

32nd Italian Symposium on Advanced Database Systems (SEBD 2024)

Villasimius, Italy
23-26 June 2024

ISBN: 979-8-3313-3332-4

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© for individual papers are retained by the authors.
All rights reserved.

Printed by Curran Associates, Inc. (2026)

Edited by:
Maurizio Atzori
Paolo Ciaccia
Federica Mandreoli
Donato Malerba
Manuela Sanguinetti
Antonio Pellicani
Federico Motta

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
Email: curran@proceedings.com
Web: www.proceedings.com

SEBD 2024

Symposium of Advanced Database Systems 2024

Proceedings of the 32nd Symposium of Advanced Database Systems

Villasimius, Italy, June 23rd to 26th, 2024.

Edited by

Maurizio Atzori *

Paolo Ciaccia **

Michelangelo Ceci ***

Federica Mandreoli ****

Donato Malerba ***

Manuela Sanguinetti *

Antonio Pellicani ***

Federico Motta ****

* [Università degli Studi di Cagliari](#), Cagliari, Italy

** [Alma Mater Studiorum - Università di Bologna](#), Bologna, Italy

*** [Università degli Studi di Bari](#), Bari, Italy

**** [Università degli Studi di Modena e Reggio Emilia](#), Modena, Italy

Table of Contents

- [Preface](#) I-V

There were 76 papers submitted for peer-review to SEBD 2024 main track. Out of these, 66 were accepted for this volume (11 regular papers and 55 discussion papers). Out of them, 38 were presented as an oral

presentation and 28 as a booster and poster presentation. Additionally, 5 papers were accepted to be presented at the Doctoral Consortium.

Keynotes

- [Enhancing Data Precision with Large Language Models: Analyzing Failures and Innovating Database Curation](#) 1-2
Georg Gottlob
- [Detecting and Fixing Unfairness in Data-Driven Decision Making](#) 3
H.V. Jagadish
- [Scalable Vector Analytics: A Story of Twists and Turns](#) 4-5
Themis Palpanas

Taste-of-time Award: Winner

- [Tasting the Time: How M-Tree Broke the “27 Club” Curse](#) 6-8
Paolo Ciaccia, Marco Patella, Fausto Rabitti, Pavel Zezula

Taste-of-time Award: Honorable Mention

- [Knowledge Discovery su Schemi per l’Integrazione di Basi di Dati](#) 9-13
Massimo La Camera, Luigi Palopoli, Domenico Saccà, Domenico Ursino

Knowledge Graphs and Semantic Analysis

- [Building Taxonomies with Triplet Queries](#) 14-24
Donatella Firmani, Sainyam Galhotra, Barna Saha, Divesh Srivastava
- [Mining Validating Shape for Large Knowledge Graphs via Dynamic Reservoir Sampling](#) 25-34
Matteo Lissandrini, Kashif Rabbani, Katja Hose
- [A Provenance-Based Caching System to Speed-up SPARQL Query Answering](#) 35-50
Gianmaria Silvello, Dennis Dosso
- [Computing the Why-Provenance for Datalog Queries via SAT Solvers](#) 51-60
Marco Calautti, Ester Livshits, Andreas Pieris, Markus Schneider
- [Initial achievements in relation extraction from RNA-focused scientific papers](#) 61-69
Emanuele Cavalleri, Mauricio Soto Gomez, Ali Pashaeibarough, Dario Malchiodi, Harry Caufield, Justin Reese, Christopher J. Mungall, Peter Robinson, Elena Casiraghi, Giorgio Valentini, Marco Mesiti
- [Towards a Standard for Triggers in Property Graphs](#) 70-79
Luigi Bellomarini, Anna Bernasconi, Stefano Ceri, Alessia Gagliardi, Davide Magnanimiti, Davide Martinenghi
- [ReliK: A Reliability Measure for Knowledge Graph Embeddings](#) 80-90
Maximilian K. Egger, Wenyue Ma, Davide Mottin, Panagiotis Karras, Ilaria Bordino, Francesco Gullo, Aris Anagnostopoulos
- [The ARIADNEplus Knowledge Base: a Linked Open Data set for archaeological research](#) 91-100
Alessia Bardi, Miriam Baglioni, Michele Artini, Andrea Mannocci, Gina Pavone

Data Mining and Business Intelligence

- [Assessing Speech Model Performance: A Subgroup Perspective](#) 101-111
Alkis Koudounas, Eliana Pastor, Elena Baralis
- [An Innovative Big Temporal Data Analytics Technique over Real-Life Healthcare Datasets: The F-TBDA Approach](#) 112-121

- Alfredo Cuzzocrea, Geertruida H. de Bock, Willemijn J. Maas, Selim Soufragi, Abderraouf Hafsaoui
- [Emotional Data Querying: The BI Scenario](#) 122-130
Sandro Bimonte, Patrick Marcel, Stefano Rizzi
- [Colossal Trajectory Mining: Semantic Co-movement Pattern Mining](#) 131-141
Chiara Forresi, Matteo Francia, Enrico Gallinucci, Matteo Golfarelli, Manuele Pasini
- [Compressing Big OLAP Data Cubes over Mobile Clouds: A Hierarchy-Based Data Partitioning Approach](#) 142-151
Alfredo Cuzzocrea, Mojtaba Hajian, Abderraouf Hafsaoui
- [Max Flow Vulnerability of Undirected Planar Networks](#) 152-162
Lorenzo Balzotti, Paolo G. Franciosa

Advanced AI Techniques in Databases

- [Bootstrapping Gene Expression-Cancer Knowledge Bases with Limited Human Annotations](#) 163-173
Stefano Marchesin, Laura Menotti, Fabio Giachelle, Gianmaria Silvello, Omar Alonso
- [Speeding up Vision Transformers Through Reinforcement Learning](#) 174-184
Francesco Cauteruccio, Michele Marchetti, Davide Traini, Domenico Ursino, Luca Virgili
- [Verification of Unary Communicating Datalog Programs](#) 185-194
Cyriac Aiswarya, Diego Calvanese, Francesco Di Cosmo, Marco Montali
- [Nowcasting of the energy production of wind power plants through spatially-aware model trees](#) 195-204
Annunziata D'Aversa, Gianvito Pio
- [Data Filtering for a Sustainable Model Training](#) 205-216
Francesco Scala, Sergio Flesca, Luigi Pontieri
- [Mitigating Unfairness in Machine Learning: A Taxonomy and an Evaluation Pipeline](#) 217-226
Chiara Criscuolo, Tommaso Dolci, Mattia Salnitri
- [Refining Triplet Sampling for Improved Self-Supervised Representation Learning](#) 227-246
Manuel Goyo, Giacomo Frisoni, Gianluca Moro, Claudio Sartori
- [Using Graph Neural Networks for Heterogeneous Event Classification](#) 247-259
Valerio Bellandi, Stefano Montanelli, Darya Shlyk, Stefano Siccardi

Natural Language Processing in Databases

- [Text-to-SQL with Large Language Models: Exploring the Promise and Pitfalls](#) 260-270
Luca Sala, Giovanni Sullutrone, Sonia Bergamaschi
- [From Product Sheet to Text and Video: A NLG Pipeline to Transform Structured Data into Comprehensive Descriptions](#) 271-280
Andrea Avignone, Alessandro Fiori, Silvia Chiusano, Giuseppe Rizzo
- [Efficient and Effective Multi-Vector Dense Retrieval with EMVB](#) 281-289
Franco Maria Nardini, Cosimo Rulli, Rossano Venturini
- [CAMEO: Fostering Joint Conversational Search and Recommendation](#) 290-301
Tommaso Di Noia, Guglielmo Faggioli, Marco Ferrante, Nicola Ferro, Fedelucio Narducci, Raffaele Perego, Giuseppe Santucci
- [A Framework for the Generation of Training Examples from Tabular Data](#) 302-311
Jean-Flavien Bussotti, Paolo Papotti, Donatello Santoro, Enzo Veltri
- [Clustering Amendments with Semantic Embeddings](#) 312-320
Alessandro Sajeve, Stefano Iannucci, Carlo Marchetti, Paolo Merialdo, Riccardo Torlone
- [Impact of Data Augmentation on Hate Speech Detection in Roman Urdu](#) 321-330
Fariha Maqbool, Blerina Spahiu, Andrea Maurino
- [Named Entity Recognition using context similarity data augmentation](#) 331-338
Ilaria Bartolini, Angelo Chianese, Vincenzo Moscato, Marco Postiglione, Giancarlo Sperli, Andrea Vignali

Process Mining

- [Attacking Maritime Control Systems Through Process Mining](#) 339-348
Giacomo Longo, Francesco Lupia, Andrea Pugliese, Enrico Russo
- [Enhancing Next Activity Prediction with Adversarial Training of Vision Transformers](#) 349-358
Vincenzo Pasquadibisceglie, Annalisa Appice, Giovanna Castellano, Donato Malerba
- [Process-level Model Repair through Instance Graphs Representations](#) 359-367
Laura Genga, Claudia Diamantini, Emanuele Storti, Domenico Potena
- [Exploring Large Language Models for Procedure Extraction from Documents](#) 368-376
Anisa Rula, Jennifer D'Souza

Applications

- [Evaluating status and value assortativity in Threads](#) 377-386
Gianluca Bonifazi, Enrico Corradini, Domenico Ursino
- [Design of a Telemedicine Infrastructure for Rural and Remote Areas](#) 387-391
Pietro Hiram Guzzi, Patrizia Vizza, Giuseppe Tradigo, Sergio Greco Pierangelo Veltri
- [Food Certification through Collaborative Sensory Analysis Methods and Tools](#) 392-400
Ada Bagozi, Devis Bianchini
- [“Dead or Alive, we can deny it”. A Differentially Private Approach to Survival Analysis](#) 401-411
Francesco L. De Faveri, Guglielmo Faggioli, Nicola Ferro, Riccardo Spizzo
- [ASYDE: An Argumentation-based System for classifying Driving behaviors](#) 412-421
Bettina Fazzinga, Sergio Flesca, Filippo Furfaro, Giuseppina Monterosso, Irina Trubitsyna
- [Infantile Predictors of Functional Gastrointestinal Disorders: A Machine Learning Approach to Risk Assessment](#) 422-430
Flavia Indrio, Elio Masciari, Flavia Marchese, Matteo Rinaldi, Gianfranco Maffei, Enea Vincenzo Napolitano, Isadora Beghetti, Luigi Corvaglia, Arianna Aceti
- [Machine Learning-Augmented Ontology-Based Data Access for Renewable Energy Data](#) 431-445
Marco Calautti, Damiano Duranti, Paolo Giorgini
- [Automated Knowledge Extraction from Legal Texts using ASKE](#) 446-455
Silvana Castano, Alfio Ferrara, Stefano Montanelli, Sergio Picascia, Davide Riva
- [Integrating Brain Networks and Multi-Modal Data for Early Detection of Alzheimer’s Disease](#) 456-467
Carmela Comito, Clara Pizzuti, Marcello Sammarra, Annalisa Socievole
- [PNRRorienta: A Web Application for Managing Schools, Courses, and Students Involved in the PNRR Orientation Initiative](#) 468-476
Andrea Pasin, Lorenza Da Re, Andrea Gerosa, Lidia Pezzuoli, Silvia Preciso, Nicola Ferro
- [TypoAlert: a browser extension against typosquatting](#) 477-485
Francesco Blefari, Angelo Furfaro, Giovambattista Ianni, Alessandro Viscomi

Data Quality, Preprocessing and Data Analysis

- [The Future of Sustainable Data Preparation](#) 486-497
Barbara Pernici, Cinzia Cappiello, Edoardo Ramalli, Matteo Palmonari, Federico Belotti, Flavio De Paoli, Angelo Mozzillo, Luca Zecchini, Giovanni Simonini, Sonia Bergamaschi, Tiziana Catarci, Matteo Filosa, Marco Angelini, Dario Benvenuti
- [Comparing Incomplete Database Instances](#) 498-507
Boris Glavic, Giansalvatore Mecca, Renée J. Miller, Paolo Papotti, Donatello Santoro, Enzo Veltri
- [From why-provenance to why+provenance: Towards addressing deep data explanations in Data-Centric AI](#) 508-517
Paolo Missier, Riccardo Torlone
- [Autonomous Intelligent Systems: From Illusion of Control to Inescapable Delusion](#) 518-527
Stéphane Grumbach, Giorgio Resta, Riccardo Torlone
- [POLARIS: A framework to guide the development of Trustworthy AI systems](#) 528-537
Maria Teresa Baldassarre, Danilo Caivano, Domenico Gigante, Azzurra Ragone

