25th International Conference on Database Theory

ICDT 2022, March 29–April 1, 2022, Edinburgh, UK
(Virtual Conference)

Edited by

Dan Olteanu

Nils Vortmeier

Editors

Dan Olteanu
University of Zurich, Switzerland
dan.olteanu@uzh.ch

Nils Vortmeier
University of Zurich, Switzerland
nils.vortmeier@uzh.ch

ACM Classification 2012
Information systems → Data management systems; Information systems → Database design and models; Information systems → Database query processing; Information systems → Database transaction processing; Information systems → Join algorithms; Information systems → Query languages; Information systems → Query optimization; Information systems → Query planning; Information systems → Relational database model; Information systems → Parallel and distributed DBMSs; Information systems → Semantic web description languages; Information systems → Stream management; Theory of computation → Approximation algorithms analysis; Theory of computation → Complexity theory and logic; Theory of computation → Constraint and logic programming; Theory of computation → Incomplete, inconsistent, and uncertain databases; Theory of computation → Database theory; Theory of computation → Data structures and algorithms for data management; Theory of computation → Description logics; Theory of computation → Finite Model Theory; Theory of computation → Logic and databases; Theory of computation → Problems, reductions and completeness; Theory of computation → Streaming, sublinear and near linear time algorithms; Mathematics of computing → Random number generation

ISBN 978-3-95977-223-5

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at https://www.dagstuhl.de/dagpub/978-3-95977-223-5.

Publication date
March, 2022

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 https://portal.dnb.de.

Digital Object Identifier: 10.4230/LIPIcs.ICDT.2022.0
ISBN 978-3-95977-223-5 **ISSN 1868-8969** **https://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.

ISSN 1868-8969

https://www.dagstuhl.de/lipics

Contents

Preface
Dan Olteanu and Nils Vortmeier 0:vii

Organization
..... 0:ix

External Reviewers
..... 0:xi

Contributing Authors
..... 0:xiii

The ICDT 2022 Test-of-Time Award
..... 0:xv

Invited Talks

On an Information Theoretic Approach to Cardinality Estimation
Hung Q. Ngo 1:1–1:21

Counting the Solutions to a Query
Marcelo Arenas 2:1–2:1

Answering Unions of Conjunctive Queries with Ideal Time Guarantees
Nofar Carmeli 3:1–3:1

Regular Papers

On the Hardness of Category Tree Construction
Shay Gershtein, Uri Avron, Ido Guy, Tova Milo, and Slava Novgorodov 4:1–4:17

Linear Programs with Conjunctive Queries
Florent Capelli, Nicolas Crosetti, Joachim Niehren, and Jan Ramon 5:1–5:19

Certifiable Robustness for Nearest Neighbor Classifiers
Austen Z. Fan and Paraschos Koutris 6:1–6:20

Improved Approximation and Scalability for Fair Max-Min Diversification
Raghavendra Addanki, Andrew McGregor, Alexandra Meliou, and Zafeiria Moumoulidou 7:1–7:21

Rewriting with Acyclic Queries: Mind Your Head
Gaetano Geck, Jens Keppeler, Thomas Schwentick, and Christopher Spinrath 8:1–8:20

Parallel Acyclic Joins with Canonical Edge Covers
Yufei Tao 9:1–9:19

Splitting Spanner Atoms: A Tool for Acyclic Core Spanners
Dominik D. Freydenberger and Sam M. Thompson 10:1–10:18

Practical Relational Calculus Query Evaluation
Martin Raszyk, David Basin, Srđan Krstić, and Dmitriy Traytel 11:1–11:21

25th International Conference on Database Theory (ICDT 2022).
Editors: Dan Olteanu and Nils Vortmeier
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Characterising Fixed Parameter Tractability for Query Evaluation over Guarded TGDs
Cristina Feier 12:1–12:20

Tuple-Generating Dependencies Capture Complex Values
Maximilian Marx and Markus Krötzsch 13:1–13:20

Inference of Shape Graphs for Graph Databases
Benoît Groz, Aurélien Lemay, Sławek Staworko, and Piotr Wieczorek 14:1–14:20

Expressiveness of SHACL Features
Bart Bogaerts, Maxime Jakubowski, and Jan Van den Bussche 15:1–15:16

Robustness Against Read Committed for Transaction Templates with Functional Constraints
Brecht Vandevoort, Bas Ketsman, Christoph Koch, and Frank Neven 16:1–16:17

