Tran (IBM Silicon Valley Lab, USA), Peter J. Haas (IBM Almaden Research Center, USA), Utkarsh Srivastava (Stanford University, USA)

385 **Efficiently Processing Queries on Interval-and-Value Tuples in Relational Databases**

Jost Enderle, Nicole Schneider, Thomas Seidl (RWTH Aachen University, Germany)

Table of Contents

- 397 **Selectivity Estimation for Fuzzy String Predicates in Large Data Sets**
Liang Jin, Chen Li (University of California, Irvine, USA)

Summarization

- 409 **Space Efficiency in Synopsis Construction Algorithms**
Sudipto Guha (University of Pennsylvania, USA)
- 421 **One-Pass Wavelet Synopses for Maximum-Error Metrics**
Panagiotis Karras, Nikos Mamoulis (University of Hong Kong, China)
- 433 **MDL Summarization with Holes**
Shaofeng Bu, Laks V.S. Lakshmanan, Raymond Ng (University of British Columbia, Canada)

Views and Cache Management

- 445 **View Matching for Outer-Join Views**
Per-Åke Larson, Jingren Zhou (Microsoft Research, USA)
- 457 **Caching with “Good Enough” Currency, Consistency, and Completeness**
Hongfei Guo (University of Wisconsin – Madison, USA), Per-Åke Larson (Microsoft Research, USA), Raghu Ramakrishnan (University of Wisconsin – Madison, USA)
- 469 **Query Caching and View Selection for XML Databases**
Bhushan Mandhani, Dan Suciu (University of Washington, USA)

Query Optimization #2

- 481 **Optimizing Nested Queries with Parameter Sort Orders**
Ravindra Guravannavar (Indian Institute of Technology, Bombay, India), Ramanujam H.S. (Sybase Software India, Pune, India), S Sudarshan (Indian Institute of Technology, Bombay, India)
- 493 **Stack-based Algorithms for Pattern Matching on DAGs**
Li Chen, Amarnath Gupta, M. Erdem Kurul (University of California, San Diego, USA)
- 505 **Bidirectional Expansion For Keyword Search on Graph Databases**
Varun Kacholia, Shashank Pandit (University of California Berkeley, USA), Soumen Chakrabarti, S Sudarshan (Indian Institute of Technology, Bombay, India), Rushi Desai, Hrishikesh Karambelkar (University of California Berkeley, USA)

New Applications

- 517 **Link Spam Alliances**
Zoltán Gyöngyi, Hector Garcia-Molina (Stanford University, USA)

- 529 **The SphereSearch Engine for Unified Ranked Retrieval of Heterogeneous XML and Web Documents**
Jens Graupmann, Ralf Schenkel, Gerhard Weikum (Max-Planck Institut für Informatik, Germany)

- 541 **Hubble: An Advanced Dynamic Folder Technology for XML**
Ning Li, Joshua Hui, Hui-I Hsiao, Kevin S. Beyer (IBM Almaden Research Center, USA)

Architectural Issues

- 553 **C-Store: A Column-oriented DBMS**
Michael Stonebraker, Daniel Abadi (MIT CSAIL, USA), Adam Batkin (Brandeis University, USA), Xuedong Chen (University of Massachusetts, Boston, USA), Mitch Cherniack (Brandeis University, USA), Miguel Ferreira, Edmond Lau, Amerson Lin, Samuel Madden (MIT CSAIL, USA), Elizabeth O’Neil, Pat O’Neil (University of Massachusetts, Boston, USA), Alex Rasin (Brown University, USA), Nga Tran (Brandeis University, USA), Stan Zdonik (Brown University, USA)
- 565 **Fine-Grained Replication and Scheduling with Freshness and Correctness Guarantees**
Fuat Akal, Can Türker, Hans-Jörg Schek (ETH Zurich, Switzerland), Yuri Breitbart (Kent State University, USA), Torsten Grabs (Microsoft Corporation, USA), Lourens Veen (University of Twente, The Netherlands)
- 577 **Cache-conscious Frequent Pattern Mining on a Modern Processor**
Amol Ghoting, Gregory Buehrer, Srinivasan Parthasarathy (The Ohio State University, USA), Daehyun Kim, Anthony Nguyen, Yen-Kuang Chen, Pradeep Dubey (Intel Corporation, USA)

