18th International Conference on Database Theory

ICDT'15, March 23–27, 2015, Brussels, Belgium

Edited by

Marcelo Arenas

Martín Ugarte

LIPIcs – Vol. 31 – ICDT'15 www.dagstuhl.de/lipics

Editors

Marcelo Arenas
Pontificia Universidad Católica de Chile
Santiago, Chile
marenas@ing.puc.cl

Martín Ugarte
Pontificia Universidad Católica de Chile
Santiago, Chile
martinugarte@puc.cl

ACM Classification 1998
H.2: Database Management, H.2.1 Normal forms, H.2.2 Schema and subschema, H.2.3 Query languages, H.2.4 Query processing, H.2.4 Relational databases, H.2.4 Distributed databases, H.2.5 Heterogeneous Databases, H.3.5 Online Information Services, H.1: Miscellaneous – Privacy, H.4.1 Office Automation: Workflow management, B.4.4 Performance Analysis and Design Aids: Formal models, Verification, F.1.3 Complexity measures and classes, F.4.1 Computational Logic, Model Theory, G.2.2 Graph Theory – Hypergraphs

ISBN 978-3-939897-79-8

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at http://www.dagstuhl.de/dagpub/978-3-939897-79-8.

Publication date
March, 2015

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

License
This work is licensed under a Creative Commons Attribution 3.0 Unported license (CC-BY 3.0): http://creativecommons.org/licenses/by/3.0/legalcode.

In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Digital Object Identifier: 10.4230/LIPIcs.ICDT.2015.1

ISBN 978-3-939897-79-8 **ISSN 1868-8969** **http://www.dagstuhl.de/lipics**

LIPIcs – Leibniz International Proceedings in Informatics

LIPIcs is a series of high-quality conference proceedings across all fields in informatics. LIPIcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

Editorial Board

- Susanne Albers (TU München)
- Chris Hankin (Imperial College London)
- Deepak Kapur (University of New Mexico)
- Michael Mitzenmacher (Harvard University)
- Madhavan Mukund (Chennai Mathematical Institute)
- Catuscia Palamidessi (INRIA)
- Wolfgang Thomas (RWTH Aachen)
- Pascal Weil (*Chair*, CNRS and University Bordeaux)
- Reinhard Wilhelm (Saarland University)

ISSN 1868-8969

http://www.dagstuhl.de/lipics

Contents

Preface vii

ICDT 2015 Test of Time Award ix

Organization xi

External Reviewers xiii

List of Authors xv

Invited Talks

The Confounding Problem of Private Data Release
Graham Cormode 1

Using Locality for Efficient Query Evaluation in Various Computation Models
Nicole Schweikardt 13

Large-Scale Similarity Joins With Guarantees
Rasmus Pagh 15

Awards Session

A Declarative Framework for Linking Entities
Douglas Burdick, Ronald Fagin, Phokion G. Kolaitis, Lucian Popa, and Wang-Chiew Tan 25

Asymptotic Determinacy of Path Queries using Union-of-Paths Views
Nadime Francis 44

(Regular Paper)
Games for Active XML Revisited
Martin Schuster and Thomas Schwentick 60

Query Evaluation

Answering Conjunctive Queries with Inequalities
Paraschos Koutris, Tova Milo, Sudeepa Roy, and Dan Suciu 76

SQL's Three-Valued Logic and Certain Answers
Leonid Libkin 94

A Trichotomy in the Complexity of Counting Answers to Conjunctive Queries
Hubie Chen and Stefan Mengel 110

Data Examples and Learning

Learning Tree Patterns from Example Graphs
Sara Cohen and Yaacov Y. Weiss 127

18th International Conference on Database Theory (ICDT'15).
Editors: Marcelo Arenas and Martín Ugarte
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Characterizing XML Twig Queries with Examples
Sławek Staworko and Piotr Wieczorek 144

The Product Homomorphism Problem and Applications
Balder ten Cate and Victor Dalmau 161

Graph Databases and Semantic Web

Regular Queries on Graph Databases
Juan L. Reutter, Miguel Romero, and Moshe Y. Vardi 177

Complexity and Expressiveness of ShEx for RDF
Sławek Staworko, Iovka Boneva, Jose E. Labra Gayo, Samuel Hym, Eric G. Prud'hommeaux, and Harold Solbrig 195

CONSTRUCT Queries in SPARQL
Egor V. Kostylev, Juan L. Reutter, and Martín Ugarte 212

Separability by Short Subsequences and Subwords
Piotr Hofman and Wim Martens 230

Algorithms and Workflows

Process-Centric Views of Data-Driven Business Artifacts
Adrien Koutsos and Victor Vianu 247

