

2024 IEEE 40th International Conference on Data Engineering Workshops (ICDEW 2024)

**Utrecht, Netherlands
13-16 May 2024**



**IEEE Catalog Number: CFP2445A-POD
ISBN: 979-8-3503-8404-8**

**Copyright © 2024 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP2445A-POD
ISBN (Print-On-Demand):	979-8-3503-8404-8
ISBN (Online):	979-8-3503-8403-1
ISSN:	1943-2895

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2024 IEEE 40th International Conference on Data Engineering Workshops (ICDEW) **ICDEW 2024**

Table of Contents

7th Int. Workshop on Data Engineering Meets Intelligent Food and Cooking Recipes (DECOR'24)

DECOR 2024: Where Data Engineering Meets Intelligent Food and Cooking Innovation	1
<i>Frederic Andres (National Institute of Informatics, Japan), Farshad Fotouhi (Wayne State University, USA), Gheorghita Ghinea (Brunel University London, UK), and Steve Tanimoto (University of Washington, USA)</i>	
Analyzing Socioeconomic Status through Culinary Ingredients: A Large-Scale Study of Pita and Pizza Dishes	5
<i>Oz Kilic (Carleton University, Canada; Middle East Technical University, Türkiye) and Tuğba Taşkaya Temizel (Middle East Technical University, Türkiye)</i>	
From Text to Taste: Advancing Smart Appliances with Multilingual Recipe Interpretation	13
<i>Vlad Andrei Negru (Technical University of Cluj-Napoca, Romania), Robert Vacareanu (Technical University of Cluj-Napoca, Romania; University of Arizona, USA), Camelia Lemnaru (Technical University of Cluj-Napoca, Romania), Mihai Surdeanu (University of Arizona, USA), and Rodica Potolea (Technical University of Cluj-Napoca, Romania)</i>	
Using LLMs to Extract Food Entities from Cooking Recipes	21
<i>Vasiliki Pitsilou ("Athena" Research Center, Greece), George Papadakis (National and Kapodistrian University of Athens, Greece), and Dimitrios Skoutas ("Athena" Research Center, Greece)</i>	
Food Computing for Nutrition and Health	29
<i>Shuqiang Jiang (Chinese Academy of Sciences; University of Chinese Academy of Sciences)</i>	

International Workshop on Data-Driven Smart Cities (DASC '24)

Experimental Probing of Graph Convolutional Neural Networks Architectures for Traffic Analysis	32
<i>Bahare Salehi (Université Libre de Bruxelles (ULB), Belgium) and Mahmoud Sakr (Université Libre de Bruxelles (ULB), Belgium)</i>	

On the Ecosystem of High-Definition (HD) Maps	40
<i>Yuanjie Zhu (University of California Riverside, USA), Hussah Alrashid (University of California Riverside, USA), Song Bai (University of California Riverside, USA), Chunhan Zhang (University of California Riverside, USA), Ziliang Zhang (University of California Riverside, USA), Zhengyi Qu (University of California Riverside, USA), Reem Y. Ali (University of California Riverside, USA), and Amr Magdy (University of California Riverside, USA)</i>	
Graph-Based Optimisation of Network Expansion in a Dockless Bike Sharing System	48
<i>Mark Roantree (Dublin City University, Ireland), Niamh Murphy (Dublin City University, Ireland), Dinh Viet Cuong (Dublin City University, Ireland), and Vuong M. Ngo (Ho Chi Minh City Open University, Vietnam)</i>	
Evaluating the Relationship Between Urban Structure and Mobility Through Clustering Analysis: A Case Study	56
<i>Emilia Lenzi (Politecnico di Milano, Italy), Carlo Andrea Biraghi (Politecnico di Milano, Italy), Giuseppe Giova (Politecnico di Milano, Italy), and Gianmarco Naro (Politecnico di Milano, Italy)</i>	
A Pipeline for Effective Monitoring and Improvement of Energy Metrics in Public Buildings	64
<i>Alex Vakaloudis (Munster Technological University, Ireland), Ronan Roche (Munster Technological University, Ireland), Derry Crowley (Cork County Council, Ireland), and Tadhg O'Meara (Cork County Council, Ireland)</i>	
Intelligent Customer Behaviour Analysis in the Norwegian Market	71
<i>Kristian Brathovde (OsloMet University and Coop Norge, Norway), Youcef Djenouri (University of South-Eastern Norway, Norway; IDEAS NCBR, Poland), Anis Yazidi (OsloMet University and Coop Norge, Norway), and Gautam Srivastava (Brandon University, Canada)</i>	
Building Trustworthy Smart Cities: a Systems Engineering Approach to Data Engineering at the Smart Metrology Campus	78
<i>Michael Benedict Ulbig (Physikalisch-Technische Bundesanstalt, Germany), Daniel Hutzschenreuter (Physikalisch-Technische Bundesanstalt, Germany), and Barbara Jung (Physikalisch-Technische Bundesanstalt, Germany)</i>	

International Workshop on Multivariate Time Series Analytics (MulTiSA'24)

Subset Models for Multivariate Time Series Forecast	86
<i>Raphael Saldanha (University of Montpellier, France), Victor Ribeiro (DEXL, LNCC, Petrópolis, Brazil), Eduardo H.M. Pena (DACOM, UTFPR, Campo Mourão, Brazil), Marcel Pedrosa (PCDaS, LIS, ICICT, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil), Reza Akbarinia (University of Montpellier, France), Patrick Valduries (University of Montpellier, France; DEXL, LNCC, Petrópolis, Brazil), and Fabio Porto (DEXL, LNCC, Petrópolis, Brazil)</i>	
Data-Hungry Fault Detection Algorithms Can Try Transfer Learning for Starters	91
<i>Jurgen van den Hoogen (Osnabrück University, Germany), Dan Hudson (Osnabrück University, Germany), and Martin Atzmüller (Osnabrück University, Germany)</i>	

Anomaly Detectors for Multivariate Time Series: The Proof of the Pudding is in the Eating	96
<i>Phillip Wenig (University of Potsdam, Germany), Sebastian Schmidl (University of Potsdam, Germany), and Thorsten Papenbrock (Philipps University of Marburg, Germany)</i>	
Parameter-Free Streaming Distance-Based Outlier Detection	102
<i>Apostolos Giannoulidis (Aristotle University of Thessaloniki, Greece), Nikodimos Nikolaidis (Atlantis Engineering, Greece), and Anastasios Gounaris (Aristotle University of Thessaloniki, Greece)</i>	
Beyond the Dimensions: A Structured Evaluation of Multivariate Time Series Distance Measures	107
<i>Jens E. d'Hondt (Eindhoven University of Technology, the Netherlands), Odysseas Papapetrou (Eindhoven University of Technology, the Netherlands), and John Paparrizos (Ohio State University, USA)</i>	
Time Series Problems in the Energy Sector	113
<i>Christos Dalamagkas (Public Power Corporation, Greece), Angelos Georgakis (Public Power Corporation, Greece), Ioannis Papadopoulos (Public Power Corporation, Greece), Kostas Hrissagis-Chrysagis (Public Power Corporation, Greece), and George Papadakis (Public Power Corporation, Greece)</i>	
MultiCast: Zero-Shot Multivariate Time Series Forecasting Using LLMs	119
<i>Georgios Chatzigeorgakidis ("Athena" Research Center, Greece), Konstantinos Lentzos ("Athena" Research Center, Greece), and Dimitrios Skoutas ("Athena" Research Center, Greece)</i>	
Data Augmentation for Multivariate Time Series Classification: An Experimental Study	128
<i>Romain Ilbert (Huawei Paris Research Center & LIPADE, France), Thai V. Hoang (TH Consulting, France), and Zonghua Zhang (CRSC R&D Institute Group Co. Ltd., China)</i>	
Challenges in Modeling Drug Shortage Events in the Pharmaceutical Domain	140
<i>Laura-Maria Tološi-Halacheva (Teva Pharmaceuticals, Bulgaria), Radoslav Andreev (Teva Pharmaceuticals, Bulgaria), Oleg Shcherbakov (Teva Pharmaceuticals, Bulgaria), and Eran Nevo (Teva Pharmaceuticals, Bulgaria)</i>	
Extended Framework and Evaluation for Multivariate Streaming Anomaly Detection with Machine Learning	144
<i>Andreas Koch (Airbus Defence and Space GmbH, Germany; Technical University of Munich), Michael Petry (Airbus Defence and Space GmbH, Germany; Technical University of Munich), and Martin Werner (Technical University of Munich, Germany)</i>	
Towards Ptolemaic Metric Properties of the Z-Normalized Euclidean Distance for Multivariate Time Series Indexing	153
<i>Max Pernklau (University in Hagen, Germany) and Christian Beecks (University in Hagen, Germany)</i>	
Exploiting Individual Graph Structures to Enhance Ecological Momentary Assessment (EMA) Forecasting	158
<i>Mandani Ntekouli (Maastricht University, The Netherlands), Gerasimos Spanakis (Maastricht University, The Netherlands), Lourens Waldorp (University of Amsterdam, The Netherlands), and Anne Roefs (Maastricht University, The Netherlands)</i>	