Correctness and Performance

- 589 **Parallel Execution of Test Runs for Database Application Systems**
Florian Haftmann (i-TV-T AG, Germany), Donald Kossmann (i-TV-T AG, Germany, ETH Zurich, Switzerland), Eric Lo (ETH Zurich, Switzerland)
- 601 **Query Execution Assurance for Outsourced Databases**
Radu Sion (Stony Brook University, USA)
- 613 **Automatic Composition of Transition-based Semantic Web Services with Messaging**
Daniela Berardi (Università di Roma “La Sapienza”, Italy), Diego Calvanese (Libera Università di Bolzano/Bozen, Italy), Giuseppe De Giacomo (Università di Roma “La Sapienza”, Italy), Richard Hull (Bell Labs Research, Lucent Technologies, USA), Massimo Mecella (Università di Roma “La Sapienza”, Italy)

Table of Contents

DB and IR #1

- 625 **An Efficient and Versatile Query Engine for TopX Search**
Martin Theobald, Ralf Schenkel, Gerhard Weikum (Max-Planck Institute for Informatics, Germany)
- 637 **KLEE: A Framework for Distributed Top-k Query Algorithms**
Sebastian Michel (Max-Planck Institute for Informatics, Germany), Peter Triantafyllou (University of Patras, Greece), Gerhard Weikum (Max-Planck Institute for Informatics, Germany)
- 649 **Scaling and Time Warping in Time Series Querying**
Ada Wai-chee Fu (The Chinese University of Hong Kong, China), Eamonn Keogh (University of California, Riverside, USA), Leo Yung Hang Lau (The Chinese University of Hong Kong, China), Chotirat Ann Ratanamahatana (University of California, Riverside, USA)

Peer-to-Peer and Distributed Computing

- 661 **BATON: A Balanced Tree Structure for Peer-to-Peer Networks**
Hosagrahar Visvesvaraya Jagadish (University of Michigan, USA, National University of Singapore, Singapore), Beng Chin Ooi, Quang Hieu Vu (National University of Singapore, Singapore)
- 673 **Client Assignment in Content Dissemination Networks for Dynamic Data**
Shetal Shah, Krithi Ramamritham (Indian Institute of Technology, Bombay, India), Chinya Ravishankar (University of California, Riverside, USA)
- 685 **Indexing Data-oriented Overlay Networks**
Karl Aberer, Anwitaman Datta, Manfred Hauswirth, Roman Schmidt (Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland)

Data Mining

- 697 **Streaming Pattern Discovery in Multiple Time-Series**
Spiros Papadimitriou, Jimeng Sun, Christos Faloutsos (Carnegie Mellon University, USA)
- 709 **Mining Compressed Frequent-Pattern Sets**
Dong Xin, Jiawei Han, Xifeng Yan, Hong Cheng (University of Illinois at Urbana-Champaign, USA)
- 721 **Discovering Large Dense Subgraphs in Massive Graphs**
David Gibson, Ravi Kumar, Andrew Tomkins (IBM Almaden Research Center, USA)

Query Optimization and Summarization

- 733 **General Purpose Database Summarization**
Regis Saint-Paul, Guillaume Raschia, Noureddine Mouaddib (LINA - Polytech 'Nantes, France)
- 745 **Online Estimation For Subset-Based SQL Queries**
Christopher Jermaine, Alin Dobra, Abhijit Pol, Shantanu Joshi (University of Florida, USA)
- 757 **Content-Based Routing: Different Plans for Different Data**
Pedro Bizarro (University of Wisconsin – Madison, USA), Shivnath Babu (Stanford University, USA), David J. DeWitt (University of Wisconsin – Madison, USA), Jennifer Widom (Stanford University, USA)
- 769 **REED: Robust, Efficient Filtering and Event Detection in Sensor Networks**
Daniel Abadi, Samuel Madden, Wolfgang Lindner (MIT CSAIL, USA)

