Editorial

Kai-Uwe Sattler¹ · Theo Härder²

Angenommen: 7. Februar 2022 / Online publiziert: 1. März 2022
© Der/die Autor(en) 2022

1 Schwerpunktthema „Best Papers of BTW 2021“

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 und war als 19. Auflage in der Woche vom 8. bis 12. März 2021 an der TU Dresden geplant.

Jedoch fand die BTW pandemiebedingt erstmals virtuell statt, und zwar in Form von mehreren wöchentlichen, jeweils 90-minütigen Online-Sitzungen im Zeitraum vom April bis Juni 2021. Den Abschluss bildete im September die Präsentation von Ergebnissen der Data Science Challenge, die bereits zum dritten Mal erfolgreich durchgeführt wurde.

Die Auswahl der Forschungsbeiträge erfolgte im Rahmen eines regulären Auswahlprozesses durch ein Programmkomitee unter Leitung von Melanie Herschel (Univ. Stuttgart) und Kai-Uwe Sattler (TU Ilmenau). Zudem wurden zwei DBIS-Dissertationspreise an Frau Asia Joanna Biega (Univ. des Saarlandes) und Herrn Felix Gessert (Univ. Hamburg) verliehen. Die virtuelle Tagung wurde vom Organisationskomitee der TU Dresden unter Koordinierung des Tagungsleiters Wolfgang Lehner hervorragend organisiert und fand mit bis zu 180 Teilnehmenden pro Termin regen Zuspruch. Einen detaillierten Überblick über den Ablauf der Konferenz gibt in diesem Heft der Beitrag *BTW2021*

erstmals als digitale Vortragsreihe. In diesem Community-Beitrag wird auch über die Erfahrungen berichtet, die bei dieser Konferenz erstmalig mit dem Aspekt *Wiederholbarkeit bzw. Nachvollziehbarkeit von Forschungsergebnissen (Reproducibility)* gemacht wurden.

Wir haben die Organisatoren und das Sieger-Team der Data Science Challenge gebeten, einen Beitrag über den Wettbewerb und die beste Lösung vorzubereiten. Sie beschreiben im Beitrag *Data Science Meets High-Tech Manufacturing – The BTW 2021 Data Science Challenge* die Aufgabenstellung, die Ablaufbedingungen und die zur Verfügung gestellten Datensammlungen des Wettbewerbs, für den Lösungen zum Problem des vorausschauenden Energie-Managements in großen Produktionsanlagen gesucht wurden. Dazu stellte der Halbleiterhersteller GlobalFoundries (Dresden) reale Produktions- und Energiedaten zur Verfügung.

Außerdem wurden von den Organisatoren für das Schwerpunktthema dieses Heftes die fünf am besten bewerteten Beiträge vorgeschlagen, die nach einer Erweiterung und erneuten Begutachtung in dieses Heft des Datenbank-Spektrums aufgenommen werden sollten. Leider haben es nur drei Beiträge als „Best Papers“ in dieses Heft geschafft, da nicht alle eingeladenen Autoren den Aufwand zur Erweiterung und Verbesserung ihrer Beiträge übernehmen wollten.

Der B^2 -Baum setzt sich aus einer äußeren Struktur, die einem herkömmlichen B^+ -Baum entspricht, und inneren seitenbasierten Strukturen zusammen, in denen jede Seite einen Trie-basierten Suchbaum enthält. Im ersten Beitrag *B^2 -Tree: Page-Based String Indexing in Concurrent Environments* untersuchen Josef Schmeißer, Maximilian Schüle (TU München), Viktor Leis (Univ. Erlangen-Nürnberg), Thomas Neumann und Alfons Kemper (TU München) den Einsatz solcher Bäume für nicht rein Hauptspeicher-basierte DBS, die auch Zugriff auf seitenbasierte Externspeicher unterstützen. Dadurch lassen sich Cache-Zugriffe optimieren und für bestimmte Operationen im Vergleich zu B^+ -Bäumen eine verbesserte Leistung erzielen.

✉ Kai-Uwe Sattler
kus@tu-ilmenau.de

Theo Härder
haerder@cs.uni-kl.de

¹ FG Datenbanken und Informationssysteme, TU Ilmenau, 98684 Ilmenau, Deutschland

² AG Datenbanken und Informationssysteme, TU Kaiserslautern, 67663 Kaiserslautern, Deutschland

ML-Modelle werden häufig benutzt, um in bestimmten Systemen Empfehlungen oder sogar Entscheidungen abzuleiten. Der Verbesserung von Fairness und Genauigkeit in diesem Kontext widmet sich der zweite Beitrag *Metrics and Algorithms for Locally Fair and Accurate Classifications using Ensembles* mit den Autoren Nico Lässig, Sarah Oppold und Melanie Herschel (Univ. Stuttgart).