A Dyadic Simulation Approach to Efficient Range-Summability
Jingfan Meng, Huayi Wang, Jun Xu, and Mitsunori Ogihara 17:1–17:18

Discovering Event Queries from Traces: Laying Foundations for Subsequence-Queries with Wildcards and Gap-Size Constraints
Sarah Kleest-Meißner, Rebecca Sattler, Markus L. Schmid, Nicole Schweikardt, and Matthias Weidlich 18:1–18:21

Streaming Enumeration on Nested Documents
Martín Muñoz and Cristian Riveros 19:1–19:18

Preface

The 25. International Conference on Database Theory (ICDT 2022) was held from March 29 to April 1, 2022, as an online event.

The Program Committee has selected 16 research papers out of 41 submissions for publication at the conference. It has further decided to give the Best Paper Award to *On the Hardness of Category Tree Construction* by Shay Gershtein, Uri Avron, Ido Guy, Tova Milo, and Slava Novgorodov, and the Best Newcomer Paper Award to *Linear programs with conjunctive queries* by Florent Capelli, Nicolas Crosetti, Joachim Niehren, and Jan Ramon. We congratulate the winners!

Apart from the 16 regular papers, these proceedings include abstracts for the invited (shared) EDBT/ICDT keynote by Marcelo Arenas (Pontificia Universidad Católica de Chile) and for the ICDT invited tutorial by Nofar Carmeli (École Normale Supérieure Paris), as well as the invited paper associated with the (shared) EDBT/ICDT keynote by Hung Ngo (RelationalAI Inc.).

A committee formed by Antoine Amarilli, Alin Deutsch, and Emanuel Sallinger has decided to give the Test-of-Time Award for ICDT 2022 to the ICDT 2012 paper *Factorised Representations of Query Results: Size Bounds and Readability* by Dan Olteanu and Jakub Závodný.

We would like to thank all people who contributed to the success of ICDT 2022, including the authors of all submitted papers, keynote and invited tutorial speakers, and, of course, all members of the Program Committee as well as the external reviewers, for the very substantial work that they have invested over the two submission cycles of ICDT 2022. Their commitment and sagacity were crucial to ensure that the final program of the conference satisfies the highest standards. We would also like to thank the ICDT Council members for their support on a wide variety of matters, and the local organizers of the EDBT/ICDT 2022 conference, led by General Chairs Paolo Guagliardo, Milos Nikolic, and Andreas Pieris, for the great job they did in organizing the conference and co-located events. Finally, we wish to acknowledge Dagstuhl Publishing for their support with the publication of the proceedings in the LIPIcs (Leibniz International Proceedings in Informatics) series.

Dan Olteanu and Nils Vortmeier
March 2022

25th International Conference on Database Theory (ICDT 2022).
Editors: Dan Olteanu and Nils Vortmeier
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Organization

General Chairs

Paolo Guagliardo (University of Edinburgh)

Milos Nikolic (University of Edinburgh)

Andreas Pieris (University of Edinburgh)

Program Chair

Dan Olteanu (University of Zurich)

Program Committee

Mahmoud Abo Khamis (RelationalAI Inc.)

Suman Bera (UC Santa Cruz)

Nofar Carmeli (École Normale Supérieure Paris)

Ismail Ilkan Ceylan (University of Oxford)

Daniel Deutch (Tel Aviv University)

Cristina Feier (Bremen University)

Xiao Hu (Duke University)

Ahmet Kara (University of Zurich)

Egor Kostylev (University of Oslo)

Andrew McGregor (University of Massachusetts, Amherst)

Mikaël Monet (Inria Lille)

Hung Ngo (RelationalAI Inc.)

Liat Peterfreund (École Normale Supérieure Paris)

Reinhard Pichler (Vienna University of Technology)

Cristian Riveros (Pontificia Universidad Católica de Chile)

Thomas Schwentick (TU Dortmund University)

Dan Suciu (University of Washington)

Yufei Tao (The Chinese University of Hong Kong)

Szymon Toruńczyk (University of Warsaw)

Proceedings Chair

Nils Vortmeier (University of Zurich)

25th International Conference on Database Theory (ICDT 2022).
Editors: Dan Olteanu and Nils Vortmeier
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

