

Robert Wrembel · Johann Gamper ·
Gabriele Kotsis · A Min Tjoa ·
Ismail Khalil (Eds.)

LNCS 13428

Big Data Analytics and Knowledge Discovery

24th International Conference, DaWaK 2022
Vienna, Austria, August 22–24, 2022
Proceedings

DaWaK 2022



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

Moti Yung 

Columbia University, New York, NY, USA


More information about this series at <https://link.springer.com/bookseries/558>

Robert Wrembel · Johann Gamper ·
Gabriele Kotsis · A Min Tjoa ·
Ismail Khalil (Eds.)

Big Data Analytics and Knowledge Discovery

24th International Conference, DaWaK 2022
Vienna, Austria, August 22–24, 2022
Proceedings

Editors

Robert Wrembel 
Poznań University of Technology
Poznań, Poland

Johann Gamper 
Free University of Bozen-Bolzano
Bozen-Bolzano, Italy

Gabriele Kotsis
Johannes Kepler University of Linz
Linz, Austria

A Min Tjoa
Vienna University of Technology
Vienna, Austria

Ismail Khalil
Johannes Kepler University of Linz
Linz, Austria

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-031-12669-7 ISBN 978-3-031-12670-3 (eBook)
<https://doi.org/10.1007/978-3-031-12670-3>

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to 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

DaWaK was established in 1999 as the International Conference on Data Warehousing and Knowledge Discovery. It ran continuously under this name until its 16th edition (2014, Munich, Germany). In 2015 (Valencia, Spain) it was renamed the International Conference on Big Data Analytics and Knowledge Discovery to better reflect new research directions in the broad and dynamically developing area of data analytics, but it retained its DaWaK acronym. In 2022, the 24th edition of DaWaK took place during August 22–24 in Vienna, Austria, back as a physical event following online editions in 2020 and 2021.

Since the very beginning, the DaWaK conference has been a high-quality forum for researchers, practitioners, and developers in the field of data integration, data warehousing, data analytics, and, recently, big data analytics. The main objectives of this event are to explore, disseminate, and exchange knowledge in these fields through scientific and industry talks. Big data analytics and knowledge discovery remain hot research areas for both academia and the software industry. They are continuously evolving, fueled by advances in hardware and software. Important research topics associated with these major areas include data lakes (schema-free repositories of heterogeneous data), conceptual/logical/physical database design, theoretical foundations for data engineering, data integration (especially linking structured and semi-structured data sources), big data management (mixing relational tables, text, and any types of files), query languages (beyond SQL), scalable analytical algorithms, parallel storage and computing systems (cloud, parallel database systems, Spark, MapReduce, HDFS), graph processing, stream and time series processing, IoT architectures, artificial intelligence/machine learning algorithms, and applications of these solutions in industry.

DaWaK 2022 attracted 57 papers, from which the Program Committee selected 12 regular papers and 12 short papers, yielding an acceptance rate of 21% for the regular paper category, of 27% for the short paper category, and of 42% overall. Each paper was reviewed by at least three reviewers and in some cases up to four. The accepted papers cover a variety of research topics on both theoretical and practical aspects. The program included the following topics: (1) text analytics, (2) data warehousing and OLAP, (3) feature selection algorithms, (4) time series processing, (5) schema discovery and construction, (6) pattern discovery, and (7) machine learning algorithms. Thanks to the reputation of DaWaK, selected best papers of DaWaK 2022 will be invited for a special issue of the Data & Knowledge Engineering (DKE, Elsevier) journal. Therefore, the PC chairs would like to thank the DKE Editor-in-Chief, Carson Woo, for his approval of the special issue.

We would like to express our sincere gratitude to all Program Committee members and the external reviewers who reviewed the papers thoroughly and in a timely manner. Finally, we would like to thank the DEXA conference organizers for their continuous

support and guidance especially Ismail Khalil for providing a great deal of assistance (as always), putting his experience at our disposal.

August 2022

Johann Gamper
Robert Wrembel

Organization

Program Committee Chairs

Robert Wrembel
Johann Gamper

Poznan University of Technology, Poland
Free University of Bozen-Bolzano, Italy

Steering Committee

Gabriele Kotsis
A Min Tjoa
Robert Wille
Bernhard Moser
Alfred Taudes
Ismail Khalil

Johannes Kepler University Linz, Austria
TU Wien, Austria
Software Competence Center Hagenberg, Austria
Software Competence Center Hagenberg, Austria
Vienna University of Economics and Business
and Austrian Blockchain Center, Austria
Johannes Kepler University Linz, Austria

