

# 24th International Conference on Database Theory

ICDT 2021, March 23–26, 2021, Nicosia, Cyprus

Edited by

Ke Yi

Zhewei Wei



*Editors*

**Ke Yi** 

The Hong Kong University of Science and Technology, Hong Kong  
yike@ust.hk

**Zhewei Wei** 

Renmin University of China, China  
zhewei@ruc.edu.cn

*ACM Classification 2012*

Information systems → Data management systems; Information systems → Database design and models; Information systems → Database query processing; Information systems → Query languages; Information systems → Relational database model; Information systems → Parallel and distributed DBMSs; Information systems → Information integration; Information systems → Stream management; Theory of computation → Incomplete, inconsistent, and uncertain databases; Theory of computation → Complexity theory and logic; Theory of computation → Database theory

**ISBN 978-3-95977-179-5**

*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/978-3-95977-179-5>.

*Publication date*

March, 2021

*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>.

*License*

This work is licensed under a Creative Commons Attribution 4.0 International license (CC-BY 4.0):

<https://creativecommons.org/licenses/by/4.0/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.2021.0

**ISBN 978-3-95977-179-5**

**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.

### *Editorial Board*

- Luca Aceto (*Chair*, Gran Sasso Science Institute and Reykjavik University)
- Christel Baier (TU Dresden)
- Mikolaj Bojanczyk (University of Warsaw)
- Roberto Di Cosmo (INRIA and University Paris Diderot)
- Javier Esparza (TU München)
- Meena Mahajan (Institute of Mathematical Sciences)
- Dieter van Melkebeek (University of Wisconsin-Madison)
- Anca Muscholl (University Bordeaux)
- Luke Ong (University of Oxford)
- Catuscia Palamidessi (INRIA)
- Thomas Schwentick (TU Dortmund)
- Raimund Seidel (Saarland University and Schloss Dagstuhl – Leibniz-Zentrum für Informatik)

**ISSN 1868-8969**

**<https://www.dagstuhl.de/lipics>**



## ■ Contents

Preface	
<i>Ke Yi and Zhewei Wei</i> .....	0:vii
Organization	
.....	0:ix
External Reviewers	
.....	0:xi
Authors	
.....	0:xiii
ICDT 2021 Test of Time Award	
.....	0:xv

### Invited Talks

Explainability Queries for ML Models and its Connections with Data Management Problems	
<i>Pablo Barceló</i> .....	1:1–1:1
Comparing Apples and Oranges: Fairness and Diversity in Ranking	
<i>Julia Stoyanovich</i> .....	2:1–2:1

### Regular Papers

Box Covers and Domain Orderings for Beyond Worst-Case Join Processing	
<i>Kaleb Alway, Eric Blais, and Semih Salihoglu</i> .....	3:1–3:23
A Purely Regular Approach to Non-Regular Core Spanners	
<i>Markus L. Schmid and Nicole Schweikardt</i> .....	4:1–4:19
Ranked Enumeration of Conjunctive Query Results	
<i>Shaleen Deep and Paraschos Koutris</i> .....	5:1–5:19
Towards Optimal Dynamic Indexes for Approximate (and Exact) Triangle Counting	
<i>Shangqi Lu and Yufei Tao</i> .....	6:1–6:23
Grammars for Document Spanners	
<i>Liat Peterfreund</i> .....	7:1–7:18
Input–Output Disjointness for Forward Expressions in the Logic of Information Flows	
<i>Heba Aamer and Jan Van den Bussche</i> .....	8:1–8:18
Conjunctive Queries: Unique Characterizations and Exact Learnability	
<i>Balder ten Cate and Victor Dalmau</i> .....	9:1–9:24
The Complexity of Aggregates over Extractions by Regular Expressions	
<i>Johannes Doleschal, Noa Bratman, Benny Kimelfeld, and Wim Martens</i> .....	10:1–10:20

24th International Conference on Database Theory (ICDT 2021).