External Reviewers

Sabyasachi Basu

Michael Benedikt

Arnaud Durand

Antonia Kormpa

Victor Marsault

Stefan Mengel

Frank Neven

Matthias Niewerth

Milos Nikolic

Florin Rusu

Luc Segoufin

Christopher Spinrath

Sławek Staworko

Nils Vortmeier

Domagoj Vrgoč

Chenghong Wang

Yisu Wang

25th International Conference on Database Theory (ICDT 2022).
Editors: Dan Olteanu and Nils Vortmeier
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Contributing Authors

Raghavendra Addanki

Marcelo Arenas

Uri Avron

David Basin

Bart Bogaerts

Florent Capelli

Nofar Carmeli

Nicolas Crosetti

Austen Z. Fan

Cristina Feier

Dominik D. Freydenberger

Gaetano Geck

Shay Gershtein

Benoît Groz

Ido Guy

Maxime Jakubowski

Jens Keppeler

Bas Ketsman

Sarah Kleest-Meißner

Christoph Koch

Paraschos Koutris

Markus Krötzsch

Srđan Krstić

Aurélien Lemay

Maximilian Marx

Andrew McGregor

Alexandra Meliou

Jingfan Meng

Tova Milo

Zafeiria Moumoulidou

Martín Muñoz

Frank Neven

Hung Q. Ngo

Joachim Niehren

Slava Novgorodov

Mitsunori Ogihara

Jan Ramon

Martin Raszyk

Cristian Riveros

Rebecca Sattler

Markus L. Schmid

Nicole Schweikardt

Thomas Schwentick

Christopher Spinrath

Sławek Staworko

Yufei Tao

Sam M. Thompson

Dmitriy Traytel

Jan Van den Bussche

Brecht Vandevoort

Huayi Wang

Matthias Weidlich

Piotr Wieczorek

Jun Xu

25th International Conference on Database Theory (ICDT 2022).
Editors: Dan Olteanu and Nils Vortmeier
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The ICDT 2022 Test-of-Time Award

In 2013, the International Conference on Database Theory (ICDT) began awarding the ICDT Test-of-Time (ToT) award, with the goal of recognizing one paper, or a small number of papers, presented at earlier ICDT conferences that have best met the "test of time". In 2022, the award recognizes a paper selected from the proceedings of the ICDT 2012 conference that has had the highest impact in terms of research, methodology, conceptual contribution, or transfer to practice over the past decade. The award was presented during the EDBT/ICDT 2022 Joint Conference, March 29 – April 1st, 2022.

The 2022 ToT Committee consists of Antoine Amarilli, Alin Deutsch, and Emanuel Sallinger. After careful consideration and soliciting external assessments, the committee has chosen the following contribution for the 2022 ICDT Test-of-Time Award:

Factorised Representations of Query Results: Size Bounds and Readability
Dan Olteanu and Jakub Závodný

This paper introduced the fundamental concept of *factorized data representations*, subsequently studied as *factorized databases* in database theory. Factorized representations avoid redundancy in relations by building them up from singleton tuples using the union and product operators. The work by Olteanu and Závodný gave a formal definition of these factorized representations and showed their numerous benefits: they are more succinct than flat relations, can be used to evaluate queries more efficiently, and their tuples can be enumerated with linear-time preprocessing and constant delay. Their work further shows size bounds on factorized representations of query results depending on graph-theoretic parameters of the query, and also considers many extensions, such as factorized provenance representations, connections to readability width, and aggregation over factorized representations.

Since then, this concept of factorized databases has had profound impact on several areas of database theory, database systems research, and neighboring areas. Applications include query evaluation over graph databases, improved enumeration results, factorized computation of aggregates, and factorized machine learning. Their work has successfully bridged the gap between theory and practice: it has prompted implementations of its core ideas in subsequent works, and it has also sparked practical research across several independent areas.

Factorized databases are acknowledged as an inspiration within many lines of practical and theoretical research, including the work on aggregate queries by Abo Khamis et al. (PODS 2016 best paper award), the work on SPARQL by Abul-Basher et al. (EDBT 2021 best short paper award), and many others. This conclusively demonstrates the lasting influence of this highly cited paper and of its full version titled "Size Bounds for Factorised Representations of Query Results" which later appeared in *ACM Transactions on Database Systems*.

Antoine Amarilli, Télécom Paris
Alin Deutsch, UC San Diego
Emanuel Sallinger, TU Wien and Oxford

The ICDT Test-of-Time Award Committee for 2022

25th International Conference on Database Theory (ICDT 2022).
Editors: Dan Olteanu and Nils Vortmeier
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany