2014 IEEE 30th International Conference on Data Engineering Workshops

(ICDEW 2014)

Chicago, Illinois, USA
31 March – 4 April 2014
GDM 2014 — 5th International Workshop on Graph Data Management: Techniques and Applications

1  GDM 2014 Preface Pages

SESSION 1

4  BIIG: Enabling Business Intelligence with Integrated Instance Graphs
   (André Petermann, Martin Junghanns, Robert Müller, Erhard Rahm)

12  Graph Databases for Large-Scale Healthcare Systems: A Framework for Efficient Data Management and Data Services
   (Yubin Park, Mallikarjun Shankar, Byung-Hoon Park, Joydeep Ghosh)

SESSION 2

20  On Isomorphic Matching of Large Disk-Resident Graphs Using an XQuery Engine
    (Carlos R. Rivero, Hasan M. Jamil)

28  Typing Query Languages for Data Graphs
    (Dario Colazzo, Carlo Sartiani)

32  Privacy-Preserving Reachability Query Services for Sparse Graphs
    (Peipei Yi, Zhe Fan, Shuxiang Yin)

IIWeb 2014 — 10th International Workshop on Information Integration on the Web

36  IIWeb 2014 Preface Pages

SESSION 1

39  A Hashtags Dictionary from Crowdsourced Definitions
    (Mérième Ghenname, Julien Subercaze, Christophe Gravier, Frédérique Laforest, Mounia Abik, Rachida Ajhoun)

45  In Schema Matching, Even Experts are Human: Towards Expert Sourcing in Schema Matching
    (Tomer Sagi, Avigdor Gal)

50  Semantic Management of Enterprise Integration Patterns: A Use Case in Smart Grids
    (Om P. Patri, Anand V. Panangadan, Vikrambhai S. Sorathia, Viktor K. Prasanna)

SESSION 2

56  Bootstrapping Wikipedia to Answer Ambiguous Person Name Queries
    (Toni Gruetze, Gjergji Kasneci, Zhe Zuo, Felix Naumann)

62  Mapping Abstract Queries to Big Data Web Resources for On-the-Fly Data Integration and Information Retrieval
    (Hasan M. Jamil)

68  Scholarly Big Data Information Extraction and Integration in the CiteSeer\textsuperscript{X} Digital Library
    (Kyle Williams, Jian Wu, Sagnik Ray Choudhury, Madian Khabsa, C. Lee Giles)
### Session 3

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>74</td>
<td>Aggregation of Similarity Measures in Schema Matching Based on Generalized Mean</td>
<td>Faten A. Elshwimy, Alsayed Algergawy, Amany Sarhan, Elsayed A. Sallam</td>
</tr>
<tr>
<td>80</td>
<td>A Tool for Personal Data Extraction</td>
<td>Daniela Vianna, Alicia-Michelle Yong, Chaojun Xia, Amélie Marian, Thu Nguyen</td>
</tr>
<tr>
<td>84</td>
<td>Reconciling Malware Labeling Discrepancy via Consensus Learning</td>
<td>Ting Wang, Xin Hu, Shicong Meng, Reiner Sailer</td>
</tr>
</tbody>
</table>

### BDCA 2014 — Workshop on Big Data Customer Analytics

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>BDCA 2014 Preface Pages</td>
<td></td>
</tr>
</tbody>
</table>

**Session I: In-store/POS Analytics**

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>97</td>
<td>Leveraging In-Memory Technology for Interactive Analyses of Point-of-Sales Data</td>
<td>David Schwalb, Martin Faust, Jens Krueger, Hasso Plattner</td>
</tr>
<tr>
<td>103</td>
<td>Overlap versus Partition: Marketing Classification and Customer Profiling in Complex Networks of Products</td>
<td>Diego Pennacchioli, Michele Coscia, Dino Pedreschi</td>
</tr>
</tbody>
</table>

**Session II: Online Analytics**

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>Harnessing the Crowds for Multi-Channel Marketing Monitoring</td>
<td>Haggai Roitman, Gilad Barkai, David Konopnicki, Michal Shmueli-Scheuer</td>
</tr>
<tr>
<td>115</td>
<td>Characterizing Comparison Shopping Behavior: A Case Study</td>
<td>Mona Gupta, Happy Mittal, Parag Singla, Amitabha Bagchi</td>
</tr>
</tbody>
</table>
### LOPS 2014 — Workshop on Long Term Preservation for Big Scientific Data

123  LOPS 2014 Preface Pages

**Session 1**

126  A Provenance-Based Approach to Manage Long Term Preservation of Scientific Data  
*(Renato Beserra Sousa, Daniel Cintra Cugler, Joana Esther Gonzales Malaverri, Claudia Bauzer Medeiros)*

134  PDS4: A Model-Driven Planetary Science Data Architecture for Long-Term Preservation  
*(John S. Hughes, Daniel Crichton, Sean Hardman, Emily Law, Ronald Joyner, Paul Ramirez)*

142  Data and Software Preservation for Open Science (DASPOS)  
*(Michael D. Hildreth)*

**Session 2**

147  Method and Components for Creating Scientific Workflow  
*(Yuan Lin, Isabelle Mougenot, Thérèse Libourel)*

154  A Validation Framework for the Long Term Preservation of High Energy Physics Data  
*(Dmitri Ozerov, David M. South)*

### CloudDB 2014 — 6th International Workshop on Cloud Data Management

159  CloudDB 2014 Preface Pages

**Session 1: Quality of Services**

162  Towards Improvements on the Quality of Service for Multi-Tenant RDBMS in the Cloud  

170  PolarDBMS: Towards a Cost-Effective and Policy-Based Data Management in the Cloud  
*(Ilir Fetai, Filip-M. Brinkmann, Heiko Schuldt)*

178  SLA-Driven Workload Management for Cloud Databases  
*(Dimokritos Stamatakis, Olga Papaemmanouil)*

**Session 2: Query Processing**

182  Parallel Join Executions in RAMCloud  
*(Christian Tinnfeld, Donald Kossmann, Joos-Hendrik Boese, Hasso Plattner)*

191  Neighbor-Base Similarity Matching for Graphs  
*(Hang Zhang, Hongzhi Wang, Jianzhong Li, Hong Gao)*

199  Data Stream Partitioning Re-Optimization Based on Runtime Dependency Mining  
*(Emeric Viel, Haruyasu Ueda)*

**Session 3: System Architecture**

207  Curracurrong Cloud: Stream Processing in the Cloud  
*(Vasvi Kakkad, Akon Dey, Alan Fekete, Bernhard Scholz)*

215  ORESTES: A Scalable Database-as-a-Service Architecture for Low Latency  
*(Felix Gessert, Florian Bücklers, Norbert Ritter)*
SESSION 4: BENCHMARKING

223  YCSB+T: Benchmarking Web-Scale Transactional Databases
     (Akon Dey, Alan Fekete, Raghunath Namibiar, Uwe Röhm)

231  Benchmarking Cloud-Based Tagging Services
     (Tanu Malik, Kyle Chard, Ian Foster)

DESWeB 2014 — 5th International Workshop on Data Engineering Meets the Semantic Web

239  DESWeB 2014 Preface Pages

SESSION 1

242  Extending Contexts with Ontologies for Multidimensional Data Quality Assessment
     (Mostafa Milani, Leopoldo Bertossi, Sina Ariyan)

248  Assigning Global Relevance Scores to DBpedia Facts
     (Philipp Langer, Patrick Schulze, Stefan George, Matthias Kohnen, Tobias Metzke,
      Ziawasch Abedjan, Gjergji Kasneci)

254  Balloon Fusion: SPARQL Rewriting Based on Unified Co-Reference Information
     (Kai Schlegel, Florian Stegmaier, Sebastian Bayerl, Michael Granitzer, Harald Kosch)

SESSION 2

260  LODHub — A Platform for Sharing and Integrated Processing of Linked Open Data
     (Stefan Hagedorn, Kai-Uwe Sattler)

SESSION 3

263  RQ-RDF-3X: Going Beyond Triplestores
     (Jyoti Leeka, Srikanta Bedathur)

269  On Reflection in Linked Data Management
     (George H.L. Fletcher)

272  How to Generate Query Parameters in RDF Benchmarks?
     (Andrey Gubichev, Renzo Angles, Peter Boncz)
SMDB 2014 — 9th International Workshop on Self-Managing Database Systems

275  SMDB 2014 Preface Pages

Session 1: Adaptive and Personalized Data Management

278  Towards Automated Personalized Data Storage
    (John Lange, Alexandros Labrinidis, Panos K. Chrysanthis)

284  Cinderella — Adaptive Online Partitioning of Irregularly Structured Data
    (Kai Herrmann, Hannes Voigt, Wolfgang Lehner)

292  Interactive Data Exploration Based on User Relevance Feedback
    (Kyriaki Dimitriadou, Olga Papaemmanouil, Yanlei Diao)

Session 2: Autonomous Data Stream Processing

296  Auto-Scaling Techniques for Elastic Data Stream Processing
    (Thomas Heinze, Valerio Pappalardo, Zbigniew Jerzak, Christof Fetzer)

303  Execution and Optimization of Continuous Windowed Aggregation Queries
    (Harold Lim, Shivnath Babu)

310  Automated Operator Placement in Distributed Data Stream Management Systems
    Subject to User Constraints
    (Cory Thoma, Alexandros Labrinidis, Adam J. Lee)

ICDE 2014 PhD Symposium

317  ICDE 2014 PhD Symposium Preface Pages

Session 1: Data Exploration

320  Automatic User Steering for Interactive Data Exploration
    (Kyriaki Dimitriadou)

325  2B or Not 2B and Everything in Between — Novel Evaluation Methods for Matching Problems
    (Tomer Sagi)

330  Trip Similarity Computation for Context-Aware Travel Recommendation Exploiting Geotagged Photos
    (Zhenxing Xu)

Session 2: Query Processing

335  Towards Optimization of RDF Analytical Queries on MapReduce
    (Padmashree Ravindra)

340  Predictive Query Processing On Moving Objects
    (Abdeltawab M. Hendawi)

345  Aggregates Caching for Enterprise Applications
    (Stephan Müller)

Session 3: Towards Graduation

350  Analysis and Detection of Low Quality Information in Social Networks
    (De Wang)

355  User-Driven Refinement of Imprecise Queries
    (Bahar Qarabaqi)