
Advances in Database Technology — EDBT 2015

18th International Conference
on Extending Database Technology
Brussels, Belgium, March 23–27, 2015
Proceedings

Editors

Gustavo Alonso
Floris Geerts
Lucian Popa
Pablo Barceló
Jens Teubner
Martín Ugarte
Jan Van den Bussche
Jan Paredaens



Advances in Database Technology — EDBT 2015
Proceedings of the 18th International Conference
on Extending Database Technology
Brussels, Belgium, March 23–27, 2015

Editors

Gustavo Alonso, ETH Zurich, Switzerland
Floris Geerts, University of Antwerp, Belgium
Lucian Popa, IBM Research, USA
Pablo Barceló, University of Chile, Chile
Jens Teubner, TU Dortmund, Germany
Martín Ugarte, Pontificia Universidad Católica de Chile, Chile
Jan Van den Bussche, Hasselt University, Belgium
Jan Paredaens, University of Antwerp, Belgium



OpenProceedings.org
University of Konstanz
University Library
78457 Konstanz, Germany

COPYRIGHT NOTICE: Copyright © 2015 by the authors of the individual papers.

Distribution of all material contained in this volume is permitted under the terms of the Creative Commons license
CC-by-nc-nd 4.0

OpenProceedings ISBN: 978-3-89318-067-7

DOI of this front matter: 10.5441/002/edbt.2015.01

Foreword

The 2015 International Conference on Extending Database Technology (EDBT) took place between the 23rd and the 27th of March in Brussels, Belgium. In its 18th edition, EDBT 2015 continued its long tradition of offering an outstanding research venue for the database community where to present and discuss recent contributions.

This year, there were 184 submissions to the research track, 19 to the industrial track, and 25 to the demo track. While these numbers are somewhat lower than in recent editions, the quality of the submissions was very high, which made the job of the Program Committee quite difficult. At the end, the program committee selected for publication 13 demos, 9 industrial papers, and 47 research papers. Of the latter, 5 of them were Vision Papers, shorter papers proposing radically new ideas, which were presented in their own session at the conference.

I would like to take this opportunity to thank all those that have made the 2015 EDBT edition such a success. First and foremost, all the authors of papers submitted to the conference, thereby providing the basis for a strong program, as well as the program committee members for their effort and dedication to study the submissions in detail and engaging in many interesting discussions about the papers, their contributions, their merits, and how to create the best possible program. Special mention should be made of the small committee in charge of deciding the Test of Time Award, in this occasion covering 4 editions of the conference (from 1988 to 1994): Martin Kersten, Christoph Koch, and Guido Moerkotte. They selected the following paper for the award:

Geo-Relational Algebra: A Model and Query Language for Geometric Database Systems, Ralf Hartmut Güting, University of Hagen, Germany, from EDBT 1988.

I also would like to thank Lucian Popa and Jens Teubner for the great work they have done in running the industrial and demo tracks. Martín Ugarte has done an excellent job with the proceedings, as have Pablo Barceló with the tutorials and Peter Fischer with the workshops. The Conference Chair—Floris Geerts—and the local organizers have also been instrumental in coordinating all the efforts of what is a very complex and demanding endeavor. As a joint EDBT/ICDT Conference, we have four keynote talks, two of which have been proposed by the EDBT community. Thanks go to Christoph Koch and Wolfgang Lehner for their acceptance of our invitation. Finally, Christine Collet and Norman Paton have been instrumental in coordinating the overall effort with the Executive Board of EDBT.