Program Committee

Alberto Abello
Cristina D. Aguiar
Syed Muhammad Fawad Ali
Witold Andrzejewski
Sylvio Barbon
Ladjel Bellatreche
Sadok Ben Yahia
Fadila Bentayeb
Jorge Bernardino
Vasudha Bhatnagar
Sandro Bimonte
Pawel Boinski
Kamel Boukhalfa
Omar Boussaid
Stephane Bressan
Wellington Cabrera
Sharma Chakravarthy
Silvia Chiusano
Alfredo Cuzzocrea

Universitat Politècnica de Catalunya, Spain
USP São Carlos, Brazil
Poznan University of Technology, Poland
Poznan University of Technology, Poland
University of Trieste, Italy
LIAS/ENSMA, France
University of Tunis El Manar, Tunisia
ERIC Lab, University of Lyon, France
ISEC - Polytechnic Institute of Coimbra, Portugal
University of Delhi, India
INRAE, France
Poznan University of Technology, Poland
University of Science and Technology Houari
Boumediene, Algeria
ERIC Lab, University of Lyon, France
National University of Singapore, Singapore
University of Houston, USA
University of Texas at Arlington, USA
Politecnico di Torino, Italy
University of Calabria, Italy

Laurent D’Orazio	University of Rennes, CNRS, IRISA, France
Jérôme Darmont	Université Lyon 2, France
Soumyava Das	Teradata Labs, USA
Karen Davis	Miami University, USA
Claudia Diamantini	Università Politecnica delle Marche, Italy
Alin Dobra	University of Florida, USA
Christos Doulkeridis	University of Piraeus, Greece
Markus Endres	University of Passau, Germany
Leonidas Fegaras	University of Texas at Arlington, USA
Philippe Fournier-Viger	Shenzhen University, China
Matteo Francia	University of Bologna, Italy
Filippo Furfaro	University of Calabria, Italy
Pedro Furtado	University of Coimbra, Portugal
Luca Gagliardelli	University of Modena and Reggio Emilia, Italy
Enrico Gallinucci	University of Bologna, Italy
Kazuo Goda	University of Tokyo, Japan
Maria Teresa Gómez López	University of Seville, Spain
Marcin Gorawski	Silesian University of Technology and Wrocław University of Technology, Poland
Sven Groppe	University of Lübeck, Germany
Frank Höppner	Ostfalia University of Applied Sciences, Germany
Mirjana Ivanovic	University of Novi Sad, Serbia
Stephane Jean	LISI/ENSMA and University of Poitiers, France
Selma Khouri	Ecole Nationale Supérieure d’Informatique, Algeria
Uday Kiran	University of Tokyo, Japan
Jens Lechtenbörgner	University of Münster, Germany
Young-Koo Lee	Kyung Hee University, South Korea
Carson Leung	University of Manitoba, Canada
Sebastian Link	University of Auckland, New Zealand
Sofian Maabout	LaBRI, University of Bordeaux, France
Patrick Marcel	Université de Tours, LIFAT, France
Alejandro Maté	University of Alicante, Spain
Paolo Missier	Newcastle University, UK
Jun Miyazaki	Tokyo Institute of Technology, Japan
Anirban Mondal	University of Tokyo, Japan
Rim Moussa	ENICarthage, Tunisia
Kjetil Nørnvåg	Norwegian University of Science and Technology, Norway
Boris Novikov	National Research University Higher School of Economics, Russia
Makoto Onizuka	Osaka University, Japan

Carlos Ordonez	University of Houston, USA
Jaroslav Pokorný	Charles University in Prague, Czech Republic
Praveen Rao	University of Missouri, USA
Franck Ravat	IRIT, Université de Toulouse, France
Oscar Romero	Universitat Politècnica de Catalunya, Spain
Keun Ho Ryu	Chungbuk National University, South Korea
Ilya Safro	University of Delaware, USA
Abhishek Santra	University of Texas at Arlington, USA
Kai-Uwe Sattler	Ilmenau University of Technology, Germany
Alkis Simitsis	HP Labs, USA
Günther Specht	University of Innsbruck, Austria
Kostas Stefanidis	Tampere University, Finland
Emanuele Storti	Università Politecnica delle Marche, Italy
Olivier Teste	IRIT, France
Dimitri Theodoratos	New Jersey Institute of Technology, USA
Maik Thiele	HTW Dresden, Germany
Juan Trujillo	University of Alicante, Spain
Maurice van Keulen	University of Twente, The Netherlands
Panos Vassiliadis	University of Ioannina, Greece
Isabelle Wattiau	ESSEC and CNAM, France
Lena Wiese	Goethe University Frankfurt, Germany
Szymon Wilk	Poznan University of Technology, Poland
Yinuo Zhang	Teradata Labs, USA
Yongjun Zhu	Yonsei University, South Korea

External Reviewers

Nabil El Malki
Chiara Forresi
Joseph Giovanelli
Alex Mircoli
Uchechukwu Njoku
Shivika Prasanna
Arun George Zachariah

