



Editorial

Ralf Krestel¹ · Udo Kruschwitz² · Michael Wiegand³ · Theo Härder⁴

Accepted: 9 March 2023 / Published online: 22 March 2023
© The Author(s) 2023

1 Schwerpunktthema „Trends in Social Media Analysis to Address Fake News, Hate Speech, or Bias“

Social media has many benefits: from staying in contact with close and not-so-close friends, over exercising the right to voice one's opinion, to communicating with many like-minded people all over the world and providing an additional channel for information exchange.

Unfortunately, social media has also been abused and misused ever since its inception. Hate speech is prevalent on many sites alienating trusting users and hindering fruitful discussions. Fake news are distributed through social media platforms with dangerous effects. But even without malicious intention, social media can be misleading due to various biases in the system.

In this special issue of *Datenbank-Spektrum*, we will explore and present current trends in the field of automatically detecting and managing hate speech, fake news, bias and other toxic content in the context of social media. We solicited novel research contributions and received eight sub-

missions in total. Each submission was reviewed by two independent reviewers and then discussed among the editors. Four papers were finally accepted covering a broad range of topics.

The first article entitled *Automated multilingual detection of Pro-Kremlin propaganda in newspapers and Telegram posts* by Veronika Solopova, Oana-Juliana Popescu, Christoph Benzmüller and Tim Landgraf explores the specific natural language processing problem of propaganda detection and looks at it both with a quantitative as well as a qualitative angle. Transformer-based and linguistically motivated approaches are investigated and compared. This leads to some surprising findings including the use of some seemingly neutral lexical items that turn out to be strong indicators of propaganda in this context. The authors hope to make a contribution towards understanding patterns of fake news or propaganda detection and in particular offering some steps towards addressing the problem in less resourced languages.

The next contribution, *Generalizability of Abusive Language Detection Models on Homogeneous German Datasets* by Nina Seemann, Yeong Su Lee, Julian Höllig and Michaela Geierhos compares the compatibility of different German datasets for abusive language detection from the GermEval and HASOC evaluation campaigns. The authors find that a combination of different datasets is not always beneficial. Its effectiveness not only depends on the similarity of annotation schemes but also on the source from which the respective text samples have been drawn. The authors take into consideration different types of learning methods. An error analysis of the output of the strongest classifier provides more insights into the properties of the different datasets.

The third article *Moving Beyond Benchmarks and Competitions: Towards Addressing Social Media Challenges in an Educational Context* by Dimitri Ognibene, Gregor Donabauer, Emily Theophilou, Sathya Bursic, Francesco Lomonaco, Rodrigo Wilkens, Davinia Hernandez-Leo, and Udo Kruschwitz emphasizes the importance of not just chasing state-of-the-art performance measures when tack-

✉ Ralf Krestel
r.krestel@zbw.eu

Udo Kruschwitz
Udo.Kruschwitz@ur.de

Michael Wiegand
michael.wiegand@aau.at

Theo Härder
haerder@cs.uni-kl.de

¹ ZBW Leibniz-Informationszentrum Wirtschaft & CAU Kiel, Düsternbrooker Weg 120, 24105 Kiel, Germany

² Informationswissenschaft, Universität Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany

³ Digital Age Research Center, Alpen-Adria-Universität Klagenfurt, Universitätsstr. 65–67, 9020 Klagenfurt, Austria

⁴ AG DBIS, RPTU Kaiserslautern-Landau, 67663 Kaiserslautern, Germany

ling fake news, filter bubbles, or cyberbullying. Instead they argue that such social media threats should be addressed by educating users with a focus on teenagers. Hence, the efforts developed as part of the COURAGE project are twofold: building multi-modal threat detectors and content analyzers on the one hand and educating users in dealing with social media threats and the output of machine learning methods active on social media sites.

The final article of this special issue *Avoiding Bias when Capturing Illegal Hate Speech* by Johannes Schäfer exemplifies the creation of a dataset sampled from Twitter for hate speech detection with a fine-grained class inventory bearing in mind the definitions of German law code (“Strafgesetzbuch (StGB)”). The annotation scheme distinguishes between 4 subclasses of hate speech: malicious gossip/defamation, incitement to commit offenses, incitement of masses, and insults. In their classification experiments, the authors also focus on the role of identity terms. While such terms are often biased towards specific classes (“identity term bias”), the authors also argue that the removal of such terms, which can be regarded as “brute-force” bias mitigation, omits essential information for a classifier to make a correct prediction.

We would like to thank the authors of all submissions for their contribution. In addition to that, we are particularly grateful to the reviewers who had to work towards a particularly tight schedule, namely Josef Ruppenhofer (IDS, Mannheim), Thomas Mandl (University of Hildesheim), Melanie Siegel (Hochschule Darmstadt), Julia Maria Struß (FH Potsdam), Supriyo Mandal (ZBW Kiel), Gregor Donabauer (University of Regensburg), Marco Viviani (University of Milan-Bicocca), Andrew MacFarlane (City University), Sean MacAvaney (University of Glasgow), Maik Fröbe (University of Jena), Elisabeth Eder (Alpen-Adria-Universität Klagenfurt), Michael Granitzer (University of Passau), Seid Muhie Yimam (Universität Hamburg), Julian Risch (deepset.ai), and Betty van Aken (BHT Berlin).

