

Table of Contents

Keynotes

1 **Data Access Patterns in The Amazon.com Technology Platform**
Werner Vogels (Amazon.com, USA)

2 **Technology for Developing Regions**
Eric Brewer (UC Berkeley, USA)

10 Year Best Paper Award

3 **Self-Tuning Database Systems: A Decade of Progress**
Surajit Chaudhuri, Vivek Narasayya (Microsoft Research, USA)

Research Sessions

Research Session 1: Uncertain and Probabilistic Data

15 **Probabilistic Skylines on Uncertain Data**
Jian Pei (Simon Fraser University, Canada), Bin Jiang, Xuemin Lin, Yidong Yuan (University of New South Wales, Australia)

27 **Matching Twigs in Probabilistic XML**
Benny Kimelfeld, Yehoshua Sagiv (Hebrew University, Israel)

39 **OLAP over Imprecise Data with Domain Constraints**
Doug Burdick, AnHai Doan (University of Wisconsin, USA), Raghu Ramakrishnan (Yahoo! Research, USA), Shivakumar Vaithyanathan (IBM Almaden Research Center, USA)

51 **Materialized Views in Probabilistic Databases for Information Exchange and Query Optimization**

Christopher Re, Dan Suciu (University of Washington, USA)

Research Session 2: XML Query Processing

63 **LCS-TRIM: Dynamic Programming Meets XML Indexing and Querying**
Shirish Tatikonda, Srinivasan Parthasarathy, Matthew Goyder (Ohio State University, USA)

75 **Extending XQuery with Window Functions**
Irina Botan, Peter M. Fischer (ETH Zurich, Switzerland), Dana Florescu (Oracle Corporation, USA), Donald Kossmann, Tim Kraska, Rokas Tamosevicius (ETH Zurich, Switzerland)

87 **Structured Materialized Views for XML Queries**
Andrei Arion (Inria Futurs and LRI, Universite Paris Sud, France), Veronique Benzaken (LRI, Universite Paris Sud, France), Ioana Manolescu (INRIA, France), Yannis Papakonstantinou (UC San Diego, USA)

99 **XRPC: Interoperable and Efficient Distributed XQuery**
Ying Zhang, Peter Boncz (CWI, the Netherlands)

Research Session 3: Outsourcing and Authentication

111 **Security in Outsourcing of Association Rule Mining**
Wai Kit Wong, David W. Cheung (The University of Hong Kong, China), Edward Hung (The Hong Kong Polytechnic University, China), Ben Kao, Nikos Mamoulis (The University of Hong Kong, China)

123 **Over-encryption: Management of Access Control Evolution on Outsourced Data**

Sabrina De Capitani di Vimercati, Sara Foresti (DTI - University of Milan, Italy), Sushil Jajodia (George Mason University, USA), Stefano Paraboschi (University of Bergamo, Italy), Pierangela Samarati (Univ. of Milan, Italy)

135 **CADS: Continuous Authentication on Data Streams**

Stavros Papadopoulos, Yin Yang, Dimitris Papadias (HKUST, China)

147 **Proof-Infused Streams: Enabling Authentication of Sliding Window Queries On Streams**

Feifei Li (Boston University, USA), Ke Yi, Marios Hadjieleftheriou (AT&T Labs, USA), George Kollios (Boston University, USA)

Research Session 4: Data Stream Processing

159 **Staying FIT: Efficient Load Shedding Techniques for Distributed Stream Processing**

Nesime Tatbul (ETH Zurich, Switzerland), Ugur Cetintemel, Stan Zdonik (Brown University, USA)

171 **A Simple and Efficient Estimation Method for Stream Expression Cardinalities**

Aiyou Chen, Jin Cao, Tian Bu (Bell Labs, Alcatel-Lucent, USA)

183 **Ad-hoc Top-k Query Answering for Data Streams**

Gautam Das (University of Texas at Arlington, USA), Dimitrios Gunopulos (University of California – Riverside, USA), Nick Koudas, Nikos Sarkas (University of Toronto, Canada)

Research Session 5: Text Databases

195 **Extending Q-Grams to Estimate Selectivity of String Matching with Low Edit Distance**

Hongrae Lee, Raymond Ng (Univ. of British Columbia, Canada), Kyuseok Shim (Seoul National University)

207 **Fast nGram-Based String Search Over Data Encoded Using Algebraic Signatures**