Gustavo Alonso
EDBT 2015 Program Chair

Program Committee Members

Ashraf Aboulmaga (U Waterloo)
Walid Aref (Purdue)
Roger Barga (Microsoft)
Spyros Blanas (Ohio State U)
Steven Blott (DCU)
Michael Böhlen (U Zurich)
Klemens Böhm (KIT)
Philippe Bonnet (ITU)
Luc Bouganim (INRIA)
Nieves Brisaboa (UDC)
Alejandro Buchmann (TU Darmstadt)
K. Selcuk Candan (ASU)
Michael Carey (UC Irvine)
Stefano Ceri (Politecnico Milano)
Lei Chen (HKUST)
Beng Chin Ooi (NUS)
Graham Cormode (U Warwick)
Isabel Cruz (UI Chicago)
Philippe Cudre-Mauroux (U Freiburg)
Sudipto Das (Microsoft)
Khuzaima Daudjee (U Waterloo)
Jens Dittrich (U Saarland)
Michael Duller (Oracle)
Wenfei Fan (U Edinburgh)
Alan Fekete (U Sydney)
Peter Fischer (Uni Freiburg)
Christoph Freytag (HU Berlin)
Johann Gamper (UNIBZ)
Minos Garofalakis (TU Crete)
Boris Glavic (IIT Illinois)
Jiawei Han (UI Urbana Champaign)
Thomas Heinis (Imperial College)
Melanie Herschel (INRIA)
Stratos Idreos (Harvard)
Arantza Illarramendi (UPV/EHU)
U Kang (KAIST)
Arijit Khan (ETH Zurich)
Nick Koudas (U Toronto)
Georg Lausen (Uni Freiburg)
Wolfgang Lehner (TU Dresden)
Hong-Va Leong (PolyU Hong Kong)
Justin Levandoski (Microsoft)
Roy Levin (IBM)
Ninghui Li (Purdue)
Feifei Li (U Utah)
Eric Lo (PolyU Hong Kong)
Guy Lohman (IBM)
Bertram Ludaescher (UC Davis)
Nikos Mamoulis (HKU Hong Kong)
Stefan Manegold (CWI)
Norman May (SAP)
Sharad Mehrotra (UC Irvine)
Rene Müller (IBM)
Thomas Neumann (TU München)
Esther Pacitti (U Montpellier 2)
Themis Palpanas (Paris Descartes)
Olga Papaemmanouil (Brandeis U)
Marta Patiño (Politécnica Madrid)
Torben B. Pedersen (U Aalborg)
Peter Pietzuch (Imperial College)
Evgelia Pitoura (U Ioannina)
Calton Pu (Georgia Tech)
Philippe Pucheral (INRIA)
Krithi Ramamrithan (IIT Bombay)
Maya Ramanath (IIT Delhi)
Rodolfo Resende (UFMG)
Tore Risch (Uppsala U)
Kenneth Ross (Columbia)
Pierangela Samarati (U Milan)
Marc H. Scholl (U Konstanz)
Heiko Schuldt (U Basel)
Assaf Schuster (Technion IIT)
Thomas Seidl (RWTH Aachen)
Timos Sellis (RMIT)
Liuba Shriram (Brandeis U)
Jianwen Su (UCSB)
Letizia Tanca (Politecnico Milano)
Ernest Teniente (U Poli de Catalunya)
Peter Triantafyllou (U Glasgow)
Anthony K.H. Tung (NUS)
Vasilis Vassalos (AUEB)
Marcos Vaz Salles (U Copenhagen)
Yannis Velegrakis (U Trento)
Gottfried Vossen (U Münster)
Kyu-Young Whang (KAIST)
Jef Wijsen (UMONS)
Yoshitaka Yamamoto (U Yamanashi)
Carlo Zaniolo (UCLA)
Wenjie Zhang (UNSW)
Esteban Zimanyi (ULB)

EDBT 2015 Test of Time Award

In 2014, EDBT began awarding the EDBT test-of-time (ToT) award, with the goal of recognizing one paper, or a small number of papers, presented at EDBT earlier and that have best met the “test of time”, i.e. that has had the most impact in terms of research, methodology, conceptual contribution, or transfer to practice over the past decade(s). The EDBT ToT award for 2015 will be presented during the EDBT/ICDT 2015 Joint Conference, March 23-27, in Brussels (Belgium). The EDBT 2015 Test of Time Award committee was formed by Martin Kersten (CWI, The Netherlands), Guido Moerkotte (Uni Mannheim, Germany), Christoph Koch (EPFL, Switzerland), all members of the EDBT 2015 PC and chaired by Gustavo Alonso, the EDBT 2015 PC chair.

The committee was charged with selecting a paper or a small number of papers from the proceedings of the following 4 editions: EDBT’88 - Venice, EDBT’90 - Venice, EDBT’92 - Vienna, EDBT’94 - Cambridge.

After careful consideration, the committee has decided to select the following paper, as the EDBT ToT Award winner for 2015:

Geo-Relational Algebra: A Model and Query Language for Geometric Database Systems

by Prof. Dr. Ralf Hartmut Güting, University of Hagen, Germany
Published in the EDBT 1988 proceedings, 506-527

The paper addresses the user’s conceptual model of a database system for geometric data. It proposes to extend relational database management systems by integrating geometry at all levels: At the conceptual level, relational algebra is extended to include geometric data types and operators. At the implementation level, the wealth of algorithms and data structures for geometric problems developed in the past decade in the field of Computational Geometry is exploited. The paper starts from a view of relational algebra as a many-sorted algebra which allows to easily embed geometric data types and operators. A concrete algebra for two-dimensional applications is developed. It can be used as a highly expressive retrieval and data manipulation language for geometric as well as standard data. Also, geo-relational database systems and their implementation strategy are discussed.

