Lecture Notes in Computer Science 3896

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

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
New York University, NY, USA

Doug Tygar
University of California, Berkeley, CA, USA

Moshe Y. Vardi
Rice University, Houston, TX, USA

Gerhard Weikum
Max-Planck Institute of Computer Science, Saarbruecken, Germany

Yannis Ioannidis Marc H. Scholl
Joachim W. Schmidt Florian Matthes
Mike Hatzopoulos Klemens Boehm
Alfons Kemper Torsten Grust
Christian Boehm (Eds.)

Advances in Database Technology – EDBT 2006

10th International Conference on Extending Database Technology
Munich, Germany, March 26-31, 2006
Proceedings

Volume Editors

Yannis Ioannidis
E-mail: yannis@di.uoa.gr
Marc H. Scholl
E-mail: Marc.Scholl@uni-konstanz.de
Joachim W. Schmidt
E-mail: j.w.schmidt@tu-harburg.de
Florian Matthes
E-mail: matthes@in.tum.de
Mike Hatzopoulos
E-mail: mike@di.uoa.gr
Klemens Boehm
E-mail: boehm@ipd.uka.de
Alfons Kemper
E-mail: Alfons.Kemper@in.tum.de
Torsten Grust
E-mail: grust@in.tum.de
Christian Boehm
E-mail: boehm@dbs.ifi.lmu.de

Library of Congress Control Number: 2006922026

CR Subject Classification (1998): H.2, H.4, H.3, C.2.4, K.4.4

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

ISSN 0302-9743
ISBN-10 3-540-32960-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-32960-2 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 2006
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11687238 06/3142 5 4 3 2 1 0

Preface

The series of International Conferences on Extending Database Technology (EDBT) is an established and prestigious forum for the exchange of the latest research results in data management. It provides unique opportunities for database researchers, practitioners, developers, and users to explore new ideas, techniques, and tools, and to exchange experiences. This volume contains the proceedings of the 10th EDBT Conference, held in Munich, Germany, March 27-29, 2006. The conference included 3 keynote talks, 56 full-size and 4 half-size research papers in 20 sessions, 8 industrial presentations in 3 sessions, 1 panel session, 5 tutorials in 7 sessions, and 20 demonstrations in 4 sessions. All of the research papers as well as papers and abstracts from most of the other sessions are included here.

Distinguished members of the database and information-retrieval communities delivered the three keynotes, which were all in the spirit of the banner theme chosen for EDBT 2006: "From Database Systems to Universal Data Management." Martin Kersten, a pioneer in the area of database support for ambient application environments and the investigator of several kernel database architectures, discussed various hard issues that arise in organic database systems, i.e., systems that can be embedded in several hardware applications and have autonomic behavior. Alan Smeaton, a leader in content-based retrieval of information in a wide variety of media, introduced us to the world of digital video libraries and challenged us with several open problems associated with their effective management. Finally, David Maier, whose introduction of object-oriented concepts to the field has been pivotal for the establishment of object-relational databases as the current state of the art, used the Lewis and Clark expedition as an analogy to discuss the problems faced when trying to deploy dataspace systems, i.e., systems that manage enterprise data in the entire spectrum from fully structured to completely unstructured.

The Research Program Committee consisted of 76 members and was chaired by Yannis Ioannidis (University of Athens). It accepted 60 papers (56 regular-size and 4 half-size) out of 352 submissions, both of which are the largest numbers ever for EDBT. Papers were submitted from 37 countries. The reviewing process was managed by the Microsoft Conference Management Toolkit, developed and supported by Surajit Chaudhuri and the CMT team (from Microsoft Research).

The Industrial and Applications Program was assembled by a small committee under Alfons Kemper (Technical University of Munich). The 8 short papers that appear in the proceedings were selected from 28 submissions.

One panel session was solicited by Klemens Böhm (University of Karlsruhe), who also selected five tutorials out of ten submissions.

Christian Böhm (University of Munich) led a 14-member committee in putting together the Demonstrations Program, consisting of 20 demonstrations selected from 52 submitted proposals.

Mike Hatzopoulos (University of Athens) edited the proceedings. He also produced an electronic version for inclusion in the SIGMOD digital library and for posting on the Web prior to the conference.

The program and social activities of EDBT 2006 are the result of a huge effort by many hundreds of authors, reviewers, presenters, and organizers. We thank them all for helping to make the conference a success. In particular, we want to thank Gisela

Krügel and Josef Lankes (both from the Technical University of Munich) for the smooth local organization.

Finally, this 10th edition of the EDBT Conference series coincides with the year of retirement of Joachim Schmidt, who was one of the "fathers" of EDBT. The success of the EDBT Association, its conferences, and services to the community is largely due to the effort and expertise that its founders invested. As a tribute to Joachim's commitment and dedication to EDBT, the organizers invited Joachim as an Honorary Chair of this year's conference. Thank you Joachim for all these years!

Yannis Ioannidis	Program Chair
Marc H. Scholl	General Chair
Florian Matthes	Executive Chair
Michael Hatzopoulos	Proceedings Chair
Klemens Böhm	Panel and Tutorial Chair
Alfons Kemper	Industrial and Applications Chair
Torsten Grust	Workshop Chair
Christian Böhm	Demonstrations Chair

Table of Contents

Invited Lectures

Database Architecture Fertilizers: Just-in-Time, Just-Enough, and Autonomous Growth
Martin Kersten 1

Digital Video: Just Another Data Stream?
Alan F. Smeaton 2

Charting a Dataspace: Lessons from Lewis and Clark
David Maier 3

Data Streams

Fast Approximate Wavelet Tracking on Streams
Graham Cormode, Minos Garofalakis, Dimitris Sacharidis 4

Resource Adaptive Periodicity Estimation of Streaming Data
Michail Vlachos, Deepak S. Turaga, Philip S. Yu 23

On Futuristic Query Processing in Data Streams
Charu C. Aggarwal 41

Semantic Heterogeneity

Detecting Similarities in Ontologies with the SOQA-SimPack Toolkit
Patrick Ziegler, Christoph Kiefer, Christoph Sturm, Klaus R. Dittrich, Abraham Bernstein 59

Holistic Schema Matching for Web Query Interfaces
Weifeng Su, Jiying Wang, Frederick Lochovsky 77

Data Mapping as Search
George H.L. Fletcher, Catharine M. Wyss 95

Distributed Databases

Parallelizing Skyline Queries for Scalable Distribution
Ping Wu, Caijie Zhang, Ying Feng, Ben Y. Zhao, Divyakant Agrawal, Amr El Abbadi 112

Replication, Load Balancing and Efficient Range Query Processing in DHTs
Theoni Pitoura, Nikos Ntarmos, Peter Triantafillou 131

IQN Routing: Integrating Quality and Novelty in P2P Querying and Ranking
Sebastian Michel, Matthias Bender, Peter Triantafillou, Gerhard Weikum 149

Multidimensionality and Nearest-Neighbor Searches

Efficient Quantile Retrieval on Multi-dimensional Data
Man Lung Yiu, Nikos Mamoulis, Yufei Tao 167

Fast Nearest Neighbor Search on Road Networks
Haibo Hu, Dik Lun Lee, Jianliang Xu 186

Approximation Techniques to Enable Dimensionality Reduction for Voronoi-Based Nearest Neighbor Search
Christoph Brochhaus, Marc Wichterich, Thomas Seidl 204

Privacy and Security

Authorization-Transparent Access Control for XML Under the Non-Truman Model
Yaron Kanza, Alberto O. Mendelzon, Renée J. Miller, Zheng Zhang 222

On Honesty in Sovereign Information Sharing
Rakesh Agrawal, Evimaria Terzi 240

Temporal Data Management

Multi-dimensional Aggregation for Temporal Data
Michael Böhlen, Johann Gamper, Christian S. Jensen 257

Similarity Search on Time Series Based on Threshold Queries
Johannes Aßfalg, Hans-Peter Kriegel, Peer Kröger, Peter Kunath, Alexey Pryakhin, Matthias Renz 276

Supporting Temporal Slicing in XML Databases
Federica Mandreoli, Riccardo Martoglia, Enrico Ronchetti 295

Text Databases and Information Retrieval

Indexing Shared Content in Information Retrieval Systems
Andrei Z. Broder, Nadav Eiron, Marcus Fontoura, Michael Herscovici, Ronny Lempel, John McPherson, Runping Qi, Eugene Shekita 313

Feedback-Driven Structural Query Expansion for Ranked Retrieval of XML Data
Ralf Schenkel, Martin Theobald 331

Expressiveness and Performance of Full-Text Search Languages
Chavdar Botev, Sihem Amer-Yahia, Jayavel Shanmugasundaram 349

Schema Management

Model-Independent Schema and Data Translation
Paolo Atzeni, Paolo Cappellari, Philip A. Bernstein 368

Physical Design Refinement: The "Merge-Reduce" Approach
Nicolas Bruno, Surajit Chaudhuri 386

Online, Non-blocking Relational Schema Changes
Jørgen Løland, Svein-Olaf Hvasshovd 405

Approximation and Estimation

Deferred Maintenance of Disk-Based Random Samples
Rainer Gemulla, Wolfgang Lehner 423

Exploiting Cluster Analysis for Constructing Multi-dimensional Histograms on Both Static and Evolving Data
Filippo Furfaro, Giuseppe M. Mazzeo, Cristina Sirangelo 442

HASE: A Hybrid Approach to Selectivity Estimation for Conjunctive Predicates
Xiaohui Yu, Nick Koudas, Calisto Zuzarte 460

Data and Query Patterns

On High Dimensional Skylines
Chee-Yong Chan, H.V. Jagadish, Kian-Lee Tan, Anthony K.H. Tung, Zhenjie Zhang 478

From Analysis to Interactive Exploration: Building Visual Hierarchies from OLAP Cubes
Svetlana Vinnik, Florian Mansmann 496

DPTree: A Distributed Pattern Tree Index for Partial-Match Queries in Peer-to-Peer Networks
Dyce Jing Zhao, Dik Lun Lee, Qiong Luo 515

XML Queries and Updates

A Decomposition-Based Probabilistic Framework for Estimating the Selectivity of XML Twig Queries
Chao Wang, Srinivasan Parthasarathy, Ruoming Jin 533

Conflicting XML Updates
Mukund Raghavachari, Oded Shmueli 552

Improving the Efficiency of XPath Execution on Relational Systems
Haris Georgiadis, Vasilis Vassalos 570

Data Streams and Pub/Sub Systems

Bridging Physical and Virtual Worlds: Complex Event Processing for RFID Data Streams
Fusheng Wang, Shaorong Liu, Peiya Liu, Yijian Bai 588

On Concurrency Control in Sliding Window Queries over Data Streams
Lukasz Golab, Kumar Gaurav Bijay, M. Tamer Özsu 608

Towards Expressive Publish/Subscribe Systems
Alan Demers, Johannes Gehrke, Mingsheng Hong, Mirek Riedewald, Walker White 627

Data Mining and Knowledge Discovery

Finding Data Broadness Via Generalized Nearest Neighbors
Jayendra Venkateswaran, Tamer Kahveci, Orhan Camoglu 645

TrajPattern: Mining Sequential Patterns from Imprecise Trajectories of Mobile Objects
Jiong Yang, Meng Hu 664

On Exploring the Power-Law Relationship in the Itemset Support Distribution
Kun-Ta Chuang, Jiun-Long Huang, Ming-Syan Chen 682

Images, Multimedia, and User Interfaces

Fast Query Point Movement Techniques with Relevance Feedback for Content-Based Image Retrieval
Danzhou Liu, Kien A. Hua, Khanh Vu, Ning Yu 700

On Fast Non-metric Similarity Search by Metric Access Methods
Tomáš Skopal 718

Constructing a Generic Natural Language Interface for an XML Database
Yunyao Li, Huahai Yang, H.V. Jagadish 737