DB and IR #2

- 781 **Shuffling a Stacked Deck: The Case for Partially Randomized Ranking of Search Engine Results**
Sandeep Pandey (Carnegie Mellon University, USA), Sourashis Roy (University of California, Los Angeles, USA), Christopher Olston (Carnegie Mellon University, USA), Junghoo Cho (University of California, Los Angeles, USA), Soumen Chakrabarti (Indian Institute of Technology, Bombay, India)
- 793 **Indexing Mixed Types for Approximate Retrieval**
Liang Jin (University of California, Irvine, USA), Nick Koudas (University of Toronto, Canada), Chen Li (University of California, Irvine, USA), Anthony K. H. Tung (National University of Singapore, Singapore)
- 805 **Answering Queries from Statistics and Probabilistic Views**
Nilesh Dalvi, Dan Suciu (University of Washington, USA)

DB Algorithms

- 817 **Inspector Joins**
Shimin Chen, Anastassia Ailamaki (Carnegie Mellon University, USA), Phillip Gibbons (Intel Research Pittsburgh, USA), Todd Mowry (Carnegie Mellon University, Intel Research Pittsburgh, USA)
- 829 **Revisiting Pipelined Parallelism in Multi-Join Query Processing**
Bin Liu, Elke Rundensteiner (Worcester Polytechnic Institute, USA)
- 841 **Early Hash Join: A Configurable Algorithm for the Efficient and Early Production of Join Results**
Ramon Lawrence (University of Iowa, USA)

Table of Contents

Spatio-Temporal Query Processing

- 853 **On Map-Matching Vehicle Tracking Data**
Sotiris Brakatsoulas, Dieter Pfoser (RA Computer Technology Institute, Athens, Greece), Randall Salas, Carola Wenk (University of Texas at San Antonio, USA)
- 865 **An Efficient and Scalable Approach to CNN Queries in a Road Network**
Hyung-Ju Cho, Chin-Wan Chung (Korea Advanced Institute of Science & Technology, Korea)
- 877 **Complex Spatio-Temporal Pattern Queries**
Marios Hadjieleftheriou, George Kollios (Boston University, USA), Petko Bakalov, Vassilis Tsotras (University of California, Riverside, USA)

Privacy

- 889 **Distributed Privacy Preserving Information Sharing**
Nan Zhang, Wei Zhao (Texas A&M University, USA)
- 901 **On k-Anonymity and the Curse of Dimensionality**
Charu Aggarwal (IBM T. J. Watson Research Center, USA)
- 910 **Checking for k-Anonymity Violation by Views**
Chao Yao (George Mason University, USA), X. Sean Wang (The University of Vermont, USA), Sushil Jajodia (George Mason University, USA)

Spatial and Temporal Databases

- 922 **Indexing Multi-Dimensional Uncertain Data with Arbitrary Probability Density Functions**
Yufei Tao (City University of Hong Kong, China), Reynold Cheng (The Hong Kong Polytechnic University, China), Xiaokui Xiao (City University of Hong Kong, China), Wang Kay Ngai, Ben Kao (University of Hong Kong, China), Sunil Prabhakar (Purdue University, USA)
- 934 **A Trajectory Splitting Model for Efficient Spatio-Temporal Indexing**
Slobodan Rasetic, Jörg Sander, James Elding, Mario Nascimento (University of Alberta, Canada)
- 946 **On Computing Top-t Most Influential Spatial Sites**
Tian Xia, Donghui Zhang, Evangelos Kanoulas, Yang Du (Northeastern University, USA)