The committee members unanimously agreed that this paper clearly stands out in terms of relevance, impact, and influence in databases. Of all the papers considered, this is the one that had had the most and longest lasting impact with results that are still relevant today and whose influence can be traced to many real systems and a significant amount of follow up work.

The paper pioneered an important application area well before it became mainstream and did it in a systematic and clean way that has been very influential in both research and practice. Modern commercial systems all support geographic data types that are nowadays used in a wide range of applications and use cases (maps, locations based services, geographic information systems, mobility, etc.).

The selection committee also appreciated very much the cleanliness, completeness, insights, formalism, and systematic treatment of the problem as well as the approach followed by the author in selecting and solving a research problem.

Table of Contents

Foreword	i
Program Committee Members	ii
Test-of-Time Award	iii
Table of Contents	iv
Online Data Partitioning in Distributed Database Systems <i>Kaiji Chen, Yongluan Zhou, Yu Cao</i>	1
Chariots: A Scalable Shared Log for Data Management in Multi-Datacenter Cloud Environments <i>Faisal Nawab, Vaibhav Arora, Divyakant Agrawal, Amr El Abbadi</i>	13
Hermes: Dynamic Partitioning for Distributed Social Network Graph Databases <i>Daniel Nicoara, Shahin Kamali, Khuzaima Daudjee, Lei Chen</i>	25
On Debugging Non-Answers in Keyword Search Systems <i>Akanksha Baid, Wentao Wu, Chong Sun, AnHai Doan, Jeffrey F. Naughton</i>	37
Elevating Annotation Summaries To First-Class Citizens In InsightNotes <i>Karim Ibrahim, Dongqing Xiao, Mohamed Eltabakh</i>	49
Estimating Data Integration and Cleaning Effort <i>Sebastian Kruse, Paolo Papotti, Felix Naumann</i>	61
Mining Frequent Co-occurrence Patterns across Multiple Data Streams <i>Ziqiang Yu, Xiaohui Yu, Yang Liu, Wenzhu Li, Jian Pei</i>	73
Extracting Aggregate Answer Statistics for Integration <i>Zainab Zolaktaf, Jian Xu, Rachel Pottinger</i>	85
Discovering Recurring Patterns in Time Series <i>R. Uday Kiran, Haichuan Shang, Masashi Toyoda, Masaru Kitsuregawa</i>	97
Learning Path Queries on Graph Databases <i>Angela Bonifati, Radu Ciucanu, Aurélien Lemay</i>	109
TimeReach: Historical Reachability Queries on Evolving Graphs <i>Konstantinos Semertzidis, Evaggelia Pitoura, Kostas Lillis</i>	121
Efficiently Computing Top-K Shortest Path Join <i>Lijun Chang, Xuemin Lin, Lu Qin, Jeffrey Xu Yu, Jian Pei</i>	133
SIEF: Efficiently Answering Distance Queries for Failure Prone Graphs <i>Yongrui Qin, Quan Z. Sheng, Wei Emma Zhang</i>	145
A Selectivity based approach to Continuous Pattern Detection in Streaming Graphs <i>Sutanay Choudhury, Lawrence Holder, George Chin, Khushbu Agarwal, John Feo</i>	157
Scaling Unbound-Property Queries on Big RDF Data Warehouses using MapReduce <i>Padmashree Ravindra, Kemafor Anyanwu</i>	169
Reaching a desired set of users via different paths: an online advertising technique on micro-blogging platforms <i>Milad Eftekhari, Nick Koudas, Yashar Ganjali</i>	181
On Optimality of Jury Selection in Crowdsourcing <i>Yudian Zheng, Reynold Cheng, Silviu Maniu, Luyi Mo</i>	193