2 Fachbeitrag

Das SIMD-Paradigma wurde zum Kernprinzip für die Optimierung der Anfrageverarbeitung in spaltenorientierten Datenbanksystemen, wobei nur die LOAD/STORE-Instruktionen als effizient genug erachtet wurden, um den erwarteten Speedup zu erzielen. In ihrem Beitrag *Partition-based SIMD Processing and its Application to Columnar Database Systems* jedoch zeigen Juliana Hildebrandt, Johannes Pietrzyk, Alexander Krause, Dirk Habich und Wolfgang Lehner (TU Dresden), dass durch geeigneten Einsatz der GATHER-Instruktion die gleiche Performanz wie mit Hilfe der LOAD-Instruktion erreicht werden kann. Die Autoren entwickeln dazu ein neues Zugriffsmuster

und demonstrieren dessen Anwendbarkeit und Effizienz experimentell an zwei repräsentativen Beispielen.

3 Community-Beiträge

Die Rubrik „Dissertationen“ enthält in diesem Heft 22 Kurzfassungen von Dissertationen aus der deutschsprachigen DBIS-Community, die im vergangenen Jahr erfolgreich abgeschlossen wurden.

Die Rubrik „Community“ berichtet unter „News“ über aktuelle Informationen, welche die DBIS-Gemeinde betreffen.

4 Künftige Schwerpunktthemen

4.1 Managing Data and Metadata in Complex Enterprise Landscapes

The digital transformation generates huge amounts of heterogeneous data, across the entire lifecycle of all kinds of products and services and across all kinds of businesses. Extracting insights from these data by applying data analytics and AI constitutes a critical success factor for enterprises, e.g., to optimize processes and reinvent business models. Comprehensive analytics efforts and vast amounts of data have made enterprise data landscapes far more complex revealing globally distributed, federated and hybrid deployed structures of analytical and operational data systems. This poses new challenges to both data management and metadata management: new kinds of data platforms have emerged, e.g., data lakes, data catalogs and data marketplaces, semantic techniques for managing data and metadata are increasingly becoming popular in industry practice, data governance and data strategy concepts are developed to ensure the compliant and economically beneficial use of data.

In this special issue of Datenbank-Spektrum, we call for contributions on technical and organizational aspects of data management and metadata management in complex enterprise landscapes, interpreted broadly. We welcome original contributions – including technical papers, interdisciplinary and application-oriented papers, case studies and survey papers – relating to the following areas, but not limited to:

- Data platform architectures and technologies, e.g., data lakes, data catalogs, data marketplaces, feature stores
- Architecting and modeling data and metadata in data platforms, e.g., semantic data modeling for data lakes and data catalogs, reference data models, data model management, data model evolution

- Data engineering and metadata management for analytics and AI, e.g., for data pipelines and MLOps
- Data integration and data quality in complex enterprise landscapes, e.g., federated data integration, semantic data integration, distributed data quality assessments
- Enterprise data architecture: organizing data and metadata across the enterprise landscape, e.g., across several data lakes, data catalogs and operational systems
- Data governance and data strategy, e.g., data ownership and data stewardship across operational and analytical systems, organizational roles for data governance and data analytics, data offense and data defense concepts

Paper format: 8–10 pages, double-column (cf. author guidelines at <http://www.springer.com/13222>). Contributions either in German or in English are welcome.

Deadline for submissions: February 1st, 2023
Issue delivery: DASP-2-2023 (July 2023)

Guest editors:

Christoph Gröger, Robert Bosch GmbH, Stuttgart
christoph.groeger@de.bosch.com
Holger Schwarz, University of Stuttgart
holger.schwarz@ipvs.uni-stuttgart.de

4.2 Best Workshop Papers of BTW 2023

This special issue of the “Datenbank-Spektrum” is dedicated to the Best Papers of the Workshops, Demos and Data Science Challenge running at the BTW 2023 at the TU Dresden. The selected Workshop contributions should be extended to match the format of regular DASP papers.

Paper format: 8–10 pages, double-column
Selection of the Best Papers by the Workshop chairs and the guest editor: April 1st, 2023

Deadline for submissions: June 1st, 2023
Issue delivery: DASP-3-2023 (Nov. 2023)

Guest editor:

Uta Störl, FernUni Hagen
uta.stoerl@fernuni-hagen.de

4.3 Data Management on Quantum Hardware

With the recent availability of cloud-hosted quantum hardware, the potential of this technology for the field of data management is starting to be explored by a growing community of researchers. The topics of interest include machine learning on quantum computers, as well as core tasks of the database management system, such as query opti-

mization and transaction scheduling. This enumeration is by no means final, since research in this area is at a very early stage, and the potential availability of future hardware, such as QRAM, may offer completely new opportunities.

However, quantum software engineering for data management, as well as dealing with a completely new family of hardware, holds its challenges: Learning how to program quantum computers requires a fresh mindset and has a steep learning curve. A further challenge is dealing with the limitations of today’s early prototype hardware.

In this special issue of Datenbank-Spektrum, we will provide a dedicated forum for exploring and presenting current trends in the intersection between data management research, quantum software engineering, and architecting database systems using quantum hardware.

We welcome original contributions, including technical papers, application-oriented papers, case studies, survey papers, and position papers. In particular, we also welcome contributions from industry.

Topics of interest include, but are not limited to:

- Machine learning and quantum computing
- Natural language processing and quantum computing
- Data management applications and quantum computing
- Database systems technology and quantum computing
- Database systems architecture and quantum hardware
- Teaching quantum computing in the context of data management

Paper format: 8–10 pages, double-column (cf. author guidelines at <http://www.springer.com/13222>). We welcome contributions in both German and English.

Deadline for submissions: Oct. 1st, 2023
Publication of special issue: DASP-1-2024 (March 2024)

Guest editors:

Stefanie Scherzinger, University of Passau
stefanie.scherzinger@uni-passau.de
Uta Störl, University of Hagen
uta.stoerl@fernuni-hagen.de

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.