PODS’09
Proceedings of the Twenty-Eighth
ACM SIGMOD-SIGACT-SIGART Symposium on
Principles of Database Systems

Sponsored by:
ACM SIGMOD, ACM SIGACT, & ACM SIGART

Supported by:
Microsoft, Oracle, SAP, Sybase, Google, IBM, Netezza, Yahoo! Labs, AT&T, Hewlett-Packard,
and the University of Massachusetts–Amherst
Foreword

This volume contains the proceedings of the Twenty-eighth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2009), held in Providence, Rhode Island, on June 29–July 1, 2009, in conjunction with the 2009 ACM SIGMOD International Conference on Management of Data. The proceedings include 3 invited papers and 26 contributed papers. One invited paper is based on the keynote address by Raghu Ramakrishnan, while the other two are based on the invited tutorials by Leonid Libkin and Lars Arge, respectively. In addition, the announcement of the 2009 ACM PODS Alberto O. Mendelzon Test-of-Time Award also appears in the proceedings.

The program committee selected 26 contributed papers for presentation at the conference from 97 submissions by authors from 21 countries. Most of the selected papers are preliminary reports on work in progress. While they have been read by the program committee members, they have not been formally refereed. It is expected that many of them will eventually appear in more polished and detailed form in scientific journals. In addition, the program committee also selected the paper “Size and Treewidth Bounds for Conjunctive Queries” by Georg Gottlob, Stephanie Lee, and Gregory Valiant for the PODS 2009 Best Paper Award, and the paper “XPath Evaluation in Linear Time with Polynomial Combined Complexity” by Pawel Parys for the PODS 2009 Best Student Paper Award, a new award for the PODS conferences. Warmest congratulations to the authors of these papers.

I thank all authors who submitted papers to the symposium, and the members of the program committee for the enormous amount of work they did, especially during the review process. The program committee did not meet in person, but carried out extensive discussions during the two-and-a-half-week electronic PC meeting. I am grateful to the two PODS General Chairs, Jan Paredaens (2009) and Phokion G. Kolaitis (2008) for their tremendous help with many details related to the organization of the symposium, to Maurizio Lenzerini and Leonid Libkin, the previous PODS Program Chairs, for very useful advice and suggestions on many issues, and to Yi Chen, who served as both Publicity Chair and Proceedings Chair for PODS 2009. I also thank many colleagues involved in the organization of the conference for fruitful collaboration: Donald Kossman (SIGMOD Program Chair); Ugur Cetintemel and Stan Zdonik (SIGMOD General Co-Chairs). Finally, I thank the SIGMOD/PODS conference sponsors for their support and Andrei Voronkov for his help and support for the EasyChair system, which made it really easy to manage and coordinate the discussions during the electronic PC meeting.

Jianwen Su

PODS 2009 Program Chair
# Table of Contents

PODS 2009 Symposium Organization ........................................................................................................viii

PODS 2009 External Reviewers ......................................................................................................................ix

PODS 2009 Sponsors & Supporters ................................................................................................................x

**Session 1: Opening and Invited Talk**
Session Chair: Jan Paredaens (*University of Antwerp*)

- **A Web of Concepts** .................................................................................................................................1
  Nilesh Dalvi, Ravi Kumar, Bo Pang, Raghu Ramakrishnan, Andrew Tomkins, Philip Bohannon, Sathiya Keerthi, Srujana Merugu (*Yahoo! Research*)

**Session 2: Schema Mappings**
Session Chair: Jianwen Su (*University of California, Santa Barbara*)

- **Generalized Schema-Mappings: From Termination To Tractability** ......................................................13
  Bruno Marnette (*Oxford University*)

- **Reverse Data Exchange: Coping with Nulls** ..........................................................................................23
  Ronald Fagin (*IBM Almaden Research Center*), Phokion G. Kolaitis (*University of California, Santa Cruz & IBM Almaden Research Center*), Lucian Popa (*IBM Almaden Research Center*), Wang-Chiew Tan (*University of California, Santa Cruz*)

- **XML Schema Mappings** .........................................................................................................................33
  Shun’ichi Amano, Leonid Libkin, Filip Murlak (*University of Edinburgh*)

