

# Advances in Database Technology – EDBT 2011

---

14th International Conference  
on Extending Database Technology  
Uppsala, Sweden, March 22–24, 2011  
Proceedings

Editors:

Anastasia Ailamaki (EPFL, Switzerland)  
Sihem Amer-Yahia (Yahoo! Research, USA)  
Jignesh Patel (University of Wisconsin-Madison, USA)  
Tore Risch (Uppsala University, Sweden)  
Pierre Senellart (Télécom ParisTech, France)  
Julia Stoyanovich (University of Pennsylvania, USA)

Advances in Database Technology – EDBT 2011  
Proceedings of the 14th International Conference  
on Extending Database Technology  
Uppsala, Sweden, March 22–24, 2011

Editors:  
Anastasia Ailamaki  
Sihem Amer-Yahia  
Jignesh Patel  
Tore Risch  
Pierre Senellart  
Julia Stoyanovich

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

ACM COPYRIGHT NOTICE. Copyright © 2011 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: 978-1-4503-0528-0

# Table of Contents

Foreword .....	viii
Program Committee Members .....	ix–x
<b>Invited Papers</b>	
Map-Reduce Extensions and Recursive Queries <i>Foto N. Afrati, Vinayak Borkar, Michael Carey, Neoklis Polyzotis, and Jeffrey D. Ullman</i> .....	1–8
Database Researchers: Plumbers or Thinkers? <i>Gerhard Weikum</i> .....	9–10
<b>Research Sessions</b>	
<b>Mining and Complex Events</b>	
Novel Techniques to Reduce Search Space in Multiple Minimum Supports-Based Frequent Pattern Mining Algorithms <i>R. Uday Kiran and P. Krishna Reddy</i> .....	11–20
Mining Closed Discriminative Dyadic Sequential Patterns <i>David Lo, Hong Cheng, and Lucia</i> .....	21–32
Sequenced Event Set Pattern Matching <i>Bruno Cadonna, Johann Gamper, and Michael H. Böhlen</i> .....	33–44
<b>Data Streams</b>	
GPX-Matcher: A Generic Boolean Predicate-based XPath Expression Matcher <i>Mohammad Sadoghi, Ioana Burcea, and Hans-Arno Jacobsen</i> .....	45–56
An Optimal Strategy for Monitoring Top-k Queries in Streaming Windows <i>Di Yang, Avani Shastri, Elke Rundensteiner, and Matthew Ward</i> .....	57–68
Fast and Accurate Computation of Equi-Depth Histograms over Data Streams <i>Hamid Mousavi and Carlo Zaniolo</i> .....	69–80
<b>Energy and Performance</b>	
Energy Proportionality for Disk Storage Using Replication <i>Jinoh Kim and Doron Rotem</i> .....	81–92
Caching Query-biased Snippets for Efficient Retrieval <i>Diego Ceccarelli, Claudio Lucchese, Salvatore Orlando, Raffaele Perego, and Fabrizio Silvestri</i> .....	93–104
Efficient and Scalable Data Evolution with Column Oriented Databases <i>Ziyang Liu, Bin He, Hui-I Hsiao, and Yi Chen</i> .....	105–116

## Data in the Cloud

Native Support of Multi-tenancy in RDBMS for Software as a Service <i>Oliver Schiller, Benjamin Schiller, Andreas Brodt, and Bernhard Mitschang</i> . . . . .	117–128
SLA-Tree: A Framework for Efficiently Supporting SLA-based Decisions in Cloud Computing <i>Yun Chi, Hyun Moon, Hakan Hacigümüş, and Junichi Tatemura</i> . . . . .	129–140
On Enhancing Scalability for Distributed RDF/S Stores <i>George Tsatsanifos, Dimitris Sacharidis, and Timos Sellis</i> . . . . .	141–152

