

Table of Contents

Keynote Address I

Navigating Oceans of Data	1
<i>David Maier, V.M. Megler, António M. Baptista, Alex Jaramillo, Charles Seaton, and Paul J. Turner</i>	

Uncertain and Probabilistic Data

Probabilistic Range Monitoring of Streaming Uncertain Positions in GeoSocial Networks	20
<i>Kostas Patroumpas, Marios Papamichalis, and Timos Sellis</i>	
Probabilistic Frequent Pattern Growth for Itemset Mining in Uncertain Databases	38
<i>Thomas Bernecker, Hans-Peter Kriegel, Matthias Renz, Florian Verhein, and Andreas Züfle</i>	
Evaluating Trajectory Queries over Imprecise Location Data	56
<i>Xike Xie, Reynold Cheng, and Man Lung Yiu</i>	
Efficient Range Queries over Uncertain Strings	75
<i>Dongbo Dai, Jiang Xie, Huiran Zhang, and Jiaqi Dong</i>	
Continuous Probabilistic Sum Queries in Wireless Sensor Networks with Ranges	96
<i>Nina Hubig, Andreas Züfle, Tobias Emrich, Mario A. Nascimento, Matthias Renz, and Hans-Peter Kriegel</i>	

Parallel and Distributed Data Management

Partitioning and Multi-core Parallelization of Multi-equation Forecast Models	106
<i>Lars Dannecker, Matthias Böhm, Wolfgang Lehner, and Gregor Hackenbroich</i>	
Integrating GPU-Accelerated Sequence Alignment and SNP Detection for Genome Resequencing Analysis	124
<i>Mian Lu, Yuwei Tan, Jiuxin Zhao, Ge Bai, and Qiong Luo</i>	
Discovering Representative Skyline Points over Distributed Data	141
<i>Akrivi Vlachou, Christos Doulkeridis, and Maria Halkidi</i>	

SkyQuery: An Implementation of a Parallel Probabilistic Join Engine
for Cross-Identification of Multiple Astronomical Databases 159
*László Dobos, Tamás Budavári, Nolan Li, Alexander S. Szalay, and
István Csabai*

Efficient Filtering in Micro-blogging Systems: We Won't Get Flooded
Again 168
Ryadh Dahimene, Cedric Du Mouza, and Michel Scholl

Graph Processing

Regular Path Queries on Large Graphs 177
André Koschmieder and Ulf Leser

Sampling Connected Induced Subgraphs Uniformly at Random 195
Xuesong Lu and Stéphane Bressan

Discovery of Top-k Dense Subgraphs in Dynamic Graph Collections 213
Elena Valari, Maria Kontaki, and Apostolos N. Papadopoulos

On the Efficiency of Estimating Penetrating Rank on Large Graphs 231
Weiren Yu, Jiajin Le, Xuemin Lin, and Wenjie Zhang

Towards Efficient Join Processing over Large RDF Graph Using
MapReduce 250
Xiaofei Zhang, Lei Chen, and Min Wang

Panel

Panel on “Data Infrastructures and Data Management Research:
Close Relatives or Total Strangers?” 260
Yannis Ioannidis

Mining Multidimensional Data

Efficient Similarity Search in Very Large String Sets 262
*Dandy Fenz, Dustin Lange, Astrid Rheinländer,
Felix Naumann, and Ulf Leser*

Substructure Clustering: A Novel Mining Paradigm for Arbitrary Data
Types 280
Stephan Günnemann, Brigitte Boden, and Thomas Seidl

BT* – An Advanced Algorithm for Anytime Classification 298
Philipp Kranen, Marwan Hassani, and Thomas Seidl

Finding the Largest Empty Rectangle Containing Only a Query Point in Large Multidimensional Databases	316
<i>Gilberto Gutiérrez and José R. Paramá</i>	
Sensitivity of Self-tuning Histograms: Query Order Affecting Accuracy and Robustness	334
<i>Andranik Khachatryan, Emmanuel Müller, Christian Stier, and Klemens Böhm</i>	

Provenance and Workflows

Database Support for Exploring Scientific Workflow Provenance Graphs	343
<i>Manish Kumar Anand, Shawn Bowers, and Bertram Ludäscher</i>	
(Re)Use in Public Scientific Workflow Repositories	361
<i>Johannes Starlinger, Sarah Cohen-Boulakia, and Ulf Leser</i>	
Aggregating and Disaggregating Flexibility Objects	379
<i>Laurynas Šikšnys, Mohamed E. Khalefa, and Torben Bach Pedersen</i>	
Fine-Grained Provenance Inference for a Large Processing Chain with Non-materialized Intermediate Views	397
<i>Mohammad Rezwanaul Haq, Peter M.G. Apers, and Andreas Wombacher</i>	
Automatic Conflict Resolution in a CDSS	406
<i>Fayez Khazalah, Zaki Malik, and Brahim Medjahed</i>	