Linear-Trend Normalization for Multivariate Subsequence Similarity Search	167
<i>Thibaut Germain (Université Paris Saclay, Université Paris Cité, France), Charles Truong (Université Paris Saclay, Université Paris Cité, France), and Laurent Oudre (Université Paris Saclay, Université Paris Cité, France)</i>	

International Workshop on Fairness in AI (FAIR'24)

Closed-Loop View of the Regulation of AI: Equal Impact Across Repeated Interactions	176
<i>Quan Zhou (Imperial College London), Ramen Ghosh (Atlantic Technological University), Robert Shorten (Imperial College London), and Jakub Mareček (Czech Technical University in Prague)</i>	
Examining and Explaining Individual Fairness in Dynamic Pricing	182
<i>Wei Guo (Southeast University, China), Yan Lyu (Southeast University, China), Xueyong Xu (North Information Control Research Academy Group Co., Ltd., China), and Weiwei Wu (Southeast University, China)</i>	
Inspecting and Measuring Fairness of Unlabeled Image Datasets	191
<i>Rebeka Görge (Fraunhofer IAIS, Germany), Michael Mock (Fraunhofer IAIS, Germany), and Maram Akila (Fraunhofer IAIS & Lamarr, Germany)</i>	
Fairness in Ranking: Robustness Through Randomization Without the Protected Attribute	201
<i>Andrii Kliachkin (Università degli Studi di Padova, Italy), Eleni Psaroudaki (National Technical University of Athens, Greece; Athena Research & Innovation Center in Information Communication & Knowledge Technologies, Greece), Jakub Mareček (Czech Technical University in Prague, the Czech Republic), and Dimitris Fotakis (National Technical University of Athens, Greece; Athena Research & Innovation Center in Information Communication & Knowledge Technologies, Greece)</i>	
Pricefair: On Fair Scheduling of Heterogeneous Resources	209
<i>Aristotelis Peri (Athens University of Economics and Business, Greece), Dimitrios Tomaras (Athens University of Economics and Business, Greece), Vana Kalogeraki (Athens University of Economics and Business, Greece), and Dimitrios Gunopulos (National and Kapodistrian University of Athens, Greece)</i>	
Fairness in AI: Challenges in Bridging the gap Between Algorithms and law	217
<i>Giorgos Giannopoulos (Athena Research Center, Greece), Maria Psalla (Athena Research Center, Greece), Loukas Kavouras (Athena Research Center, Greece), Dimitris Sacharidis (Université Libre de Bruxelles, Belgium), Jakub Marecek (Czech Technical University, Czech Republic), Germán M Matilla (Czech Technical University, Czech Republic), and Ioannis Emiris (Athena Research Center, Greece)</i>	
On Explaining Unfairness: An Overview	226
<i>Christos Fragkathoulas (University of Ioannina and Archimedes / Athena RC, Greece), Vasiliki Papanikou (University of Ioannina and Archimedes / Athena RC, Greece), Danae Pla Karidi (Archimedes / Athena RC, Greece), and Evaggelia Pitoura (University of Ioannina and Archimedes / Athena RC, Greece)</i>	

Optimal Transport for Fairness: Archival Data Repair Using Small Research Data Sets	237
<i>Abigail Langbridge (Imperial College, London), Anthony Quinn (Imperial College, London; Trinity College, Dublin), and Robert Shorten (Imperial College, London)</i>	
Exploring Fairness Interpretability with FairnessFriend: A Chatbot Solution	246
<i>Chiara Criscuolo (Politecnico di Milano, Italy) and Tommaso Dolci (Politecnico di Milano, Italy)</i>	
A Framework for Feasible Counterfactual Exploration Incorporating Causality, Sparsity and Density	254
<i>Kleopatra Markou (National and Kapodistrian University of Athens, Greece), Dimitrios Tomaras (Athens University of Economics and Business, Greece), Vana Kalogeraki (Athens University of Economics and Business, Greece), and Dimitrios Gunopulos (National and Kapodistrian University of Athens, Greece)</i>	

