

# **2010 IEEE 26th International Conference on Data Engineering Workshops**

**(ICDEW 2010)**

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



**IEEE Catalog Number: CFP1045A-PRT  
ISBN: 978-1-4244-6522-4**

---

## 4th International Workshop on Ranking in Databases (DBRank'10)

---

- 1        **DBRank'10 Preface Pages**
- 4        **Subspace Similarity Search Using the Ideas of Ranking and Top-k Retrieval**  
*Thomas Bernecker, Tobias Emrich, Franz Graf, Hans-Peter Kriegel, Peer Kröger, Matthias Renz, Erich Schubert, Arthur Zimek, LMU München, Germany*
- 10       **Efficient k-Nearest Neighbor Queries with the Signature Quadratic Form Distance**  
*Christian Beecks, Merih Seran Uysal, Thomas Seidl, RWTH Aachen University, Germany*
- 16       **Top-k Pipe Join**  
*Davide Martinenghi, Marco Tagliasacchi, Politecnico di Milano, Italy*
- 20       **On Novelty in Publish/Subscribe Delivery**  
*Dimitris Souravlias, Marina Drosou, Kostas Stefanidis, Evaggelia Pitoura, University of Ioannina, Greece*
- 23       **Ranking for Data Repairs**  
*Mohamed Yakout, Ahmed K. Elmagarmid, Jennifer Neville, Purdue University, USA*
- 

## 2nd International Workshop on Information & Software as Services (WISS'10)

---

- 29       **WISS'10 Preface Pages**
- 33       **End-to-End Confidentiality for a Message Warehousing Service Using Identity-Based Encryption**  
*Yuecel Karabulut<sup>1</sup>, Harald Weppner<sup>1</sup>, Ike Nassi<sup>1</sup>, Anusha Nagarajan<sup>2</sup>, Yash Shroff<sup>2</sup>, Nishant Dubey<sup>2</sup>, Tyelisa Shields<sup>2</sup>*  
*<sup>1</sup>SAP, USA; <sup>2</sup>Carnegie Mellon University, USA*
- 41       **The HiBench Benchmark Suite: Characterization of the MapReduce-Based Data Analysis**  
*Shengsheng Huang, Jie Huang, Jinquan Dai, Tao Xie, Bo Huang, Intel, China*
- 52       **Towards Enterprise Software as a Service in the Cloud**  
*Jan Schaffner<sup>1</sup>, Dean Jacobs<sup>2</sup>, Benjamin Eckart<sup>1</sup>, Jan Brunnert<sup>1</sup>, Alexander Zeier<sup>1</sup>*  
*<sup>1</sup>HPI, Germany; <sup>2</sup>SAP, Germany*

---

## 5th International Workshop on Self-Managing Database Systems (SMDB'10)

---

- 60      **SMDB'10 Preface Pages**
- 63      **A Generic Auto-Provisioning Framework for Cloud Databases**  
*Jennie Rogers<sup>1</sup>, Olga Papaemmanouil<sup>2</sup>, Ugur Cetintemel<sup>1</sup>*  
<sup>1</sup>Brown University, USA; <sup>2</sup>Brandeis University, USA
- 69      **Adaptive Indexing for Relational Keys**  
*Goetz Graefe, Harumi Kuno, HP, USA*
- 75      **Autonomic Workload Execution Control Using Throttling**  
*Wendy Powley<sup>1</sup>, Patrick Martin<sup>1</sup>, Mingyi Zhang<sup>1</sup>, Paul Bird<sup>2</sup>, Keith McDonald<sup>2</sup>*  
<sup>1</sup>Queen's University, Canada; <sup>2</sup>IBM, Canada
- 81      **On the Use of Query-Driven XML Auto-Indexing**  
*Karsten Schmidt, Theo Härder, University of Kaiserslautern, Germany*
- 87      **Statistics-Driven Workload Modeling for the Cloud**  
*Archana Ganapathi, Yanpei Chen, Armando Fox, Randy Katz, David Patterson, University of California at Berkeley, USA*
- 93      **Vertical Partitioning of Relational OLTP Databases Using Integer Programming**  
*Rasmus Resen Amossen, IT University of Copenhagen, Denmark*
- 99      **Automatic Tuning of the Multiprogramming Level in Sybase SQL Anywhere**  
*Mohammed Abouzour<sup>1</sup>, Kenneth Salem<sup>2</sup>, Peter Bumbulis<sup>1</sup>*  
<sup>1</sup>Sybase iAnywhere, Canada; <sup>2</sup>University of Waterloo, Canada
- 105     **Caching All Plans with Just One Optimizer Call**  
*Debabrata Dash, Ioannis Alagiannis, Cristina Maier, Anastasia Ailamaki, EPFL, Switzerland*
- 111     **Towards Workload-Aware Self-Management: Predicting Significant Workload Shifts**  
*Marc Holze, Ali Haschimi, Norbert Ritter, University of Hamburg, Germany*