## XML and Semistructured Databases

Answering Tree Pattern Queries Using Views: A Revisit <i>Junhu Wang, Jiang Li, and Jeffery Xu Yu</i> . . . . .	153–164
Dynamic Reasoning on XML Updates <i>Federico Cavaliere, Giovanna Guerrini, and Marco Mesiti</i> . . . . .	165–176
Algebraic Incremental Maintenance of XML Views <i>Angela Bonifati, Martin Goodfellow, Domenica Sileo, and Ioana Manolescu</i> . . . . .	177–188

## Personalization

Keyword-based, Context-aware Selection of Natural Language Query-patterns <i>Giorgio Orsi, Letizia Tanca, and Eugenio Zimeo</i> . . . . .	189–200
Unified Structure and Content Search for Personal Information Management Systems <i>Wei Wang, Amélie Marian, and Thu Nguyen</i> . . . . .	201–212
TopRecs: Top-k Algorithms and Data Structures for Item-based Collaborative Filtering <i>Mohammad Khabbazzhayej Tajer and Laks V. S. Lakshmanan</i> . . . . .	213–224

## Indexing

Efficient Answering of Set Containment Queries for Skewed Item Distributions <i>Manolis Terrovitis, Panagiotis Bouros, Panos Vassiliadis, Nikos Mamoulis, and Timos Sellis</i> . . . . .	225–236
Subspace Clustering for Indexing High Dimensional Data: A Main Memory Index based on Local Reductions and Individual Multi-Representations <i>Stephan Günnemann, Hardy Kremer, Dominik Lenhard, and Thomas Seidl</i> . . . . .	237–248
On (Not) Indexing Quadratic Form Distance by Metric Access Methods <i>Tomáš Skopal, Tomáš Bartoš, and Jakub Lokoč</i> . . . . .	249–258

## Spatial and Temporal Databases

SeMiTri: A Framework for Semantic Annotation of Heterogeneous Trajectories <i>Zhixian Yan, Dipanjan Chakraborty, Christine Parent, Stefano Spaccapietra, and Karl Aberer</i> . . . . .	259–270
Efficient Execution Plans for Distributed Skyline Query Processing <i>João B. Rocha-Junior, Akrivi Vlachou, Christos Doukeridis, and Kjetil Nørkvåg</i> . . . . .	271–282
Probabilistic Range Queries for Uncertain Trajectories on Road Networks <i>Kai Zheng, Goce Trajcevski, Xiaofang Zhou, and Peter Scheuermann</i> . . . . .	283–294

## Query Processing and Optimization

TAGs: Scalable Threshold-Based Algorithms for Proximity Computation in Graphs <i>Apostolos Lyritsis, Apostolos Papadopoulos, and Yannis Manolopoulos</i> .....	295–306
Projection for XML Update Optimization <i>Amine Baazizi, Nicole Bidoit-Tollu, Dario Colazzo, Noor Malla, and Marina Sahakyan</i> .....	307–318
Efficient Reverse Skyline Retrieval with Arbitrary Non-Metric Similarity Measures <i>Prasad Deshpande and Deepak P.</i> .....	319–330

## Graph Databases

Fast Random Graph Generation <i>Sadegh Nobari, Xuesong Lu, Panagiotis Karras, and Stéphane Bressan</i> .....	331–342
Symmetrizations for Clustering Directed Graphs <i>Venu Satuluri and Srinivasan Parthasarathy</i> .....	343–354
Efficient Discovery of Frequent Subgraph Patterns in Uncertain Graph Databases <i>Odysseas Papapetrou, Ekaterini Ioannou, and Dimitrios Skoutas</i> .....	355–366

## Application-driven Processing

Finding Closed Frequent Item Sets by Intersecting Transactions <i>Christian Borgelt, Xiaoqian Yang, Ruben Nogales-Cadenas, Pedro Carmona-Saez, and Alberto Pascual-Montano</i> .....	367–376
Aspect-Oriented Relational Algebra <i>Curtis Dyreson</i> .....	377–388
Synopses for Probabilistic Data Over Large Domains <i>Nicholas Larusso and Ambuj Singh</i> .....	389–400