International Workshop on Data Platform Design, Management, and Optimization (DataPlat'24)

Towards an End-to-End Data Quality Optimizer	262
<i>Valerie Restat (University of Hagen, Germany), Meike Klettke (University of Regensburg, Germany), and Uta Störl (University of Hagen, Germany)</i>	
Data Science Tasks Implemented with Scripts Versus GUI-Based Workflows: The Good, the Bad, and the Ugly	267
<i>Alexander K. Taylor (UC Los Angeles, USA), Yicong Huang (UC Irvine, USA), Junheng Hao (UC Los Angeles, USA), Xinyuan Lin (UC Irvine, USA), Xiusi Chen (UC Los Angeles, USA), Wei Wang (UC Los Angeles, USA), and Chen Li (UC Irvine, USA)</i>	
CASA: Classification-Based Adjusted Slot Admission Control for Query Processing Engines	278
<i>Tim Zeyl (Huawei Cloud, Canada), Harshwin Venugopal (Huawei Cloud, Canada), Calvin Sun (Huawei Cloud, Canada), and Paul Larson (Huawei Cloud, Canada)</i>	
Design and Development of a Provenance Capture Platform for Data Science	285
<i>Luca Gregori (Roma Tre University, Italy), Paolo Missier (University of Birmingham, UK), Matthew Stidolph (Newcastle University, UK), Riccardo Torlone (Roma Tre University, Italy), and Alessandro Wood (Roma Tre University, Italy)</i>	
Collaboration Management for Federated Learning	291
<i>Marius Schlegel (TU Ilmenau, Germany), Daniel Scheliga (TU Ilmenau, Germany), Kai-Uwe Sattler (TU Ilmenau, Germany), Marco Seeland (TU Ilmenau, Germany), and Patrick Mäder (TU Ilmenau, Germany)</i>	

International Workshop on Databases and Machine Learning (DBML'24)

Directions Towards Efficient and Automated Data Wrangling with Large Language Models	301
<i>Zeyu Zhang (University of Amsterdam; Amsterdam UMC), Paul Groth (University of Amsterdam), Iacer Calixto (Amsterdam UMC; University of Amsterdam), and Sebastian Schelter (University of Amsterdam)</i>	
Relationalizing Tables with Large Language Models: The Promise and Challenges	305
<i>Zezhou Huang (Columbia University, USA) and Eugene Wu (Columbia University, USA)</i>	
ChimeraTL: Transfer Learning in DBMS with Fewer Samples	310
<i>Tatsuhiko Nakamori (Keio University), Shohei Matsuura (LY Corporation), Takashi Miyazaki (LY Corporation), Sho Nakazono (LY Corporation), Taiki Sato (LY Corporation), Takashi Hoshino (Cybozu Labs), and Hideyuki Kawashima (Keio University)</i>	
Will Sharing Metadata Leak Privacy?	317
<i>Danning Zhan (TUDelft) and Rihan Hai (TUDelft)</i>	
ReClean: Reinforcement Learning for Automated Data Cleaning in ML Pipelines	324
<i>Mohamed Abdelaal (Software AG, Germany), Anil Bora Yayak (Friedrich-Alexander-University of Erlangen-Nuerenberg, Germany), Kai Klede (Friedrich-Alexander-University of Erlangen-Nuerenberg, Germany), and Harald Schoening (Software AG, Germany)</i>	
OPTWIN: Drift Identification with Optimal Sub-Windows	331
<i>Mauro D. L. Tosi (University of Luxembourg, Luxembourg) and Martin Theobald (University of Luxembourg, Luxembourg)</i>	
Evaluating Ambiguous Questions in Semantic Parsing	338
<i>Simone Papicchio (Politecnico di Torino, Italy), Paolo Papotti (EURECOM, France), and Luca Cagliero (Politecnico di Torino, Italy)</i>	

3rd Workshop on Search, Exploration, and Analysis in Heterogenous Datastore, Graph Data Edition (SEAGraph'24)