Witold Litwin (CERIA, France), Riad Mokadem (CERIA, Paris Dauphine University, France), Philippe Rigaux (Univ. Paris-Dauphine and INRIA-Orsay, Equipe Gemo, France), Thomas Schwarz (Univ. Santa Clara, USA)

219 **Effective Phrase Prediction**

Arnab Nandi, H.V. Jagadish (University of Michigan, USA)

Research Session 6: Relational Models and Views

231 **Lazy Maintenance of Materialized Views**

Jingren Zhou, Paul Larson (Microsoft Research, USA), Hicham G. Elmongui (Purdue University, USA)

243 **Extending Dependencies with Conditions**

Loreto Bravo (University of Edinburgh, United Kingdom), Wenfei Fan (University of Edinburgh/Bell Labs, United Kingdom), Shuai Ma (University of Edinburgh, United Kingdom)

255 **Unifying Data and Domain Knowledge Using Virtual Views**

Lipyew Lim, Haixun Wang, Min Wang (IBM Watson Research Center, USA)

Research Session 7: Skyline Query Processing

- 267 **Efficient Skyline Computation over Low-Cardinality Domains**
Michael Morse, Jignesh Patel, H.V. Jagadish (University of Michigan, USA)
- 279 **Approaching the Skyline in Z Order**
Ken Lee (Penn State University, USA), Baihua Zheng (Singapore Management University), Huajing Li, Wang-Chien Lee (Penn State University, USA)
- 291 **Efficient Computation of Reverse Skyline Queries**
Evangelos Dellis, Bernhard Seeger (University of Marburg, Germany)

Research Session 8: Data Quality

- 303 **VGRAM: Improving Performance of Approximate Queries on String Collections Using Variable-Length Grams**
Chen Li (University of California, Irvine, USA), Bin Wang, Xiaochun Yang (Northeastern University, China)
- 315 **Improving Data Quality: Consistency and Accuracy**
Gao Cong (Microsoft Research, Asia, China), Wenfei Fan (University of Edinburgh/Bell Labs, United Kingdom), Floris Geerts (University of Edinburgh/University of Hasselt/Transnational Univ. Limburg, United Kingdom), Xibei Jia, Shuai Ma (University of Edinburgh, United Kingdom)
- 327 **Example-driven design of efficient record matching queries**
Surajit Chaudhuri (Microsoft Research, USA), Bee-Chung Chen (University of Wisconsin - Madison, USA), Venkatesh Ganti, Raghav Kaushik (Microsoft Research, USA)

Research Session 9: Novel Architectures

- 339 **Adaptive Aggregation on Chip Multiprocessors**
John Cieslewicz, Kenneth Ross (Columbia University, USA)
- 351 **To Share or Not To Share?**
Ryan Johnson, Nikos Hardavellas, Ippokratis Pandis, Naju Mancheril (Carnegie Mellon University, USA), Stavros Harizopoulos (MIT, USA), Kivanc Sabirli (Carnegie Mellon University, USA), Anastasia Ailamaki (Carnegie Mellon University, USA, and EPFL, Switzerland), Babak Falsafi (EPFL, Switzerland)
- 363 **Executing Stream Joins on the Cell Processor**
Bugra Gedik, Philip Yu, Rajesh Bordawekar (IBM Watson Research Center, USA)

Research Session 10: Web Data Management and Search

- 375 **RankMass Crawler: A Crawler with High PageRank Coverage Guarantee**
Junghoo Cho, Uri Schonfeld (UCLA, USA)
- 387 **EntityRank: Searching Entities Directly and Holistically**
Tao Cheng, Xifeng Yan, Kevin Chen-Chuan Chang (University of Illinois at Urbana-Champaign, USA)
- 399 **Building Structured Web Community Portals: A Top-Down, Compositional, and Incremental Approach**
Pedro DeRose, Warren Shen, Fei Chen, An-Hai Doan (University of Wisconsin, USA), Raghu Ramakrishnan (Yahoo! Research, USA)
- 411 **Scalable Semantic Web Data Management Using Vertical Partitioning**
Daniel Abadi, Adam Marcus, Samuel Madden, Katherine Hollenbach (MIT, USA)

Research Session 11: Time-Series Data Mining

423 **Ranked Subsequence Matching in Time-Series Databases**

Wook-Shin Han, Jinsoo Lee (Kyungpook National University), Yang-Sae Moon (Kangwon National University, Republic of Korea), Haifeng Jiang (IBM Almaden Research Center, USA)