Editors: Ke Yi and Zhewei Wei



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

Answer Counting Under Guarded TGDs <i>Cristina Feier, Carsten Lutz, and Marcin Przybyłko</i> .....	11:1–11:22
Maximum Coverage in the Data Stream Model: Parameterized and Generalized <i>Andrew McGregor, David Tench, and Hoa T. Vu</i> .....	12:1–12:20
Diverse Data Selection under Fairness Constraints <i>Zafeiria Mousoulidou, Andrew McGregor, and Alexandra Meliou</i> .....	13:1–13:25
Enumeration Algorithms for Conjunctive Queries with Projection <i>Shaleen Deep, Xiao Hu, and Paraschos Koutris</i> .....	14:1–14:17
The Shapley Value of Inconsistency Measures for Functional Dependencies <i>Ester Livshits and Benny Kimelfeld</i> .....	15:1–15:19
Database Repairing with Soft Functional Dependencies <i>Nofar Carmeli, Martin Grohe, Benny Kimelfeld, Ester Livshits, and Muhammad Tibi</i> .....	16:1–16:17
Uniform Reliability of Self-Join-Free Conjunctive Queries <i>Antoine Amarilli and Benny Kimelfeld</i> .....	17:1–17:17
Efficient Differentially Private $F_0$ Linear Sketching <i>Rasmus Pagh and Nina Mesing Stausholm</i> .....	18:1–18:19
Fine-Grained Complexity of Regular Path Queries <i>Katrin Casel and Markus L. Schmid</i> .....	19:1–19:20
Ranked Enumeration of MSO Logic on Words <i>Pierre Bourhis, Alejandro Grez, Louis Jachiet, and Cristian Riveros</i> .....	20:1–20:19
Approximate Similarity Search Under Edit Distance Using Locality-Sensitive Hashing <i>Samuel McCauley</i> .....	21:1–21:22
Locality-Aware Distribution Schemes <i>Bruhathi Sundarmurthy, Paraschos Koutris, and Jeffrey Naughton</i> .....	22:1–22:25

## ■ Preface

The 24. International Conference on Database Theory (ICDT 2021) was held in Nicosia, Cyprus, from March 23 to 26, 2021. The Program Committee has selected 20 research papers out of 42 submissions for publication at the conference. It has further decided to give the Best Paper Award to *Answer Counting Under Guarded TGDs* by Cristina Feier, Carsten Lutz, and Marcin Przybyłko. We congratulate the winners! Apart from the 20 regular papers, these proceedings include abstracts for the invited (shared) EDBT/ICDT keynotes by Pablo Barceló (Pontificia Universidad Católica de Chile) and by Julia Stoyanovich (New York University).

A committee formed by Yael Amsterdamer, Rasmus Pagh, and Pierre Senellart has decided to give the Test of Time Award for ICDT 2021 to the ICDT 2011 paper *Knowledge compilation meets database theory: compiling queries to decision diagrams* by Abhay Jha and Dan Suciu. We congratulate also the winners of this award!

We would like to thank all people who contributed to the success of ICDT 2021, including the authors of all submitted papers, keynote and invited talk 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 2021. 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, the local organizers of the EDBT/ICDT 2021 conference, led by General Chairs Demetris Zeinalipour and Panos K. Chrysanthis, 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.

Ke Yi and Zhewei Wei  
March 2021







## ■ Organization

### General Chairs

Demetris Zeinalipour (University of Cyprus)

Panos K. Chrysanthis (University of Cyprus and University of Pittsburgh)

### Program Chair

Ke Yi (The Hong Kong University of Science and Technology)

### Proceedings Chair

Zhewei Wei (Renmin University of China)

### Program Committee

Yael Amerdamer (Bar Ilan University)

Meghyn Bienvenu (CNRS, University of Bordeaux)

Vladimir Braverman (Johns Hopkins University)

Marco Calautti (University of Trento)

Hubie Chen (Birkbeck, University of London)

Sara Cohen (The Hebrew University)

Martin Grohe (RWTH Aachen University)

Benny Kimelfeld (Technion, Israel Institute of Technology)

Paraschos Koutris (University of Wisconsin-Madison)

Domenico Lembo (Sapienza University of Rome)

Stefan Mengel (CNRS, CRIL)

Matthias Niewerth (University of Bayreuth)

Dan Olteanu (University of Oxford)