On The I/O Complexity of Dynamic Distinct Counting
Xiaocheng Hu, Yufei Tao, Yi Yang, Shengyu Zhang, and Shuigeng Zhou 265

Shared-Constraint Range Reporting
Sudip Biswas, Manish Patil, Rahul Shah, and Sharma V. Thankachan 277

Distributed Query Processing

Optimal Broadcasting Strategies for Conjunctive Queries over Distributed Data
Bas Ketsman and Frank Neven 291

Datalog Queries Distributing over Components
Tom Ameloot, Bas Ketsman, Frank Neven, and Daniel Zinn 308

Distributed Streaming with Finite Memory
Frank Neven, Nicole Schweikardt, Frédéric Servais, and Tony Tan 324

Consistency and Repairs

From Causes for Database Queries to Repairs and Model-Based Diagnosis and Back
Babak Salimi and Leopoldo Bertossi 342

On the Relationship between Consistent Query Answering and Constraint Satisfaction Problems
Carsten Lutz and Frank Wolter 363

On the Data Complexity of Consistent Query Answering over Graph Databases
Pablo Barceló and Gaëlle Fontaine 380

Preface

The 18th International Conference on Database Theory (ICDT 2015) was held in Brussels, Belgium, March 23–27, 2015. Originally biennial, the ICDT conference has been held annually and jointly with EDBT ("Extending Database Technology") since 2009.

The proceedings of ICDT 2015 include a paper by Graham Cormode (University of Warwick) based on the keynote address by him, a paper by Rasmus Pagh (IT University of Copenhagen) based on the keynote address by him, an overview of an invited lecture by Nicole Schweikardt (Humboldt University of Berlin), a laudation concerning the ICDT 2015 Test of Time Award, and 22 research papers that were selected by the Program Committee from 50 submissions.

Out of the 22 accepted papers, the Program Committee selected the paper *A Declarative Framework for Linking Entities* by Douglas Burdick, Ronald Fagin, Phokion G. Kolaitis, Lucian Popa and Wang-Chiew Tan for the ICDT 2015 Best Paper Award. Furthermore, the Program Committee selected the paper *Asymptotic Determinacy of Path Queries using Union-of-Paths Views* by Nadime Francis for the ICDT 2015 Best Student Paper Award. The ICDT 2015 Test of Time Award is given to the paper *Efficient Computation of Frequent and Top-k Elements in Data Streams* by Ahmed Metwally, Divyakant Agrawal, and Amr El Abbadi, which originally appeared in the proceedings of ICDT 2005. Warmest congratulations to the authors of these award winning papers!

I thank all authors who submitted papers to ICDT 2015. I would also like to thank all members of the Program Committee, and the external reviewers, for the enormous amount of work they have done. The Program Committee did not meet in person, but carried out extensive discussions during the electronic PC meeting. I thank Andrei Voronkov for his EasyChair system, which made it easy to manage and coordinate the discussion.

I thank the ICDT Council members for their help in selecting the Program Committee and their advice on issues of policy during the conference. Special thanks also go to Martín Ugarte, the Proceedings Chair of EDBT/ICDT 2015. I thank many colleagues involved in the organisation of the conference for fruitful collaboration, in particular, Floris Geerts (EDBT/ICDT 2015 Conference Chair).

Marcelo Arenas
January 2015

18th International Conference on Database Theory (ICDT'15).
Editors: Marcelo Arenas and Martín Ugarte
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

ICDT 2015 Test of Time Award

In 2013, the International Conference on Database Theory (ICDT) began awarding the ICDT test-of-time award, with the goal of recognising one paper, or a small number of papers, presented at ICDT a decade earlier that have best met the "test of time". In 2015, the award will recognise a paper from the ICDT 2005 proceedings that has had the most impact in terms of research, methodology, conceptual contribution, or transfer to practise over the past decade. The award will be presented during the EDBT/ICDT 2015 Joint Conference, March 23–27, 2015 in Brussels, Belgium.

The committee consisting of Serge Abiteboul, Sudeepa Roy, and chaired by Leonid Libkin (2005 PC co-chair) has chosen the following recipient of the 2015 ICDT Test of Time Award:

"Efficient Computation of Frequent and Top-k Elements in Data Streams"

by Ahmed Metwally, Divyakant Agrawal, and Amr El Abbadi

The paper studies the problem of finding items which occur most frequently in a data stream. This is a basic algorithmic problem of great practical importance. The paper proposed a SpaceSaving algorithm which was shown to both provide theoretical guarantees and to perform significantly better than others in practical scenarios. Since its publication, the paper has made impact in both algorithms research and practical implementations of streaming algorithms. Several implementations of the algorithm have been made available, and they are used both in industry and as benchmarks to compare against other streaming algorithms. The paper has been highly cited: many papers have made use of the data structure, either directly, or to solve new problems. The algorithm itself is easy to motivate and state, and consequently it is taught in a number of algorithms courses.