OLAP

- 958 **Efficient Implementation of Large-Scale Multi-Structural Databases**
Ronald Fagin, Phokion Kolaitis, Ravi Kumar, Jasmine Novak, D. Sivakumar, Andrew Tomkins (IBM Almaden Research Center, USA)

- 970 **OLAP Over Uncertain and Imprecise Data**
Doug Burdick (University of Wisconsin – Madison, USA), Prasad Deshpande, T.S. Jayram (IBM Almaden Research Center, USA), Raghu Ramakrishnan (University of Wisconsin – Madison, USA), Shivakumar Vaithyanathan (IBM Almaden Research Center, USA)

- 982 **Prediction Cubes**
Bee-Chung Chen, Lei Chen, Yi Lin, Raghu Ramakrishnan (University of Wisconsin – Madison, USA)

Integration and Mapping #2

- 994 **Tuning Schema Matching Software using Synthetic Scenarios**
Mayssam Sayyadian, Yoonkyong Lee, AnHai Doan (University of Illinois, USA), Arnon Rosenthal (The MITRE Corporation, USA)
- 1006 **Semantic Adaptation of Schema Mappings when Schemas Evolve**
Cong Yu (University of Michigan, USA), Lucian Popa (IBM Almaden Research Center, USA)
- 1018 **Mapping Maintenance for Data Integration Systems**
Robert McCann, Bedoor K. AlShebli, Quoc Le, Hoa Nguyen, Long Vu, AnHai Doan (University of Illinois, USA)

Industrial Sessions

Data Warehousing and Data Mining

- 1031 **Bridging the Gap between OLAP and SQL**
Jens-Peter Dittrich (ETH Zurich, Switzerland), Donald Kossmann (ETH Zurich, i-TV-T AG, Switzerland), Alexander Kreutz (i-TV-T AG, Switzerland)
- 1043 **Optimizing Refresh of a Set of Materialized Views**
Nathan Folkert, Abhinav Gupta, Andrew Witkowski, Sankar Subramanian, Srikanth Bellamkonda, Shrikanth Shankar, Tolga Bozkaya, Lei Sheng (Oracle Corporation USA)
- 1055 **Large Scale Data Warehouses on Grid: Oracle Database 10g and HP ProLiant Systems**
Meikel Poess (Oracle Corporation, USA), Raghunath Othayoth (Hewlett-Packard Company, Houston, USA)

Applications

- 1067 **The Integrated Microbial Genomes (IMG) System: A Case Study in Biological Data Management**
Victor M. Markowitz, Frank Korzeniewski, Krishna Palaniappan, Ernest Szeto (Lawrence Berkeley National Lab, USA), Natalia Ivanova, Nikos Kyrpides (Microbial Genome Analysis Program, Joint Genome Institute, USA)

Table of Contents

1079 **A Heartbeat Mechanism and Its Application in Gigascope**

Theodore Johnson (AT&T Labs – Research, USA), S. Muthukrishnan, Vladislav Shkapenyuk (Rutgers University, USA), Oliver Spatscheck (AT&T Labs – Research, USA)

1089 **Using a Fuzzy Classification Query Language for Customer Relationship Management**

Andreas Meier, Nicolas Werro (University of Fribourg, Switzerland), Martin Albrecht, Miltiadis Sarakinos (Swisscom Fixnet AG, Switzerland)

Potpourri

1097 **Flexible Database Generators**

Nicolas Bruno, Surajit Chaudhuri (Microsoft Corporation, USA)

1108 **Recovery Principles in MySQL Cluster 5.1**

Mikael Ronström, Jonas Orelund (MySQL AB, Sweden)

1116 **Getting Priorities Straight: Improving Linux Support for Database I/O**

Christoffer Hall, Philippe Bonnet (University of Copenhagen, Denmark)

New Data Types and Algorithms

1128 **Temporal Management of RFID Data**

Fusheng Wang, Peiya Liu (Siemens Corporate Research, USA)

1140 **Supporting RFID-based Item Tracking Applications in Oracle DBMS Using a Bitmap Datatype**

