Silvia Chiusano · Tania Cerquitelli · Robert Wrembel · Kjetil Nørvåg · Barbara Catania · Genoveva Vargas-Solar · Ester Zumpano (Eds.)

**Communications in Computer and Information Science** 

1652

# New Trends in Database and Information Systems

ADBIS 2022 Short Papers, Doctoral Consortium and Workshops: DOING, K-GALS, MADEISD, MegaData, SWODCH, Turin, Italy September 5–8, 2022, Proceedings





### **Communications** in Computer and Information Science

1652

#### **Editorial Board Members**

Joaquim Filipe

Polytechnic Institute of Setúbal, Setúbal, Portugal

Ashish Ghosh

Indian Statistical Institute, Kolkata, India

Raquel Oliveira Prates

Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

Lizhu Zhou

Tsinghua University, Beijing, China

More information about this series at https://link.springer.com/bookseries/7899

Silvia Chiusano · Tania Cerquitelli · Robert Wrembel · Kjetil Nørvåg · Barbara Catania · Genoveva Vargas-Solar · Ester Zumpano (Eds.)

## New Trends in Database and Information Systems

ADBIS 2022 Short Papers, Doctoral Consortium and Workshops: DOING, K-GALS, MADEISD, MegaData, SWODCH, Turin, Italy September 5–8, 2022, Proceedings



Editors
Silvia Chiusano
Politecnico di Torino
Turin, Italy

Robert Wrembel Department of Technology Poznań. Poland

Barbara Catania D University of Genoa Genoa, Italy

Ester Zumpano D University of Calabria Rende, Italy Tania Cerquitelli D Politecnico di Torino Turin. Italy

Kjetil Nørvåg (1) Norwegian University of Science and Technology Trondheim, Norway

Genoveva Vargas-Solar CNRS
Villeurbanne Cedex, France

ISSN 1865-0929 ISSN 1865-0937 (electronic) Communications in Computer and Information Science ISBN 978-3-031-15742-4 ISBN 978-3-031-15743-1 (eBook) https://doi.org/10.1007/978-3-031-15743-1

#### © Springer Nature Switzerland AG 2022

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 year ADBIS – the Conference on Advances in Databases and Information Systems – celebrated its 26th anniversary.

The first ADBIS conference was held in Saint Petersburg, Russia (1997). Since then, ADBIS has taken place annually, with previous editions were held in: Poznan, Poland (1998); Maribor, Slovenia (1999); Prague, Czech Republic (2000); Vilnius, Lithuania (2001); Bratislava, Slovakia (2002); Dresden, Germany (2003); Budapest, Hungary (2004); Tallinn, Estonia (2005); Thessaloniki, Greece (2006); Varna, Bulgaria (2007); Pori, Finland (2008); Riga, Latvia (2009); Novi Sad, Serbia (2010); Vienna, Austria (2011); Poznan, Poland (2012); Genoa, Italy (2013); Ohrid, North Macedonia (2014); Poitiers, France (2015); Prague, Czech Republic (2016); Nicosia, Cyprus (2017); Budapest, Hungary (2018); Bled, Slovenia (2019); Lyon, France (2020); and Tartu, Estonia (2021).

The official ADBIS portal – http://adbis.eu – provides up to date information on all ADBIS conferences, committees, publications, and issues related to the ADBIS community.

The 26th ADBIS was held in Turin, Italy, during September 5–8, 2022, as a hybrid event. It received significant attention from both the research and industrial communities, as 90 papers were submitted to the conference. The papers were reviewed by an international Program Committee consisting of 85 members. The Program Committee selected 28 short papers from the main ADBIS conference, an acceptance rate of 42%, for inclusion in this volume. These papers cover the following topics: data modeling and visualization, data processing (fairness, cleaning, pre-processing), information and process retrieval, data access optimization, graph data, data science, and machine learning.

