

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Bertram Ludäscher Nikos Mamoulis (Eds.)

Scientific and Statistical Database Management

20th International Conference, SSDBM 2008
Hong Kong, China, July 9-11, 2008
Proceedings

Volume Editors

Bertram Ludäscher
University of California
One Shields Avenue, Davis, CA 95616, USA
E-mail: ludaesch@ucdavis.edu

Nikos Mamoulis
University of Hong Kong
Pokfulam Road, Hong Kong
E-mail: nikos@cs.hku.hk

Library of Congress Control Number: 2008929534

CR Subject Classification (1998): H.2.8, H, G.3, I.3.5, E.1-2, J.1-3

LNCS Sublibrary: SL 3 – Information Systems and Application,
incl. Internet/Web and HCI

ISSN	0302-9743
ISBN-10	3-540-69476-5 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-69476-2 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com

© Springer-Verlag Berlin Heidelberg 2008
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12323043 06/3180 5 4 3 2 1 0

Preface

The International Conference on Scientific and Statistical Database Management (SSDBM) is an established forum for the exchange of the latest research results on concepts, tools, and techniques for scientific and statistical database applications. The 2008 meeting marked the 20th time that scientific domain experts, databases researchers, practitioners, and developers came together to share their new insights and to discuss in a stimulating environment future research directions.

This volume contains the proceedings of the 20th SSDBM Conference, held in Hong Kong, China, July 9–11, 2008. The conference included 3 keynote talks, 28 long and 7 short papers in 9 sessions, and 8 posters and demonstrations in a single session.

Distinguished members of the community delivered the three keynotes, which were about the past, present, and future management of scientific and statistical data. Alex Szalay, an expert in large-scale scientific data management, discussed “New Challenges in Petascale Scientific Databases,” managing huge scientific data repositories, with a focus on particular examples taken from astronomy. Nick Koudas, a leader in semi-structured text management, talked about “Adventures in the Blogosphere,” a huge network of textual data (including blogs, social networks, wikis), and BlogScope, a system that collects and analyzes such data. Finally, Per Svensson, a pioneer in database systems development for scientific applications, provided a historical review on “The Evolution of Vertical Database Architectures” from the perspective and performance needs of a scientific or statistical large-scale data analyst user.

The Program Committee, consisting of 37 members, accepted 43 papers (28 long, 7 short, and 8 posters/demos) from a total of 84 submissions. The reviewing process was managed by the EasyChair Conference System, an excellent free conference management system, developed by Andrei Voronkov.

The program and activities of SSDBM 2008 were the result of a large effort by the authors, reviewers, presenters, and organizers. We thank them all for helping to make this conference a success. In particular, we would like to thank Max Egenhofer for his great help in the early stages of the organization. We are grateful to the Department of Computer Science of Hong Kong University, especially to Maria Lam and the student helpers for their great help with the local organization. We believe that SSDBM 2008 continued the successful tradition of the series, providing an interesting program and lively discussions in a pleasant environment.

Davis, California
Hong Kong, China
April 2008

Bertram Ludäscher
Nikos Mamoulis

Conference Organization

General and Program Chairs

Nikos Mamoulis (General Chair)
Bertram Ludäscher (PC Co-chair)
Max Egenhofer (PC Co-chair)

Program Committee

Ken Barker	Mohamed Mokbel
Elisa Bertino	Kyriakos Mouratidis
Claudio Bettini	Mario Nascimento
Shawn Bowers	Silvia Nittel
Gilberto Camara	Moirra Norrie
Lei Chen	Beng Chin Ooi
Reynold Cheng	Dimitris Papadias
Ian Davidson	John Pfaltz
Jim Frew	Philippe Rigaux
Michael Gertz	Doron Rotem
Carole Goble	Bernhard Seeger
Moustafa Hammad	Kurt Stockinger
Panos Kalnis	Yufei Tao
Larry Kerschberg	Stijn Vansummeren
Martin Kersten	Marianne Winslett
George Kollios	Man Lung Yiu
Feifei Li	Jeffrey Xu Yu
Xiaosong Ma	Xiaofang Zhou
Francesco Malvestuto	

External Reviewers

Spiridon Bakiras	Maria Damiani
Ilaria Bartolini	Ke Deng
Jeffrey Bergamini	Alexandre de Spindler
Angelo Brayner	Cédric du Mouza
Alexander Brodsky	Fatima Farag
Lijun Chang	Conny Franke
Jinchuan Chen	Dario Freni
Zhihong Chong	Gabriel Fung
Chi-Yin Chow	Antoon Goderis