---

## 2nd Workshop on Management and Mining of Uncertain Data (MOUND'10)

---

- 117     **MOUND'10 Preface Pages**
- 120     **Constrained Frequent Itemset Mining from Uncertain Data Streams**  
*Carson Kai-Sang Leung, Boyu Hao, Fan Jiang, University of Manitoba, Canada*
- 128     **Cleansing Uncertain Databases Leveraging Aggregate Constraints**  
*Haiquan Chen<sup>1</sup>, Wei-Shinn Ku<sup>1</sup>, Haixun Wang<sup>2</sup>*  
<sup>1</sup>Auburn University, USA; <sup>2</sup>Microsoft, China
- 136     **U-DBSCAN : A Density-Based Clustering Algorithm for Uncertain Objects**  
*Apinya Tepwankul, Songrit Maneewongwattana, King Mongkut's University of Technology Thonburi, Thailand*

---

## 2nd International Workshop on New Trends in Information Integration (NTII'10)

---

- 144      **NTII'10 Preface Pages**
- 147      **A First Step Towards Integration Independence**  
*Laura M. Haas<sup>1</sup>, Renée J. Miller<sup>2</sup>, Donald Kossmann<sup>3</sup>, Martin Hentschel<sup>3</sup>*  
<sup>1</sup>IBM, USA; <sup>2</sup>University of Toronto, Canada; <sup>3</sup>ETH Zürich, Switzerland
- 151      **Towards Best-Effort Merge of Taxonomically Organized Data**  
*David Thau<sup>1</sup>, Shawn Bowers<sup>2</sup>, Bertram Ludäscher<sup>1</sup>*  
<sup>1</sup>University of California at Davis, USA; <sup>2</sup>Gonzaga University, USA
- 155      **Streaming Data Integration: Challenges and Opportunities**  
*Nesime Tatbul, ETH Zürich, Switzerland*
- 159      **Partitioning Real-Time ETL Workflows**  
*Alkis Simitsis, Chetan Gupta, Song Wang, Umeshwar Dayal, HP, USA*
- 163      **Midas for Government: Integration of Government Spending Data on Hadoop**  
*Antonio Sala<sup>1</sup>, Calvin Lin<sup>2</sup>, Howard Ho<sup>3</sup>*  
<sup>1</sup>University of Modena & Reggio Emilia, Italy; <sup>2</sup>University of California at Berkeley, USA;  
<sup>3</sup>IBM, USA
- 167      **Coordination of Data in Heterogenous Domains**  
*Michael Lawrence, Rachel Pottinger, Sheryl Staub-French, University of British Columbia, Canada*
- 171      **BI-Style Relation Discovery Among Entities in Text**  
*Wojciech M. Barczyński, Falk Brauer, Adrian Mocan, Marcus Schramm, Jan Froemberg, SAP, Germany*
- 175      **Profiling Linked Open Data with ProLOD**  
*Christoph Böhm, Felix Naumann, Ziawasch Abedjan, Dandy Fenz, Toni Grütze, Daniel Hefenbrock, Matthias Pohl, David Sonnabend, HPI, Germany*
- 179      **Duplicate Detection in Probabilistic Data**  
*Fabian Panse<sup>1</sup>, Maurice van Keulen<sup>2</sup>, Ander de Keijzer<sup>2</sup>, Norbert Ritter<sup>1</sup>*  
<sup>1</sup>University of Hamburg, Germany; <sup>2</sup>University of Twente, The Netherlands
- 183      **Complement Union for Data Integration**  
*Jens Bleiholder<sup>1</sup>, Sascha Szott<sup>2</sup>, Melanie Herschel<sup>3</sup>, Felix Naumann<sup>1</sup>*  
<sup>1</sup>HPI, Germany; <sup>2</sup>ZIB, Germany; <sup>3</sup>Universität Tübingen, Germany

---

# 1st International Workshop on Data Engineering Meets the Semantic Web (DESWeb'10)

---

- 187      **DESWeb'10 Preface Pages**
- 191      **Semantic Flooding: Search Over Semantic Links**  
*Fausto Giunchiglia<sup>1</sup>, Uladzimir Kharkevich<sup>1</sup>, Alethia Hume<sup>1</sup>, Piyatat Chatvorawit<sup>2</sup>*  
<sup>1</sup>University of Trento, Italy; <sup>2</sup>Asian Institute of Technology, Thailand
- 197      **An Ontology-Based Retrieval System Using Semantic Indexing**  
*Soner Kara<sup>1</sup>, Özgür Alan<sup>1</sup>, Orkunt Sabuncu<sup>1</sup>, Samet Akpınar<sup>2</sup>, Nihan K. Çiçekli<sup>2</sup>, Ferda N. Alpaslan<sup>2</sup>*  
<sup>1</sup>Orbim Corp., Turkey; <sup>2</sup>METU, Turkey
- 203      **Keyword Based Search Over Semantic Data in Polynomial Time**  
*Paolo Cappellari<sup>1</sup>, Roberto De Virgilio<sup>2</sup>, Antonio Maccioni<sup>2</sup>, Michele Miscione<sup>2</sup>*  
<sup>1</sup>University of Alberta, Canada; <sup>2</sup>Università Roma Tre, Italy
- 209      **Towards Better Entity Resolution Techniques for Web Document Collections**  
*Surender Reddy Yerva, Zoltán Miklós, Karl Aberer, EPFL, Switzerland*
- 215      **Processing Online News Streams for Large-Scale Semantic Analysis**  
*Miloš Krstajić<sup>1</sup>, Florian Mansmann<sup>1</sup>, Andreas Stoffel<sup>1</sup>, Martin Atkinson<sup>2</sup>, Daniel A. Keim<sup>1</sup>*  
<sup>1</sup>University of Konstanz, Germany; <sup>2</sup>EC Joint Research Centre, Italy
- 221      **DIVERSUM: Towards Diversified Summarisation of Entities in Knowledge Graphs**  
*Marcin Sydow<sup>1</sup>, Mariusz Piłkuła<sup>1</sup>, Ralf Schenkel<sup>2</sup>*  
<sup>1</sup>Polish-Japanese Institute of Information Technology, Poland; <sup>2</sup>Max-Planck Institute for Informatics, Germany
- 227      **The Entity Name System: Enabling the Web of Entities**  
*Heiko Stoermer, Themis Palpanas, George Giannakopoulos, University of Trento, Italy*
- 233      **Ontology Alignment Argumentation with Mutual Dependency Between Arguments and Mappings**  
*Paulo Maio, Nuno Silva, Politécnico do Porto, Portugal*
- 239      **Summarizing Ontology-Based Schemas in PDMS**  
*Carlos Eduardo Pires<sup>1</sup>, Paulo Sousa<sup>2</sup>, Zoubida Kedad<sup>3</sup>, Ana Carolina Salgado<sup>2</sup>*  
<sup>1</sup>UFCEG, Brazil; <sup>2</sup>UFPE, Brazil; <sup>3</sup>UVSQ, France
- 245      **A Framework for Automatic Schema Mapping Verification Through Reasoning**  
*Paolo Cappellari<sup>1</sup>, Denilson Barbosa<sup>1</sup>, Paolo Atzeni<sup>2</sup>*  
<sup>1</sup>University of Alberta, Canada; <sup>2</sup>Università Roma Tre, Italy
- 251      **Extensions to the Pig Data Processing Platform for Scalable RDF Data Processing Using Hadoop**  
*Yusuke Tanimura, Akiyoshi Matono, Steven Lynden, Isao Kojima, AIST, Japan*
- 257      **Optimized Data Access for Efficient Execution of Semantic Services**  
*Thorsten Möller, Heiko Schuldt, University of Basel, Switzerland*

---

## 2nd International Workshop on Modeling, Managing and Mining Evolving Social Networks (M3SN'10)

---

- 263 **M3SN'10 Preface Pages**
- 266 **Privometer: Privacy Protection in Social Networks**  
*Nilothpal Talukder, Mourad Ouzzani, Ahmed K. Elmagarmid, Hazem Elmeleegy, Mohamed Yakout, Purdue University, USA*
- 270 **On the Influence of Social Factors on Team Recommendations**  
*Michele Brocco, Georg Groh, Christian Kern, Technische Universität München, Germany*
- 278 **Towards Discovery of Eras in Social Networks**  
*Michele Berlingerio<sup>1</sup>, Michele Coscia<sup>1</sup>, Fosca Giannotti<sup>1</sup>, Anna Monreale<sup>1</sup>, Dino Pedreschi<sup>2</sup>*  
<sup>1</sup>CNR-ISTI, Italy; <sup>2</sup>University of Pisa, Italy
- 282 **Mining and Representing Recommendations in Actively Evolving Recommender Systems**  
*Ira Assent, Aalborg University, Denmark*
- 

## ICDE 2010 PhD Workshop

---

- 286 **ICDE 2010 PhD Workshop Preface Pages**
- 289 **Maximizing Visibility of Objects**  
*Muhammed Miah, University of Texas at Arlington, USA*
- 293 **Improving Product Search with Economic Theory**  
*Beibei Li, New York University, USA*
- 297 **Toward Large Scale Data-Aware Search: Ranking, Indexing, Resolution and Beyond**  
*Tao Cheng, University of Illinois at Urbana-Champaign, USA*
- 301 **Graphical Models for Dependencies and Queries in Uncertain Data**  
*Ruiwen Chen, University of Ottawa, Canada*
- 305 **Privacy-Preserving Data Publishing**  
*Ruilin Liu, Stevens Institute of Technology, USA*
- 309 **Flash-Enabled Database Storage**  
*Ioannis Koltsidas, University of Edinburgh, UK*
- 313 **A Database Server for Next-Generation Scientific Data Management**  
*Mohamed Y. Eltabakh, Purdue University, USA*
- 317 **CareDB: A Context and Preference-Aware Location-Based Database System**  
*Justin J. Levandoski, University of Minnesota, USA*
- 321 **Graph Indexing for Reachability Queries**  
*Hilmi Yildirim, Rensselaer Polytechnic Institute, USA*
- 325 **Evaluating Path Queries Over Route Collections**  
*Panagiotis Bouros, National Technical University of Athens, Greece*
- 329 **Advances in Constrained Clustering**  
*Zijie Qi, University of California at Davis, USA*
- 333 **Towards a Task-Based Search and Recommender Systems**  
*Gabriele Tolomei, Università Ca' Foscari Venezia, Italy*
- 337 **On Dynamic Data Clustering and Visualization Using Swarm Intelligence**  
*Esin Saka, University of Louisville, USA*
- 341 **Fast Algorithms for Time Series Mining**  
*Lei Li, Carnegie Mellon University, USA*

DFC[fUa '6cc\_

'5i h cf' bXYI