XML Data Management

A New Design for a Native XML Storage and Indexing Manager
Jihad Boulos, Shant Karakashian 755

XML Duplicate Detection Using Sorted Neighborhoods
Sven Puhlmann, Melanie Weis, Felix Naumann 773

Handling Interlinked XML Instances on the Web
Erik Behrends, Oliver Fritzen, Wolfgang May 792

Query Optimization

Query Planning in the Presence of Overlapping Sources
Jens Bleiholder, Samir Khuller, Felix Naumann, Louiqa Raschid, Yao Wu 811

Optimizing Monitoring Queries over Distributed Data
Frank Neven, Dieter Van de Craen 829

Progressive Query Optimization for Federated Queries
Stephan Ewen, Holger Kache, Volker Markl, Vijayshankar Raman 847

Data Structures and Indexing

Indexing Spatially Sensitive Distance Measures Using Multi-resolution Lower Bounds
Vebjorn Ljosa, Arnab Bhattacharya, Ambuj K. Singh 865

Indexing Incomplete Databases
Guadalupe Canahuate, Michael Gibas, Hakan Ferhatosmanoglu 884

FlexInd: A Flexible and Parameterizable Air-Indexing Scheme for Data Broadcast Systems
André Seifert, Jen-Jou Hung 902

Nontraditional Query Processing

Multi-query SQL Progress Indicators
Gang Luo, Jeffrey F. Naughton, Philip S. Yu 921

Finding Equivalent Rewritings in the Presence of Arithmetic Comparisons
Foto Afrati, Rada Chirkova, Manolis Gergatsoulis, Vassia Pavlaki 942

Fast Computation of Reachability Labeling for Large Graphs
Jiefeng Cheng, Jeffrey Xu Yu, Xuemin Lin, Haixun Wang, Philip S. Yu 961

Spatial Data Management

Distributed Spatial Clustering in Sensor Networks
Anand Meka, Ambuj K. Singh 980

SCUBA: Scalable Cluster-Based Algorithm for Evaluating Continuous Spatio-temporal Queries on Moving Objects
Rimma V. Nehme, Elke A. Rundensteiner 1001

Caching Complementary Space for Location-Based Services
Ken C.K. Lee, Wang-Chien Lee, Baihua Zheng, Jianliang Xu 1020

Extending Data Base Technology

Evolving Triggers for Dynamic Environments
Goce Trajcevski, Peter Scheuermann, Oliviu Ghica, Annika Hinze, Agnes Voisard 1039

A Framework for Distributed XML Data Management
Serge Abiteboul, Ioana Manolescu, Emanuel Taropa 1049

Querying and Updating Probabilistic Information in XML
Serge Abiteboul, Pierre Senellart 1059

An ECA Rule Rewriting Mechanism for Peer Data Management Systems
Dan Zhao, John Mylopoulos, Iluju Kiringa, Verena Kantere 1069

Industrial Session

Business Intelligence

A Metric Definition, Computation, and Reporting Model for Business Operation Analysis
Fabio Casati, Malu Castellanos, Umeshwar Dayal, Ming-Chien Shan 1079

BISON: Providing Business Information Analysis as a Service
Hakan Hacıgümüş, James Rhodes, Scott Spangler, Jeffrey Kreulen 1084

The Design and Architecture of the τ-Synopses System
Yossi Matias, Leon Portman, Natasha Drukh 1088

Database System Enhancements

Integrating a Maximum-Entropy Cardinality Estimator into DB2 UDB
Marcel Kutsch, Peter J. Haas, Volker Markl, Nimrod Megiddo, Tam Minh Tran 1092

Improving DB2 Performance Expert - A Generic Analysis Framework
Laurent Mignet, Jayanta Basak, Manish Bhide, Prasan Roy, Sourashis Roy, Vibhuti S. Sengar, Ranga R. Vatsavai, Michael Reichert, Torsten Steinbach, D.V.S. Ravikant, Soujanya Vadapalli 1097

Managing Collections of XML Schemas in Microsoft SQL Server 2005
Shankar Pal, Dragan Tomic, Brandon Berg, Joe Xavier 1102

Business Data Processing

Enabling Outsourced Service Providers to Think Globally While Acting Locally
Kevin Wilkinson, Harumi Kuno, Kannan Govindarajan, Kei Yuasa, Kevin Smathers, Jyotirmaya Nanda, Umeshwar Dayal 1106

Another Example of a Data Warehouse System Based on Transposed Files
Antonio Albano, Luca De Rosa, Lucio Goglia, Roberto Goglia, Vincenzo Minei, Cristian Dumitrescu 1110

Demonstration Papers

XG: A Grid-Enabled Query Processing Engine
Radu Sion, Ramesh Natarajan, Inderpal Narang, Thomas Phan 1115

Managing and Querying Versions of Multiversion Data Warehouse
Robert Wrembel, Tadeusz Morzy 1121

Natix Visual Interfaces
A. Böhm, M. Brantner, C-C. Kanne, N. May, G. Moerkotte 1125

Hermes - A Framework for Location-Based Data Management
Nikos Pelekis, Yannis Theodoridis, Spyros Vosinakis, Themis Panayiotopoulos 1130

TeNDaX, a Collaborative Database-Based Real-Time Editor System
Stefania Leone, Thomas B. Hodel-Widmer, Michael Boehlen, Klaus R. Dittrich 1135

Synopses Reconciliation Via Calibration in the τ-Synopses System
Yariv Matia, Yossi Matias, Leon Portman 1139

X-Evolution: A System for XML Schema Evolution and Document Adaptation
Marco Mesiti, Roberto Celle, Matteo A. Sorrenti, Giovanna Guerrini 1143

TQuEST: Threshold Query Execution for Large Sets of Time Series
Johannes Aßfalg, Hans-Peter Kriegel, Peer Kröger, Peter Kunath, Alexey Pryakhin, Matthias Renz 1147

VICO: Visualizing Connected Object Orderings
Stefan Brecheisen, Hans-Peter Kriegel, Matthias Schubert, Michael Gruber 1151

XQueryViz: An XQuery Visualization Tool
Jihad Boulos, Marcel Karam, Zeina Koteiche, Hala Ollaic 1155

SAT: Spatial Awareness from Textual Input
Dmitri V. Kalashnikov, Yiming Ma, Sharad Mehrotra, Ramaswamy Hariharan, Nalini Venkatasubramanian, Naveen Ashish 1159

MUSCLE: Music Classification Engine with User Feedback
Stefan Brecheisen, Hans-Peter Kriegel, Peter Kunath, Alexey Pryakhin, Florian Vorberger 1164

*i*MONDRIAN: A Visual Tool to Annotate and Query Scientific Databases
Floris Geerts, Anastasios Kementsietsidis, Diego Milano 1168

The SIRUP Ontology Query API in Action
Patrick Ziegler, Christoph Sturm, Klaus R. Dittrich 1172

Querying Mediated Geographic Data Sources
Mehdi Essid, François-Marie Colonna, Omar Boucelma, Abdelkader Betari 1176

FIS-by-Step: Visualization of the Fast Index Scan for Nearest Neighbor Queries
Elke Achtert, Dominik Schwald 1182

ArHeX: An Approximate Retrieval System for Highly Heterogeneous XML Document Collections
Ismael Sanz, Marco Mesiti, Giovanna Guerrini, Rafael Berlanga Llavori 1186

MonetDB/XQuery—Consistent and Efficient Updates on the Pre/Post Plane
Peter Boncz, Jan Flokstra, Torsten Grust, Maurice van Keulen, Stefan Manegold, Sjoerd Mullender, Jan Rittinger, Jens Teubner 1190

STRIDER: A Versatile System for Structural Disambiguation
Federica Mandreoli, Riccardo Martoglia, Enrico Ronchetti 1194

An Extensible, Distributed Simulation Environment for Peer Data Management Systems
Katja Hose, Andreas Job, Marcel Karnstedt, Kai-Uwe Sattler 1198

Panel

Data Management in the Social Web
Karl Aberer 1203

Author Index 1205