Additionally, this volume includes papers accepted at the five workshops and Doctoral Consortium, which accompanied ADBIS 2022. Each workshop had its own international Program Committee, whose members served as the reviewers of papers included in this volume. The maximum paper acceptance rate at each of these events did not exceed 50%. In total, 53 papers were submitted to these workshops, out of which 26 were selected for presentation and publication in this volume, giving an overall acceptance rate of 49%.

The following five workshops were run at ADBIS 2022:

DOING: 3rd Workshop on Intelligent Data – From Data to Knowledge, chaired by Mirian Halfeld-Ferrari (Université d'Orléans, France) and Carmem S. Hara (Universidade Federal do Parana, Brazil). The idea underlying DOING is to gather researchers in natural language processing, databases, and artificial intelligence to discuss two main problems: (1) how to extract information from textual data and represent it in knowledge bases, and (2) how to propose intelligent methods for handling and maintaining these databases with new forms of requests, including efficient, flexible, and secure analysis mechanisms, adapted to the user, and with quality and privacy preservation guarantees.

K-GALS: 1st Workshop on Knowledge Graphs Analysis on a Large Scale, chaired by Mariella Bonomo (University of Palermo, Italy) and Simona E. Rombo (University of Palermo, Italy). The goal of K-GALS is to provide participants with the opportunity to introduce and discuss new methods, theoretical approaches, algorithms, and software tools that are relevant to the knowledge graphs based research, especially when it is focused on a large scale. Relevant issues include how knowledge graphs can be used to represent knowledge, how systems managing knowledge graphs work, and which applications can be provided on top of a knowledge graphs, in a distributed environment.

MADEISD: 4th Workshop on Modern Approaches in Data Engineering and Information System Design, chaired by Ivan Luković (University of Belgrade, Serbia), Slavica Kordić (University of Novi Sad, Serbia), and Sonja Ristić (University of Novi Sad, Serbia). The main goal of MADEISD is to address open questions and real potentials for various applications of modern approaches and technologies in data engineering and information system design so as to develop and implement effective software services in support of information management in various organization systems.

MegaData: 2nd Workshop on Advanced Data Systems Management, Engineering, and Analytics, chaired by Feras M. Awaysheh (University of Tartu, Estonia), Fahed Alkhabbas (Malmö University, Sweden), and Said Alawadi (Uppsala University, Sweden). MegaData aims to report on the advances and trends in database deployment models and environments from both the infrastructure and application levels. This year, questions related to how to develop data as the foundation for advanced databases and information systems and what the next generation of data systems will look like led the discussion on the latest trends in modern data systems.

SWODCH: 2nd Workshop on Semantic Web and Ontology Design for Cultural Heritage, chaired by Antonis Bikakis (University College London, UK), Roberta Ferrario (ISTC-CNR, Italy), Stéphane Jean (University of Poitiers and ISAE-ENSMA, France), Béatrice Markhoff (University of Tours, France), Alessandro Mosca (Free University of Bozen-Bolzano, Italy), and Marianna Nicolosi Asmundo (University of Catania, Italy). The foundational purpose of SWODCH is to gather original research work about both application and foundational issues emerging from the design of conceptual models, ontologies, and Semantic Web technologies for the digital humanities, here understood according to its broader definition including cultural heritage, digital history, archaeology, and related fields. The application-oriented focus of SWODCH, on the other hand, aims at bringing together stakeholders from various scientific fields involved in the development or deployment of Semantic Web solutions, including computer scientists, data scientists, and digital humanists.

The ADBIS 2022 Doctoral Consortium (DC) provided a forum for PhD students to present their research projects to the scientific community. DC papers describe the status of PhD students' research, a comparison with relevant related work, their results, and plans on how to experiment, validate and consolidate their contribution. Through the Doctoral Consortium, PhD students receive feedback from a selected group of mentors about potential opportunities and perspectives and suggestions to pursue their work. The PhD Consortium allows PhD students to establish international collaborations with members and participants of the ADBIS community.

