

# Advances in Database Technology – EDBT 2008

---

11th International Conference  
on Extending Database Technology  
Nantes, France, March 25–29, 2008  
Proceedings

Editors:

Alfons Kemper (Technische Universität München, Germany)

Patrick Valduriez (INRIA and LINA, France)

Noureddine Mouaddib (INRIA and LINA, France)

Jens Teubner (IBM T.J. Watson Research Center, USA)

Mokrane Bouzeghoub (Univ. of Versailles, France)

Volker Markl (IBM Almaden Research Center, USA)

Laurent Amsaleg (IRISA, France)

Ioana Manolescu (INRIA-Futurs, France)

Advances in Database Technology – EDBT 2008  
Proceedings of the 11th International Conference  
on Extending Database Technology  
Nantes, France, March 25–29, 2008

Editors:  
Alfons Kemper  
Patrick Valduriez  
Noureddine Mouaddib  
Jens Teubner  
Mokrane Bouzeghoub  
Volker Markl  
Laurent Amsaleg  
Ioana Manolescu

The Association for Computing Machinery  
2 Penn Plaza, Suite 701  
New York, NY, 10121-0701

ACM COPYRIGHT NOTICE. Copyright © 2008 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Publications Dept., ACM, Inc., fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, +1-978-750-8400, +1-978-750-4470 (fax).

ACM ISBN: 978-1-59593-926-5

# Table of Contents

Foreword .....	ix
Program Committee Members .....	x–xi

## Invited Papers

Reality Check: A Case Study of an EII Research Prototype Encountering Customer Needs <i>Eric Simon</i> .....	1
Biocomputational Puzzles: Data, Algorithms, and Visualization <i>Dennis Shasha</i> .....	2
Building Web Applications without a Database System <i>Donald Kossmann</i> .....	3

## Research Sessions

### P2P

P2P Systems with Transactional Semantics <i>Shyam Antony, Divyakant Agrawal, and Amr El Abbadi</i> .....	4–15
Summary Management in P2P systems <i>Rabab Hayek, Guillaume Raschia, Patrick Valduriez, and Noureddine Mouaddib</i> ....	16–25
Semantic Peer, Here are the Neighbors You Want! <i>Wilma Penzo, Stefano Lodi, Federica Mandreoli, Simona Sassatelli, and Riccardo Martoglia</i> .....	26–37

### XML

Scalable XQuery Type Matching <i>Jens Teubner</i> .....	38–48
Engineering Succinct DOM <i>O’Neil Delpratt, Rajeev Raman, and Naila Rahman</i> .....	49–60
Revisiting Redundancy and Minimization in an XPath Fragment <i>Benny Kimelfeld and Yehoshua Sagiv</i> .....	61–72

### Data Integration

Schema Merging and Mapping Creation for Relational Sources <i>Rachel Pottinger and Philip Bernstein</i> .....	73–84
Schema Mapping Verification: The Spicy Way <i>Angela Bonifati, Giansalvatore Mecca, Alessandro Pappalardo, Salvatore Raunich, and Gianvito Summa</i> .....	85–96