Organizers



Contents

Text Analytics

An Integration of TextGCN and Autoencoder into Aspect-Based Sentiment Analysis	3
<i>Yi-Hang Tsai, Chia-Ming Chang, Kun-Hsiang Chen, and San-Yih Hwang</i>	
OpBerg: Discovering Causal Sentences Using Optimal Alignments	17
<i>Justin Wood, Nicholas Matiasz, Alcino Silva, William Hsu, Alexej Abyzov, and Wei Wang</i>	
Text-Based Causal Inference on Irony and Sarcasm Detection	31
<i>Recep Firat Cekinel and Pinar Karagoz</i>	
Sarcastic RoBERTa: A RoBERTa-Based Deep Neural Network Detecting Sarcasm on Twitter	46
<i>Maciej Hercog, Piotr Jaroński, Jan Kolanowski, Paweł Mieczynski, Dawid Wisniewski, and Jędrzej Potoniec</i>	
A Fast NDFA-Based Approach to Approximate Pattern-Matching for Plagiarism Detection in Blockchain-Driven NFTs	53
<i>Dariusz Galiş, Ciprian Pungilă, and Viorel Negru</i>	

Data Warehousing and OLAP

On Decisive Skyline Queries	61
<i>Akrivi Vlachou, Christos Doulkeridis, Joăo B. Rocha-Junior, and Kjetil Nęrvåg</i>	
Safeness: Suffix Arrays Driven Materialized View Selection Framework for Large-Scale Workloads	74
<i>Mohamed Kechar and Ladjel Bellatreche</i>	
A Process Warehouse for Process Variants Analysis	87
<i>Lisana Berberi</i>	

Feature Selection Algorithms

Unsupervised Features Ranking via Coalitional Game Theory for Categorical Data	97
<i>Chiara Balestra, Florian Huber, Andreas Mayr, and Emmanuel Müller</i>	

Multi-label Online Streaming Feature Selection Algorithms via Extending Alpha-Investing Strategy 112
Tianqi Ji, Xizhi Guo, Yunqian Li, Dan Li, Jun Li, and Jianhua Xu

Feature Selection Under Fairness and Performance Constraints 125
Ginel Dorleon, Imen Megdiche, Nathalie Bricon-Souf, and Olivier Teste

Time Series Processing

Interpretable Input-Output Hidden Markov Model-Based Deep Reinforcement Learning for the Predictive Maintenance of Turbofan Engines 133
Ammar N. Abbas, Georgios C. Chasparis, and John D. Kelleher

Pathology Data Prioritisation: A Study Using Multi-variate Time Series 149
Jing Qi, Girvan Burnside, and Frans Coenen

Outlier/Anomaly Detection of Univariate Time Series: A Dataset Collection and Benchmark 163
David Muhr and Michael Affenzeller

Schema Discovery and Construction

Automatic Machine Learning-Based OLAP Measure Detection for Tabular Data 173
Yuzhao Yang, Fatma Abdelhédi, Jérôme Darmont, Franck Ravat, and Olivier Teste

Discovering Overlapping Communities Based on Cohesive Subgraph Models over Graph Data 189
Said Jabbour, Mourad Kmimech, and Badran Raddaoui

Discovery of Keys for Graphs 202
Morteza Alipourlangouri and Fei Chiang

OPTIMA: Framework Selecting Optimal Virtual Model to Query Large Heterogeneous Data 209
Chahrazed B. Bachir Belmehdi, Abderrahmane Khiat, and Nabil Keskes

Pattern Discovery

Q-VIPER: Quantitative Vertical Bitwise Algorithm to Mine Frequent Patterns 219
Thomas J. Czubryt, Carson K. Leung, and Adam G. M. Pazdor

Enhanced Sliding Window-Based Periodic Pattern Mining from Dynamic Streams 234
Evan W. Madill, Carson K. Leung, and Justin M. Gouge

Explainable Recommendations for Wearable Sensor Data 241
Niccoló Marastoni, Barbara Oliboni, and Elisa Quintarelli

Machine Learning

SLA-Aware Cloud Query Processing with Reinforcement Learning-Based Multi-objective Re-optimization 249
Chenxiao Wang, Le Gruenwald, and Laurent d’Orazio

Mahalanobis Distance Based K-Means Clustering 256
Paul O. Brown, Meng Ching Chiang, Shiqing Guo, Yingzi Jin, Carson K. Leung, Evan L. Murray, Adam G. M. Pazdor, and Alfredo Cuzzocrea

Grapevine Phenology Prediction: A Comparison of Physical and Machine Learning Models 263
Francisco J. Lacueva-Pérez, Sergio Ilarri, Juan J. Barriuso, Joaquín Balduque, Gorka Labata, and Rafael del-Hoyo

Author Index 271