

Ladjel Bellatreche  
Marlon Dumas  
Panagiotis Karras  
Raimundas Matulevičius (Eds.)

LNCS 12843

# Advances in Databases and Information Systems

25th European Conference, ADBIS 2021  
Tartu, Estonia, August 24–26, 2021  
Proceedings

ADBIS



Springer

## Founding Editors

Gerhard Goos

*Karlsruhe Institute of Technology, Karlsruhe, Germany*

Juris Hartmanis

*Cornell University, Ithaca, NY, USA*


## Editorial Board Members

Elisa Bertino 

*Purdue University, West Lafayette, IN, USA*

Wen Gao

*Peking University, Beijing, China*

Bernhard Steffen 

*TU Dortmund University, Dortmund, Germany*

Gerhard Woeginger 

*RWTH Aachen, Aachen, Germany*

Moti Yung

*Columbia University, New York, NY, USA*

More information about this subseries at <http://www.springer.com/series/7409>


Ladjel Bellatreche · Marlon Dumas ·  
Panagiotis Karras · Raimundas Matulevičius (Eds.)


# Advances in Databases and Information Systems


25th European Conference, ADBIS 2021  
Tartu, Estonia, August 24–26, 2021  
Proceedings

*Editors*

Ladjet Bellatreche   
LIAS/ISAE-ENSMA  
Futuroscope, Chasseneuil Cedex, France

Panagiotis Karras   
Aarhus University  
Aarhus, Denmark

Marlon Dumas   
University of Tartu  
Tartu, Estonia

Raimundas Matulevičius   
University of Tartu  
Tartu, Estonia

ISSN 0302-9743                      ISSN 1611-3349 (electronic)  
Lecture Notes in Computer Science  
ISBN 978-3-030-82471-6              ISBN 978-3-030-82472-3 (eBook)  
<https://doi.org/10.1007/978-3-030-82472-3>

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

© Springer Nature Switzerland AG 2021

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

This volume contains a selection of the papers presented at the 25th European Conference on Advances in Databases and Information Systems (ADBIS 2021), held during August 24–26, 2021, at Tartu, Estonia.

The ADBIS series of conferences aims at providing a forum for the presentation and dissemination of research on database and information systems, the development of advanced data storage and processing technologies, and designing data-enabled systems/software/applications. ADBIS 2021 in Tartu continues after St. Petersburg (1997), Poznań (1998), Maribor (1999), Prague (2000), Vilnius (2001), Bratislava (2002), Dresden (2003), Budapest (2004), Tallinn (2005), Thessaloniki (2006), Varna (2007), Pori (2008), Riga (2009), Novi Sad (2010), Vienna (2011), Poznań (2012), Genoa (2013), Ohrid (2014), Poitiers (2015), Prague (2016), Nicosia (2017), Budapest (2018), Bled (2019), and Lyon (2020). This edition has been totally managed during the COVID-19 pandemic.

The program of ADBIS 2021 includes keynotes, research papers, thematic workshops, and a doctoral consortium. The conference attracted 70 paper submissions from 261 authors in 39 countries from all continents. After rigorous reviewing by the Program Committee (73 reviewers from 28 countries), the 18 papers included in this LNCS proceedings volume were accepted as full contributions, making an acceptance rate of 26%.

Furthermore, the Program Committee selected 8 more papers as short contributions and 21 papers from the five workshops and doctoral consortium which are published in a companion volume entitled *New Trends in Databases and Information Systems* in Springer's *Communications in Computer and Information Science (CCIS)* series. All papers were evaluated by at least three reviewers and some by four reviewers. The selected papers span a wide spectrum of topics in databases and related technologies, tackling challenging problems and presenting inventive and efficient solutions. In this volume, these papers are organized in seven sections: (1) High-dimensional Data and Data Streams, (2) Social Media and Text Mining, (3) Advanced Query Processing, (4) Patterns and Events, (5) Data Integration, (6) Complex Data, and (7) Database Internals and Processes.

For this edition of ADBIS 2021, we had three keynote talks by experts from three continents: America, Asia, and Europe. The first keynote was given by Divesh Srivastava, Head of Database Research at AT&T, on "*Towards High-Quality Big Data: Lessons from FIT*". The second one by Sanjay Chawla, Research Director of the Qatar Computing Research Institute (QCRI) Data Analytics department, on "*A perspective on prescriptive and reinforcement learning*". The third keynote by Dirk Draheim, Head of the Information Systems Group at Tallinn University of Technology, Estonia, addressed "*Data exchange for Digital Government: Where are we heading?*".

ADBIS 2021 strived to create conditions for more experienced researchers to share their knowledge and expertise with young researchers. In addition, the following five

workshops and the doctoral consortium associated with ADBIS were co-allocated with the main conference:

- Intelligent Data - from data to knowledge (DOING 2021), organized by Mirian Halfeld Ferrari (Université d'Orléans, France) and Carmem H. Hara (Universidade Federal do Paraná, Curitiba, Brazil).
- Data-Driven Process Discovery and Analysis (SIMPDA 2021), organized by Paolo Ceravolo (Università degli Studi di Milano, Italy), Maurice van Keulen (University of Twente, The Netherlands), and Maria Teresa Gomez Lopez (University of Seville, Spain),
- Modern Approaches in Data Engineering and Information System Design (MADEISD 2021), organized by Ivan Luković (University of Novi Sad, Serbia), Slavica Kordić (University of Novi Sad, Serbia), and Sonja Ristić (University of Novi Sad, Serbia).
- Advances in Data Systems Management, Engineering, and Analytics (MegaData 2021), organized by Yaser Jararweh (Duquesne University, USA), Tomás F. Pena (University of Santiago de Compostela, Spain) and Feras M. Awaysheh (University of Tartu, Estonia).
- Computational Aspects of Network Science (CAoNS 2021), organized by Dimitrios Katsaros (University of Thessaly, Greece) and Yannis Manolopoulos (Open University of Cyprus and Aristotle University of Thessaloniki, Greece).
- Doctoral Consortium (DC), co-chaired by Mirjana Ivanović (University of Novi Sad, Serbia) and Olaf Hartig (Linköping University, Sweden).

Each workshop and the DC has its own international Program Committee. The accepted papers were published by Springer in CCIS.

The best papers from the main conference and workshops were invited for submission to special issues of the following journals: Information Systems (Elsevier), Information Systems Frontiers (Springer), and Computer Science and Information Systems (ComSIS Consortium).

We would like to express our gratitude to every individual who contributed to the success of ADBIS 2021. First, we thank all authors for submitting their research papers to the conference. We are also indebted to the members of the community who offered their precious time and expertise in performing various roles ranging from organizing to reviewing - their efforts, energy, and degree of professionalism deserve the highest commendations. Special thanks to the Program Committee members and the external reviewers for evaluating papers submitted to ADBIS 2021, ensuring the quality of the scientific program, despite the COVID-19 pandemic. A special thanks to our keynote speakers who honored us with their exciting talks at ADBIS 2021. Thanks also to all the colleagues, secretaries, and engineers involved in the conference organization, as well as the workshop organizers. Special thanks are due to the members of the Steering Committee, in particular, its chair, Yannis Manolopoulos, for all their help and guidance. A particular thank you to the University of Tartu's Institute of Computer Science for hosting and supporting the conference despite the uncertainties created by the ongoing pandemic.

Finally, we thank Springer for publishing the proceedings containing invited and research papers in the LNCS series. The Program Committee work relied on EasyChair, and we thank its development team for creating and maintaining it; it offered a great support throughout the different phases of the reviewing process.

Last but not least, we thank the participants of ADBIS 2021 for sharing their work and presenting their achievements, thus providing a lively, fruitful and constructive forum, and giving us the pleasure of knowing that our work was purposeful.

June 2021

Ladjel Bellatreche  
Marlon Dumas  
Panagiotis Karras  
Raimundas Matulevičius



# Organization

## General Chairs

Marlon Dumas                      University of Tartu, Estonia  
Raimundas Matulevičius        University of Tartu, Estonia

## Program Committee Co-chairs

Ladjel Bellatreche                ISAE-ENSMA Poitiers, France  
Panagiotis Karras                Aarhus University, Denmark

## Workshop Co-chairs

Ahmed Awad                        University of Tartu, Estonia  
Matthias Weidlich                Humboldt University of Berlin, Germany

## Doctoral Consortium Co-chairs

Mirjana Ivanović                 University of Novi Sad, Serbia  
Olaf Hartig                         Linköping University, Sweden

## Webmaster

Mubashar Iqbal                    University of Tartu, Estonia

## Proceedings Technical Editor

Abasi-Amefon Obot Affia        University of Tartu, Estonia

## Technical Arrangements

Orlenys Lopez-Pintado         University of Tartu, Estonia

## Financial and Local Arrangements

Anneli Vainumae                 University of Tartu, Estonia

## Steering Committee

Yannis Manolopoulos            Open University of Cyprus, Cyprus  
    (Chair)  
Ladjel Bellatreche                ISAE-ENSMA Poitiers, France

Maria Bielikova	Kempelen Institute of Intelligent Technologies, Slovakia
Barbara Catania	University of Genoa, Italy
Jérôme Darmont	University of Lyon 2, France
Johann Eder	Alpen Adria Universität Klagenfurt, Austria
Tomáš Horváth	Eötvös Loránd University, Hungary
Mirjana Ivanović	University of Novi Sad, Serbia
Marite Kirikova	Riga Technical University, Latvia
Rainer Manthey	University of Bonn, Germany
Manuk Manukyan	Yerevan State University, Armenia
Tadeusz Morzy	Poznan University of Technology, Poland
Kjetil Nørvgå	Norwegian University of Science and Technology, Norway
Boris Novikov	National Research University Higher School of Economics, Saint Petersburg, Russia
George Papadopoulos	University of Cyprus, Cyprus
Jaroslav Pokorný	Charles University in Prague, Czech Republic
Bernhard Thalheim	Christian Albrechts University, Kiel, Germany
Goce Trajcevski	Iowa State University, USA
Valentino Vranić	Slovak University of Technology in Bratislava, Slovakia
Tatjana Welzer	University of Maribor, Slovenia
Robert Wrembel	Poznan University of Technology, Poland
Ester Zumpano	University of Calabria, Italy

## Program Committee

Alberto Abello	Universitat Politècnica de Catalunya, Barcelona, Spain
Reza Akbarinia	Inria, France
Bernd Amann	Sorbonne Université, Paris, France
Hassan Badir	ENSA Tanger, Morocco
Amin Beheshti	Macquarie University, Australia
Andreas Behrend	Technical University of Cologne, Germany
Sadok Ben Yahia	Tallinn University of Technology, Estonia
Soumia Benkrid	ESI Algiers, Algeria
Djamal Benslimane	University of Lyon 1, France
Fadila Bentayeb	University of Lyon 2, France
Miklos Biro	Software Competence Center Hagenberg, Austria
Kamel Boukhafa	USTHB, Algeria
Barbara Catania	University of Genoa, Italy
Tania Cerquitelli	Politecnico di Torino, Italy
Richard Chbeir	University of Pau and Pays de l'Adour, France
Isabelle Comyn-Wattiau	ESSEC Business School, France
Ajantha Dahanayake	Lappeenranta University of Technology, Finland
Jérôme Darmont	University of Lyon 2, France
Christos Doukeridis	University of Piraeus, Greece

Cedric Du Mouza	CNAM, France
Markus Endres	University of Passau, Germany
Philippe Fournier-Viger	Harbin Institute of Technology, Shenzhen, China
Johann Gamper	Free University of Bozen-Bolzano, Italy
Gabriel Ghinita	University of Massachusetts at Boston, USA
Olga Gkoutouna	George Mason University, USA
Giancarlo Guizzardi	Free University of Bozen-Bolzano, Italy, and University of Twente, The Netherlands
Hele-Mai Haav	Tallinn University of Technology, Estonia
Dirk Habich	TU Dresden, Germany
Mirian Halfeld-Ferrari	Université d'Orléans, France
Tomáš Horvách	Eötvös Loránd University, Hungary
Mirjana Ivanović	University of Novi Sad, Serbia
Stefan Jablonski	University of Bayreuth, Germany
Stéphane Jean	Poitiers University, France
Zoubida Kedad	University of Versailles, France
Marite Kirikova	Riga Technical University, Latvia
Attila Kiss	Eötvös Loránd University, Hungary
Sergio Lifschitz	Pontificia Universidade Católica do Rio de Janeiro, Brazil
Sebastian Link	The University of Auckland, New Zealand
Ivan Luković	University of Novi Sad, Serbia
Zakaria Maamar	Zayed University, Dubai, United Arab Emirates
Wojciech Macyna	Wroclaw University of Technology, Poland
Federica Mandreoli	University of Modena, Italy
Yannis Manolopoulos	Open University of Cyprus, Cyprus
Patrick Marcel	Université de Tours, France
Pascal Molli	University of Nantes, France
Anirban Mondal	Ashaka University, India
Tadeusz Morzy	Poznan University of Technology, Poland
Kjetil Nørvåg	Norwegian University of Science and Technology, Norway
Boris Novikov	National Research University Higher School of Economics, Saint Petersburg, Russia
Andreas Oberweis	Karlsruhe Institute of Technology, Germany
Carlos Ordonez	University of Houston, USA
Oscar Pastor	Universidad Politécnica de Valencia, Spain
Jaroslav Pokorný	Charles University in Prague, Czech Republic
Franck Ravat	IRIT and Université Toulouse 1 Capitole, France
Stefano Rizzi	University of Bologna, Italy
Oscar Romero	Universitat Politècnica de Catalunya, Spain
Carmem S. Hara	Universidade Federal do Paraná, Curitiba, Brazil
Gunter Saake	University of Magdeburg, Germany
Kai-Uwe Sattler	Technical University Ilmenau, Germany
Milos Savic	University of Novi Sad, Serbia
Kostas Stefanidis	Tampere University, Finland

Sergey Stupnikov	Russian Academy of Sciences, Russia
Olivier Teste	Université de Toulouse, France
Maik Thiele	TU Dresden, Germany
Goce Trajcevski	Iowa State University, USA
Anton Tsitsulin	University of Bonn, Germany
Panos Vassiliadis	University of Ioannina, Greece
Thanasis Vergoulis	“Athena” Research Center, Greece
Tatjana Welzer	University of Maribor, Slovenia
Marek Wojciechowski	Poznan University of Technology, Poland
Robert Wrembel	Poznan University of Technology, Poland
Vladimir Zadorozhny	University of Pittsburgh, USA

### **Additional Reviewers**

Petar Jovanovic	Universitat Politècnica de Catalunya, Spain
Elio Mansour	University of Pau and Pays de l’Adour, France
Sergi Nadal	Universitat Politècnica de Catalunya, Spain
Rediana Koçi	Universitat Politècnica de Catalunya, Spain
Nadia Yacoubi	Ayadi Institut Supérieur de Gestion de Tunis, Tunisia
Serafeim Chatzopoulos	“Athena” Research Center, Greece

# **ADBIS'2021 Keynotes**

# Towards High-Quality Big Data: A Focus on Time

Divesh Srivastava

AT&T Chief Data Office, USA

**Abstract.** Data are being generated, collected, and analyzed today at an unprecedented scale, and data-driven decision making is sweeping through all aspects of society. As the use of big data has grown, so too have concerns that poor-quality data, prevalent in large data sets, can have serious adverse consequences on data-driven decision making. Responsible data science thus requires a recognition of the importance of veracity, the fourth “V” of big data. In this talk, we first present a vision of high-quality big data, with a focus on time, and highlight the substantial challenges that the first three V’s, volume, velocity, and variety, bring to dealing with veracity in long data. We present the FIT Family of adaptive, data-driven statistical tools that we have designed, developed, and deployed at AT&T for continuous data quality monitoring of a large and diverse collection of continuously evolving data. These tools monitor data movement to discover missing, partial, duplicated, and delayed data; identify changes in the content of spatiotemporal streams; and pinpoint anomaly hotspots based on persistence, pervasiveness, and priority. We conclude with lessons relevant to long data quality that are cause for optimism.

# A Perspective on Prescriptive Learning

Sanjay Chawla

Qatar Computing Research Institute, Hamad Bin Khalifa University  
schawla@hbku.edu.qa

**Abstract.** We provide a brief overview of the emerging area of prescriptive learning which combines supervised learning with optimization. Prescriptive Learning is most active in operations research but is now finding applications in diverse areas ranging from database optimization to chip design. Reinforcement Learning (RL) is the most developed form of PL for sequential and stochastic optimization problems. We will give an example of how RL can be applied to a traditional and well-studied join ordering problem for query optimization.

# Data Exchange for Digital Government: Where Are We Heading?

Dirk Draheim

Information Systems Group, Tallinn University of Technology, Estonia  
dirk.draheim@taltech.ee

**Abstract.** In all countries, we currently see major efforts in digital transformation initiatives. The United Nations e-Government Survey 2020 puts a strong emphasis on data, which makes sense, given the huge progress in big data and data science in the last decade. The UN survey distinguishes between data-informed, data-driven and data-centric approaches to digital government. Actually, Gartner defined the notion of data-centric government already in 2014. Still, today, we are far away from such data-centric government. How comes? How to shape the next generation of e-government technologies? In service of such and similar questions, we walk through a series of important data exchange technologies: the Estonian data exchange layer X-Road, the European federated data infrastructure GAIA-X, the European Blockchain Services Infrastructure (EBSI), and the IoT data management solution FIWARE. Finally, based on the notion of data governance architecture, we give an overview of our proposed digital government architecture framework that is intended to help in large-scale digital government design efforts.



