

Lecture Notes in Computer Science

16048

Founding Editors


Gerhard Goos


Juris Hartmanis

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*

The series Lecture Notes in Computer Science (LNCS), including its subseries Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics (LNBI), has established itself as a medium for the publication of new developments in computer science and information technology research, teaching, and education.


LNCS enjoys close cooperation with the computer science R & D community, the series counts many renowned academics among its volume editors and paper authors, and collaborates with prestigious societies. Its mission is to serve this international community by providing an invaluable service, mainly focused on the publication of conference and workshop proceedings and postproceedings. LNCS commenced publication in 1973.

Carson K. Leung · Anton Dignös ·
Gabriele Kotsis · A. Min Tjoa · Ismail Khalil
Editors

Big Data Analytics and Knowledge Discovery

27th International Conference, DaWaK 2025
Bangkok, Thailand, August 25–27, 2025
Proceedings

Editors

Carson K. Leung 
University of Manitoba
Winnipeg, MB, Canada

Anton Dignös 
Free University of Bozen-Bolzano
Bolzano, Italy

Gabriele Kotsis 
Johannes Kepler University Linz
Linz, Austria

A. Min Tjoa 
Vienna University of Technology
Vienna, Austria

Ismail Khalil 
Johannes Kepler University Linz
Linz, Austria

ISSN 0302-9743

ISSN 1611-3349 (electronic)

Lecture Notes in Computer Science

ISBN 978-3-032-02214-1

ISBN 978-3-032-02215-8 (eBook)

<https://doi.org/10.1007/978-3-032-02215-8>

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2025

This work is subject to copyright. All rights are solely and exclusively licensed 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

If disposing of this product, please recycle the paper.

Preface

DAWAK was established in 1999 as the International Conference on Data Warehousing and Knowledge Discovery. It was run continuously under this name until its 16th edition in 2014. Then, it was renamed in 2015 to the International Conference on Big Data Analytics and Knowledge Discovery, while retaining its DAWAK acronym. This name change was made to better reflect new research directions in the broad and rapidly growing areas of big data analytics and knowledge discovery from massive data.

Since its inception, the DAWAK conference has provided a high-quality platform for researchers, practitioners, and developers worldwide in the areas of database systems, data integration, data warehousing, cloud computing, programming languages, both traditional and big data analytics, and more recently, artificial intelligence (AI) and data science. The main objectives of the conference are to explore, disseminate, and exchange knowledge in these fields through scientific and industrial talks. With big data analytics, knowledge discovery, and AI becoming important research frontiers in both academia and the IT industry, DAWAK continues to evolve alongside innovations in algorithms, data processing architectures, and high-performance computing technologies.

Important research topics associated with these major areas include: models for big data; parallel processing and DBMS technology; distributed system architectures; scalability and parallelization; query languages, processing and optimization; semantics for big data intelligence; data warehouse and data lake architectures; data mining and text mining; AI, machine learning and deep learning; pre-processing and data cleaning; poly-store and multistore architectures; NoSQL storage systems; cloud infra-structures and metadata for big data; big data storage, indexing, quality, provenance, search, discovery and management; big data analytics on various data (e.g., text, graph, unstructured data, spatio-temporal data, sensor data, real-time events); privacy and security in analytics; data visualisation; big data applications; data science; data engineering, management and analytics.

This book comprises the Proceedings of the 27th International Conference on Big Data Analytics and Knowledge Discovery (DAWAK 2025), held August 25–27, 2025, in Bangkok, Thailand. This marked the second time the DAWAK conference was host-ed outside of Europe—the first being DAWAK 2023 in Penang, Malaysia.

DAWAK 2025 received 62 paper submissions. Of these, the Program Committee selected 12 as regular papers, resulting in an acceptance rate of 19%. Additionally, 14 papers were accepted as short papers, aimed at showcasing pioneering research and innovative projects across various big data analytics and knowledge discovery disciplines. These short papers highlight early-stage research, emerging concepts and pre-liminary findings, fostering meaningful discussions and potential collaborations.