435 **Indexable PLA for Efficient Similarity Search**

Qiuxia Chen, Lei Chen, Xiang Lian, Yunhao Liu (HKUST, China), Jeffrey Xu Yu (The Chinese University of Hong Kong, China)

447 **Mining Approximate Top-K Subspace Anomalies in Multi-Dimensional Time-Series Data**

Xiaolei Li, Jiawei Han (UIUC, USA)

459 **Time Series Compressibility and Privacy**

Spiros Papadimitriou (IBM Watson Research Center, USA), Feifei Li, George Kollios (Boston University, USA), Philip Yu (IBM Watson Research Center, USA)

Research Session 12: Top-k Queries and Ranking I

471 **A Bayesian Method for Guessing the Extreme Values in a Data Set**

Mingxi Wu, Chris Jermaine (University of Florida, USA)

483 **Efficient Processing of Top-k Dominating Queries on Multi-Dimensional Data**

Man Lung Yiu (Aalborg University, Denmark), Nikos Mamoulis (University of Hong Kong, China)

495 **Best Position Algorithms for Top-k Queries**

Reza Akbarinia (ATLAS group, INRIA, France), Esther Pacitti, Patrick Valduriez (ATLAS group, INRIA and LINA, France)

507 **Sum-Max Monotonic Ranked Joins for Evaluating Top-K Twig Queries on Weighted Data Graphs**

Yan Qi, K. Selcuk Candan (Arizona State University, USA), Maria Luisa Sapino (University of Torino, Italy)

Research Session 13: Private and Secure Databases

519 **Answering Aggregation Queries in a Secure System Model**

Tingjian Ge, Stan Zdonik (Brown University, USA)

531 **The Boundary Between Privacy and Utility in Data Publishing**

Vibhor Rastogi, Sungho Hong, Dan Suciu (University of Washington, USA)

543 **Minimality Attack in Privacy Preserving Data Publishing**

Raymond Chi-Wing Wong, Ada Wai-Chee Fu (The Chinese University of Hong Kong, China), Ke Wang, Jian Pei (Simon Fraser University, Canada)

555 **On the Correctness Criteria of Fine-Grained Access Control in Relational Databases**

Qihua Wang (Purdue University, USA), Ting Yu (North Carolina State University, USA), Ninghui Li (Purdue University, USA), Jorge Lobo (IBM Watson Research Center, USA), Elisa Bertino (Purdue University, USA), Keith Irwin (North Carolina State University, USA), Ji-Won Byun (Purdue University, USA)

Research Session 14: Spatial Databases

567 **Modeling and Querying Vague Spatial Objects Using Shapelets**

Daniel Zinn, Jim Bosch, Michael Gertz (University of California at Davis, USA)

579 **On Efficient Spatial Matching**
Raymond Chi-Wing Wong, Yufei Tao, Ada Wai-Chee Fu, Xiaokui Xiao (The Chinese University of Hong Kong, China)

591 **Main-Memory Operation Buffering for Efficient R-Tree Update**
Laurynas Biveinis, Simonas Saltenis, Christian S. Jensen (Aalborg University, Denmark)

Research Session 15: Business and Web Services

603 **Monitoring Business Processes with Queries**
Catriel Beeri (Hebrew University, Israel), Alon Pilberg, Tova Milo, Anat Eyal (Tel Aviv University, Israel)

615 **An Approach to Optimize Data Processing in Business Processes**
Marko Vrhovnik, Holger Schwarz (University of Stuttgart, Germany), Oliver Suhre (IBM Deutschland Entwicklung GmbH, Germany), Bernhard Mitschang (University of Stuttgart, Germany), Volker Markl (IBM Almaden Research Center, USA), Albert Maier (IBM Deutschland Entwicklung GmbH, Germany), Tobias Kraft (University of Stuttgart, Germany)

627 **Reasoning about the Behavior of Semantic Web Services with Concurrent Transaction Logic**
Dumitru Roman (University of Innsbruck / DERI Innsbruck, Austria), Michael Kifer (Stony Brook University, USA)

Research Session 16: Information Integration I

639 **Randomized Algorithms for Data Reconciliation in Wide Area Aggregate Query Processing**
Fei Xu, Chris Jermaine (University of Florida, USA)

651 **Query Processing over Incomplete Autonomous Databases**
Garrett Wolf, Hemal Kathri, Bhaumik Chokshi, Jianchun Fan, Yi Chen, Subbarao Kambhampati (Arizona State University, USA)