**Session 3: Awards**
Session Chair: Catriel Beeri (*Hebrew University*)

- **The ACM PODS Alberto O. Mendelzon Test-of-Time-Award 2009** ......................................................43
  Catriel Beeri (*Hebrew University*), Phokion G. Kolaitis (*University of California, Santa Cruz & IBM Research*), Christos H. Papadimitriou (*University of California, Berkeley*), (The Alberto O. Mendelzon Test-of-Time Award Committee for 2009)

- **Size and Treewidth Bounds for Conjunctive Queries** ..........................................................................45
  Georg Gottlob (*Oxford Man Institute of Quantitative Finance & University of Oxford*), Stephanie Tien Lee (*University of Oxford*), Gregory J. Valiant (*University of California, Berkeley*)

- **XPath Evaluation in Linear Time with Polynomial Combined Complexity** .....................................55
  Pawel Parys (*Warsaw University*)

**Session 4: Invited Tutorial 1**
Session Chair: Christoph Koch (*Cornell University*)

- **The Finite Model Theory Toolbox of a Database Theoretician** .............................................................65
  Leonid Libkin (*University of Edinburgh*)

**Session 5: Extended Models**
Session Chair: Dirk Van Gucht (*Indiana University*)

- **A General Datalog-Based Framework for Tractable Query Answering over Ontologies** ..................77
  Andrea Calì, Georg Gottlob, Thomas Lukasiewicz (*University of Oxford*)

- **Satisfiability and Relevance for Queries over Active Documents** ......................................................87
  Serge Abiteboul, Pierre Bourhis, Bogdan Marinoiu (*INRIA Saclay - Île-de-France, University Paris Sud*)
• Relative Information Completeness
  Wenfei Fan (University of Edinburgh & Bell Laboratories), Floris Geerts (University of Edinburgh)

• Relationship Privacy: Output Perturbation for Queries with Joins
  Vibhore Rastogi (University of Washington), Michael Hay, Gerome Miklau (University of Massachusetts, Amherst), Dan Suciu (University of Washington)

Session 6: Data Analysis and Optimization
Session Chair: Divesh Srivastava (AT&T Laboratories – Research)

• An Efficient Rigorous Approach for Identifying Statistically Significant Frequent Itemsets
  Adam Kirsch, Michael Mitzenmacher (Harvard University), Andrea Pietracaprina, Geppino Pucci (University of Padova), Eli Upfal (Brown University), Fabio Vandin (University of Padova)

• Similarity Caching
  Flavio Chierichetti (Sapienza University of Rome), Ravi Kumar, Sergei Vassilvitskii (Yahoo! Research)

• Indexing Uncertain Data
  Pankaj K. Agarwal (Duke University), Siu-Wing Cheng (Hong Kong University of Science and Technology), Yufei Tao (CUHK), Ke Yi (Hong Kong University of Science and Technology)

Session 7: Stream Processing
Session Chair: Yufei Tao (Chinese University of Hong Kong)

• Optimal Sampling from Sliding Windows
  Vladimir Braverman, Rafail Ostrovsky, Carlo Zaniolo (University of California, Los Angeles)

• Space-optimal Heavy Hitters with Strong Error Bounds
  Radu Berinde (Massachusetts Institute of Technology), Graham Cormode (AT&T Laboratories - Research), Piotr Indyk (Massachusetts Institute of Technology), Martin J. Strauss (University of Michigan)

• Optimal Tracking of Distributed Heavy Hitters and Quantiles
  Ke Yi, Qin Zhang (Hong Kong University of Science and Technology)

Session 8: Invited Tutorial 2
Session Chair: Subhash Suri (University of California, Santa Barbara)

• Worst-Case Efficient Range Search Indexing
  Lars Arge (Aarhus University),

Session 9: Indexing
Session Chair: Lars Arge (University of Aarhus)

• Secondary Indexing in One Dimension: Beyond B-trees and Bitmap Indexes
  Rasmus Pagh (IT University of Copenhagen), Srinivasa Rao Satti (Seoul National University)

• Dynamic Indexability and Lower Bounds for Dynamic One-Dimensional Range Query Indexes
  Ke Yi (Hong Kong University of Science and Technology)

Session 10: Query Evaluation and Optimization
Session Chair: Lucian Popa (IBM Almaden Research Center)

• Satisfiability of Downward XPath with Data Equality Tests
  Diego Figueira (INRIA Saclay, LSV)

• Equivalence of Nested Queries with Mixed Semantics
  David DeHaan (University of Waterloo)