The 2022 edition of the ADBIS DC included seven papers accepted after a very selective peer-review process, which are published in this volume. Additionally, the ADBIS 2022 DC program included a panel about strategies for preparing for post-PhD life with professionals with accomplished careers in academia, industry, and entrepreneurship.

The ADBIS 2022 Organizing Committee supported diversity and inclusion by offering some grants, supporting a few researchers to participate in the conference and become part of the ADBIS community. All the grants were assigned based on the underrepresented community, gender, and role/position. The grants include:

- two free regular registrations, assigned to researchers from Argentina and Brazil,
- three regular registration fee discounts of 200 Euros, assigned to researchers from Estonia, Lebanon, and Italy, and
- four regular registration fee discounts of 150 Euros, assigned to researchers from Croatia, France, and Italy.

We would like to wholeheartedly thank all participants, authors, PC members, workshop organizers, session chairs, doctoral consortium chairs, workshop chairs, volunteers, and co-organizers for their contributions in making ADBIS 2022 a great success. We would also like to thank the ADBIS Steering Committee and all sponsors.

July 2022

Silvia Chiusano Tania Cerquitelli Robert Wrembel Kjetil Nørvåg Barbara Catania Genoveva Vargas-Solar Ester Zumpano

#### **Organization**

#### **General Chair**

Silvia Chiusano Polytechnic University of Turin, Italy

#### **Program Committee Chairs**

Tania Cerquitelli Polytechnic University of Turin, Italy Robert Wrembel Poznan University of Technology, Poland

#### **Workshop Chairs**

Kjetil Nørvåg Norwegian University of Science and Technology,

Norway

Barbara Catania University of Genoa, Italy

#### **Doctoral Consortium Chairs**

Genoveva Vargas-Solar CNRS, LIRIS, France
Ester Zumpano University of Calabria, Italy

#### **Local Organization Committee Chairs**

Luca CaglieroPolytechnic University of Turin, ItalyPaolo GarzaPolytechnic University of Turin, ItalyBartolomeo VacchettiPolytechnic University of Turin, ItalyGiovanni MalnatiPolytechnic University of Turin, Italy

#### **Publicity Chair**

Oscar Romero Polytechnic University of Catalonia -

BarcelonaTech, Spain

#### **Proceedings Chair**

Khalid Belhajjame Paris Dauphine University - PSL, France

#### **Special Issue Chair**

Ladjel Bellatreche ISAE-ENSMA, France

#### **Diversity and Inclusion Chair**

Jérôme Darmont University of Lyon 2, France

#### **Steering Committee**

Andreas Behrend TH Köln, Germany Ladjel Bellatreche ISAE-ENSMA, France

Maria Bielikova Kempelen Institute of Intelligent Technologies,

Slovakia

Barbara Catania University of Genoa, Italy
Jérôme Darmont University of Lyon 2, France
Johann Eder Universität Klagenfurt, Austria

Johann Gamper Free University of Bozen-Bolzano, Italy
Tomáš Horváth Eötvös Loránd University, Hungary
Mirjana Ivanović University of Novi Sad, Serbia
Marite Kirikova Riga Technical University, Latvia
Manuk Manukyan Yerevan State University, Armenia
Raimundas Matulevicius University of Tartu, Estonia

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

George Papadopoulos University of Cyprus, Cyprus

Jaroslav Pokorný Charles University in Prague, Czech Republic

Oscar Romero Polytechnic University of Catalonia -

BarcelonaTech, Spain

Sergey Stupnikov Russian Academy of Sciences, Russia

Bernhard Thalheim University of 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 Unviersity of Technology, Poland

Ester Zumpano University of Calabria, Italy

#### **Program Committee**

Damiano Carra

Jérôme Darmont

Alberto Abelló Polytechnic University of Catalonia -

BarcelonaTech, Catalonia

Cristina D. Aguiar University of São Paulo, Brasil

Syed Muhammad Fawad Ali IBM, Germany

Bernd Amann Sorbonne University, France

Poznan University of Technology, Poland Witold Andrzeiewski Daniele Apiletti Polytechnic University of Turin, Italy Costin Badica University of Craiova, Romania Svlvio Barbon Junior University of Trieste, Italy