Rasmus Pagh (IT University of Copenhagen)

Sudeepa Roy (Duke University)

Atri Rudra (University at Buffalo, SUNY)

Francesco Scarcello (DIMES, University of Calabria)

Srikanta Tirthapura (Apple Inc., Iowa State University)

Stijn Vansummeren (Université Libre de Bruxelles)

Jef Wijsen (University of Mons)





## ■ External Reviewers

Antoine Amarill

Mohammad Javad Amiri

Alexandr Andoni

Marcelo Arenas

Anton Belyy

Andrea Calí

Nofar Carmeli

Shaleen Deep

Cibele Freire

Dominik D. Freydenberger

Filippo Furfaro

Gianluigi Greco

Montserrat Hermo

Xiao Hu

Raj Jayaram

Zhengjie Miao

Cristian Molinaro

Frank Neven

Milos Nikolic

Francesco Parisi

Tina Popp

Andrea Pugliese

Juan L. Reutter

Cristian Riveros

Domenico Saccà

Uri Stemmer

Philip Wellnitz

Samson Zhou





## ■ Contributing Authors

Heba Aamer	Alejandro Grez	Jeffrey Naughton
Kaleb Alway	Martin Grohe	Rasmus Pagh
Antoine Amarilli	Xiao Hu	Liat Peterfreund
Eric Blais	Louis Jachiet	Marcin Przybyłko
Pierre Bourhis	Benny Kimelfeld	Cristian Riveros
Noa Bratman	Paraschos Koutris	Semih Salihoglu
Jan Van den Bussche	Ester Livshits	Markus L. Schmid
Nofar Carmeli	Shangqi Lu	Nicole Schweikardt
Katrin Casel	Carsten Lutz	Nina Mesing Stausholm
Balder ten Cate	Wim Martens	Bruhathi Sundarmurthy
Victor Dalmau	Samuel McCauley	Yufei Tao
Shaleen Deep	Andrew McGregor	David Tench
Johannes Doleschal	Alexandra Meliou	Muhammad Tibi
Cristina Feier	Zafeiria Mounoulidou	Hoa T. Vu





## ■ ICDT 2021 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 ICDT a decade earlier that have best met the “test of time”. In 2021, the award recognizes a paper from the ICDT 2011 proceedings that has had the most impact in terms of research, methodology, conceptual contribution, or transfer to practice over the past decade. The award is to be presented during the EDBT/ICDT 2021 Joint Conference, March 23–26, 2021 in Nicosia, Cyprus.

The ICDT 2021 Test of Time Award committee consists of Yael Amsterdamer (Chair), Rasmus Pagh, and Pierre Senellart. After careful consideration and soliciting external assessments, the committee has chosen the following recipient of the 2021 ICDT Test of Time Award:

*Knowledge compilation meets database theory: compiling queries to decision diagrams*  
**Abhay Jha and Dan Suciu**

There are two main approaches to computing the probability of a query result over probabilistic databases: the extensional approach exploits the structure of the query for efficient evaluation for some classes of queries; the intensional approach first tractably computes a representation of the lineage of the query and then attempts to compute the probability of this Boolean function. This paper shows that a number of cases known to be tractable in the extensional method lead to tractability in the intensional method because lineages can be produced in specific tractable formalisms (such as OBDDs, FBDDs, d-DNNFs) which are well-studied target compilation classes in knowledge compilation, and for which weighted model counting is tractable. The paper leaves open the major question of whether all tractable cases can be explained in the same manner.

With their work, Jha and Suciu established a strong connection between the fields of knowledge compilation and probabilistic databases, which was both foundational and entirely original. This has sparked research in and across different areas: in database theory in the form of further refinements of the results and progress towards the resolution of the question left open; in database systems by demonstrating that the intensional approach and the use of knowledge compilation techniques are viable for probabilistic query evaluation; and in knowledge compilation by further motivating and reviving interest for the study of the weighted variant of the model counting problem.

Yael Amsterdamer  
Bar-Ilan University

Rasmus Pagh  
University of Copenhagen

Pierre Senellart  
ENS, PSL University

*The ICDT Test-of-Time Award Committee for 2021*