• Equivalence of SQL Queries in Presence of Embedded Dependencies
  Rada Chirkova (North Carolina State University), Michael R. Genesereth (Stanford University)
Session 11: XML
Session Chair: Tova Milo (Tel Aviv University)

- Running Tree Automata on Probabilistic XML .................................................................227
  Sara Cohen (The Hebrew University), Benny Kimelfeld (IBM Almaden Research Center),
  Yehoshua Sagiv (The Hebrew University)

- XML with Incomplete Information: Models, Properties, and Query Answering ............237
  Pablo Barceló (University of Chile), Leonid Libkin (University of Edinburgh),
  Antonella Poggi (Sapienza University of Rome), Cristina Sirangelo (LSV, ENS-Cachan, CNRS and INRIA)

- Distributed XML Design .........................................................................................................247
  Serge Abiteboul (INRIA Saclay, Île-de-France & University Paris Sud),
  Georg Gottlob (University of Oxford), Marco Manna (University of Calabria)

Session 12: Uncertain Data
Session Chair: Andrea Cali (University of Oxford)

- Consensus Answers for Queries over Probabilistic Databases ........................................259
  Jian Li, Amol Deshpande (University of Maryland)

- Exceeding Expectations and Clustering Uncertain Data ................................................269
  Sudipto Guha (University of Pennsylvania), Kamesh Munagala (Duke University)

- Computing All Skyline Probabilities for Uncertain Data ..................................................279
  Mikhail J. Atallah, Yinian Qi (Purdue University)

Author Index ...................................................................................................................................288
PODS 2009 Symposium Organization

**Executive Committee:** Jan Paredaens *(Chair, University of Antwerp, Belgium)*  
Georg Gottlob *(University of Oxford, UK)*  
Phokion Kolaitis *(Univ. of California, Santa Cruz & IBM Research, USA)*  
Maurizio Lenzerini *(Sapienza University of Rome, Italy)*  
Leonid Libkin *(University of Edinburgh, UK)*  
Jianwen Su *(University of California, Santa Barbara, USA)*

**General Chair:** Jan Paredaens *(University of Antwerp, Belgium)*

**Program Chair:** Jianwen Su *(University of California, Santa Barbara, USA)*

**Program Committee:** Gustavo Alonso *(ETH Zürich, Switzerland)*  
Pablo Barceló *(University of Chile, Chile)*  
Toon Calders *(Eindhoven University of Technology, The Netherlands)*  
Andrea Cali *(University of Oxford, UK)*  
Anirban Dasgupta *(Yahoo! Research, USA)*  
Giuseppe De Giacomo *(Sapienza University of Rome, Italy)*  
Wenfei Fan *(University of Edinburgh & Bell Labs, UK & USA)*  
Floris Geerts *(University of Edinburgh, UK)*  
Michael Kifer *(SUNY Stony Brook, USA)*  
Wim Martens *(Dortmund University of Technology, Germany)*  
Frank McSherry *(Microsoft Research, USA)*  
Nina Mishra *(Microsoft Research & University of Virginia, USA)*  
Sunil Prabhakar *(Purdue University, USA)*  
Nicole Schweikardt *(Frankfurt University, Germany)*  
Luc Segoufin *(INRIA, France)*  
Jianwen Su *(University of California, Santa Barbara, USA)*  
V.S. Subrahmanian *(University of Maryland, USA)*  
Subhash Suri *(University of California, Santa Barbara, USA)*  
Wang-Chiew Tan *(University of California, Santa Cruz, USA)*  
Balder ten Cate *(University of Amsterdam, The Netherlands)*  
Dirk Van Gucht *(Indiana University, USA)*  
Victor Vianu *(University of California, San Diego, USA)*