# Contents

## Keynotes Talks

- A Perspective on Prescriptive Learning ADBIS'2021 Keynote . . . . . 3  
*Sanjay Chawla*
- Data Exchange for Digital Government: Where Are We Heading?  
ADBIS2021 Keynote. . . . . 7  
*Dirk Draheim*

## Patterns and Events

- Maximal Mixed-Drove Co-Occurrence Patterns. . . . . 15  
*Witold Andrzejewski and Paweł Boinski*
- Efficiently Mining Large Gradual Patterns Using Chunked Storage Layout. . . 30  
*Dickson Odhiambo Owuor and Anne Laurent*
- A General Method for Event Detection on Social Media . . . . . 43  
*Yihong Zhang, Masumi Shirakawa, and Takahiro Hara*
- 5W1H Aware Framework for Representing and Detecting Real Events  
from Multimedia Digital Ecosystem. . . . . 57  
*Siraj Mohammed, Fekade Getahun, and Richard Chbeir*

## Social Media and Text Mining

- MONITOR: A Multimodal Fusion Framework to Assess Message Veracity  
in Social Networks . . . . . 73  
*Abderrazek Azri, Cécile Favre, Nouria Harbi, Jérôme Darmont,  
and Camille Noûs*
- Joint Management and Analysis of Textual Documents and Tabular Data  
Within the AUDAL Data Lake . . . . . 88  
*Pegdwendé N. Sawadogo, Jérôme Darmont, and Camille Noûs*
- Aggregation and Summarization of Thematically Similar Twitter  
Microblog Messages . . . . . 102  
*Markus Endres, Lena Rudenko, and Dominik Gröninger*

**Indexes, Queries and Constraints**

Inserting Keys into the Robust Content-and-Structure (RCAS) Index . . . . . 121  
*Kevin Wellenzohn, Luka Popovic, Michael Böhlen, and Sven Helmer*

Optimizing Execution Plans in a Multistore . . . . . 136  
*Chiara Forresi, Matteo Francia, Enrico Gallinucci,  
and Matteo Golfarelli*

Integrity Constraints for Microcontroller Programming in Datalog . . . . . 152  
*Stefan Brass and Mario Wenzel*

Chance Constraint as a Basis for Probabilistic Query Model . . . . . 167  
*Maksim Goman*

**High-Dimensional Data and Data Streams**

Unsupervised Feature Selection for Efficient Exploration of High  
Dimensional Data . . . . . 183  
*Arnab Chakrabarti, Abhijeet Das, Michael Cochez, and Christoph Quix*

MuLOT: Multi-level Optimization of the Canonical Polyadic Tensor  
Decomposition at Large-Scale . . . . . 198  
*Annabelle Gillet, Éric Leclercq, and Nadine Cullot*

From Large Time Series to Patterns Movies: Application to Airbus  
Helicopters Flight Data . . . . . 213  
*Benjamin Chazelle , Pierre-Loic Maisonneuve, Ammar Mechouche,  
Jean-Marc Petit, and Vasile-Marian Scuturici*

**Data Integration**

Experimental Evaluation Among Reblocking Techniques Applied  
to the Entity Resolution . . . . . 229  
*Laís Soares Caldeira, Guilherme Dal Bianco, and Anderson A. Ferreira*

FiLiPo: A Sample Driven Approach for Finding Linkage Points Between  
RDF Data and APIs . . . . . 244  
*Tobias Zeimet, and Ralf Schenkel*

SMAT: An Attention-Based Deep Learning Solution to the Automation  
of Schema Matching . . . . . 260  
*Jing Zhang, Bonggun Shin, Jinho D. Choi, and Joyce C. Ho*

Towards a Cloud-WSDL Metamodel: A New Extension of WSDL  
for Cloud Service Description. . . . . 275  
*Souad Ghazouani, Anis Tissaoui, and Richard Chbeir*

**Author Index** . . . . . 289