Finding the Most Diverse Products using Preference Queries <i>Orestis Gkorgkas, Akrivi Vlachou, Christos Doulkeridis, Kjetil Norvag</i>	205
Index Design for Enforcing Partial Referential Integrity Efficiently <i>Mozhgan Memari, Sebastian Link</i>	217
Query Optimization over Cloud Data Market <i>Yu Li, Eric Lo, Man Lung Yiu, Wenjian Xu</i>	229
Learning to Rank Adaptively for Scalable Information Extraction <i>Pablo Barrio, Gonçalo Simões, Helena Galhardas, Luis Gravano</i>	241
The Sweet Spot between Inverted Indices and Metric-Space Indexing for Top-K-List Similarity Search <i>Evica Milchevski, Avishek Anand, Sebastian Michel</i>	253
Optimizing Reformulation-based Query Answering in RDF <i>Damian Bursztyn, François Goasdoué, Ioana Manolescu</i>	265
Resolving XML Semantic Ambiguity <i>Nathalie Charbel, Joe Tekli, Richard Chbeir, Gilbert Tekli</i>	277
SpMacho - Optimizing Sparse Linear Algebra Expressions with Probabilistic Density Estimation <i>David Kernert, Frank Köhler, Wolfgang Lehner</i>	289
Efficient evaluation of threshold queries of derived fields in a numerical simulation database <i>Kalin Kanov, Randal Burns, Cristian C. Lalescu</i>	301
Identifying Converging Pairs of Nodes on a Budget <i>Konstantina Lazaridou, Konstantinos Semertzidis, Evaggelia Pitoura, Panayiotis Tsaparas</i>	313
Query-Based Outlier Detection in Heterogeneous Information Networks <i>Jonathan Kuck, Honglei Zhuang, Xifeng Yan, Hasan Cam, Jiawei Han</i>	325
Efficient caching for constrained skyline queries <i>Michael Lind Mortensen, Sean Chester, Ira Assent, Matteo Magnani</i>	337
Explanations for Skyline Query Results <i>Sean Chester, Ira Assent</i>	349
Efficient Processing of Hamming-Distance-Based Similarity-Search Queries Over MapReduce <i>Mingjie Tang, Yongyang Yu, Walid G. Aref, Qutaibah M. Malluhi, Mourad Ouzzani</i>	361
Joins for Hybrid Warehouses: Exploiting Massive Parallelism in Hadoop and Enterprise Data Warehouses <i>Yuanyuan Tian, Tao Zou, Fatma Ozcan, Romulo Goncalves, Hamid Pirahesh</i>	373
Benchmarking Smart Meter Data Analytics <i>Xiufeng Liu, Lukasz Golab, Wojciech Golab, Ihab F. Ilyas</i>	385
Crowd-Selection Query Processing in Crowdsourcing Databases: A Task-Driven Approach <i>Zhou Zhao, Furu Wei, Ming Zhou, Weikeng Chen, Wilfred Ng</i>	397
Dismantling Complicated Query Attributes with Crowd <i>Matan Laadan, Tova Milo</i>	409
Group Recommendation with Temporal Affinities <i>Sihem Amer-Yahia, Behrooz Omidvar-Tehrani, Senjuti Basu Roy, Nafiseh Shabib</i>	421
On Processing Top-k Spatio-Textual Preference Queries <i>George Tsatsanifos, Akrivi Vlachou</i>	433
Probabilistic Resource Route Queries with Reappearance <i>Gregor Jossé, Klaus Arthur Schmid, Matthias Schubert</i>	445