Als dritter Beitrag findet sich in diesem Heft *Aggregate-based Training Phase for ML-based Cardinality Estimation* von Lucas Woltmann, Claudio Hartmann, Dirk Habich und Wolfgang Lehner (TU Dresden). Die Autoren wenden in diesem Beitrag ML-Methoden zur Kardinalitätsabschätzung, einer zentralen Aufgabe bei der Verarbeitung und Optimierung von Anfragen, an und können zeigen, dass sich dadurch Genauigkeit und Leistung verbessern lassen.

2 Fachbeitrag

Für alle Aspekte von Data Science wachsen mit der rapiden Zunahme der Datenvolumina in vielen Bereichen die Anforderungen für das Data Engineering. Eine immer wichtiger werdende Aufgabe ist dabei die Entwicklung von Tools, die eine automatische Verwaltung und Organisation der Daten unterstützen und von Domänen-Experten eingesetzt werden können. Meike Klettke (Univ. Rostock) und Uta Störl (Univ. Hagen) führen im Fachbeitrag *Four Generations in Data Engineering for Data Science – The Past, Presence and Future of a Field of Science* vier Generationen von Data-Engineering-Ansätzen zur Klassifikation der dazu vorhandenen Technologien ein. Außerdem zeigen sie, welche Tools für die wissenschaftliche Landschaft der nächsten Dekade benötigt werden.

3 Community-Beiträge

Die Task Force / Arbeitsgruppe „Data Science“ der GI e.V. hat Vorschläge für die Inhalte von Data-Science-Masterstudiengängen ausgearbeitet, die als Diskussionsgrundlage für die Akkreditierung entsprechender Studiengänge dienen sollen. Im Beitrag *GI entwickelt Empfehlungen zur Gestaltung von Data-Science-Masterstudiengängen* stellen Jörg Desel (Univ. Hagen), Daniel Krupka und Julia Meisner (GI e.V. Berlin) den Hintergrund und die Inhalte dieser Empfehlungen genauer vor.

Die Rubrik „Dissertationen“ enthält in diesem Heft elf 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 Berlin Institute for the Foundations of Learning and Data (BIFOLD)

Das Berlin Institute for the Foundations of Learning and Data (BIFOLD) ist ein von BMBF und dem Land Berlin gefördertes Kompetenzzentrum, das aus der Fusion des Berlin Big Data Center (BBDC) und dem Berliner Zentrum für Maschinelles Lernen (BZML) hervorgegangen ist. BIFOLD hat sich zum Ziel gesetzt, hochinnovative Technologien zu entwickeln, die riesige Datenmengen organisieren und mit deren Hilfe fundierte Entscheidungen getroffen werden können, um wirtschaftlichen und gesellschaftlichen Mehrwert zu schaffen. Zu diesem Zweck werden die bislang isoliert voneinander existierenden Gebiete Datenmanagement und Maschinelles Lernen verschmolzen. Die Technologien des Zentrums sollen den Stand der Technik in der Erforschung von Methoden des Datenmanagements, des maschinellen Lernens und deren Schnittstelle vorantreiben und die führende Stellung Deutschlands in Wissenschaft und Wirtschaft im Bereich der KI ausbauen. Als Technologietreiber stehen mehrere wirtschaftlich, wissenschaftlich und gesellschaftlich relevante Anwendungsbereiche im Fokus: Fernerkundung, digitalisierte Geisteswissenschaften, die Medizin sowie Informationsmarktplätze.

Aufbauend auf weltweit anerkannten Forschungsergebnissen sollen eine automatische Optimierung, Parallelisierung sowie eine skalierbare und adaptive Verarbeitung von Algorithmen in heterogenen, verteilten Umgebungen unter Einsatz moderner Rechnerarchitekturen ermöglicht werden. Daneben stehen Erklärbarkeit, verantwortungsvolles Datenmanagement und innovative Anwendungen der Datenanalyse im Fokus. Behandelt werden dabei Bereiche des Datenmanagements, des maschinellen Lernens, der linearen Algebra, der Statistik, der Wahrscheinlichkeitstheorie, der Computerlinguistik sowie der Signalverarbeitung. Durch Entwicklung und Bereitstellung von Open-Source-Systemen sowie von Algorithmen und Methoden zur Datenanalyse wird das Zentrum die Ausbildung, Forschung, Entwicklung, Innovation und kommerzielle Nutzung von Big Data Analytics und KI-Anwendungen in Deutschland fördern und so deutschen Firmen einen Wettbewerbsvorteil sichern.