## Web and Services

Data Integration with Dependent Sources <i>Anish Das Sarma, Luna Dong, and Alon Halevy</i> .....	401–412
Constructing Concept Relation Network and its Application to Personalized Web Search <i>Kenneth Wai-Ting Leung, Hing Yuet Fung, and Dik Lun Lee</i> .....	413–424
Effective and Efficient Sampling Methods for Deep Web Aggregation Queries <i>Fan Wang and Gagan Agrawal</i> .....	425–436
Making Interval-Based Clustering Rank-Aware <i>Julia Stoyanovich, Sihem Amer-Yahia, and Tova Milo</i> .....	437–448

## Prediction and Discovery

Predicting Completion Times of Batch Query Workloads using Interaction-aware Models and Simulation <i>Mumtaz Ahmad, Songyun Duan, Ashraf Aboulnaga, and Shivnath Babu</i> .....	449–460
Memory-Efficient Frequent-Itemset Mining <i>Benjamin Schlegel, Rainer Gemulla, and Wolfgang Lehner</i> .....	461–472
Link-based Hidden Attribute Discovery for Objects on Web <i>Jiuming Huang, Haixun Wang, Yan Jia, and Ariel Fuxman</i> .....	473–484

Real-time Approximate Range Motif Discovery & Data Redundancy Removal Algorithm <i>Ankur Narang and Souvik Bhattacharjee</i> .....	485–496
---	---------

## Industrial and Applications Sessions

### The Design of Industrial Database Systems

Queries on Dates: Fast yet not Blind <i>Jaroslav Szlichta, Parke Godfrey, Jarek Gryz, Wenbin Ma, Przemyslaw Pawluk, and Calisto Zuzarte</i> .....	497–502
Designing Integration Flows Using Hypercubes <i>Kevin Wilkinson and Alkis Simitsis</i> .....	503–508
Experience in Continuous analytics as a Service (CaaS) <i>Qiming Chen, Meichun Hsu, and Hans Zeller</i> .....	509–514

### Data Mining Applications

SocialSearch: Enhancing Entity Search with Social Network Matching <i>Gae-won You, Seung-won Hwang, Zaiqing Nie, and Ji-Rong Wen</i> .....	515–520
Road Crash Proneness Prediction using Data Mining <i>Richi Nayak, Daniel Emerson, Justin Weligamage, and Noppadol Piyatrapoomi</i> ....	521–526

## Tutorials

Schema Matching and Mapping: From Usage to Evaluation <i>Angela Bonifati and Yannis Velegarakis</i> .....	527–529
Big Data and Cloud Computing: Current State and Future Opportunities <i>Divyakant Agrawal, Sudipto Das, and Amr El Abbadi</i> .....	530–533
The Hidden Web, XML and the Semantic Web: Scientific Data Management Perspectives <i>Fabian Suchanek, Aparna Varde, Richi Nayak, and Pierre Senellart</i> .....	534–537

## Demonstrations

A Probabilistic XML Merging Tool <i>Talel Abdessalem, Mouhamadou Lamine Ba, and Pierre Senellart</i> .....	538–541
Taking the OXPath down the Deep Web <i>Andrew Sellers, Tim Furche, Georg Gottlob, Giovanni Grasso, and Christian Schallhart</i> .....	542–545
SPRINT: Ranking Search Results by Paths <i>Christoph Böhm, Felix Naumann, Eyk Kny, Benjamin Emde, and Ziawasch Abedjan</i> .....	546–549
A Query Optimization Assistant for XPath <i>Haris Georgiadis, Minas Charalambidis, and Vasilis Vassalos</i> .....	550–553
TPM: Supporting Pattern Matching Queries for Road-Network Trajectory Data <i>Gook-Pil Roh and Seung-won Hwang</i> .....	554–557
QueryViz: Helping Users Understand SQL Queries and their Patterns <i>Wolfgang Gatterbauer and Jonathan Danaparamita</i> .....	558–561
True Language-Level SQL Debugging <i>Torsten Grust, Fabian Kliebhan, Jan Rittinger, and Tom Schreiber</i> .....	562–565