The accepted papers cover a variety of research topics on both theoretical and practical aspects. The program included among others the following topics: (1) data management and indices, (2) data mining and knowledge discovery, (3) graph data processing

and analytics, (4) large language models (LLMs), (5) neural networks, as well as (6) sequential data analytics and recommendation systems.

Owing to the reputation of DAWAK, selected best papers from DAWAK 2025 will be invited for a special issue of the Data & Knowledge Engineering (DKE, Elsevier) journal. We extend our sincere thanks to Carson Woo, Editor-in-Chief of DKE, for approving the special issue.

We are also grateful to the members of the Program Committee and external reviewers for their thorough and timely evaluations of the submitted papers. Finally, we express our heartfelt appreciation to the DEXA conference organizers for their continuous support and guidance, with special thanks to Ismail Khalil for his invaluable help in all tasks leading to having the proceedings published.

August 2025

Carson K. Leung
Anton Dignös
Gabriele Kotsis
A. Min Tjoa
Ismail Khalil

Organization

Program Committee Chairs

Carson K. Leung
Anton Dignös

University of Manitoba, Canada
Free University of Bozen-Bolzano, Italy

Publicity Chair

Kanda Saikaew

Khon Kaen University, Thailand

Steering Committee

Gabriele Kotsis Johannes
A Min Tjoa
Lukas Fischer
Bernhard Moser
Christine Strauss
Ismail Khalil

Kepler University Linz, Austria
Vienna University of Technology, Austria
Software Competence Center Hagenberg, Austria
Software Competence Center Hagenberg, Austria
University of Vienna, Austria
Johannes Kepler University Linz, Austria

Program Committee Members

Abhishek Santra
Alberto Abello
Alejandro Maté
Andrea Kö
Besim Bilalli
Boris Novikov
Carlos Ordonez
Chaiporn Jaikaeo
Chitsutha Soomlek
Chotirat Ratanamahatana
Christos Doulkeridis
Chutiporn Anutariya
Cristina Dutra de Aguiar
Darja Solodovnikova

University of Texas at Arlington, USA
Universitat Politècnica de Catalunya, Spain
University of Alicante, Spain
Corvinus University of Budapest, Hungary
Universitat Politècnica de Catalunya, Spain
Finland
University of Houston, USA
Kasetsart University, Thailand
Khon Kaen University, Thailand
Chulalongkorn University, Thailand
University of Piraeus, Greece
Asian Institute of Technology, Thailand
Universidade de São Paulo, Brazil
University of Latvia, Latvia

Dimitri Theodoratos	New Jersey Institute of Technology, USA
Ela Pustulka	University of Applied Sciences and Arts Northwestern Switzerland, Switzerland
Elisa Quintarelli	Università di Verona, Italy
Emanuele Storti	Università Politecnica delle Marche, Italy
Enrico Gallinucci	University of Bologna, Italy
Fadila Bentayeb	Université Lyon 2, France
Franck Ravat, IRIT	Université de Toulouse, France
Frank Höppner	Ostfalia University of Applied Sciences, Germany
Genoveva Vargas-Solar	CNRS-LIRIS, France
Hwan-Seung Yong	Ewha Womans University, South Korea
Isabelle Wattiau	ESSEC & CNAM, France
Jaroslav Pokorný	Charles University in Prague, Czech Republic
Jens Lechtenbörger	University of Münster, Germany
Jérôme Darmont	Université Lyon 2, France
Jorge Bernardino	Polytechnic Institute of Coimbra, Portugal
Jun Miyazaki	Tokyo Institute of Technology, Japan
Kamel Boukhalfa	USTHB University, Algeria
Kanda Saikaew	Khon Kaen University, Thailand
Kazuo Goda	University of Tokyo, Japan
Keun Ho Ryu	Chungbuk National University, South Korea
Kjetil Nørvåg	Norwegian University of Science and Technology, Norway
Ladjel Bellatreche	ENSMA, France
Laurent D’Orazio	University of Rennes, CNRS, IRISA, France
Leonidas Fegaras	University of Texas at Arlington, USA
Luca Gagliardelli	University of Modena and Reggio Emilia, Italy
Maik Thiele	Technische Universität Dresden, Germany
Marcin Gorawski	Silesian University of Technology, Poland
Markus Endres	University of Passau, Germany
Marut Buranarach	National Electronics and Computer Technology Center - NECTEC, Thailand
Matteo Francia	University of Bologna, Italy
Michele Linardi	CY Cergy Paris University, France
Mirjana Ivanovic	University of Novi Sad, Serbia
Mourad Khayati	University of Fribourg, Switzerland
Olivier Teste	IRIT, France
Omar Boussaid	Université Lyon 2, France
Oscar Romero	Universitat Politècnica de Catalunya, Spain
Patrick Marcel	Université d’Orléans, France
Rim Moussa	ENI Carthage, Tunisia
Sana Sellami	Aix-Marseille University, France