663 **iTrails: Pay-as-you-go Information Integration in Dataspaces**
Marcos Antonio Vaz Salles, Jens-Peter Dittrich, Shant Kirakos Karakashian, Olivier René Girard, Lukas Blunzsch (ETH Zurich, Switzerland)

Research Session 17: Information Integration II

675 **Update Exchange with Mappings and Provenance**
Todd Green, Grigoris Karvounarakis, Zachary Ives, Val Tannen (University of Pennsylvania, USA)

687 **Data Integration with Uncertainty**
Xin Dong (University of Washington, USA), Alon Halevy (Google Inc., USA), Cong Yu (University of Michigan, USA)

699 **Context-Aware Wrapping: Synchronized Data Extraction**
Shui-Lung Chuang, Kevin Chen-Chuan Chang, ChengXiang Zhai (University of Illinois at Urbana-Champaign, USA)

Research Session 18: Query Processing

711 **Processing Forecasting Queries**
Songyun Duan, Shivnath Babu (Duke University, USA)

723 **Cooperative Scans: Dynamic Bandwidth Sharing in a DBMS**
Marcin Zukowski, Sándor Héman, Niels Nes, Peter Boncz (CWI, the Netherlands)

735 **Stop-and-Restart Style Execution for Long Running Decision Support**

Queries

Surajit Chaudhuri, Raghav Kaushik, Ravishankar Ramamurthy (Microsoft Research, USA), Abhijit Pol (University of Florida, USA)

Research Session 19: Data Privacy, Anonymization, and Outsourcing

746 K-Anonymization as Spatial Indexing: Toward Scalable and Incremental Anonymization

Tochukwu Iwuchukwu, Jeffrey Naughton (University of Wisconsin - Madison, USA)

758 Fast Data Anonymization with Low Information Loss

Gabriel Ghinita (NUS), Panagiotis Karras (HKU), Panos Kalnis (NUS), Nikos Mamoulis (University of Hong Kong, China)

770 Privacy Skyline: Privacy with Multidimensional Adversarial Knowledge

Bee-Chung Chen (University of Wisconsin - Madison, USA), Raghu Ramakrishnan (Yahoo! Research, USA), Kristen LeFevre (University of Wisconsin - Madison, USA)

782 Integrity Auditing of Outsourced Data

Min Xie (Renmin University of China, China), Haixun Wang, Jian Yin (IBM Watson Research Center, USA), Xiaofeng Meng (Renmin University of China)

Research Session 20: Novel Data Mining Applications

794 Adaptive Fastest Path Computation on a Road Network: A Traffic Mining Approach

Hector Gonzalez, Jiawei Han, Xiaolei Li, Margaret Myslinska, John Paul Sondag (UIUC, USA)

806 Seeking Stable Clusters in the Blogosphere

Nilesh Bansal, Fei Chiang, Nick Koudas (University of Toronto, Canada), Frank Wm. Tompa (University of Waterloo, Canada)

818 On Dominating Your Neighborhood Profitably

Cuiping Li (Renmin University of China), Anthony Tung (National University of Singapore), Wen Jin, Martin Ester (Simon Fraser University)

830 Detecting Attribute Dependencies from Query Feedback

Peter Haas (IBM Almaden Research Center, USA), Fabian Hueske (University of Ulm, Germany), Volker Markl (IBM Almaden Research Center, USA)

Research Session 21: Sensor Networks and Information Dissemination

842 Making Sense of Suppressions and Failures in Sensor Data: A Bayesian Approach

Adam Silberstein, Alan Gelfand, Kamesh Munagala, Gavino Puggioni, Jun Yang (Duke University, USA)

854 MIST: Distributed Indexing and Querying in Sensor Networks using Statistical Models

Arnab Bhattacharya, Anand Meka, Ambuj Singh (UCSB, USA)

866 Early Profile Pruning on XML-aware Publish/Subscribe Systems

Mirella Moro, Petko Bakalov, Vassilis Tsotras (University of California, Riverside, USA)

878 Value-Based Notification Conditions in Large-Scale Publish/Subscribe Systems

Badrish Chandramouli, Jeff Phillips, Jun Yang (Duke University, USA)

Research Session 22: Top-k Queries and Ranking II

890 Efficiently Answering Top-k Typicality

Queries on Large Databases

Ming Hua, Jian Pei (Simon Fraser University), Ada Wai-Chee Fu (The Chinese University of Hong Kong, China), Xuemin Lin (University of New South Wales, Australia), Hong-fung Leung (The Chinese University of Hong Kong, China)

902 Depth Estimation for Ranking Query Optimization

Karl Schnaitter (UC Santa Cruz, USA), Joshua Spiegel (BEA Systems, Inc.), Neoklis Polyzotis (UC Santa Cruz, USA)

914 Anytime Measures for Top-k Algorithms

Benjamin Arai (University of California, Riverside, USA), Gautam Das (University of Texas at Arlington, USA), Dimitrios Gunopulos (University of California, Riverside,), Nick Koudas (University of Toronto, Canada)

Research Session 23: Indexing and Search

926 Towards Graph Containment Search and Indexing

Chen Chen (UIUC, USA), Xifeng Yan, Philip Yu (IBM Watson Research Center, USA), Jiawei Han (UIUC, USA), Dong-Qing Zhang (Thomson Images & Beyond), Xiaohui Gu (IBM Watson Research Center, USA)

938 Graph Indexing: Tree + Delta \geq Graph

Peixiang Zhao, Jeffrey Xu Yu (The Chinese University of Hong Kong, China), Philip Yu (IBM Watson Research Center, USA)

950 Multi-Probe LSH: Efficient Indexing for High-Dimensional Similarity Search

Qin Lv, William Josephson, Zhe Wang, Moses Charikar, Kai Li (Princeton University, USA)

Research Session 24: Distributed Data Management

962 STAR: Self-Tuning Aggregation for Scalable Monitoring

Navendu Jain, Mike Dahlin, Yin Zhang, Dmitry Kit, Prince Mahajan (University of Texas at Austin, USA), Praveen Yalagandula (HP Labs, USA)

974 SQLB: A Query Allocation Framework for Autonomous Consumers and Providers

Jorge-Arnulfo Quiané-Ruiz (ATLAS group, INRIA and LINA, France), Philippe Lamarre (LINA, Univ. de Nantes, France), Patrick Valduriez (ATLAS group, INRIA and LINA, France)

986 Peer-to-Peer Similarity Search in Metric Spaces

Christos Doukeridis, Akrivi Vlachou, Yannis Kotidis, Michalis Vazirgiannis (AUEB, Greece)

Research Session 25: Schema and Structure Management

998 Inferring XML Schema Definitions from XML Data

Geert Jan Bex, Frank Neven, Stijn Vansummeren (Hasselt University/Transnational University of Limburg, Belgium)

1010 Querying Complex Structured Databases

Cong Yu, H.V. Jagadish (University of Michigan, USA)

1022 Measuring the Structural Similarity of Semistructured Documents Using Entropy

Sven Helmer (University of London, Birkbeck, UK)

Research Session 26: Information Extraction and Text

- 1033 **Declarative Information Extraction Using Datalog with Embedded Extraction Predicates**
Warren Shen, AnHai Doan, Jeffrey Naughton (University of Wisconsin - Madison, USA), Raghu Ramakrishnan (Yahoo! Research, USA)
- 1045 **A Relational Approach to Incrementally Extracting and Querying Structure in Unstructured Data**
Eric Chu, Akanksha Baid, Ting Chen, AnHai Doan, Jeffrey Naughton (University of Wisconsin - Madison, USA)
- 1057 **Efficient Keyword Search over Virtual XML Views**
Feng Shao, Lin Guo, Chavdar Botev, Anand Bhaskar, Muthiah Chettiar, Fan Yang (Cornell University, USA), Jayavel Shanmugasundaram (Yahoo! Research, USA)

Research Session 27: Query Optimization for Novel Applications

- 1069 **A General Framework for Modeling and Processing Optimization Queries**
Michael Gibas, Ning Zheng, Hakan Ferhatosmanoglu (Ohio State University, USA)
- 1081 **On the Production of Anorexic Plan Diagrams**
Harish D, Pooja Darera, Jayant Haritsa (Indian Institute of Science)
- 1093 **Efficient Use of the Query Optimizer for Automated Database Design**
Stratos Papadomanolakis, Debabrata Dash (Carnegie Mellon University, USA), Anastasia Ailamaki (Carnegie Mellon University, USA, and EPFL, Switzerland)