- [Data Pipelines Assessment: The Role of Data Engine Deployment Models](#) 538-550
Claudio A. Ardagna, Valerio Bellandi, Marco Luzzara, Antongiaco Polimeno
- [Schema Decomposition via Transformation Patterns](#) 551-564
Théo Abgrall
- [Improving Malicious Accounts Discrimination through a New Feature Engineering Approach Using Relaxed Functional Dependencies](#) 565-573
Loredana Caruccio, Gaetano Cimino, Stefano Cirillo, Domenico Desiato, Giuseppe Polesea, Genoveffa Tortora

Interpretable AI

- [A Comparative Assessment of eXplainable AI Tools in Predicting Hard Disk Drive Health](#) 574-584
Flora Amato, Antonino Ferraro, Antonio Galli, Valerio La Gatta, Francesco Moscato, Vincenzo Moscato, Marco Postiglione, Carlo Sansone, Giancarlo Sperli
- [Causal Mediation Analysis for Interpreting Large Language Models](#) 585-594
Elisabetta Rocchetti, Alfio Ferrara
- [A Clustering-based Approach for Interpreting Black-box Models](#) 595-603
Luca Ferragina, Simona Nisticò
- [Symbolic Regression for Transparent Clinical Decision Support: A Data-Centric Framework for Scoring System Development](#) 604-614
Veronica Guidetti, Federica Mandreoli
- [Data and System Traceability for Transparent AI in Medical Imaging](#) 615-623
Sara Colantonio, Andrea Berti, Gianluca Carloni, Claudia Caudai, Giulio Del Corso, Danila Germanese, Eva Pachetti, Maria Antonietta Pascali, Varvara Kalokyri, Haridimos Kondylakis, Charalampos Kalantzopoulos, Nikolaos Tachos, Dimitris Fotiadis, Valentina Giannini, Simone Mazzetti, Daniele Regge, Nickolas Papanikolaou, Konstantinos Marias, Manolis Tsiknakis

Entity Resolution and Data Integration

- [FDup framework: A General-purpose solution for Efficient Entity Deduplication of Record Collections](#) 624-632
Michele De Bonis, Claudio Atzori, Sandro La Bruzzo, Paolo Manghi
- [Combining Entity Resolution and Query Answering in Ontologies: A Formal Conceptual Framework](#) 633-642
Ronald Fagin, Phokion G. Kolaitis, Domenico Lembo, Lucian Popa, Federico Scafoglieri
- [Overlap-Based Duplicate Table Detection](#) 643-652
Luca Zecchini, Tobias Bleifuß, Giovanni Simonini, Sonia Bergamaschi, Felix Naumann
- [Personalised Exploration Graphs on top of Data Lakes](#) 653-661
Devis Bianchini, Valeria De Antonellis, Massimiliano Garda
- [How Transformers Are Revolutionizing Entity Matching](#) 662-670
Matteo Paganelli, Donato Tiano, Francesco Del Buono, Andrea Baraldi, Riccardo Benassi, Giacomo Guiduzzi, Francesco Guerra
- [Privacy-Preserving Data Integration for Health: Adhering to OMOP-CDM Standard](#) 671-680
Lisa Trigiante, Domenico Beneventano
- [A Minimum Metadataset for Data Lakes Supporting Healthcare Research](#) 681-691
Davide Piantella, Pierluigi Reali, Priyansh Kumar, Letizia Tanca
- [Identifying key factors in designing data spaces for Urban Digital Twin Platforms: a data driven approach](#) 692-703
Cristian Martella, Angelo Martella, Amro Issam Hamed Attia Ramadan, Antonella Longo

Doctoral Consortium

- [Understanding Emerging Cyber Attacks and Vulnerabilities Targeting Maritime Systems](#) 704-709
Giacomo Longo
 - [Data/Process Analysis for Advanced Interoperable Cyber Ranges](#) 710-716
Giuseppe Salerno
 - [Large Language Models integration in Digital Humanities](#) 717-723
Giovanni Sullutrone
 - [Symbolic Regression for Medical Scoring Systems: a Bayesian and Multi-Objective Approach](#) 724-730
Mattia Billa
 - [A Lakehouse-based platform for Data-driven Sustainability Monitoring in Energy-Intensive Production](#)
731-737
Paola Magrino
-