VIII Organization

Hoyoung Jeung
Mohamed Khalefa
Jay Kola
Ashish Kundu
Justin Levandoski
Jiangtian Li
Xiang Lian
Erietta Liarou
Jessica Lin
Dan Lin
Heshan Lin
Dekang Lin
Yimin Lin
Aretusa Lopes
Saravanan Muthaiyah
Francesco M. Malvestuto
Alex Markowetz
Sergio Mascetti
Rimma Nehme
Niels Nes
Moirra Norrie
Stavros Papadopoulos

Linda Pareschi
Michalis Potamias
Belen Prados Suarez
Daniele Riboni
Carols Rueda
Lefteris Sidiourgos
Shaoxu Song
Kurt Stockinger
Arash Termehchy
Sonny Vaupel
Quang Hieu Vu
Changliang Wang
Leanne Wu
Xike Xie
Jongpil Yoon
Ting Yu
Cammie Zhuang
Charles Zhang
Jianting Zhang
Yang Zhang
Zhe Zhang
Lei Zou

Table of Contents

Keynote Talks

New Challenges in Petascale Scientific Databases.....	1
<i>Alexander Szalay</i>	
Adventures in the Blogosphere.....	2
<i>Nick Koudas</i>	
The Evolution of Vertical Database Architectures – A Historical Review	3
<i>Per Svensson</i>	

Query Optimization in Scientific Databases

Linked Bernoulli Synopses: Sampling along Foreign Keys.....	6
<i>Rainer Gemulla, Philipp Rösch, and Wolfgang Lehner</i>	
Query Planning for Searching Inter-dependent Deep-Web Databases	24
<i>Fan Wang, Gagan Agrawal, and Ruoming Jin</i>	
Summarizing Two-Dimensional Data with Skyline-Based Statistical Descriptors.....	42
<i>Graham Cormode, Flip Korn, S. Muthukrishnan, and Divesh Srivastava</i>	
Query Selectivity Estimation for Uncertain Data	61
<i>Sarveet Singh, Chris Mayfield, Rahul Shah, Sunil Prabhakar, and Susanne Hambrusch</i>	

Privacy

Disclosure Risks of Distance Preserving Data Transformations	79
<i>E. Onur Turgay, Thomas B. Pedersen, Yücel Saygın, Erkay Savaş, and Albert Levi</i>	
Privacy-Preserving Publication of User Locations in the Proximity of Sensitive Sites	95
<i>Bharath Krishnamachari, Gabriel Ghinita, and Panos Kalnis</i>	
A Probabilistic Framework for Building Privacy-Preserving Synopses of Multi-dimensional Data	114
<i>Filippo Furfaro, Giuseppe M. Mazzeo, and Domenico Saccà</i>	

Searching and Mining Graphs

Efficient Similarity Search for Tree-Structured Data	131
<i>Guoliang Li, Xuhui Liu, Jianhua Feng, and Lizhu Zhou</i>	
Hierarchical Graph Embedding for Efficient Query Processing in Very Large Traffic Networks	150
<i>Hans-Peter Kriegel, Peer Kröger, Matthias Renz, and Tim Schmidt</i>	
Monitoring Aggregate k-NN Objects in Road Networks	168
<i>Lu Qin, Jeffrey Xu Yu, Bolin Ding, and Yoshiharu Ishikawa</i>	
RAM: Randomized Approximate Graph Mining	187
<i>Shijie Zhang and Jiong Yang</i>	

Data Streams

Finding Frequent Items over General Update Streams	204
<i>Sumit Ganguly, Abhayendra N. Singh, and Satyam Shankar</i>	
Efficiently Discovering Recent Frequent Items in Data Streams	222
<i>Ferry Irawan Tanton, Nishad Manerikar, and Themis Palpanas</i>	
Prioritized Evaluation of Continuous Moving Queries over Streaming Locations	240
<i>Kostas Patroumpas and Timos Sellis</i>	

Scientific Database Applications

A Comparative Evaluation of XML Difference Algorithms with Genomic Data	258
<i>Cornelia Hedeler and Norman W. Paton</i>	
Adaptive Request Scheduling for Parallel Scientific Web Services	276
<i>Heshan Lin, Xiaosong Ma, Jiangtian Li, Ting Yu, and Nagiza Samatova</i>	
ViP: A User-Centric View-Based Annotation Framework for Scientific Data	295
<i>Qinglan Li, Alexandros Labrinidis, and Panos K. Chrysanthis</i>	
Ontology Database: A New Method for Semantic Modeling and an Application to Brainwave Data	313
<i>Paea LePend, Dejing Dou, Gwen A. Frishkoff, and Jiawei Rong</i>	