18th International Conference on Database Theory (ICDT'15).
Editors: Marcelo Arenas and Martín Ugarte

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Organization

Conference Chair

Floris Geerts (U. of Antwerp)

Program Chair

Marcelo Arenas (PUC Chile)

Program Committee

Marcelo Arenas (PUC Chile)
Pankaj Agarwal (Duke U.)
Angela Bonifati (U. of Lille 1 & Inria)
Edith Cohen (Microsoft Research)
Giuseppe De Giacomo (Sapienza U. di Roma)
Daniel Deutch (Tel Aviv U.)
Gaelle Fontaine (U. of Chile)
Todd Green (LogicBlox & UC Davis)
Sebastian Maneth (U. of Edinburgh)
Filip Murlak (U. of Warsaw)
S Muthukrishnan (Rutgers U.)
Reinhard Pichler (Vienna U. of Technology)
Christopher Re (Stanford U.)
Cristian Riveros (PUC Chile)
Sudeepa Roy (U. of Washington)
Cristina Sirangelo (LSV, ENS-Cachan)
Yufei Tao (Chinese U. of Hong Kong)
Balder Ten Cate (LogicBlox & UC Santa Cruz)
Jan Van Den Bussche (Hasselt U.)
Stijn Vansummeren (U. Libre de Bruxelles)
Victor Vianu (UC San Diego)
David Woodruff (IBM Almaden)

Proceedings Chair

Martín Ugarte (PUC Chile)

18th International Conference on Database Theory (ICDT'15).
Editors: Marcelo Arenas and Martín Ugarte

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

External Reviewers

Foto Afrati
Peter Alvaro
Tom Ameloot
Guillaume Bagan
Pablo Barceló
Manuel Bodirsky
Iovka Boneva
Pierre Bourhis
Paolo Ciaccia
Radu Ciucanu
Claire David
Alin Deutsch
Esther Ezra
Wenfei Fan
George H. L. Fletcher
Olivier Gauwin
Sam Haney
Piotr Hofman
Hossein Jowhari
Phokion G. Kolaitis
Paraschos Koutris
Kasper Larsen
Domenico Lembo
Maurizio Lenzerini
Jerry Li
Leonid Libkin
Katja Losemann
Wim Martens
Dániel Marx
Filip Mazowiecki
Andrew McGregor
Marco Montali
Jelani Nelson
Frank Neven
Jorge Pérez
Andreas Pieris
Chung Keung Poon
Miguel Romero
Riccardo Rosati
Emanuel Sallinger
Vadim Savenkov
Evgeny Sherkhonov
Mantas Simkus
Sebastian Skritek
Wang-Chiew Tan
Srikanta Tirthapura
Sophie Tison
Domagoj Vrgoc
Johannes Wallner
Adam Witkowski
Peter Wood
Zhilin Wu
You Wu
Xiaokui Xiao
Ke Yi
Marc Zeitoun
Thomas Zeume
Wuzhou Zhang

18th International Conference on Database Theory (ICDT'15).
Editors: Marcelo Arenas and Martín Ugarte
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

List of Authors

Tom Ameloot
Pablo Barceló
Leopoldo Bertossi
Sudip Biswas
Iovka Boneva
Douglas Burdick
Balder ten Cate
Hubie Chen
Sara Cohen
Graham Cormode
Victor Dalmau
Ronald Fagin
Gaelle Fontaine
Nadime Francis
Piotr Hofman
Xiaocheng Hu
Samuel Hym
Bas Ketsman
Phokion G. Kolaitis
Egor V. Kostylev
Paraschos Koutris
Adrien Koutsos
Jose Emilio Labra Gayo
Leonid Libkin
Carsten Lutz
Wim Martens
Stefan Mengel
Tova Milo
Frank Neven
Rasmus Pagh
Manish Patil
Lucian Popa
Eric Prud'Hommeaux
Juan L. Reutter
Miguel Romero
Sudeepa Roy
Babak Salimi
Martin Schuster
Nicole Schweikardt
Thomas Schwentick
Frédéric Servais
Rahul Shah
Harold Solbrig
Sławek Staworko
Dan Suciu
Tony Tan
Wang-Chiew Tan
Yufei Tao
Sharma Thankachan
Martín Ugarte
Moshe Y. Vardi
Victor Vianu
Yaacov Y. Weiss
Piotr Wieczorek
Frank Wolter
Yi Yang
Shengyu Zhang
Shuigeng Zhou
Daniel Zinn

18th International Conference on Database Theory (ICDT'15).
Editors: Marcelo Arenas and Martín Ugarte

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany