Lernen, Wissen, Daten, Analysen (LWDA) Conference Proceedings

LWDA'23

October 09-11, 2023
Marburg, Germany

CEUR-WS.org/Vol-3630/Preface.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

Editors:
Michael Leyer
Philipps-University of Marburg, Germany/Queensland University of Technology, Australia

Johannes Wichmann
Philipps-University of Marburg, Germany

© 2023 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

Preface

LWDA 2023 conference provides a joint forum for experienced and young researchers, to bring insights to recent trends, technologies and applications and to promote interaction in the research field of big data and beyond.

The acronym LWDA expands in German to "Lernen. Wissen. Daten. Analysen." (English: "Learning. Knowledge. Data. Analytics."). Recent research in the field is presented and discussed from the viewpoint of machine learning, data mining, knowledge extraction, knowledge management, information retrieval, personalization, database management, information systems, big data management and big data analytics to name a few.

The LWDA conference series comprises the workshops BIA, DB, FGWM, IR and KDML which are organized by the respective special interest groups within the German Computer Science Society:

- FG-BIA 2023 – Business Intelligence & Analytics
- FG-DB 2023 - Database Systems
- FG-FGWM 2023 - Knowledge Management
- FG-IR 2023 - Information Retrieval
- FG-KDML 2023 - Knowledge Discovery, Data Mining and Machine Learning

The papers published in LWDA 2023 proceedings have been selected by independent program committees from the respective fields. The program consists of four invited keynotes and two joint research sessions as well as the community meetings of the special interest groups. In addition to these joint sessions, there are five parallel research sessions for each of the workshops focusing on more specific topics. A joint poster session gives all presenters the opportunity to discuss their work in a broader context. This year's social program includes a city tour for further interaction on the second evening.

Our distinguished keynote speakers are:

- Prof. Dr. Hannes Mühleisen - Radboud Universiteit Nijmegen
- Prof. Dr. Erhard Rahm – University of Leipzig
- Prof. Dr. Michael Granitzer – University of Passau
- Dr. Dietrich Alexander Herberg

The working group for Digitization & Process Management at the Philipps-University of Marburg is proud to host the LWDA 2023 conference. For the technical program the organizer would like to thank the workshop chairs and their programme committees for their hard work as well as the keynote speakers for their inspiring talks. We hope the participants will keep the venue as an inspiring event with fruitful discussions in mind and the readers will enjoy studying the scientific contributions in this proceedings volume. The proceedings are published with CEUR and can be found here. http://ceur-ws.org/Vol-1917

Marburg, Germany, October 2023

Conference Organization

General Chair

Michael Leyer	Philipps-University of Marburg/Queensland University of Technology

Program Chairs

Tanja Auge	University of Regensburg
Henning Baars	University of Stuttgart
Andreas Henrich	Otto-Friedrich-University of Bamberg
Thomas Mandl	University of Hildesheim
Thorsten Papenbrock	Philipps-University of Marburg
Pascal Reuss	University of Hildesheim
Jakob Schönborn	University of Hildesheim
Helge Spieker	Simula Research Laboratory
Felix Stamm	Rheinisch-Westfälische Technische Hochschule Aachen

Program Committee

Bernhard Seeger	Philipps-University of Marburg
Hazar Harmouch	Hasso-Plattner-Institute
Uta Störl	Fernuniversität Hagen
Stefan Schulte	TUHH Institute for Data Engineering
Fabian Panse	Hasso-Plattner-Institute
Benjamin Hättasch	TU Darmstadt
Annett Ungethüm	TU Dresden
Marina Tropmann-Frick	HAW Hamburg
Hannes Grunert	University of Rostock
Carsten Felden	TU Bergakademie Freiberg
Ralf Finger	Information Works
Sebastian Olbrich	Deloitte
Martin Atzmueller	University of Osnabrück
Christian Bauckhage	Fraunhofer IAIS / University of Bonn
Ulf Brefeld	Leuphana University of Lüneburg
Mirko Bunse	TU Dortmund
Dennis Groß	Radboud University
Andreas Hotho	University of Würzburg
Eyke Hüllermeier	LMU Munich
Robert Jäschke	HU Berlin
Christian Kühnert	Fraunhofer IOSB
Thomas Liebig	TU Dortmund
Thomas Seidl	LMU Munich

Pascal Welke	TU Wien
Stefan Wrobel	Fraunhofer IAIS / University of Bonn
Kerstin Bach	Norwegian University of Science and Technology
Joachim Baumeister	University of Würzburg
Ralph Bergmann	University of Trier
Viktor Eisenstadt	Universiy of Applied Sciences of Hanover
Jörg Cassens	University of Hildesheim
Lisa Grumbach	University of Trier
Andrea Kohlhase	University of Applied Sciences Neu-Ulm
Michael Kohlhase	University of Nuremberg-Erlangen
Ulrich Reimer	University of Applied Sciences of St. Gallen
Christian Severin Sauer	University of Hildesheim
Christian Zeyen	DFKI
Andreas Korger	denkbares
Johannes Wichmann	Philipps-University of Marburg
Mirjam Minor	University of Frankfurt
Carsten Wenzel	University of Hildesheim
Anna Faust	HU Berlin
Klaus Berberich	University of Applied Sciences of Saarbrücken
Norbert Fuhr	University of Duisburg-Essen
Ralf Krestel	Christian-Albrechts-University of Kiel
Christin Katharina Kreutz	Technical Applied University of Cologne
Jochen L. Leidner	Applied University of Coburg
Dirk Lewandowski	HAW Hamburg
Philipp Schaer	Technical Applied University of Cologne
Ralf Schenkel	University of Trier

Table of Contents

Risk Identification of Data Science Projects: A Literature Review 1
Maike Holtkemper, Maria Potanin, Alexander Oberst and Christian Beecks

Designing an Analytical Control Chart System with ML-predicted Quality Characteristics 14
Till Carlo Schelhorn, Jonas Gunklach and Alexander Maedche

Exploiting Foundation Models for Spoken Language Identification 28
Benedikt Augenstein and Darjan Salaj

Accelerating literature screening for systematic literature reviews with Large Language Models – development, application, and first evaluation of a solution 41
Paul Herbst and Henning Baars

Datengenossenschaften als Datentreuhänder – Eine qualitative Analyse von Pilotprojekten 52
Maximilian Werling, Patrick Weber and Heiner Lasi

Governance of Artificial Intelligence – A Framework Towards Ethical AI Applications 63
Jens F. Lachenmaier, Maximilian Werling and Dominik Morar

Integrating Machine Learning into SQL with Exasol 73
Christoph Großmann and Johannes Schildgen

Database and Workflow Optimizations for Spatial-Geometric Queries in GeoMine 86
Martin Poppinga, Joel Graef, Konrad Diedrich, Matthias Rarey and Norbert Ritter

SKYSHARK: A Benchmark with Real-world Data for Line-rate Stream Processing with FPGAs 98
Maximilian Langohr, Tim Vogler and Klaus Meyer-Wegener

SuMExplorer: Summarisation-based Frequent Subgraph Mining for Visual Exploratory Subgraph Searching 110
Chimi Wangmo and Lena Wiese

Enhancing Data Acquisition and Fault Analysis for Large-Scale Facilities: A Case Study on the Laser-Based Synchronization System at the European X-Ray Free-Electron Laser 121
Arne Grünhagen, Maximilian Schütte, Annika Eichler, Marina Tropmann-Frick and Görschwin Fey

Heterogeneity in NoSQL Databases —Challenges of Handling schema-less Data 134
Mark Lukas Möller, Dominique Hausler, Sebastian Strasser, Tanja Auge and Meike Klettke

Pythagoras: Semantic Type Detection of Numerical Data Using Graph Neural Networks 146
Sven Langenecker, Christoph Sturm, Christian Schalles and Carsten Binnig

Patient trajectory visualization for FHIR healthcare data: A use case on melanoma patients 153
Meijie Li, Wolfgang Galetzka, Bahadir Eryilmaz, Georg Christian Lodde, Elisabeth Livingstone, Jörg Schlötterer and Christin Seifert

CLEARNESS: Coreference Resolution for Generating and Ranking Arguments Extracted from Debate Portals for Queries 161
Johannes Weidmann, Lorik Dumani and Ralf Schenkel

The Information Retrieval Experiment Platform 175
Maik Fröbe, Jan Heinrich Reimer, Sean MacAvaney, Niklas Deckers, Janek Bevendorff, Benno Stein, Matthias Hagen and Martin Potthast

Applied Face Recognition in the Humanities 179
Martin Bullin and Andreas Henrich

Automatic Classification of Portraits: Application of Transformer and CNN Based Models for an Art Historic Dataset 192
Sebastian Diem and Thomas Mandl

Comparative Survey of German Hate Speech Datasets: Background, Characteristics and Biases 207
Markus Bertram, Johannes Schäfer and Thomas Mandl

Preliminary Results of a Scientometric Analysis of the German Information Retrieval Community 2020-2023 222
Philipp Schaer, Svetlana Myshkina and Jüri Keller

A Testbed for Dual-Entity Knowledge Panels 231
Leon Martin and Andreas Henrich

Vertical Search Scenarios within a Digital Study Planning Assistant 239
Tobias Hirmer, Michaela Ochs and Andreas Henrich

Integrating BDI Agents with the MATSim Traffic Simulation for Autonomous Mobility on Demand 247
Marcel Mauri, Ömer Ibrahim Erduran, Thu Pham Dieu Anh and Mirjam Minor

KIRETT: Knowledge-Graph-Based Smart Treatment Assistant for Intelligent Rescue Operations 259
Mubaris Nadeem, Johannes Zenkert, Lisa Bender, Christian Weber and Madjid Fathi

SKOS-Utils: Developing and Checking SKOS Knowledge Graphs (Tool Presentation) 271
Joachim Baumeister and Valentin Roß

Case-Based Sample Generation using Multi-Armed Bandits 282
Andreas Korger and Joachim Baumeister

Combining Information Retrieval and Large Language Models for a Chatbot that Generates Reliable, Natural-style Answers 298
Andreas Lommatzsch, Brandon Llanque, Vinay Srinath Rosenberg, Syed Ali Murad Tahir, Hristo Dimitrov Boyadzhiev and Maurice Walny

The Data Dilemma: Google Analytics' Untapped Potential and Web Data Literacy 311
Tom Alby

Bridging the Gap: Examining the trust dimensions of smart contracts using supply chain applications 325
Wieland Müller and Michael Leyer

A Feature-wise Comparative Assessment of the CBR-based Methodologies FLEA and SEASALT 339
Viktor Eisenstadt, Jessica Bielski, Christoph Langenhan, Klaus-Dieter Althoff and Andreas Dengel

Comparative Analysis of Text-Based CBR Algorithms for Cybercrime Profiling Investigations..... 347
Marc Krüger

Cover Song Identification in Practice with Multimodal Co-Training..... 359
Simon Hachmeier and Robert Jäschke

Higher-Order DeepTrails: Unified Approach to *Trails..... 372
Tobias Koopmann, Jan Pfister, André Markus, Astrid Carolus, Carolin Wienrich and Andreas Hotho

Fast k-Nearest-Neighbor-Consistent Clustering 387
Lars Lenssen, Niklas Strahmann and Erich Schubert

Preprocessing Ground-Based Hyperspectral Image Data for Improving CNN-based Classification 399
Andreas Schliebitz, Heiko Tapken and Martin Atzmueller

Free-Energy Advantage Functions for Policy Transfer to Noisy Environments with Safety Constraints 414
Pierre Haritz and Thomas Liebig

Automatic Speech Detection on a Smart Beehive's Raspberry Pi..... 424
Pascal Janetzky, Philip Lissmann, Andreas Hotho and Anna Krause

Comparing Humans and Algorithms in Feature Ranking: A Case-Study in the Medical Domain 430
Jonas Hanselle, Jaroslaw Kornowicz, Stefan Heid, Kirsten Thommes and Eyke Hüllermeier

Biomedical Event Extraction with Generative Language Models 442
Fabio Barth, Leon Weber-Genzel and Ulf Leser

Liquor-HGNN: A heterogeneous graph neural network for leakage detection in water distribution networks 454
Melanie Schaller, Michael Steininger, Andrzej Dulny, Daniel Schlör and Andreas Hotho

A Document Tagging Support System for Nursing Care Experts..... 470
Beat Tödtli1, Sebastian Müller, Melanie Rickenmann, Janine Vetsch and Simon Haug

Efficient Light Source Placement using Quantum Computing 478
Sascha Mücke and Thore Gerlach

Contextual Preselection Methods in Pool-based Realtime Algorithm Configuration..... 492
Jasmin Brandt, Elias Schede, Shivam Sharma, Viktor Bengs, Eyke Hüllermeier and Kevin Tierney

A Few Models to Rule Them All: Aggregating Machine Learning Models..... 506
Florian Siepe, Phillip Wenig and Thorsten Papenbrock

Applicability of Models Trained on Generated Clinical German Datasets on Out-domain Data 521
Oğuz Şerbetçi and Ulf Leser