



Editorial

Uta Störl¹

Angenommen: 2. November 2023 / Online publiziert: 9. November 2023
© The Author(s) 2023

1 Schwerpunktthema: „Best Workshop Papers of BTW 2023“

Der Fachbereich Datenbanken und Informationssysteme (DBIS) der Gesellschaft für Informatik (GI) organisiert in einem zweijährigen Turnus seit 1985 die Konferenz „Datenbanksysteme für Business, Technologie und Web (BTW)“. Diese Konferenz stellt das zentrale Forum der Datenbank-Communities in Deutschland, in Österreich und in der Schweiz dar. Nachdem die BTW 2021 pandemiebedingt virtuell stattfand, konnte nun die 20. Auflage wieder in Präsenz in der Woche vom 6. bis 10. März 2023 an der TU Dresden stattfinden. Einen ausführlichen Bericht zur BTW gab es im letzten Heft des Datenbank-Spektrums (Band 23, Heft 2).

Seit 2001 werden in der BTW-Woche stets auch Workshops zu aktuellen Themen der Datenbankforschung durchgeführt. Diese Workshops erfreuten sich in diesem Jahr einer sehr großen Beliebtheit bei den BTW-Teilnehmerinnen und -Teilnehmern. Auf der BTW 2023 fanden vier Workshops zu folgenden Themen statt:

- Big (and Small) Data in Science and Humanities (BigDS)
- Data Engineering for Data Science (DE4DS)
- Novel Data Management Ideas on Heterogeneous Hardware Architectures (NoDMC)
- A Tutorial Workshop on ML for Systems and Systems for ML

Während der vierte Workshop ein Tutorial mit eingeladenen Vorträgen war, wurden die anderen drei Workshops im klassischen Format mit Einreichungen, Reviews und Auswahl durchgeführt. Wir haben die Organisatoren der drei ersten Workshops gebeten, die jeweils zwei besten Paper ihrer Workshops zu nominieren.

Für dieses Heft des Datenbank-Spektrums haben die eingeladenen Autoren ihre Beiträge im Vergleich zur Workshop-Version substanziell erweitert und verbessert. Diese „Best Workshop Papers“ wurden erneut streng begutachtet, bevor sie nach zwei Revisionsrunden zur Publikation angenommen wurden. Fünf dieser sechs eingeladenen Beiträge erscheinen in diesem Heft.

Die ersten beiden Beiträge dieses Heftes sind aus dem BigDS-Workshop nominiert wurden: Im Beitrag *Geo Engine: Workflow-driven Geospatial Portals for Data Science* von Christian Beilschmidt et al. schlagen die Autoren von der Geo Engine GmbH und der Philips-Universität Marburg ein Framework vor, welches in der Lage ist, die notwendigen Komponenten für den Aufbau von Geodatenportalen bereitzustellen. Geodatenportale spielen eine Schlüsselrolle bei der Verbreitung und Nutzung von domain-spezifischen Geodaten. Die Autoren demonstrieren die Nutzung des Frameworks an einem realen Datenportal und einer Data Science Anwendung.

The InsightsNet Climate Change Corpus (ICCC) von Elena Volkanovska et al. (TU Darmstadt) beschäftigt sich mit der Analyse von Daten im Kommunikationsprozess zum Klimawandel. Diese liegen häufig nicht nur in Textform, sondern auch in Bildern, Videos, Tabellen etc. vor. Die Arbeit stellt den Prozess der Erstellung eines modalen Pilotkorpus für das InsightsNet Climate Change Corpus (ICCC) und die Verwendung von Werkzeugen zur Verarbeitung natürlicher Sprache (NLP) zur Anreicherung von Korpus(meta)daten vor.

Aus dem DE4DS-Workshop wurde der Beitrag *Sports-Table: A new Corpus for Semantic Type Detection* von Autoren der DHBW Moosbach und der TU Darmstadt (Sven Langenecker et al.) nominiert. Dieser Beitrag beschäftigt sich mit Tabellenkorpora, die annotierte semantische Typen pro Spalte enthalten und die wichtig für maschinelle Lernmodelle für die Aufgabe der automatischen Erkennung semantischer Typen sind. Um die große Diskrepanz zwischen existierenden Korpora und realen Data-Lakes bezüglich des großen Anteils numerischer Daten zu überwinden, stellen sie in ihrer Arbeit einen neuen Korpus vor, welcher einen

✉ Uta Störl
uta.stoerl@fernuni-hagen.de

¹ FernUniversität in Hagen, Hagen, Deutschland

viel höheren Anteil an numerischen Spalten enthält als bestehende Korpora.