SITAC: Discovering Semantically Identical Temporally Altering Concepts in Text Archives	
<i>Amal Kaluarachchi, Debjani Roychoudhury, Aparna S. Varde, and Gerhard Weikum</i> .....	566–569
Unraveling Multi-Dimensional Data using pDView	
<i>Luigi Di Caro, Maria Luisa Sapino, and K. Selçuk Candan</i> .....	570–573
RanKloud: A Scalable Ranked Query Processing Framework on Hadoop	
<i>K. Selçuk Candan, Parth Nagarkar, Mithila Nagendra, and Renwei Yu</i> .....	574–577

# Foreword

Welcome to the 2011 edition of the International Conference on Extending Database Technology (EDBT). This year, EDBT welcomes its delegates in Uppsala, Sweden on March 22–24, 2011.

In contrast to the numbers reported in previous years, the number of research paper submissions was low (about half of the previous years): EDBT 2011 received 148 submissions in the research track. Usually, EDBT aligns submission and author notification deadlines to those of the neighboring conferences VLDB, SIGMOD, and ICDE; this year, however, the aforementioned conferences changed the deadlines significantly, which made it impossible for EDBT to align and still keep the same conference dates. In the other two tracks, numbers are similar to previous years (20 demonstration proposals, 12 industrial papers, and 4 tutorial proposals).

The reviews were conducted by an expert team of program committee members and the process was managed using the Conference Management Toolkit (CMT), sponsored by Microsoft Research, which has become a de-facto standard in our community for handling such large number of papers and reviews. In the research track, ten committee members also assumed the responsibility of overseeing the review process of a large subset of the submissions as vice-chairs, and several other program committee members acted as shepherds ensuring the high quality of the camera-ready papers. The reviewers were extremely devoted, reviews were on time, and the shepherds did a magnificent job working with the authors despite the winter holidays. We feel very privileged to have had the opportunity to work with such a distinguished set of colleagues.

The program committee in the research track decided to view the low number of submissions as an opportunity to create a program of even higher quality. Indeed, we were able to (a) invite four, as opposed to the usual three, reviews per paper; (b) pay extremely high attention to each paper and ensure high-quality reviews regardless of the outcome; and (c) shepherd the camera-ready versions of the papers to ensure that the authors and the reviewers were both satisfied with the camera-ready paper. As a result, EDBT 2011 proudly presents a high-quality program with 41 research papers, 5 industrial papers, 10 demonstrations and 3 tutorials. The program is complemented with four keynote talks (joint with the ICDT conference) from distinguished researchers in academia, as well as an invited talk from the industry.

Finally, we decided not to invite panel proposals, and to use the time to accept more good research papers. Instead, there are eight workshops around the conference, which are a well-targeted opportunity for discussions and brainstorming toward new areas of research.

We hope you will enjoy EDBT 2011. See you in Uppsala!

Anastasia Ailamaki (Program Chair)

Jignesh Patel (Tutorials Chair)

Pierre Senellart (Industrial and Applications Chair)

Siham Amer-Yahia (Demonstrations Chair)

Tore Risch (General Chair)

Julia Stoyanovich (Proceedings Chair)