**Proceedings & Publicity Chair:** Yi Chen *(Arizona State University, USA)*
PODS 2009 External Reviewers

Isolde Adler
Dmitri Akatov
Marcelo Arenas
Albert Atserias
Brian Babcock
Michael Benedikt
Henrik Björklund
Jan Van den Bussche
Benjamin Bustos
Diego Calvanese
Timothy M. Chan
Amitabh Chaudhary
Sara Cohen
Edith Cohen
Graham Cormode
Alin Deutsch
Alin Dobra
Guozhu Dong
Vuk Ercegovac
Ronald Fagin
Alberto Finzi
George Fletcher
Paul Fodor
Wouter Gelade
Cagdas Gerede
Bart Goethals
Sreenivas Gollapudi
Georg Gottlob
Dimitrios Gunopulos
André Hernich
Wenchi Hou
Richard Hull
Giovambattista Ianni
T.S. Jayram
Daniel Kifer
Benny Kimelfeld
Phokion Kolaitis
Aleksandra Korolova
Ravi Kumar
Domenico Lembo
Maurizio Lenzerini
Stefano Leonardi
Senlin Liang
Leonid Libkin
Carsten Lutz
Shuai Ma
Bruno Marnette
Marcel Marquardt
Andrew McGregor
Filip Murlak
S Muthukrishnan
Veli Mäkinen
Shubha Nabar
Frank Neven
Mathias Niepert
Lorenzo Orecchia
Rasmus Pagh
Jorge Perez
Reinhard Pichler
Andreas Pieris
Antonella Poggi
Lucian Popa
Kirk Pruhs
Edward Robertson
Riccardo Rosati
Andras Salamon
Tamas Sarlos
Matthias Schubert
Zhongnan Shen
Dan Suciu
Dirk Sudholt
Val Tannen
Nikolaj Tatti
Tuncay Tekle
Jayram Thathachar
Panayiotis Tsaparas
Pascal Urso
Stijn Vansummeren
Moshe Vardi
Hui Wan
Yuqing Wu
PODS 2009 Sponsors & Supporters

Society Sponsors:

Principle Supporters:

Supporters:

Contributors:

Web Hosting:
Author Index

Abiteboul, Serge .................. 87, 247
Agarwal, Pankaj K.............. 137
Amano, Shun’ichi .................. 33
Arge, Lars ..................... 175
Atallah, Mikhail J. .............. 279
Barceló, Pablo ................. 237
Beeri, Catriel ................. 43
Bourhis, Pierre ............. 87
Braverman, Vladimir .......... 147
Cali, Andrea ................. 77
Cheng, Siu-Wing .............. 137
Chierichetti, Flavio .......... 127
Chirkova, Rada .............. 217
Cohen, Sara ................. 227
Cormode, Graham .......... 157
Dalvi, Nilesh ................... 1
Dalvi, David ................. 207
Deshpande, Amol .......... 259
Fagin, Ronald .............. 23
Fan, Wenfei ................. 97
Figueira, Diego ............. 197
Geerts, Floris .............. 97
Genesereth, Michael R...... 217
Gottlob, Georg ............ 45, 77, 247
Guha, Sudipto ................. 269
Hay, Michael ............... 107
Indyk, Piotr .................. 157
Keerthi, Sathiya ............... 1
Kimelfeld, Benny ............ 227
Kirsch, Adam ................. 117
Kolaitis, Phokion G .......... 23, 43
Kumar, Ravi ................. 1, 127
Lee, Stephanie Tien .......... 45
Li, Jian ....................... 259
Libkin, Leonid ................. 33, 65, 237
Lukaszewicz, Thomas ..... 77
Manna, Marco ................. 247
Marinoiu, Bogdan .......... 87
Marnette, Bruno ........... 13
Merugu, Srujana ............. 279
Miklau, Gerome .............. 107
Mitzenmacher, Michael .... 117
Munagala, Kamesh .......... 269
Murlak, Filip ................. 33
Ostrovy, Rafail .............. 147
Pagh, Rasmus .................. 177
Pang, Bo .................... 1
Papadimitriou, Christos H ..... 43
Parys, Pawel .................. 55
Pietracaprina, Andrea .... 117
Poppi, Antonella .......... 237
Popa, Lucian ............ 23
Pucci, Geppino ............. 117
Qi, Yinian .................. 279
Ramakrishnan, Raghu ....... 1
Rastogi, Vibhor .......... 107
Sagiv, Yehoshua .......... 227
Satti, Sriravisa Rao .... 177
Sirangelo, Cristina .......... 237
Strauss, Martin J........... 157
Suciu, Dan .................. 107
Tan, Wang-Chiew .......... 23
Tao, Yufei .................. 137
Tomkins, Andrew .......... 1
Upfal, Eli .................. 117
Valiant, Gregory J........ 45
Vandin, Fabio ............ 117
Vassilvitskii, Sergei .... 127
Yu, Ke ..................... 137, 167, 187
Zaniolo, Carlo ............ 147
Zhang, Qin .................. 167