Ying Hu, Seema Sundara, Timothy Chorma, Jagannathan Srinivasan (Oracle Corporation, USA)

1152 **SVM in Oracle Database 10g: Removing the Barriers to Widespread Adoption of Support Vector Machines**

Boriana Milenova, Joseph Yarmus, Marcos Campos (Oracle Corporation, USA)

XML Support in Relational System

1164 **Native XML Support in DB2 Universal Database**

Matthias Nicola, Bert van der Linden (IBM Silicon Valley Lab, USA)

1175 **XQuery Implementation in a Relational Database System**

Shankar Pal, Istvan Cseri, Oliver Seeliger, Michael Rys, Gideon Schaller, Wei Yu, Dragan Tomic, Adrian Baras, Brandon Berg, Denis Churin, Eugene Kogan (Microsoft Corporation, USA)

1187 **CXHist : An On-line Classification-Based Histogram for XML String Selectivity Estimation**

Lipyeow Lim, Min Wang (IBM T. J. Watson Research Center, USA), Jeffrey Vitter (Purdue University, USA)

Web Services and Databases (Invited)

1199 **Consistency for Web Services Applications**

Paul Greenfield, Dean Kuo, Surya Nepal (CSIRO ICT Centre, Australia), Alan Fekete (University of Sydney, Australia),

Query Processing and Optimization

1204 **Query By Excel**

Andrew Witkowski, Srikanth Bellamkonda, Tolga Bozkaya, Aman Naimat, Lei Sheng, Sankar Subramanian, Allison Waingold (Oracle Corp., USA)

1216 **An Efficient SQL-based RDF Querying Scheme**

Eugene Chong, Souripriya Das, George Eadon, Jagannathan Srinivasan (Oracle Corp., USA)

1228 **Analyzing Plan Diagrams of Database Query Optimizers**

Naveen Reddy, Jayant Haritsa (Indian Institute of Science, India)

Panel

1241 **Database Publication Practices**

Philip A. Bernstein (Microsoft Research, USA), David J. DeWitt (University of Wisconsin – Madison, USA), Andreas Heuer (University of Rostock, Germany), Zachary Ives (University of Pennsylvania, USA), Christian S. Jensen (Aalborg University, Denmark), Holger Meyer (University of Rostock, Germany), M. Tamer Özsu (University of Waterloo, Canada), Richard T. Snodgrass (University of Arizona, USA), Kyu-Young Whang (Korea Advanced Institute of Science & Technology (KAIST), Korea), Jennifer Widom (Stanford University, USA)

Demo Sessions

Data Fusion

1247 **PrediCalc: A Logical Spreadsheet Management System**

Michael Kassoff, Lee-Ming Zen, Ankit Garg, Michael Genesereth, (Stanford University, USA)

1251 **Automatic Data Fusion with HumMer**

Alexander Bilke (Technische Universität Berlin, Germany), Jens Bleiholder, Christoph Böhm, Karsten Draba, Felix Naumann, Melanie Weis (Humboldt-Universität zu Berlin, Germany)

1255 **Querying Business Processes with BP-QL**

Catriel Beeri (Hebrew University, Israel), Anat Eyal, Simon Kamenkovich, Tova Milo (Tel Aviv University, Israel)

Table of Contents

P2P Based Systems

- 1259 **StreamGlobe: Processing and Sharing Data Streams in Grid-Based P2P Infrastructures**
Richard Kuntschke, Bernhard Stegmaier, Alfons Kemper, Angelika Reiser (Technische Universität München, Germany)
- 1263 **MINERVA: Collaborative P2P Search**
Matthias Bender, Sebastian Michel, (Max-Planck Institute for Informatics, Germany), Peter Triantafillou (University of Patras, Greece), Gerhard Weikum, Christian Zimmer, (Max-Planck Institute for Informatics, Germany)
- 1267 **HePToX: Marrying XML and Heterogeneity in Your P2P Databases**
Angela Bonifati (ICAR - National Research Council (CNR), Italy), Elaine Qing Chang, Terence Ho, Laks V.S. Lakshmanan, Rachel Pottinger (University of British Columbia, Canada)