Wir erbitten Einreichungen in Deutsch oder Englisch mit einem Umfang von 8 bis 10 Seiten (zweispaltig) gemäß den Layoutvorgaben (siehe <https://www.springer.com/13222>).

Frist zur Einreichung: 1. Feb. 2022

Erscheinen des Themenheftes: DASP-2-2022 (Juli 2022)

Gasteditor:

Dr. Alexander Borusan, TU Berlin

alexander.borusan@tu-berlin.de

4.2 Testing and Benchmarking Database Management Systems

Today's database management systems are the result of several decades of research and engineering. These efforts have resulted in a multitude of both open-source and commercial systems that are widely deployed in production and provide the backbone of a vast range of mission-critical applications. To ensure a high software quality, development teams routinely have to apply sophisticated testing strategies for finding defects early and ensuring the robustness, performance, and scalability of their solutions.

This special issue of the Datenbank-Spektrum focuses on aspects that are related with the novel techniques for testing and benchmarking of database management systems (i.e., classical DBMSs and beyond) as well as war stories from industry-scale use of database management systems. Topics of interest include, but are not limited to:

- Testing or benchmarking of database systems, storage services, and database applications
- Testing or benchmarking of database systems using novel hardware and software technology (non-volatile memory, hardware transactional memory, ...)
- Testing or benchmarking heterogeneous systems with hardware accelerators (GPUs, FPGAs, ASICs, ...)
- Testing or benchmarking distributed and big data systems
- Testing or benchmarking machine learning systems
- Testing or benchmarking learned database component / systems
- Specific challenges of testing or benchmarking and quality assurance for cloud-based systems
- Database system benchmark development
- War stories and lessons learned
- Formal verification techniques in the context of database management systems
- Applying automated testing tools to database management systems
- Performance and scalability testing
- Testing the reliability and availability of database systems
- Algorithms and techniques for automatic program verification
- Maximizing code coverage during testing of database systems and applications
- Generation of synthetic data for test databases
- Testing the effectiveness of adaptive policies and components

- Tools for analyzing or benchmarking database management systems (e.g., profilers, debuggers)
- Workload characterization with respect to performance metrics and engine components
- Metrics for test quality, robustness, efficiency, and effectiveness
- Operational aspects such as continuous integration and delivery pipelines
- Security and vulnerability testing
- Experimental reproduction of benchmark results
- Functional and performance testing of interactive data exploration systems
- Tracability, reproducibility and reasoning for ML-based systems

We welcome traditional research articles, experience and application reports, proof-of-concept studies, as well as benchmark proposals, surveys, and experimental studies. These can either be in the form of full submissions (8–10 pages) as well as short papers/extended abstracts (not more than 4 pages) for this issue. Please find the submission guidelines at <https://www.springer.com/journal/13222/submission-guidelines>.

Submission deadline is June 1st, 2022 (the special issue will be published in November 2022).

Guest editors:

Alexander Boehm, SAP SE, Walldorf

alexander.boehm@sap.com

Carsten Binnig, TU Darmstadt

carsten.binnig@cs.tu-darmstadt.de

Tilman Rabl, HPI, Universität Potsdam

Tilman.Rabl@hpi.de

4.3 Trends in Social Media Analysis to Address Fake News, Hate Speech, or Bias

Wir erbitten Einreichungen in Deutsch oder Englisch mit einem Umfang von 8 bis 10 Seiten (zweispaltig) gemäß den Layoutvorgaben (siehe <https://www.springer.com/13222>).

Frist zur Einreichung: 1. Okt. 2022

Erscheinen des Themenheftes: DASP-1-2023 (März 2023)

Gasteditoren:

Ralf Krestel, ZBW & CAU Kiel

r.krestel@zbw.eu

Udo Kruschwitz, Universität Regensburg

Udo.Kruschwitz@ur.de

Michael Wiegand, Universität Klagenfurt

michael.wiegand@aau.at

4.4 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 meta-data 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 <https://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

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>.