Aus dem NODMC-Workshop präsentieren Alexander Baumstark, Muhamamd Attahir Jibril und Kai-Uwe Sattler von der TU Ilmenau ihren Beitrag *Accelerating Large Table Scan Using Processing-In-Memory Technology*. Im Fokus steht dabei das Problem des CPU-Engpasses in In-Memory-Datenbanksystemen. Mit dem Processing-In-Memory (PIM) Paradigma und dessen Verfügbarkeit in kommerziellen Systemen, lässt sich dieses Problem seit kurzem adressieren. In ihrer Arbeit zeigen und untersuchen die Autoren einen Ansatz, mit dem der Tabellen-Scan durch den Einsatz von PIM optimiert werden kann.

Ebenfalls aus dem NODMC-Workshop eingeladen wurde der Beitrag *An extension of DNAContainer with a small memory footprint* der Autoren Alex El-Shaikh und Bernhard Seeger (Philips-Universität Marburg). Dieser Beitrag beschäftigt sich mit DNA als neuem Speichermedium. Trotz riesigem Datenvolumen und sehr langer Haltbarkeit hat DNA als Speichermedium auch starke Einschränkungen, welche durch DNAContainer als neuartige Speicherschnittstelle für DNA adressiert werden. Der Beitrag präsentiert eine verbesserte Version von DNAContainer, wobei der Schwerpunkt auf den Aktualisierungsmöglichkeiten der Datenstrukturen und der Optimierung des Speicherbedarfs liegt.

2 Community-Beiträge

Die Rubrik „Dissertationen“ enthält in diesem Heft elf Kurzfassungen von Dissertationen aus der deutschsprachigen DBIS-Community.

Außerdem berichtet die Rubrik „Community“ unter News über die LWDA-Konferenz, die kürzlich an der Philipps-Universität Marburg stattgefunden hat und enthält Ankündigungen zu zukünftigen Veranstaltungen der DBIS-Community.

In der Rubrik „Produkt-News“ werden Neuerungen in verschiedenen DBMS vorgestellt. Schwerpunkt sind dabei Aktualisierungen von Cloud-Angeboten und die Verbesserungen im Bereich der Vektorsuche, welche eine besonders große Bedeutung für KI-Anwendungen hat.

3 Künftige Schwerpunktthemen

3.1 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 ma-

chine learning on quantum computers, as well as core tasks of the database management system, such as query optimization 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. We welcome contributions in both, German and English.

Deadline for submissions: Oct. 22, 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

3.2 Data Management for Research Data Infrastructures

Research data plays a crucial role in advancing scientific knowledge and driving innovation across various fields. Effective data management approaches are essential for ensuring the integrity, accessibility, and usability of research data. The FAIR data principles (Findability, Accessibility, Interoperability, and Reuse) are generally accepted as a goal in this context, but their comprehensive implementation re-

mains a challenge. Hence, innovative data management solutions for Research Data Infrastructures (RDIs) are needed to unlock the full potential of research data and enable collaboration, reproducibility, and knowledge exchange.

In Germany, the German National Research Data Infrastructure (NFDI) started in 2020 and comprises 26 consortia addressing a broad variety of disciplines. Currently, research data infrastructure projects inside and outside the NFDI develop data management solutions. In this special issue of *Datenbank-Spektrum*, we will provide a forum for exploring and presenting current trends and solutions in the field of data management for research data infrastructures.

We welcome original contributions, including technical papers, application-oriented papers, case studies, survey papers, and position papers addressing for example the following challenges in data management for research data infrastructures:

- Data Volume and Variety
- Data Integration and Interoperability
- Data Quality and Integrity
- Data Security and Privacy
- Long-Term Preservation
- Ethical and Legal Considerations
- Data Discoverability and Access

Paper format: 8–10 pages, double-column. We welcome contributions in both, German, and English.

Deadline for submissions: February 1st, 2024

Publication of special issue: DASP-2-2024 (July 2024)

Guest editors:

Andreas Henrich, University of Bamberg

andreas.henrich@uni-bamberg.de

Wolfgang Müller, Heidelberg Institute for Theoretical Studies

Wolfgang.Mueller@h-its.org

Bernhard Seeger, Philipps-Universität Marburg

seeger@mathematik.uni-marburg.de

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.