Change Management and Data Dissemination

- 1271 **U-DBMS: A Database System for Managing Constantly-Evolving Data**
Reynold Cheng (The Hong Kong Polytechnic University, China), Sarvjeet Singh, Sunil Prabhakar (Purdue University, USA)
- 1275 **Database Change Notifications: Primitives for Efficient Database Query Result Caching**
Cesar Galindo-Legaria, Torsten Grabs, Christian Kleinerman, Florian Waas (Microsoft Corporation, USA)
- 1279 **CMS-ToPSS: Efficient Dissemination of RSS Documents**
Milenko Petrovic, Haifeng Liu, Hans-Arno Jacobsen (University of Toronto, Canada)

Data Sharing

- 1283 **Interactive Schema Translation with Instance-Level Mappings**
Philip A. Bernstein, Sergey Melnik, Peter Mork, (Microsoft Research, USA)
- 1287 **AReNA: Adaptive Distributed Catalog Infrastructure Based On Relevance Networks**
Vladimir Zadorozhny (University of Pittsburgh, USA), Avigdor Gal (Technion - Israel Institute of Technology, Israel), Louiqa Raschid (University of Maryland, USA), Qiang Ye (University of Pittsburgh, USA)
- 1291 **Data Sharing in the Hyperion Peer Database System**
Patricia Rodriguez-Gianolli (University of Toronto, Canada), Maddalena Garzetti (University of Trento, Italy), Lei Jiang (University of Toronto, Canada), Anastasios Kementsietsidis (University of Edinburgh, UK), Iluju Kiringa, Mehedi Masud (University of Ottawa, Canada), Renée J. Miller, John Mylopoulos (University of Toronto, Canada)

Stream Processing

- 1295 **NILE-PDT: A Phenomenon Detection and Tracking Framework for Data Stream Management Systems**
Mohamed H. Ali, Walid G. Aref (Purdue University, USA), Raja Bose (University of Florida, USA), Ahmed K. Elmagarmid (Purdue University, USA), Abdelsalam Helal (University of Florida, USA), Ibrahim Kamel (Zayed University, United Arab Emirates), Mohamed F. Mokbel (Purdue University, USA)
- 1299 **Robust Real-time Query Processing with QStream**
Sven Schmidt, Thomas Legler, Sebastian Schär, Wolfgang Lehner (Dresden University of Technology, Germany)
- 1302 **Loadstar: Load Shedding in Data Stream Mining**
Yun Chi (University of California, Los Angeles, USA), Haixun Wang, Philip S. Yu (IBM T. J. Watson Research Center, USA)

Information Extraction

- 1306 **iMeMex: Escapes from the Personal Information Jungle**
Jens-Peter Dittrich, Marcos Antonio Vaz Salles, Donald Kossmann, Lukas Blunschi (ETH Zurich, Switzerland)
- 1310 **Personalizing XML Text Search in PimenT**
Sihem Amer-Yahia (AT&T Labs – Research, USA), Irini Fundulaki (Bell Labs, USA), Prateek Jain (Bell Labs, USA and Indian Institute of Technology, Kanpur, India), Laks V.S. Lakshmanan (University of British Columbia, Canada)
- 1314 **WISE-Integrator: A System for Extracting and Integrating Complex Web Search Interfaces of the Deep Web**
Hai He, Weiyi Meng (SUNY at Binghamton, USA), Clement Yu (University of Illinois at Chicago, USA), Zonghuan Wu (University of Louisiana at Lafayette, USA)