Sandro Bimonte	INRAE, France
Selma Khouri	École Nationale Supérieure d'Informatique, Algeria
Silvia Chiusano	Politecnico di Torino, Italy
Sofian Maabout	LaBRI, University of Bordeaux, France
Soror Sahri	Université Paris Cité, France
Soumia Benkrid	Ecole Nationale Supérieure d'Informatique, Algeria
Soumyava Das	Teradata Labs, USA
Stephane Jean	University of Poitiers, ISAE-ENSMA Poitiers, France
Sven Groppe	University of Luebeck, Germany
Trasapong Thaiupathump	Chiang Mai University, Thailand
Vatcharaporn Esichaikul	Asian Institute of Technology, Thailand
Witold Andrzejewski	Poznań University of Technology, Poland
Wookey Lee	Inha University, South Korea
Yinuo Zhang	Teradata, USA
Young-Koo Lee	Kyung Hee University, South Korea

External Reviewers

Imane Hocine	University of Luxembourg, Luxembourg
Javier Sanchis	University of Alicante, Spain
Manel Ben Sassi	Université de la Manouba, ENSI, RIADI, Tunisia

Organizers





Contents

Keynote Talk

Sparse Matrix Algorithms for Evolving Neural Networks	3
<i>Carlos Ordonez</i>	

Invited Talk

Data Integration in the AI Era: Research Trends and Still Open Issues	21
<i>Robert Wrembel</i>	

Tutorial

Leveraging Machine Learning Techniques for Customer Data Deduplication - Hard-Won Lessons from a Real-World Project in the Financial Industry	39
<i>Robert Wrembel, Witold Andrzejewski, Pawel Boński, and Bartosz Bębel</i>	

Data Mining and Knowledge Discovery

FairFES - Fast Exact Sampling for Fair Classification	55
<i>Manh Khoi Duong, Nina A. Liebrand, and Stefan Conrad</i>	
Autism Detection by Analyzing Handwriting Characteristics of Chinese Characters via Deep Learning Models	70
<i>Yunrui Li, Jasin Wong, Eva E. Chen, Syauki Aulia Thamrin, and Arbee L. P. Chen</i>	
FNoDe: Faulty Node Detection in Microservices Architecture	86
<i>Harsh Borse, Utkalika Satpathy, Mainack Mondal, and Bivas Mitra</i>	
An Enhanced FP-Growth Algorithm with Hybrid Adaptive Support Threshold for Association Rule Mining	102
<i>Kanda Runapongsa Saikaew, Carson K. Leung, and Kritbodin Phiwhorm</i>	

Sequential Data Analytics and Recommendation Systems

Entity Resolution for Streaming Data with Embeddings	111
<i>Zhongwei Ma, Philippe Roose, and Jiefu Song</i>	