Advanced Indexing Methods

The hB-pi* Tree: An Optimized Comprehensive Access Method for Frequent-Update Multi-dimensional Point Data	331
<i>Panfeng Zhou and Betty Salzberg</i>	
Breaking the Curse of Cardinality on Bitmap Indexes	348
<i>Kesheng Wu, Kurt Stockinger, and Arie Shoshani</i>	
A New Approach for Optimization of Dynamic Metric Access Methods Using an Algorithm of Effective Deletion	366
<i>Renato Bueno, Daniel dos Santos Kaster, Agma Juci Machado Traina, and Caetano Traina Jr.</i>	
An Ontology-Based Index to Retrieve Documents with Geographic Information	384
<i>Miguel R. Luaces, José R. Paramá, Oscar Pedreira, and Diego Seco</i>	

Data Mining

Mining Temporal Association Patterns under a Similarity Constraint . . .	401
<i>Jin Soung Yoo and Shashi Shekhar</i>	
A General Framework for Increasing the Robustness of PCA-Based Correlation Clustering Algorithms	418
<i>Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek</i>	
Searching Correlated Objects in a Long Sequence	436
<i>Ken C.K. Lee, Wang-Chien Lee, Donna Peuquet, and Baihua Zheng</i>	

Advanced Queries and Uncertain Data

Caching Dynamic Skyline Queries	455
<i>Dimitris Sacharidis, Panagiotis Bouros, and Timos Sellis</i>	
Plot Query Processing with Wavelets	473
<i>Mehrdad Jahangiri and Cyrus Shahabi</i>	
Quality-Aware Probing of Uncertain Data with Resource Constraints . . .	491
<i>Jinchuan Chen and Reynold Cheng</i>	

Short Presentations

Efficient Computation of Statistical Significance of Query Results in Databases	509
<i>Vishwakarma Singh, Arnab Bhattacharya, and Ambuj K. Singh</i>	

Analysis of Basic Data Reordering Techniques	517
<i>Tan Apaydin, Ali Şaman Tosun, and Hakan Ferhatosmanoglu</i>	
Kriging for Localized Spatial Interpolation in Sensor Networks	525
<i>Muhammad Umer, Lars Kulik, and Egemen Tanin</i>	
Scalable Ubiquitous Data Access in Clustered Sensor Networks	533
<i>Yueh-Hua Lee, Alex Thomo, Kui Wu, and Valerie King</i>	
iJoin: Importance-Aware Join Approximation over Data Streams	541
<i>Dhananjay Kulkarni and Chinya V. Ravishankar</i>	
Efficient Continuous K -Nearest Neighbor Query Processing over Moving Objects with Uncertain Speed and Direction	549
<i>Yuan-Ko Huang, Shi-Jei Liao, and Chiang Lee</i>	
ProUD: Probabilistic Ranking in Uncertain Databases	558
<i>Thomas Bernecker, Hans-Peter Kriegel, and Matthias Renz</i>	

Poster and Demonstration Papers

Flexible Scientific Workflow Modeling Using Frames, Templates, and Dynamic Embedding	566
<i>Anne H.H. Ngu, Shawn Bowers, Nicholas Haasch, Timothy McPhillips, and Terence Critchlow</i>	
Examining Statistics of Workflow Evolution Provenance: A First Study	573
<i>Lauro Lins, David Koop, Erik W. Anderson, Steven P. Callahan, Emanuele Santos, Carlos E. Scheidegger, Juliana Freire, and Cláudio T. Silva</i>	
ELKI: A Software System for Evaluation of Subspace Clustering Algorithms	580
<i>Elke Achtert, Hans-Peter Kriegel, and Arthur Zimek</i>	
IVIP – A Scientific Workflow System to Support Experts in Spatial Planning of Crop Production	586
<i>Christopher J. Tuot, Michael Sintek, and Andreas R. Dengel</i>	
A FUSE-Based Tool for Accessing Meteorological Data in Remote Servers	592
<i>Keiichirou Ui, Toshiyuki Amagasa, and Hiroyuki Kitagawa</i>	
<i>IndeGS^{RI}</i> : Efficient View-Dependent Ranking in CFD Post-processing Queries with RDBMS	598
<i>Christoph Brochhaus and Thomas Seidl</i>	

Real-Time Integration of Geospatial Raster and Point Data Streams	605
<i>Carlos Rueda and Michael Gertz</i>	
IRMA: An Image Registration Meta-algorithm: Evaluating Alternative Algorithms with Multiple Metrics	612
<i>Kelvin T. Leung, D. Stott Parker, Alexandre Cunha, Cornelius Hojatkashani, Ivo Dinov, and Arthur W. Toga</i>	
Author Index	619