Industrial, Application, and Experience Sessions

Industrial Session 1: Decision Support

- 1105 **Dynamic Workload Management for Very Large Data Warehouses: Juggling Feathers and Bowling Balls**
Stefan Krompass (Technische Universität München, Germany), Umesh Dayal, Harumi Kuno (HP Labs, USA), Alfons Kemper (Technische Universität München, Germany)
- 1116 **Tracing Lineage Beyond Relational Operators**
Mingwu Zhang, Xiangyu Zhang, Xiang Zhang, Sunil Prabhakar (Purdue University, USA)
- 1128 **A Generic solution for Warehousing Business Process Data**
Fabio Casati (University of Trento, Italy), Malu Castellanos, Umesh Dayal (HP Labs, USA), Norman Salazar
- 1138 **Why You Should Run TPC-DS: A Workload Analysis**
Meikel Poess (Oracle Corporation, USA), Raghunath Othayoth Nambiar (Hewlett-Packard Company, USA), David Walrath (Sybase)

Industrial Session 2: Invited Talks

- 1150 **The End of an Architectural Era (It's Time for a Complete Rewrite)**
Michael Stonebraker, Samuel Madden, Daniel J. Abadi, Stavros Harizopoulos (MIT, USA), Nabil Hachem (AvantGarde Consulting, USA), Pat Helland (Microsoft Corporation, USA)
- 1161 **Computer Science 2.0: A New World of Data Management**
Michael L. Brodie (Verizon, USA)

Industrial Session 3: Data Streams

1162 **RadixZip: Linear-Time Compression of Token Streams**

Binh Vo, Gurmeet Singh Manku (Google Inc., USA)

1173 **Continuous Queries in Oracle**

Sankar Subramanian, Srikanth Bellamkonda, Hua-Gang Li, Vince Liang, Lei Sheng, Wayne Smith, James Terry, Tsae-Feng Yu, Andrew Witkowski (Oracle Corporation, USA)

1185 **Challenges and Experience in Prototyping a Multi-Modal Stream Analytic and Monitoring Application on System S**

Kun-Lung Wu, Philip Yu, Bugra Gedik, Kirsten Hildrum, Charu Aggarwal, Eric Bouillet, Wei Fan, David George, Xiaohui Gu, Gang Luo, Haixun Wang (IBM Watson Research Center, USA)

Industrial Session 4: Query Processing Engines

1197 **Efficient Bulk Deletes for Multi Dimensionally Clustered Tables in DB2**

Bishwaranjan Bhattacharjee, Timothy Malkemus (IBM Watson Research Center, USA), Sherman Lau, Sean Mckeough, Jo-Anne Kirton, Robin Von Boeschoten, John Kennedy (IBM Toronto Laboratories, Canada)

1207 **Supporting Time-Constrained SQL Queries in Oracle**

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

1219 **Request Window: an Approach to Improve Throughput of RDBMS-based Data Integration System by Utilizing Data Sharing Across Concurrent Distributed Queries**

Rubao Lee, Minghong Zhou, Huaming Liao (ICT/CAS, China)

1231 **Inverse Functions in the AquaLogic Data Services Platform**

Nicola Onose (UC San Diego, USA), Vinayak Borkar, Michael Carey (BEA Systems, Inc., USA)

Industrial Session 5: Profiling

1243 **A genetic approach for random testing of database systems**

Hardik Bati, Leo Giakoumakis, Steve Herbert, Aleksandras Surna (Microsoft Corporation)

1252 **Bridging the Application and DBMS Profiling Divide for Database Application Developers**

Surajit Chaudhuri, Vivek Narasayya, Manoj Syamala (Microsoft Research, USA)

1263 **Automating the Detection of Snapshot Isolation Anomalies**

Sudhir Jorwekar (IIT Bombay, India), Alan Fekete (University of Sydney, Australia), Krithi Ramamritham, S Sudarshan (IIT Bombay, India)

Industrial Session 6: Engine Infrastructure

1275 **Optimization of Frequent Itemset Mining on Multiple-Core Processor**

Eric Li (Intel Corp.), Li Liu (Tsinghua University and Intel China Research Center, China)

1286 **CellSort: High Performance Sorting on the Cell Processor**

Bugra Gedik, Rajesh Bordawekar, Philip Yu (IBM Watson Research Center, USA)

1298 **Increasing Buffer-Locality for Multiple Index Based Scans through Intelligent Placement and Index Scan Speed Control**

Christian A. Lang, Bishwaranjan Bhattacharjee, Timothy Malkemus (IBM Watson Research Center, USA), Kwai Wong (IBM Toronto Laboratories, Canada)