An Empirical Evaluation of Variable-Length Record B+Trees on a Modern Graph Database System	343
<i>Georgios Theodorakis (Neo4j, UK), James Clarkson (Neo4j, UK), and Jim Webber (Neo4j, UK)</i>	
Integrating Complex Pangenome Graphs	350
<i>Jérôme Arnoux (Paris Saclay University), Angela Bonifati (Lyon 1 University), Alexandra Calteau (Paris Saclay University), Stefania Dumbrava (SAMOVAR/Inst. Poltech de Paris, ENSIIE), and Guillaume Gautreau (Université Paris-Saclay, INRAE)</i>	
Towards View Management in Graph Databases	355
<i>Mohanna Shahrads (McGill University), Yu Ting Gu (McGill University), Yunjia Zheng (McGill University), and Bettina Kemme (McGill University)</i>	

The Future of Graph-Based Spatial Pattern Matching (Vision Paper)	360
<i>Nicole R. Schneider (University of Maryland, USA), Kent O’Sullivan (University of Maryland, USA), and Hanan Samet (University of Maryland, USA)</i>	
Finding the PG Schema of any (semi)structured Dataset: A Tale of Graphs and Abstraction	365
<i>Nelly Barret (Inria & Institut Polytechnique de Paris, France), Tudor Enache (Ecole Polytechnique, France), Ioana Manolescu (Inria & Institut Polytechnique de Paris, France), and Madhulika Mohanty (Inria & Institut Polytechnique de Paris, France)</i>	
Graph Lenses Over any Data: the ConnectionLens Experience	370
<i>Oana Balalau (Inria & Institut Polytechnique de Paris), Nelly Barret (Inria & Institut Polytechnique de Paris), Simon Ebel (Inria & Institut Polytechnique de Paris), Théo Galizzi (Inria & Institut Polytechnique de Paris), Ioana Manolescu (Inria & Institut Polytechnique de Paris), and Madhulika Mohanty (Inria & Institut Polytechnique de Paris)</i>	
Towards User-Centric Graph Repairs	375
<i>Amedeo Pachera (University of Lyon 1, France), Angela Bonifati (University of Lyon 1, France), and Andrea Mauri (University of Lyon 1, France)</i>	
View-Based Explanations for Graph Neural Networks (Extended Abstract)	377
<i>Tingyang Chen (Zhejiang University, China), Dazhuo Qiu (Aalborg University, Denmark), Yinghui Wu (Case Western Reserve University, USA), Arijit Khan (Aalborg University, Denmark), Xiangyu Ke (Zhejiang University, China), and Yunjun Gao (Zhejiang University, China)</i>	
Compact Path Representations for Graph Database Pattern Matching	379
<i>Wim Martens (University of Bayreuth), Matthias Niewerth (University of Bayreuth), Tina Popp (University of Bayreuth), Carlos Rojas (IMFD Chile), Stijn Vansummeren (UHasselt Data Science Institute), and Domagoj Vrgoč (PUC Chile and IMFD Chile)</i>	
OBDF: OBDA + Data Federation – Extended Abstract	381
<i>Zhenzhen Gu (Nanchang University, China), Diego Calvanese (Free University of Bozen-Bolzano, Italy; Umeå University, Sweden), Marco Di Panfilo (Free University of Bozen-Bolzano, Italy), Davide Lanti (Free University of Bozen-Bolzano, Italy), Alessandro Mosca (Free University of Bozen-Bolzano, Italy), and Guohui Xiao (University of Bergen, Norway)</i>	

Joint International Workshop on Big Data Management on Emerging Hardware and Data Management on Virtualized Active Systems (HardBD & Active)

Give a JIT on GPUs: NVRTC for Code-Generating Database Systems	384
<i>Anton Sachnov (University of Bamberg), Leonard von Merzljak (TUM), and Maximilian E. Schüle (University of Bamberg)</i>	

HunIPU: Efficient Hungarian Algorithm on IPU's	388
<i>Cheng Huang (Aarhus University), Alexander Mathiasen (Graphcore), Josef Dean (Graphcore), Davide Mottin (Aarhus University), and Ira Assent (Aarhus Univeristy)</i>	
A Three-Tier Buffer Manager Integrating CXL Device Memory for Database Systems	395
<i>Niklas Riekenbrauck (Hasso Plattner Institute, Germany), Marcel Weisgut (Hasso Plattner Institute, Germany), Daniel Lindner (Hasso Plattner Institute, Germany), and Tilmann Rabl (Hasso Plattner Institute, Germany)</i>	
CPU and GPU Hash Joins on Skewed Data	402
<i>Yuzhou Cai (University of Chinese Academy of Sciences) and Shimin Chen (University of Chinese Academy of Sciences)</i>	
Author Index	409