Andreas Behrend University of Bonn, Germany Khalid Belhajjame Paris Dauphine University - PSL, France

ISAE-ENSMA, France Ladiel Bellatreche

András Benczúr Eötvös Loránd University, Hungary

Fadila Bentayeb University of Lyon 2, France

Maria Bielikova Kempelen Institute of Intelligent Technologies,

Slovakia

Sandro Bimonte INRAE, France

Pawel Boiński Poznan University of Technology, Poland

Zoran Bosnić University of Ljubljana, Slovenia Omar Boussaid University of Lyon 2, France Drazen Brdjanin University of Bania Luka. Bosnia and Herzegovina

University of Verona, Italy University of Lyon 2, France

Claudia Diamantini Marche Polytechnic University, Italy

Christos Doulkeridis University of Piraeus, Greece Johann Eder University of Klagenfurt, Austria Markus Endres University of Passau, Germany Javier A. Espinosa-Oviedo University of Lyon, France Georgios Evangelidis University of Macedonia, Greece

Flavio Ferrarotti Software Competence Centre Hagenberg, Austria

Alessandro Fiori Polytechnic University of Turin, Italy Flavius Frasincar

Erasmus University Rotterdam,

The Netherlands

Johann Gamper Free University of Bozen-Bolzano, Italy

Matteo Golfarelli University of Bologna, Italy

Marcin Gorawski Silesian University of Technology, Poland

Jānis Grabis Riga Technical University, Latvia

Francesco Guerra University of Modena and Reggio Emilia, Italy Giancarlo Guizzardi Federal University of Espirito Santo, Brazil

Tomas Horvath Eötvös Loránd University, Hungary Mirjana Ivanović University of Novi Sad, Serbia Stefan Jablonski University of Bayreuth, Germany Aida Kamišalić University of Maribor, Slovenia Zoubida Kedad University of Versailles, France Attila Kiss Eötvös Loránd University, Hungary

Julius Köpke Alpen-Adria University of Klagenfurt, Austria

Dejan Lavbič University of Ljubljana, Slovenia Yurii Litvinov St. Petersburg State University, Russia

Audrone Lupeikiene Vilnius University, Lithuania

University of Modena and Reggio Emilia, Italy Federica Mandreoli

Open University of Cyprus, Cyprus Yannis Manolopoulos

Patrick Marcel University of Tours, France Sara Migliorini University of Verona, Italy University of Udine, Italy Angelo Montanari

Polytechnic University of Turin, Italy Lia Morra Tadeusz Morzy Poznan University of Technology, Poland **Boris Novikov** National Research University, Higher School of

Economics, Russia

Kjetil Nørvåg Norwegian University of Science and Technology,

Norway

Andreas Oberweis Karlsruhe Institute of Technology, Germany

George Papadopoulos University of Cyprus, Cyprus

Budapest University of Technology and András Pataricza

Economics, Hungary

Jan Platos Technical University of Ostrava, Czech Republic Jaroslav Pokorný Charles University in Prague, Czech Republic

Giuseppe Polese University of Salerno, Italy

Alvaro E. Prieto University of Extremadura, Spain

University of Verona, Italy Elisa Ouintarelli Miloš Radovanović University of Novi Sad, Serbia Franck Ravat University of Toulouse, France University of Bologna, Italy Stefano Rizzi

Oscar Romero Polytechnic University of Catalonia, Spain Gunter Saake

University of Magdeburg, Germany

Kai-Uwe Sattler TU Ilmenau, Germany

University of Novi Sad, Serbia Milos Savic Claudia Steinberger University of Klagenfurt, Austria Sergey Stupnikov Russian Academy of Sciences, Russia

Bernhard Thalheim University of Kiel, Germany Goce Trajcevski Iowa State University, USA Raquel Trillo-Lado University of Zaragoza, Spain