Demo Sessions

Demo Group I

1310 **SciPort: An Adaptable Scientific Data Integration Platform for Collaborative Scientific Research**

Fusheng Wang (Siemens Corporate Research, USA), Cornelius Rabsch (University of Mannheim, Germany), Patrick Kling (University of Waterloo, Canada), Gerald Madlmayr (Research Center Hagenberg, Austria), Peiya Liu (Siemens Corporate Research, USA), David Kaltschmidt (Freie University Berlin, Germany), Mo Wang (University of Duisburg-Essen, Germany), John Pearson, Joe Carpinelli (Siemens Corporate Research, USA), Pierre-Emmanuel Bourgue (Universite de Technologie de Compiègne, France), Georg Hackenberg (University of Mannheim, Germany)

1314 **DataScope: Viewing Database Contents in Google Maps' Way**

Tianyi Wu, Xiaolei Li, Dong Xin, Jiawei Han, Jacob Lee, Ricardo Redder (UIUC, USA)

1318 **XBenchMatch: a Benchmark for XML Schema Matching Tools**

Fabien Duchateau, Zohra Bellahsene (Lirimm - Univ. Montpellier 2, France), Ela Hunt (ETH Zurich, Switzerland)

1322 **GeRoMeSuite: A System for Holistic Generic Model Management**

David Kensche, Christoph Quix, Xiang Li, Yong Li (RWTH Aachen University, Germany)

1326 **Semi-Automatic Schema Integration in Clio**

Laura Chiticariu (UC Santa Cruz, USA),

Mauricio Hernandez, Lucian Popa, Phokion Kolaitis (IBM Almaden Research Center, USA)

1330 **XSeek: A Semantic XML Search Engine Using Keywords**

Ziyang Liu, Jeffrey Walker, Yi Chen (Arizona State University, USA)

1334 **Self-Organizing Schema Mappings in the GridVine Peer Data Management System**

Philippe Cudre-Mauroux, Suchit Agarwal, Adriana Budura, Parisa Haghani, Karl Aberer (EPFL, Switzerland)

1338 **CallAssist: Helping Call Center Agents in Preference Elicitation**

Ullas Nambiar, Himanshu Gupta, Mukesh Mohania (IBM India Research Lab)

1342 **NS2: Networked Searchable Store with Correctness**

Radu Sion, Sumeet Bajaj (Stony Brook University, USA), Bogdan Carbunar (Motorola Labs), Stefan Katzenbeisser (Technische Universität München, Germany)

1346 **GhostDB: Hiding Data from Prying Eyes**

Christophe Salperwyck, Nicolas AnCIAUX (INRIA, France), Mehdi Benzine (INRIA, UVSQ, France), Luc Bouganim (INRIA, France), Philippe Pucheral (INRIA, UVSQ, France), Dennis Shasha (New York University, USA)

Demo Group II

1350 **FuSem - Exploring Different Semantics of Data Fusion**

Jens Bleiholder, Karsten Draba, Felix Naumann (Hasso-Plattner-Institut, Potsdam, Germany)

1354 **IndeGS: Index Supported Graphics Data Server for CFD Data Postprocessing**

Christoph Brochhaus, Thomas Seidl (RWTH Aachen University, Germany)

1358 **A STEP Towards Realizing Codd's Vision of Rendezvous with the Casual User**

Michael Minock (Umea University, Sweden)

1362 **TRAX - Real-World Tracking of Moving Objects**

Christian S. Jensen, Stardas Pakalnis (Aalborg University, Denmark)

1366 **Zoom*UserViews: Querying Relevant Provenance in Workflow Systems**

Olivier Biton, Sarah Cohen-Boulakia, Susan B. Davidson (University of Pennsylvania, USA)

1370 **DAMIA - A Data Mashup Fabric for Intranet Applications**

Mehmet Altinel Paul Brown, Susan Cline, Rajesh Kartha, Eric Louie, Volker Markl, Louis Mau, Yip-Hing Ng, David Simmen, Ashutosh Singh (IBM Almaden Research Center, USA)

1374 **UQLIPS: A Real-time Near-duplicate Video Clip Detection System**

Heng Tao Shen, Xiaofang Zhou, Zi Huang, Jie Shao, Xiangmin Zhou (University of Queensland, Australia)

1378 **The GCX System: Dynamic Buffer Minimization in Streaming XQuery Evaluation**

Christoph Koch, Stefanie Scherzinger, Michael Schmidt (Saarland University, Germany)

1382 **A Cost-Estimation Component for Statement Sequences**

Tobias Kraft (University of Stuttgart, Germany)

1386 **Eliminating Impedance Mismatch in C++**

Joseph (Yossi) Gil, Keren Lenz (Technion, Israel)

Demo Group III

1390 **Large Scale P2P Distribution of Open-Source Software**

Serge Abiteboul (INRIA-Orsay, France), Itay Dar (Tel Aviv University, Israel), Radu Pop (INRIA and Mandriva, France), Gabriel Vasile (INRIA-Orsay, France), Dan Vodislav (CEDRIC-CNAM Paris, France), Nicoleta Preda (INRIA-Orsay, France)

1394 **HiSbase: Histogram-based P2P Main Memory Data Management**

Tobias Scholl, Bernhard Bauer, Benjamin Gufler, Richard Kuntschke, Daniel Weber, Angelika Reiser, Alfons Kemper (Technische Universität München, Germany)

1398 **P2P Authority Analysis for Social Communities**

Josiane Xavier Parreira, Sebastian Michel, Matthias Bender, Tom Crecelius, Gerhard Weikum (Max-Planck Institute for Informatics, Germany)

1402 **SOR: A Practical System for Ontology Storage, Reasoning and Search**

Jing Lu (Shanghai JiaoTong University, China), Li Ma (IBM China Research Lab, China), Lei Zhang (IBM), Jean-Sebastien Brunner, Chen Wang, Yue Pan (IBM China Research Lab, China), Yong Yu (SJTU)

1406 **Periscope/SQ: Interactive Exploration of Biological Sequence Databases**

Sandeep Tata, Willis Lang, Jignesh Patel (University of Michigan, USA)

1410 **BlogScope: A System for Online Analysis of High Volume Text Streams**

Nilesh Bansal, Nick Koudas (University of Toronto, Canada)

1414 **FluxCapacitor: Efficient Time-Travel Text Search**

Klaus Berberich, Srikanta Bedathur, Thomas Neumann, Gerhard Weikum (Max-Planck Institute for Informatics, Germany)

1418 **Deadline and QoS Aware Data Warehouse**

Wen-Syan Li, Dengfeng Gao, Rafae Bhatti, Inderpal Narang (IBM Almaden Research Center, USA), Hirofumi Matsuzawa (IBM Japan, Japan), Masayuki Numao (IBM Tokyo Research Laboratory, Japan), Masahiro Ohkawa, Takeshi Fukuda (IBM Yamato Software Laboratory, Japan)

1422 **Query language support for incomplete information in the MayBMS system**

Lyublena Antova, Christoph Koch, Dan Olteanu (Saarland University, Germany)

Tutorials

1426 **Adaptive query processing: Why, How, When, and What Next?**

Zachary Ives (University of Pennsylvania, USA), Amol Deshpande (University of Maryland at Collge Park, USA), Vijayshankar Raman (IBM, USA)

1428 **Raster databases**

Peter Baumann (Jacobs University Bremen, Germany)

1429 **From Data Privacy to Location Privacy: Models and Algorithms**

Ling Liu (Georgia Institute of Technology, USA)

1431 **Secure Data Outsourcing**

Radu Sion (Stony Brook University, USA)

1433 **Regulatory-Compliant Data Management**

Radu Sion (Stony Brook University, USA), Marianne Winslett (UIUC, USA)

1435 **Probabilistic Graphical Models and their Role in Databases**

Amol Deshpande (University of Maryland at Collge Park, USA), Sunita Sarawagi (IIT Bombay, India)

1437 **XML retrieval: db/ir in theory, web in practice**

Mariano Consens (University of Toronto, Canada), Ricardo Baeza-Yates (Yahoo! Research, Spain), Mounia Lalmas (Queen Mary, University of London, UK), Sihem Amer-Yahia (Yahoo! Research, USA)

1439 **Model Management and Schema Mappings: Theory and Practice**

Phil Bernstein (Microsoft Corporation, USA), Howard Ho (IBM Almaden Research Center, USA)

Panels

1441 **Performance Evaluation and Experimental Assessment — Conscience or Curse of Database Research?**

Ioana Manolescu (INRIA, France), Stefan Manegold (CWI, The Netherlands)

1443 **What does Web 2.0 have to do with databases?**

Sihem Amer-Yahia (Yahoo! Research, USA), Alon Halevy (Google Inc., USA)

1444 **Author Index**