Processing Scientific Queries

Tracking Distributed Aggregates over Time-Based Sliding Windows	416
<i>Graham Cormode and Ke Yi</i>	
Hinging Hyperplane Models for Multiple Predicted Variables	431
<i>Anca Maria Ivanescu, Philipp Kranen, and Thomas Seidl</i>	
Optimizing Notifications of Subscription-Based Forecast Queries	449
<i>Ulrike Fischer, Matthias Böhm, Wolfgang Lehner, and Torben Bach Pedersen</i>	
Minimizing Index Size by Reordering Rows and Columns	467
<i>Elaheh Pourabbas, Arie Shoshani, and Kesheng Wu</i>	
Data Vaults: A Symbiosis between Database Technology and Scientific File Repositories	485
<i>Milena Ivanova, Martin Kersten, and Stefan Manegold</i>	

Keynote II

Usage Data in Web Search: Benefits and Limitations 495
Ricardo Baeza-Yates and Yoelle Maarek

Support for Demanding Applications

Functional Feature Extraction and Chemical Retrieval 507
Peng Tang, Siu Cheung Hui, and Gao Cong

Scalable Computation of Isochrones with Network Expiration 526
Johann Gamper, Michael Böhlen, and Markus Innerebner

A Dataflow Graph Transformation Language and Query Rewriting
System for RDF Ontologies 544
*Marianne Shaw, Landon T. Detwiler, James F. Brinkley, and
Dan Suciu*

Sensitive Label Privacy Protection on Social Network Data 562
Yi Song, Panagiotis Karras, Qian Xiao, and Stéphane Bressan

Trading Privacy for Information Loss in the Blink of an Eye 572
Alexandra Pilalidou and Panos Vassiliadis

Demonstration and Poster Papers

Extracting Hot Spots from Satellite Data 581
Hideyuki Kawashima, Chunyong Wang, and Hiroyuki Kitagawa

A Framework for Enabling Query Rewrites when Analyzing Workflow
Records 587
Dritan Bleco and Yannis Kotidis

Towards Enabling Outlier Detection in Large, High Dimensional Data
Warehouses 591
Konstantinos Georgoulas and Yannis Kotidis

Multiplexing Trajectories of Moving Objects 595
Kostas Patroumpas, Kyriakos Toumbas, and Timos Sellis

On Optimizing Workflows Using Query Processing Techniques 601
Georgia Koukka and Anastasios Gounaris

Optimizing Flows for Real Time Operations Management 607
*Alkis Simitsis, Chetan Gupta, Kevin Wilkinson, and
Umeshwar Dayal*

(In?)Extricable Links between Data and Visualization: Preliminary Results from the VISTAS Project	613
<i>Judith Cushing, Evan Hayduk, Jerilyn Walley, Lee Zeman, Kirsten Winters, Mike Bailey, John Bolte, Barbara Bond, Denise Lach, Christoph Thomas, Susan Stafford, and Nik Stevenson-Molnar</i>	
FireWatch: G.I.S.-Assisted Wireless Sensor Networks for Forest Fires ...	618
<i>Panayiotis G. Andreou, George Constantinou, Demetrios Zeinalipour-Yazti, and George Samaras</i>	
AIMS: A Tool for the View-Based Analysis of Streams of Flight Data	622
<i>Gereon Schüller, Roman Saul, and Andreas Behrend</i>	
TARCLOUD: A Cloud-Based Platform to Support miRNA Target Prediction	628
<i>Thanasis Vergoulis, Michail Alexakis, Theodore Dalamagas, Manolis Maragkakis, Artemis G. Hatzigeorgiou, and Timos Sellis</i>	
SALSA: A Software System for Data Management and Analytics in Field Spectrometry	634
<i>Baljeet Malhotra, John A. Gamon, and Stéphane Bressan</i>	
Incremental DNA Sequence Analysis in the Cloud	640
<i>Romeo Kienzler, Rémy Bruggmann, Anand Ranganathan, and Nesime Tatbul</i>	
AITION: A Scalable Platform for Interactive Data Mining	646
<i>Harry Dimitropoulos, Herald Kllapi, Omiros Metaxas, Nikolas Oikonomidis, Eva Sitaridi, Manolis M. Tsangaris, and Yannis Ioannidis</i>	
Author Index	653