Genoveva Vargas Solar CNRS, LIRIS, France Goran Velinov Ss. Cyril and Methodius University,

North Macedonia

Peter Vojtas Charles University in Prague, Czech Republic

Isabelle Wattiau ESSEC and Cnam, France

Marek Wojciechowski Poznan University of Technology, Poland

Vladimir Zadorozhny University of Pittsburgh, USA Ester Zumpano University of Calabria, Italy

#### **Additional Reviewers**

Stylianos Argyrou Joseph Giovanelli Paul Blockhaus Nicolas Labroche Miroslav Blšták Christos Mettouris Andrea Brunello Simone Monaco Gabriel Campero Durand Federico Motta Andrea Chiorrini Uchechukwu Njoku Francesco Del Buono Thomas Photiadis Chiara Forresi Matúš Pikuliak Marco Franceschetti Nicola Saccomanno Matteo Francia Vladimir A. Shekhovtsov

Verena Geist Emanuele Storti

#### Contents

Visualization	
On an Extensible Conceptual Data Model by a Non-formal Example	3
Understanding Misinformation About COVID-19 in WhatsApp Messages  Antônio Diogo Forte Martins, José Maria Monteiro, and Javam C. Machado	14
Automatic Inference of Taxonomy Relationships Among Legal Documents  Irene Benedetto, Luca Cagliero, and Francesco Tarasconi	24
Visualisation of Numerical Query Results on Industrial Data Streams  Miran Ismaiel Nadir and Kjell Orsborn	34
ADBIS Short Papers: Fairness in Data processing	
Clustering-Based Subgroup Detection for Automated Fairness Analysis  Jero Schäfer and Lena Wiese	45
Open Science in the Cloud: The CloudFAIR Architecture for FAIR-compliant Repositories  João Pedro C. Castro, Lucas M. F. Romero, Anderson Chaves Carniel, and Cristina D. Aguiar	56
Detecting Simpson's Paradox: A Step Towards Fairness in Machine Learning Rahul Sharma, Minakshi Kaushik, Sijo Arakkal Peious, Markus Bertl, Ankit Vidyarthi, Ashwani Kumar, and Dirk Draheim	67
ADBIS Short Papers: Data Management Pipeline, Information and Process Retrieval	
FusionFlow: An Integrated System Workflow for Gene Fusion Detection in Genomic Samples	79

ProSA Pipeline: Provenance Conquers the CHASE	89
A Tabular Open Data Search Engine Based on Word Embeddings for Data  Integration	99
The Use of M2P in Business Process Improvement and Optimization  Meriem Kherbouche, Yossra Zghal, Bálint Molnár, and András Benczúr	109
Embedding Process Structure in Activities for Process Mapping and Comparison	119
Andrea Chiorrini, Claudia Diamantini, Laura Genga, Martina Pioli, and Domenico Potena	119
ADBIS Short Papers: Data Access Optimization	
Bulk Loading of the Secondary Index in LSM-Based Stores for Flash	
Memory Wojciech Macyna and Michal Kukowski	133
Blink: Lightweight Sample Runs for Cost Optimization of Big Data	
Applications	144
ADBIS Short Papers: Data Pre-processing and Cleaning	
A RoBERTa Based Approach for Address Validation	157
Empirical Comparison of Semantic Similarity Measures for Technical	
Question Answering Nabil Moncef Boukhatem, Davide Buscaldi, and Leo Liberti	167
Denoising Architecture for Unsupervised Anomaly Detection	
in Time-Series Wadie Skaf and Tomáš Horváth	178
Discretizing Numerical Attributes: An Analysis of Human Perceptions  Minakshi Kaushik, Rahul Sharma, Ankit Vidyarthi, and Dirk Draheim	188