SeMap: A Generic Mapping Construction System <i>Ting Wang and Rachel Pottinger</i> .....	97–108
<b>Consistency</b>	
Deleting Index Entries from Compliance Storage <i>Soumyadeb Mitra, Marianne Winslett, and Nikita Borisov</i> .....	109–120
Online Recovery in Cluster Databases <i>WeiBin Liang and Bettina Kemme</i> .....	121–132
A Concurrency Control Protocol for Parallel B-tree Structures without Latch-Coupling for Explosively Growing Digital Content <i>Tomohiro Yoshihara, Dai Kobayashi, and Haruo Yokota</i> .....	133–144
<b>Physical Design</b>	
Robustness in Automatic Physical Database Design <i>Kareem El Gebaly and Ashraf Aboulnaga</i> .....	145–156
Self-organizing strategies for a column-store database <i>Milena Ivanova, Martin Kersten, and Niels Nes</i> .....	157–168
Load Distribution of Analytical Query Workloads for Database Cluster Architectures <i>Thomas Phan and Wen-Syan Li</i> .....	169–180
<b>Graph Databases</b>	
A Novel Spectral Coding in a Large Graph Database <i>Lei Zou, Lei Chen, Jeffrey Xu Yu, and Yansheng Lu</i> .....	181–192
Fast Computing Reachability Labelings for Large Graphs with High Compression Rate <i>Jiefeng Cheng, Jeffrey Xu Yu, Xuemin Lin, Haixun Wang, and Philip S. Yu</i> .....	193–204
Finding Time-Dependent Shortest Paths over Large Graphs <i>Bolin Ding, Jeffrey Xu Yu, and Lu Qin</i> .....	205–216
Taxonomy-Superimposed Graph Mining <i>Ali Cakmak and Gultekin Ozsoyoglu</i> .....	217–228
<b>Indexing</b>	
Compacting Music Signatures for Efficient Music Retrieval <i>Bin Cui, H. V. Jagadish, Beng Chin Ooi, and Kian-Lee Tan</i> .....	229–240
Indexing High-Dimensional Data in Dual Distance Spaces: A Symmetrical Encoding Approach <i>Yi Zhuang, Yueting Zhuang, Qing Li, Lei Chen, and Yi Yu</i> .....	241–251
The TS-Tree: Efficient Time Series Search and Retrieval <i>Ira Assent, Ralph Krieger, Farzad Afschari, and Thomas Seidl</i> .....	252–263
<b>Confidentiality</b>	
Anonymity for Continuous Data Publishing <i>Benjamin C. M. Fung, Ke Wang, Ada Wai-Chee Fu, and Jian Pei</i> .....	264–275
Ownership Protection of Shapes Datasets with Geodesic Distance Preservation <i>Michail Vlachos, Claudio Lucchese, Deepak Rajan, and Philip S. Yu</i> .....	276–286

Zerber: r-Confidential Indexing for Distributed Documents <i>Sergej Zerr, Elena Demidova, Daniel Olmedilla, Wolfgang Nejdl, Marianne Winslett, and Soumyadeb Mitra</i> .....	287–298
<b>Materialization and Caching</b>	
XCraft: Boosting the Performance of Active XML Materialization <i>Gabriela Ruberg and Marta Mattoso</i> .....	299–310
Exact and Inexact Methods for Selecting Views and Indexes for OLAP Performance Improvement <i>Rada Chirkova, Zohreh Asgharzadeh Talebi, Yahya Fathi, and Matthias Stallmann</i> ..	311–322
Providing Freshness Guarantees for Outsourced Databases <i>Min Xie, Haixun Wang, Jian Yin, and Xiaofeng Meng</i> .....	323–332
<b>Skyline, Top-k, Preferences</b>	
Dynamic Skyline Queries in Metric Spaces <i>Lei Chen and Xiang Lian</i> .....	333–343
Fast Contextual Preference Scoring of Database Tuples <i>Kostas Stefanidis and Evaggelia Pitoura</i> .....	344–355
Efficient Online Top-K Retrieval with Arbitrary Similarity Measures <i>Prasad Deshpande, Deepak P, and Krishna Kummamuru</i> .....	356–367
<b>Data Mining</b>	
Mining All Frequent Projection-Selection Queries from a Relational Table <i>Tao-Yuan Jen, Dominique Laurent, and Nicolas Spyrtatos</i> .....	368–379
Cost-based query optimization for complex pattern mining on multiple databases <i>Ruoming Jin, David Fuhry, and Abdulkareem Alali</i> .....	380–391
On-Line Discovery of Hot Motion Paths <i>Dimitris Sacharidis, Kostas Patroumpas, Manolis Terrovitis, Michalis Potamias, Verena Kantere, Kyriakos Mouratidis, and Timos Sellis</i> .....	392–403
<b>Schema Management</b>	
Schema Polynomials and Applications <i>Kenneth Ross and Julia Stoyanovich</i> .....	404–415
Expressive Query Specification through Form Customization <i>Magesh Jayapandian and H. V. Jagadish</i> .....	416–427
Synthesizing Structured Text from Logical Database Subsets <i>Alkis Simitsis, Georgia Koutrika, Yannis Alexandrakis, and Yannis Ioannidis</i> .....	428–439
<b>Multi-Dimensional Data</b>	
HISSCLU: A Hierarchical Density-Based Method for Semi-Supervised Clustering <i>Christian Böhm and Claudia Plant</i> .....	440–451
Processing Transitive Nearest Neighbor Queries in Multi-Channel Access Environments <i>Xiao Zhang, Wang-Chien Lee, Prasenjit Mitra, and Baihua Zheng</i> .....	452–463
Multi-Dimensional Search for Personal Information Management Systems <i>Christopher Peery, Wei Wang, Amélie Marian, and Thu Nguyen</i> .....	464–475

## Data Fusion

Ensuring Correctness over Untrusted Private Database <i>Sarvjeet Singh and Sunil Prabhakar</i> .....	476–486
Data Exchange in the Presence of Arithmetic Comparisons <i>Foto Afrati, Chen Li, and Vassia Pavlaki</i> .....	487–498
SPARQLing Constraints for RDF <i>Georg Lausen, Michael Meier, and Michael Schmidt</i> .....	499–510

## Query Processing

Probabilistic Ranked Queries in Uncertain Databases <i>Xiang Lian and Lei Chen</i> .....	511–522
The SBC-Tree: An Index for Run-Length Compressed Sequences <i>Mohamed Eltabakh, WIng-Kai Hon, Rahul Shah, Walid G. Aref, and Jeffrey Vitter</i> .	523–534
Efficient LCA based Keyword Search in XML Data <i>Yu Xu and Yannis Papakonstantinou</i> .....	535–546

## Streams

Querying Time-Series Streams <i>Vivekanand Gopalkrishnan</i> .....	547–558
Optimizing On-Demand Data Broadcast Scheduling in Pervasive Environments <i>Rinku Dewri, Indrakshi Ray, Indrajit Ray, and Darrell Whitley</i> .....	559–569
On the brink: Searching for drops in sensor data <i>Gong Chen, Junghoo Cho, and Mark Hansen</i> .....	570–581

## Join Processing

A Stratified Approach to Progressive Approximate Joins <i>Wee Hyong Tok, Stéphane Bressan, and Mong-Li Lee</i> .....	582–593
Continuous Multi-Way Joins over Distributed Hash Tables <i>Stratos Idreos, Erietta Liarou, and Manolis Koubarakis</i> .....	594–605
Ring-constrained Join: Deriving Fair Middleman Locations from Pointsets via a Geometric Constraint <i>Man Lung Yiu, Panagiotis Karras, and Nikos Mamoulis</i> .....	606–617

## Industrial Sessions

### Industrial 1

Why Go Logarithmic if We Can Go Linear? Towards Effective Distinct Counting of Search Traffic <i>Ahmed Metwally, Divyakant Agrawal, and Amr El Abbadi</i> .....	618–629
Automatic Content Targeting on Mobile Phones <i>Giovanni Giuffrida, Giuseppe Tribulato, and Catarina Sismeiro</i> .....	630–639
BI Batch Manager: A System for Managing Batch Workloads on Enterprise Data Warehouses <i>Abhay Mehta, Chetan Gupta, and Umeshwar Dayal</i> .....	640–651

Data Challenges at Yahoo!	
<i>Ricardo Baeza-Yates and Raghu Ramakrishnan</i> .....	652–655

## Industrial 2

OrthoCluster: A New Tool for Mining Syntenic Blocks and Applications in Comparative Genomics	
<i>Jian Pei, Xinghuo Zeng, Ismael A. Vergara, Matthew J. Nesbitt, Ke Wang, and Nansheng Chen</i> .....	656–667

Social Ties and their Relevance to Churn in Mobile Telecom Networks	
<i>Koustuv Dasgupta, Dipanjan Chakraborty, Amit A. Nanavati, Sougata Mukherjea, Rahul Singh, Balaji Viswanathan, and Anupam Joshi</i> .....	668–677

Highly Scalable Trip Grouping for Large-Scale Collective Transportation Systems	
<i>Gyozo Gidofalvi, Torben Bach Pedersen, Tore Risch, and Erik Zeitler</i> .....	678–689

## Demonstrations

### Services and Mash-Ups

Data Services in your Spreadsheet!	
<i>Regis Saint-Paul, Boualem Benatallah, and Julien Vayssiere</i> .....	690–694

BeMatch: A Platform for Matchmaking Service Behavior Models	
<i>Juan Carlos Corrales, Daniela Grigori, Mokrane Bouzeghoub, and Javier Ernesto Burbano</i> .....	695–699

The TELAR Mobile Mashup Platform for Nokia Internet Tablets	
<i>Andreas Brodt and Daniela Nicklas</i> .....	700–704

### Access Control and Security

MCSE: A Multimedia Context Based Security Engine	
<i>Bechara Albouna and Richard Chbeir</i> .....	705–709

iDataGuard: Middleware Providing a Secure Network Drive Interface to Untrusted Internet Data Storage	
<i>Ravi Chandra Jammalamadaka, Roberto Gamboni, Sharad Mehrotra, Kent Seamons, and Nalini Venkatasubramanian</i> .....	710–714

ACCOOn: Checking Consistency of XML Write-Access Control Policies	
<i>Loreto Bravo, James Cheney, and Irini Fundulaki</i> .....	715–719

### Web and Distribution

Flint: Google-Basing the Web	
<i>Lorenzo Blanco, Valter Crescenzi, Paolo Merialdo, and Paolo Papotti</i> .....	720–724

Mine Your Own Business, Mine Others' News!	
<i>Quang-Khai Pham, Regis Saint-Paul, Boualem Benatallah, Noureddine Mouaddib, and Guillaume Raschia</i> .....	725–729

BioScout: A Life-Science Query Monitoring System	
<i>Anastasios Kementsietsidis, Frank Neven, and Dieter Van de Craen</i> .....	730–734

### Languages and Models

BIBEX: a Bibliographic Exploration Tool based on the DEX Graph Query Engine	
<i>Sergio Gómez-Villamor, Gerard Soldevila-Miranda, Aleix Giménez-Vañó, Norbert Martínez-Bazan, Victor Muntés-Mulero, and Josep-L. Larriba-Pey</i> .....	735–739

An Inductive Database and Query Language in the Relational Model <i>Lothar Richter, Jörg Wicker, Kristina Kessler, and Stefan Kramer</i> .....	740–744
---	---------

## Tutorials

Streaming in a Connected World: Querying and Tracking Distributed Data Streams <i>Graham Cormode and Minos Garofalakis</i> .....	745
Virtualization and Databases: State of the Art and Research Challenges <i>Ashraf Aboulnaga, Cristiana Amza, and Kenneth Salem</i> .....	746–747
Quality of Service and Predictability in DBMS <i>Kai-Uwe Sattler and Wolfgang Lehner</i> .....	748
Author Index .....	745–747



# Foreword

The EDBT series of conferences is an established and prestigious forum for the exchange of the latest research results in data management. Held every two years in an attractive European location, the conference provides unique opportunities for database researchers, practitioners, developers, and users to explore new ideas, techniques, and tools, and to exchange experiences.

The previous EDBT events were held in Venice, Vienna, Cambridge, Avignon, Valencia, Konstanz, Prague, Heraklion and Munich. The 11th International Conference on Extending Database Technology (EDBT 2008) was held from March 25 to 29, 2008 in Nantes, France.

Data management constitutes the essential enabling technology for scientific, engineering, business, and social communities. Technological trends, new computation paradigms, novel applications, sophisticated user interactions, they all require robust and flexible database technology to be deployed in a variety of environments and for several diverse purposes. Peer-to-peer architectures, the Grid, personal information systems, pervasive and ubiquitous computing, networked sensors, biomedical informatics, virtual digital libraries, virtual communities, and trust management are just a small sample of the great challenges ahead of us that drive research and development of the next generation of database technology.

The new information paradigms and requirements will move our research community away from any narrow interpretation of databases and expand its focus to the hard problems faced by broad visions of data, information, and knowledge management.

Researchers submitted contributions that picked up on brand new challenges and explored new and exciting technical directions wherever data management issues may be found. EDBT 2008 invited submissions of original research contributions, as well as proposals for panels, tutorials and software demonstrations. The EDBT conference series covers a broad range of topics, including traditional database management as well new issues arising in any possible domain. To substantially advance the state of the art authors were encouraged to consider novel topics and approaches rather than incremental improvements of existing results.

Putting together the EDBT 2008 proceedings was a true team effort. Special thanks go to the members of the various program committees.

Thanks to the dedicated effort of the program committee members of the Research Track we were able to select 52 high-quality research papers out of the initially 311 submitted proposals. These papers were clustered into 17 sessions and presented during the three main days of the conference.