XML Data Management

- 1318 **WmXML: A System for Watermarking XML Data**
Xuan Zhou (L3S Research Center, Germany), Hwee Hwa Pang (Institute for Infocomm Research, Singapore), Kian-Lee Tan, Dhruv Mangla (National University of Singapore, Singapore)
- 1322 **Pathfinder: XQuery - The Relational Way**
Peter Boncz (CWI Amsterdam, The Netherlands), Torsten Grust (Technische Universität München, Germany), Maurice van Keulen (University of Twente, The Netherlands), Stefan Manegold (CWI Amsterdam, The Netherlands), Jan Rittinger (CWI Amsterdam, The Netherlands, University of Konstanz, Germany), Jens Teubner (Technische Universität München, Germany)

Table of Contents

1326 **MIX: A Meta-data Indexing System for XML**
*SungRan Cho (University of Hannover, Germany),
Nick Koudas (University of Toronto, Canada), Divesh
Srivastava (AT&T Labs – Research, USA)*

1330 **ULoad: Choosing the Right Storage
for Your XML Application**
*Andrei Arion (INRIA Futurs & Universite Paris Sud,
France), Véronique Benzaken (Université Paris Sud XI,
France), Ioana Manolescu (INRIA Futurs, France),
Ravi Vijay (INRIA Futurs, France, Indian Institute of
Technology, Bombay)*

Advanced Applications

1334 **A Faceted Query Engine Applied to Archaeology**
*Kenneth Ross, Angel Janevski, Julia Stoyanovich
(Columbia University, USA)*

1338 **A Dynamically Adaptive Distributed System
for Processing Complex Continuous Queries**
*Bin Liu, Yali Zhu, Mariana Jbantova, Bradley
Momberger, Elke Rundensteiner (Worcester Polytechnic
Institute, USA)*

1342 **RankSQL: Supporting Ranking Queries in
Relational Database Management Systems**
*Chengkai Li (University of Illinois at Urbana-Champaign,
USA), Mohamed A. Soliman (University of Waterloo,
Canada), Kevin Chen-Chuan Chang (University of Illinois
at Urbana-Champaign, USA), Ihab F. Ilyas (University
of Waterloo, Canada)*

1346 **PSYCHO: A Prototype System for Pattern
Management**
*Barbara Catania, Anna Maddalena, Maurizio Mazza
(University of Genoa - DISI, Italy)*

Querying Databases with Uncertainties

1350 **Answering Imprecise Queries over Web
Databases**
*Ullas Nambiar, Subbarao Kambhampati (Arizona State
University, USA)*

1354 **ConQuer: A System for Efficient Querying
Over Inconsistent Databases**
*Ariel Fuxman (University of Toronto, Canada),
Diego Fuxman (Universidad Nacional del Sur,
Argentina), Renée J. Miller (University of Toronto,
Canada)*

1358 **QoS-based Data Access and Placement
for Federated Information Systems**
*Wen-Syan Li, Vishal Batra, Vijayshankar Raman,
Wei Han, Inderpal Narang (IBM Almaden Research
Center, USA)*

Tutorials

1363 **Approximate Joins: Concepts and Techniques**
*Nick Koudas (University of Toronto, Canada),
Divesh Srivastava (AT&T Labs – Research, USA)*

1364 **Offline and Data Stream Algorithms for Efficient
Computation of Synopsis Structures**
*Sudipto Guha (University of Pennsylvania, USA),
Kyuseok Shim (Seoul National University, Korea)*

1365 **Personalized Systems: Models and Methods
from an IR and DB Perspective**
*Yannis Ioannidis, Georgia Koutrika (University of Athens,
Greece)*

1366 **Contextual Insight in Search: Enabling
Technologies and Applications**
Aleksander Øhrn (Fast Search & Transfer ASA, Norway)

1367 **Semantic Overlay Networks**
*Karl Aberer, Philippe Cudré-Mauroux, (Ecole
Polytechnique Fédérale de Lausanne (EPFL),
Switzerland)*

1368 **XML Full-Text Search:
Challenges and Opportunities**
*Sihem Amer-Yahia (AT&T Labs – Research, USA),
Jayavel Shanmugasundaram (Cornell University, USA)*

1369 **Author Index**