

Table of Contents

1. Invited Presentation

The Scientific Data Management Center: Providing Technologies for Large Scale Scientific Exploration	1
<i>Arie Shoshani</i>	

2. Improving the End-User Experience

Query Recommendations for Interactive Database Exploration	3
<i>Gloria Chatzopoulou, Magdalini Eirinaki, and Neoklis Polyzotis</i>	
Scientific Mashups: Runtime-Configurable Data Product Ensembles	19
<i>Bill Howe, Harrison Green-Fishback, and David Maier</i>	
View Discovery in OLAP Databases through Statistical Combinatorial Optimization	37
<i>Cliff Joslyn, John Burke, Terence Critchlow, Nick Hengartner, and Emilie Hogan</i>	
Designing a Geo-scientific Request Language - A Database Approach ...	56
<i>Peter Baumann</i>	
SEEDEEP: A System for Exploring and Querying Scientific Deep Web Data Sources	74
<i>Fan Wang and Gagan Agrawal</i>	
Expressing OLAP Preferences	83
<i>Matteo Golfarelli and Stefano Rizzi</i>	

3. Indexing, Physical Design, and Energy

Energy Smart Management of Scientific Data	92
<i>Ekow Otoo, Doron Rotem, and Shih-Chiang Tsao</i>	
Data Parallel Bin-Based Indexing for Answering Queries on Multi-core Architectures	110
<i>Luke J. Gosink, Kesheng Wu, E. Wes Bethel, John D. Owens, and Kenneth I. Joy</i>	
Finding Regions of Interest in Large Scientific Datasets	130
<i>Rishi Rakesh Sinha, Marianne Winslett, and Kesheng Wu</i>	

Adaptive Physical Design for Curated Archives 148
*Tanu Malik, Xiaodan Wang, Debabrata Dash, Amitabh Chaudhary,
Anastasia Ailamaki, and Randal Burns*

MLR-Index: An Index Structure for Fast and Scalable Similarity Search
in High Dimensions 167
*Rahul Malik, Sangkyum Kim, Xin Jin, Chandrasekar Ramachandran,
Jiawei Han, Indranil Gupta, and Klara Nahrstedt*

4. Application Experience

B-Fabric: An Open Source Life Sciences Data Management System 185
Can Türker, Fuat Akal, Dieter Joho, and Ralph Schlapbach

Design and Implementation of Metadata System in PetaShare 191
Xinqi Wang and Tevfik Kosar

Covariant Evolutionary Event Analysis for Base Interaction Prediction
Using a Relational Database Management System for RNA 200
Weijia Xu, Stuart Ozer, and Robin R. Gutell

5. Invited Presentation

What Makes Scientific Workflows Scientific? 217
Bertram Ludäscher

6. Workflow

Enabling Ad Hoc Queries over Low-Level Scientific Data Sets 218
David Chiu and Gagan Agrawal

Exploring Scientific Workflow Provenance Using Hybrid Queries over
Nested Data and Lineage Graphs 237
*Manish Kumar Anand, Shawn Bowers, Timothy McPhillips, and
Bertram Ludäscher*

Data Integration with the DaltOn Framework – A Case Study 255
*Stefan Jablonski, Bernhard Volz, M. Abdul Rehman,
Oliver Archner, and Olivier Curé*

Experiment Line: Software Reuse in Scientific Workflows 264
*Eduardo Ogasawara, Carlos Paulino, Leonardo Murta,
Cláudia Werner, and Marta Mattoso*

Tracking Files in the Kepler Provenance Framework 273
*Pierre Mouallem, Roselyne Barreto, Scott Klasky,
Norbert Podhorszki, and Mladen Vouk*

BioBrowsing: Making the Most of the Data Available in Entrez	283
<i>Sarah Cohen-Boulakia and Kevin Masini</i>	
Using Workflow Medleys to Streamline Exploratory Tasks	292
<i>Emanuele Santos, David Koop, Huy T. Vo, Erik W. Anderson, Juliana Freire, and Cláudio Silva</i>	

7. Query Processing

Experiences on Processing Spatial Data with MapReduce	302
<i>Ariel Cary, Zhengguo Sun, Vagelis Hristidis, and Naphtali Rishe</i>	
Optimization and Execution of Complex Scientific Queries over Uncorrelated Experimental Data	320
<i>Ruslan Fomkin and Tore Risch</i>	
Comprehensive Optimization of Declarative Sensor Network Queries	339
<i>Izent Galpin, Christian Y.A. Brenninkmeijer, Farhana Jabeen, Alvaro A.A. Fernandes, and Norman W. Paton</i>	
Efficient Evaluation of Generalized Tree-Pattern Queries with Same-Path Constraints	361
<i>Xiaoying Wu, Dimitri Theodoratos, Stefanos Souldatos, Theodore Dalamagas, and Timos Sellis</i>	
Mode Aware Stream Query Processing	380
<i>Mingrui Wei and Elke Rundensteiner</i>	
Evaluating Reachability Queries over Path Collections	398
<i>Panagiotis Bourros, Spiros Skiadopoulos, Theodore Dalamagas, Dimitris Sacharidis, and Timos Sellis</i>	

8. Similarity Search

Easing the Dimensionality Curse by Stretching Metric Spaces	417
<i>Ives R.V. Pola, Agma J.M. Traina, and Caetano Traina Jr.</i>	
Probabilistic Similarity Search for Uncertain Time Series	435
<i>Johannes Aßfalg, Hans-Peter Kriegel, Peer Kröger, and Matthias Renz</i>	
Reverse k-Nearest Neighbor Search Based on Aggregate Point Access Methods	444
<i>Hans-Peter Kriegel, Peer Kröger, Matthias Renz, Andreas Züfle, and Alexander Katzdobler</i>	
Finding Structural Similarity in Time Series Data Using Bag-of-Patterns Representation	461
<i>Jessica Lin and Yuan Li</i>	

9. Keynote Address

Cloud Computing for Science	478
<i>Kate Keahey</i>	

10. Mining

Classification with Unknown Classes	479
<i>Chetan Gupta, Song Wang, Umeshwar Dayal, and Abhay Mehta</i>	
HSM: Heterogeneous Subspace Mining in High Dimensional Data	497
<i>Emmanuel Müller, Ira Assent, and Thomas Seidl</i>	
Split-Order Distance for Clustering and Classification Hierarchies	517
<i>Qi Zhang, Eric Yi Liu, Abhishek Sarkar, and Wei Wang</i>	
Combining Multiple Interrelated Streams for Incremental Clustering	535
<i>Zaigham Faraz Siddiqui and Myra Spiliopoulou</i>	
Improving Relation Extraction by Exploiting Properties of the Target Relation	553
<i>Eric Normand, Kevin Grant, Elias Ioup, and John Sample</i>	
<i>Cor-Split</i> : Defending Privacy in Data Re-publication from Historical Correlations and Compromised Tuples	562
<i>Daniele Riboni and Claudio Bettini</i>	
A Bipartite Graph Framework for Summarizing High-Dimensional Binary, Categorical and Numeric Data	580
<i>Guanhua Chen, Xiuli Ma, Dongqing Yang, Shiwei Tang, and Meng Shuai</i>	

11. Spatial Data

Region Extraction and Verification for Spatial and Spatio-temporal Databases	598
<i>Mark McKenney</i>	
Identifying the Most Endangered Objects from Spatial Datasets	608
<i>Hua Lu and Man Lung Yiu</i>	
Constraint-Based Learning of Distance Functions for Object Trajectories	627
<i>Wei Yu and Michael Gertz</i>	
Author Index	647