# Program Committee Members

## Research

### Reviewers

Karl Aberer	Goetz Graefe (vice-chair)	Evi Pitoura
Divyakant Agrawal (vice-chair)	Torsten Grust	Alkis Polyzotis
Anastasia Ailamaki (chair)	Thomas Heinis	Alexandra Poulouvassilis
Sihem Amer-Yahia	Bill Howe	Sunil Prabhakar
Periklis Andritsos	Stratos Idreos	Philippe Pucheral
Paolo Atzeni	Milena Ivanova	Erhard Rahm
Shivnath Babu	Dean Jacobs	Maya Ramanath
Ricardo Baeza-Yates	H V Jagadish	Elke Rundensteiner
Elisa Bertino	Christian Jensen	Simonas Saltenis
Bishwaranjan Bhattacharjee	Chris Jermaine	Sunita Sarawagi
Pedro Bizarro	Verena Kantere	Kai-Uwe Sattler
Angela Bonifati	Panagiotis Karras	Peter Scheuermann
Philippe Bonnet	Bettina Kemme	Berni Schiefer
Luc Bouganim	Alfons Kemper	Thomas Seidl
Alejandro Buchmann	Martin Kersten	Timos Sellis
Fabio Casati	Masaru Kitsuregawa (vice-chair)	Oded Shmueli
Tiziana Catarci (vice-chair)	Christoph Koch (vice-chair)	Alkis Simitsis
Bogdan Cautis	Donald Kossmann (vice-chair)	Altigran Soares da Silva
Surajit Chaudhuri (vice-chair)	Yannis Kotidis	Heinz Stockinger
Panos Chrysanthis (vice-chair)	Georgia Koutrika	Kian-Lee Tan
Anish Das Sarma	Tim Kraska	Yufei Tao
Umesh Dayal (vice-chair)	Wolfgang Lehner	Nesime Tatbul
Arjen deVries	Qiong Luo	Jens Teubner
Jens Dittrich	Nikos Mamoulis	Peter Triantafillou
Gill Dobbie	Volker Markl	Maurice van Keulen
Alin Dobra	Wim Martens	Yannis Velegrakis
Sameh Elnikety	Gianni Mecca	Stratis Viglas
Alan Fekete	Claudia Medeiros	Wei Wang
Irini Fundulaki	Kjetil Nøråg	Gerhard Weikum
Johann Gamper	Beng-Chin Ooi (vice-chair)	Haruo Yokota
Minos Garofalakis (vice-chair)	Kjell Orsborn	Marcin Zukowski
Floris Geerts	Fatma Ozcan	Björn Þór Jónsson
Johannes Gehrke (vice-chair)	Norman Paton	

### External Reviewers

Tristan Allard	Yanli Guo	Wenxin Liang
David Alves	Ali Inan	Lipyeow Lim
Toshiyuki Amagasa	Xin Jin	Paulo Marques
Nicolas Anceaix	Alekh Jindal	George Mihaila
Lujun (Tony) Fang	Mouna Kacimi	Miyuki Nakano

Jorge Quiané  
Roi Ronen  
Jörg Schad  
Stefan Schuh  
Lila Shnaiderman

Masashi Toyoda  
Octavian Udrea  
Yousuke Watanabe  
Lexing Xie  
Zhenglu Yang

Shaoyi Yin  
Naoki Yoshinaga

## **Industrial and Applications**

Tasso Argyros  
Srikanta Bedathur  
Vinayak Borkar  
Matthias Brantner  
Alfredo Cuzzocrea  
Benoit Dageville

Dieter Gawlick  
Jeff Hammerbacher  
Jian Huang  
Gjergji Kasneci  
Bogdan Marinoiu  
Richi Nayak

Pierre Senellart (chair)  
Eric Simon  
Florian Waas  
Calisto Zuzarte

## **Demonstrations**

Bernd Amann  
Sihem Amer-Yahia (chair)  
Srikanta Bedathur  
Mokrane Bouzeghoub  
Carlos Castillo  
Jian Huang  
Theodore Johnson  
Mouna Kacimi

Tim Kraska  
Anne-Marie Kermarrec  
Amélie Marian  
Marco Mesiti  
Atsuyuki Morishima  
Mirella Mor  
Esther Pacitti  
Maya Ramanath

Philippe Rigaux  
Ismael Sanz  
Jayavel Shanmugasundaram  
Julia Stoyanovich  
Torsten Suel  
Vasilis Vassalos  
Cong Yu