The industrial and application track program committee selected 7 high quality papers out of 16 submissions. These papers will be presented in 2 sessions during the conference.

Overall, the PC chairs are looking forward towards an insightful and inspirational conference.

Alfons Kemper (Program Chair)  
Patrick Valduriez (General Chair)  
Noureddine Mouaddib (Executive Chair)  
Jens Teubner (Proceedings Chair)  
Mokrane Bouzeghoub (Panel & Tutorial Chair)  
Volker Markl (Industrial & Applications Chair)  
Laurent Amsaleg (Workshop Chair)  
Ioana Manolescu (Demonstrations Chair)

# Program Committee Members

## Research

Karl Aberer	Götz Graefe	Tamer Ozsu
Antonio Albano	Torsten Grust	Esther Pacitti
Sihem Amer Yahia	Ehud Gudes	Norman Paton
Walid G. Aref	Peter Gursky	Reinhard Pichler
Paolo Atzeni	Joachim Hammer	Ivana Podnar Žarko
Wolf-Tilo Balke	Theo Härder	Alexandra Poulouvassilis
Catriel Beerl	Michael Hatzopoulos	Krithi Ramamritham
Michael Benedikt	Sven Helmer	Guillaume Raschia
Elisa Bertino	Jan Hidders	Tore Risch
Michael Böhlen	Jeong-Hyon Hwang	Uwe Röhm
Klemens Böhm	H. V. Jagadish	Elke Rundensteiner
Christian Böhm	Christian S. Jensen	Ralf Schenkel
Peter Boncz	Carl Christian Kanne	Stefanie Scherzinger
Angela Bonifati	Daniel Keim	Peter Scheuermann
Matthias Brantner	Bettina Kemme	Marc Scholl
Francois Bry	Alfons Kemper (Chair)	Stefan Schönauer
Alex Buchmann	Martin Kersten	Heiko Schuldt
Fabio Casati	Daniel Kifer	Bernhard Seeger
Tiziana Catarci	Christoph Koch	Thomas Seidl
Zhiyuan Chen	Arnd Christian König	Timos Sellis
Jan Chomicki	Harald Kosch	Vladislav Shkapenyuk
Vassilis Christophidis	Donald Kossmann	Fabrizio Silvestri
Christine Collet	Nick Koudas	Jérôme Siméon
Sara Comai	Georgia Koutrika	Val Tannen
Umeshwar Dayal	Harumi Kuno	Yufei Tao
Giuseppe De Giacomo	Richard Kuntschke	Jens Teubner
Arjen P. de Vries	Alexandros Labrinidis	Bernhard Thalheim
Alex Delis	Christian A. Lang	Martin Theobald
Stefan Dessoth	Georg Lausen	David Toman
Jens Dittrich	Wolfgang Lehner	Peter Triantafillou
Klaus R. Dittrich	Stefan Manegold	Can Türker
Asuman Dogac	Christian Mathis	Jan Van den Busche
Wenfei Fan	Florian Matthes	Maurice van Keulen
Mary Fernandez	Wolfgang May	Vasilis Vassalos
Christoph Freytag	Michele Missikoff	Stratis Viglas
Ada Fu	Ami Motro	Agnès Voisard
Norbert Fuhr	Karin Murthy	Gottfried Vossen
Cesar Galindo-Legaria	John Mylopoulos	Kyu-Young Whang
Bugra Gedik	Felix Naumann	Philip S. Yu
Michael Gertz	Thomas Neumann	Pavel Zezula
Shahram Ghandeharizadeh	Frank Neven	
Giorgio Ghelli	Dan Olteanu	
Torsten Grabs	Beng Chin Ooi	

## **Industrial & Applications**

Timos Sellis  
Jingren Zhou  
Jayant Madhavan  
Harald Schoening  
Carsten Sapia

Volker Markl (Chair)  
Josep Lluís Larriba Pey  
Yasushi Sakurai  
Jayavel Shanmugasundaram  
Calisto Zuzarte

Dean Jacobs  
Christof Bornhoevd

## **Demonstrations**

Nicolas Anciaux  
Alessandro Campi  
Alexandros Labrinidis

Ioana Manolescu (Chair)  
Mirella Moro  
Michalis Petropoulos

Felix Weigel