Cost Estimation of Spatial k-Nearest-Neighbor Operators <i>Ahmed M. Aly, Walid G. Aref, Mourad Ouzzani</i>	457
Reconstruction Privacy: Enabling Statistical Learning <i>Ke Wang, Chao Han, Ada Waichee Fu, Raymond Chi Wing Wong, Philip S. Yu</i>	469
Time series anomaly discovery with grammar-based compression <i>Pavel Senin, Jessica Lin, Xing Wang, Tim Oates, Sunil Gandhi, Arnold P. Boedihardjo, Crystal Chen, Susan Frankenstein</i>	481
K-Nearest Neighbor Temporal Aggregate Queries <i>Yu Sun, Jianzhong Qi, Yu Zheng, Rui Zhang</i>	493
Interactive Path Query Specification on Graph Databases <i>Angela Bonifati, Radu Ciucanu, Aurélien Lemay</i>	505
Flexible Analysis of Plant Genomes in a Database Management System <i>Sebastian Dorok, Sebastian Brefß, Jens Teubner, Gunter Saake</i>	509
Demonstrating Transfer-Efficient Sample Maintenance on Graphics Cards <i>Max Heimel, Martin Kiefer, Volker Markl</i>	513
NoFTL for Real: Databases on Real Native Flash Storage <i>Sergey Hardock, Iliia Petrov, Robert Gottstein, Alejandro Buchmann</i>	517
Gumbo: Guarded Fragment Queries over Big Data <i>Jonny Daenen, Frank Neven, Tony Tan</i>	521
WAVEGUIDE: Evaluating SPARQL Property Path Queries <i>Nikolay Yakovets, Parke Godfrey, Jarek Gryz</i>	525
Meta-Stars: Dynamic, Schemaless, and Semantically-Rich Topic Hierarchies in Social BI <i>Enrico Gallinucci, Matteo Golfarelli, Stefano Rizzi</i>	529
“I would like to watch something like ‘The Terminator’...” Cooperative Query Personalization Based on Perceptual Similarity <i>Christoph Lofi, Christian Nieke</i>	533
Liquid Benchmarking: A Platform for Democratizing the Performance Evaluation Process <i>Sherif Sakr, Amin Shafaat, Fuad Bajaber, Ahmed Barnawi, Omar Batarfi, Abdulrahman Altalhi</i>	537
Natural Language Specification and Violation Reporting of Business Rules over ER-modeled Databases <i>Michael Minock, Daniel Oskarsson, Björn Pelzer, Mika Cohen</i>	541
POIESIS: a Tool for Quality-aware ETL Process Redesign <i>Vasileios Theodorou, Alberto Abelló, Maik Thiele, Wolfgang Lehner</i>	545
Quarry: Digging Up the Gems of Your Data Treasury <i>Petar Jovanovic, Oscar Romero, Alkis Simitsis, Alberto Abelló, Héctor Candón, Sergi Nadal</i>	549
QaRS: A User-Friendly Graphical Tool for Semantic Query Design and Relaxation <i>Géraud Fokou, Stephane Jean, Allel Hadjali, Mickaël Baron</i>	553
Using Object-Awareness to Optimize Join Processing in the SAP HANA Aggregate Cache <i>Stephan Müller, Anisoara Nica, Lars Butzmann, Stefan Klauck, Hasso Plattner</i>	557
Transactional Replication in Hybrid Data Store Architectures <i>Hojjat Jafarpour, Junichi Tatemura, Hakan Hacigumus</i>	569
SAP HANA – From Relational OLAP Database to Big Data Infrastructure <i>Norman May, Wolfgang Lehner, Shahul Hameed P., Nitesh Maheshwari, Carsten Müller, Sudipto Chowdhuri, Anil Goel</i>	581

Taxi Queue, Passenger Queue or No Queue? - A Queue Detection and Analysis System using Taxi State Transition <i>YU LU, Shili Xiang, Wei Wu</i>	593
Data Ingestion in AsterixDB <i>Raman Grover, Michael J. Carey</i>	605
The NPD Benchmark: Reality Check for OBDA Systems <i>Davide Lanti, Martin Rezk, Guohui Xiao, Diego Calvanese</i>	617
Event Recognition for Maritime Surveillance <i>Kostas Patroumpas, Alexander Artikis, Nikos Katzouris, Marios Vodas, Yannis Theodoridis, Nikos Pelekis</i>	629
Identifying User Interests within the Data Space - a Case Study with SkyServer <i>Hoang Vu Nguyen, Klemens Boehm, Florian Becker, Bertrand Goldman, Georg Hinkel, Emmanuel Müller</i>	641
Insights on a Scalable and Dynamic Traffic Management System <i>Nikolas Zygouras, Nikos Zacheilas, Vana Kalogeraki, Dermot Kinane, Dimitrios Gunopulos</i>	653
Reconsolidating Data Structures <i>Thomas Heinis, Anastasia Ailamaki</i>	665
A Generic Solution to Integrate SQL and Analytics for Big Data <i>Nick R. Katsipoulakis, Yuanyuan Tian, Fatma Ozcan, Hamid Pirahesh, Berthold Reinwald</i>	671
ECCO- A Framework for Ecological Data Collection and Management Involving Human Workers <i>Senjuti Basu Roy, Sihem Amer-Yahia, Lucas Joppa</i>	677
ligDB—Online Query Processing Without (almost) any Storage <i>Evica Milchevski, Sebastian Michel</i>	683
Procrastination Beats Prevention: Timely Sufficient Persistence for Efficient Crash Resilience <i>Faisal Nawab, Dhruva R. Charkrabarti, Terence Kelly, Charles B. Morrey III</i>	689