Cross-Modal Sequential Point-of-Interest Recommendation
with Lightweight Hybrid Fusion Strategy 126
Tianxing Wang, Can Wang, Hui Tian, and Hong Shen

Alternatives to Shallow Autoencoders for Collaborative Filtering 142
Mario Mallea, Àngela Nebot, and Francisco Mugica

Accurate Concept Drift Detection Without Updating Autoencoders 157
Taisei Takano and Hisashi Koga

Graph Data Processing and Analytics

Parallel and Distributed SQL/PGQ Query Processing for Property Graphs 167
Kosuke Yamasaki, Tadashi Masuda, and Toshiyuki Amagasa

Graph Constraint Language for Industrial Knowledge Graphs and Machine
Learning 182
*Zhuoxun Zheng, Ognjen Savković, Baifan Zhou, Antonis Klironomos,
Evgeny Kharlamov, and Ahmet Soylu*

SemViSG: Semantic Enrichment and Visualization of Software Graphs 200
*Sami M’hamdi, Hamamache Kheddouci, Damien Charlemagne,
and Olivier Bonsignour*

Data management and Indices

Certainty Attacks Using Explainability Preprocessing 209
Carina Newen, Sofia Vergara Puccini, and Emmanuel Müller

Integrating Bitcoin Transactions into Relational Databases for IoT:
Challenges and Solutions 225
Rebeca Tonu, Otilia Muntean, and Ciprian Pungilă

Effects of Response Length on User Search Experience in Spoken
Conversational Search 231
Ken Tobioka, Takehiro Yamamoto, and Hiroaki Ohshima

Fair Proportional Top-k Ranking 237
Nina A. Liebrand, Manh Khoi Duong, and Stefan Conrad

PAID: Power-Efficient AI-Optimized Databases 244
*Ayoub Bouhatous, Ladjel Bellatreche, El Hassan Abdelwahed,
and Carlos Ordonez*

On the Costs and Benefits of Learned Indexing for Dynamic High-Dimensional Data	251
<i>Terézia Slanínáková, Jaroslav Olha, David Procházka, Matej Antol, and Vlastislav Dohnal</i>	

A Bayesian Reinforcement Learning Framework for Online Index Tuning	259
<i>Md Rakibul Hasan, Xiaoying Wu, and Dimitri Theodoratos</i>	

Large Language Models (LLMs)

Explaining Recovery Trajectories of Older Adults Post Lower-Limb Fracture Using Modality-Wise Multiview Clustering and Large Language Models	271
<i>Shehroz S. Khan, Ali Abedi, and Charlene H. Chu</i>	

Parameter Drift as a Signal for Membership Inference in Overfit-Tuned LLMs	286
<i>Takuto Kitamura and Yu Suzuki</i>	

MicroSuggest: Kernel-Aware Microservice Decomposition	302
<i>Harsh Borse, Utkalika Satpathy, Mainack Mondal, and Bivas Mitra</i>	

TraceTune: Targeted Fine-Tuning of Attention Heads for Text-to-SQL	309
<i>Saba Zamankhani and Kai-Uwe Sattler</i>	

Neural Networks

ONNYX : Optimized Neural Networks Yielding eXplainable Insights from ECG Signals-Based Data Streams	317
<i>Sanket Mishra, V. Aravindan, Rajkanwar Singh, Hasita Chowdary Meka, Sandipan Maiti, and Subhrakanta Panda</i>	

SPAPOOL: Soft Partition Assignment Pooling for Graph Neural Networks	332
<i>Rodrigue Govan, Romane Scherrer, Philippe Fournier-Viger, and Nazha Selmaoui-Folcher</i>	

Prediction of Iterative Solvers' Convergence Using Pretraining by Natural Images	341
<i>Yuki Oba, Taro Tezuka, and Hidehiko Hasegawa</i>	

Local-Aware Convolutional Modulation for Short-Term Sequential
Recommendation 349
 Tianxing Wang, Can Wang, Hui Tian, and Hong Shen

Author Index 357