Ingo Schmitt

DOING: 5rd Workshop on Intelligent Data - From Data to Knowledge	
Using Provenance in Data Analytics for Seismology: Challenges and Directions  Umberto Souza da Costa, Javier Alfonso Espinosa-Oviedo, Martin A. Musicante, Genoveva Vargas-Solar, and José-Luis Zechinelli-Martini	311
Storing Feature Vectors in Relational Image Data Warehouses  Guilherme Muzzi Rocha, Piero Lima Capelo, Anderson Chaves Carniel, and Cristina Dutra Aguiar	323
Automatic Classification of Stigmatizing Articles of Mental Illness: The Case of Portuguese Online Newspapers  Alina Yanchuk, Alina Trifan, Olga Fajarda, and José Luís Oliveira	332
A Multiplex Network Framework Based Recommendation Systems for Technology Intelligence  Foutse Yuehgoh, Sonia Djebali, and Nicolas Travers	344
Relation Extraction from Clinical Cases for a Knowledge Graph	353
Privacy Operators for Semantic Graph Databases as Graph Rewriting	366
K-GALS: 1st Workshop on Knowledge Graphs Analysis on a Large Scale	
A Collaborative Filtering Approach for Drug Repurposing	381
A Review: Biological Insights on Knowledge Graphs	388
CiteRank: A Method to Evaluate Researchers Influence Based on Citation and Collaboration Networks  Fabrizio Angiulli, Fabio Fassetti, and Cristina Serrao	400
~	

MADEISD: 4th Workshop on Modern Approaches in Data Engineering and Information System Design	
Abstract Machine for Operational Semantics of Domain-Specific Language William Steingartner, Róbert Baraník, and Valerie Novitzká	413
Machine Learning-Based Model Categorization Using Textual and Structural Features  Alireza Khalilipour, Fatma Bozyigit, Can Utku, and Moharram Challenger	425
Architecture Design of a Networked Music Performance Platform for a Chamber Choir	437
Defining Software Architecture Modalities Based on Event Sourcing Architecture Pattern	450
Supporting Data Types in Neo4j	459
Systematic Literature Review for Emotion Recognition from EEG Signals Paulina Leszczełowska and Natalia Dawidowska	467
Semantically-Driven Secure Task Execution over Wireless Sensor  Networks  Niki Hrovatin, Aleksandar Tošić, and Michael Mrissa	476
MegaData: 2nd Workshop on Advanced Data Systems Management, Engineering, and Analytics	
Stored and Inherited Relations with PKN Foreign Keys	487
Implementing the Comparison-Based External Sort  Michael Polyntsov, Valentin Grigorev, Kirill Smirnov, and George Chernishev	500
Document Versioning for MongoDB	512

Théo Abgrall

Valeria Soliani

Andreas Görres

SWODCH: 2nd Workshop on Semantic Web and Ontology Design for Cultural Heritage	
Developing and Aligning a Detailed Controlled Vocabulary for Artwork	529
Using Wikibase for Managing Cultural Heritage Linked Open Data Based on CIDOC CRM	542
Megalithism Representation in CIDOC-CRM  Ivo Santos, Renata Vieira, Cassia Trojahn, Leonor Rocha, and Enrique Cerrillo Cuenca	550
Ontology for Analytic Claims in Music	559
Testing the Word-Based Model in the Ontological Analysis of Modern Greek Derivational Morphology  Nikos Vasilogamvrakis, Maria Koliopoulou, Michalis Sfakakis, and Giannoula Giannoulopoulou	572
Analyses of Networks of Politicians Based on Linked Data: Case ParliamentSampo – Parliament of Finland on the Semantic Web	585
Neural Word Sense Disambiguation to Prune a Large Knowledge Graph of the Italian Cultural Heritage	593
Doctoral Consortium	
Querying Sensor Networks Using Temporal Property Graphs	607
Formalization of Data Integration Transformations	615

Models and Query Languages for Temporal Property Graph Databases . . . . . . 623

	Contents	XXi
Towards Prescriptive Analyses of Querying Large Knowledge Graph Mohamed Ragab	s	639
Knowledge Extraction from Biological and Social Graphs		648
Digital Twins for Urban Mobility		657
Author Index		667