Editorial

Theo Härder¹

© Gesellschaft für Informatik e.V. and Springer-Verlag GmbH Germany, part of Springer Nature 2019

1 Schwerpunktthema: Ausgewählte Beiträge von den Workshops der BTW 2019

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 fand bereits als 18. Auflage in der Woche vom 4. bis 8. März 2019 an der Universität Rostock statt.

Seit 2001 werden in der BTW-Woche stets auch Workshops zu aktuellen oder sich abzeichnenden Themen der Datenbankforschung durchgeführt. Für die BTW wurde 2017 erstmalig ein Team-Wettbewerb, Data Science Challenge genannt, konzipiert. Wegen des großen Erfolgs dieser wissenschaftlichen Auseinandersetzung haben auch die Rostocker Organisatoren der BTW einen solchen Wettbewerb angeboten, für den sie IBM und SAP als Sponsoren gewinnen konnten. Sie haben die Aufgaben des Wettbewerbs definiert, deren Durchführung überwacht, die sehr unterschiedlichen Lösungen von fünf teilnehmenden Teams bewertet und schließlich an drei Teams Preise verliehen. Neben dem Studierendenprogramm fanden im Rahmen der BTW 2019 drei Workshops zu folgenden Themen statt:

- Novel Data Management Ideas on Heterogeneous (Co-) Processores (NODMC)
- Digitale Lehre im Fach Datenbanken
- Big (and Small) Data in Science and Humanities (BigDS).

Wir haben die Organisatoren und Sieger-Teams des Data Science Challenge gebeten, einen Beitrag über den Wettbewerb und die drei Lösungen vorzubereiten. Außerdem

haben wir zusammen mit den Organisatoren zwei Workshop-Beiträge und zwei Beiträge aus dem Studierendenprogramm ausgewählt, um aktuelle Forschungstrends im DBIS-Bereich aufzuzeigen. Für dieses Heft des Datenbank-Spektrums haben die eingeladenen Autoren ihre Beiträge im Vergleich zur Workshop-Version substanziell erweitert und verbessert. Diese „Best Papers“ wurden erneut streng begutachtet, bevor sie nach zwei Revisionsrunden zur Publikation angenommen wurden.

Die Organisatoren und die Mitglieder der Sieger-Teams beschreiben im Beitrag *Particulate Matter Matters—The Data Science Challenge @ BTW2019* die Aufgabenstellung, die Ablaufbedingungen und die zur Verfügung gestellten Datensammlungen des Wettbewerbs, für den Lösungen zur Integration, Analyse und Visualisierung von Daten beim Thema Feinstaubbelastung gesucht wurden. Jedes Team skizziert in einem eigenen Kapitel seinen Lösungsansatz und seine Methoden zur Visualisierung, bevor es seine Ergebnisse interpretiert und bewertet. Die verschiedenartigen Ansätze verdeutlichen dabei, wie durch Kreativität und Vielfalt bei der Analyse und Exploration großer Datenmengen neue Erkenntnisse gewonnen werden können.

Der Beitrag *Evaluating the Vector Supercomputer SX-Aurora TSUBASA as a Co-Processor for In-Memory Database Systems* von J. Pietrzyk et al. (TU Dresden) untersucht insbesondere Möglichkeiten der Unterstützung datenintensiver Anwendungen in Hauptspeicher-Datenbanksystemen durch Vektorisierungstechniken. Dabei wird festgestellt, dass aus Datenbanksicht bestimmte Vektorisierungstechniken nicht nur für die Kompression und Dekompression von Daten gewinnbringend eingesetzt werden können, sondern auch für alle Datenbankoperatoren wie Joins, Scans oder Groupings. Durch empirische Leistungsmessungen werden die Vorteile des Vektor-Computers SX-Aurora TSUBASA verglichen mit herkömmlichen SIMD-Erweiterungen belegt.

Graph-Verarbeitungssysteme erlauben meist nur die Analyse von statischen Graphen, deren Ergebnisse sich deshalb auf den Zustand eines Graphen zu einem bestimmten Zeitpunkt beschränken. In vielen Anwendungen ist es

✉ Theo Härder
haerder@cs.uni-kl.de

¹ AG Datenbanken und Informationssysteme, TU Kaiserslautern, 67663 Kaiserslautern, Deutschland

jedoch wichtig, die zeitliche Dimension eines Graphen zu betrachten, d.h., wie sich die Anzahl von Knoten und Kanten und deren Beziehungen in welcher Geschwindigkeit usw. verändern. Aus diesen Gründen schlagen C. Rost (Uni Leipzig), A. Thor (Hochschule für Telekommunikation Leipzig) und E. Rahm (Uni Leipzig) im Beitrag *Analyzing Temporal Graphs with Gradoop* eine Erweiterung von Gradoop, ein verteiltes Framework zur Graph-Analyse, vor, wobei zeitbezogene Eigenschaften für Knoten, Kanten und Graphen eingeführt werden und diese in den Graph-Operatoren genutzt werden.

Die folgenden beiden Aufsätze wurden aus dem Studierendenprogramm ausgewählt. Im Beitrag *Lock-free Data Structures for Data Stream Processing: A Closer Look* berichtet A. Baumstark zusammen mit C. Pohl (TU Ilmenau) über ihre Arbeiten zur Optimierung der Verarbeitung von Datenströmen. Bei einem Einsatz von moderner Hardware lässt sich eine hohe Parallelisierung von Stromanfragen erzielen, die jedoch effiziente Synchronisationsmechanismen beim Zugriff auf gemeinsamen Speicher erfordert. Am Beispiel von PipeFabric, ihrem eigenen System zur Stromverarbeitung, identifizieren sie exemplarisch die zeitaufwendigsten Operationen und Engpässe bei der Stromverarbeitung. Weiterhin führen sie verschiedene Entwurfsprinzipien für sperrfreie Datenstrukturen ein, mit denen solche Operationen und Situationen optimiert werden können. Schließlich zeigen sie unter verschiedenen Arbeitslasten, wie sich durch die gewonnene „Sperrfreiheit“ die Effizienz der Join-Verarbeitung und des Datenaustauschs zwischen Operatoren verbessern lässt.

Ketten, die zwei oder mehr verschiedene Cluster verknüpfen, stellen ein bekanntes Problem für Clustering-Algorithmen wie DBSCAN oder Single Linkage Clustering dar, da oft wenige Punkte zu falschen Ergebnissen führen können. Beispielsweise könnten zwei verschiedene Cluster als eines interpretiert werden. J. Held, A. Beer und T. Seidl (LMU München) beschreiben im Beitrag *Chain-detection between Clusters* einen neuen Algorithmus, der verschiedenartige Ketten zwischen Clustern identifizieren kann, um die darunter liegende Cluster-Struktur klarer enthüllen zu können. In Kombination mit DBSCAN, aber auch mit hierarchischem Single Link Clustering, zeigen sie anhand eines realen Datensatzes, der Verkehrsunfälle in Großbritannien beschreibt, dass Ketten, die von Straßen zwischen Städten und Ortschaften stammen, korrekt erkannt werden, so dass die Cluster-Struktur der verschiedenen Ortschaften erhalten bleibt.

2 Community-Beiträge

Die Rubrik „Dissertationen“ enthält in diesem Heft 7 Kurzfassungen von Dissertationen aus der deutschsprachigen DBIS-Community.

Schließlich berichtet die Rubrik „Community“ unter *News* über die LWDA-Konferenz, die kürzlich an der Humboldt-Universität zu Berlin stattgefunden hat, und über weitere aktuelle Informationen, welche die DBIS-Gemeinde betreffen.

3 Künftige Schwerpunktthemen

3.1 Trends in Information Retrieval Evaluation

Evaluation is a central aspect in the research and development of information retrieval systems. In academia, the quantitative evaluation of such systems is mostly known under the term Cranfield paradigm. This research method has been established for more than 25 years in international evaluation campaigns such as the Text Retrieval Conference (TREC) or the Conference and Labs of the Evaluation Forum (CLEF). Meanwhile industrial research has taken a completely different approach. Many companies are able to access a large number of users and their interactions, which can be recorded and evaluated. These infrastructures allow alternative evaluations like large-scale A/B experiments or other online methods. In the last years, different approaches to go beyond TREC-style evaluations emerged to close the gap and to bring together academic and industrial evaluation.

We are calling for articles that report on novel evaluation efforts, like:

- Living labs
- Evaluation as a service
- Large-scale A/B tests
- Interactive retrieval evaluation
- Session-based evaluation
- User-centered evaluation
- Counterfactual evaluation
- Novel evaluations in application domains such as cultural heritage, digital libraries, social media, expert search, health information, etc.
- Other evaluations that go beyond TREC

Expected size of the paper: 8-10 pages, double-column (cf. the author guidelines at www.springer.com/13222). Contributions either in German or in English are welcome.

Deadline for submissions: Oct. 1st, 2019

Issue delivery: DASP-1-2020 (March 2020)

Guest editors:

Philipp Schaer, Technische Hochschule Köln
 philipp.schaer@th-koeln.de
 Klaus Berberich, Hochschule für Technik und Wirtschaft
 des Saarlandes
 klaus.berberich@htwsaar.de

Guest editors:

Philipp Cimiano (Universität Bielefeld)
 Gerhard Heyer (Universität Leipzig)
 Michael Kohlhase (FAU Erlangen-Nürnberg)
 Benno Stein (Bauhaus-Universität Weimar)
 Jürgen Ziegler (Universität Duisburg-Essen)

3.2 Argumentative Intelligence (Arguing)

This special issue of the “Datenbank-Spektrum” is dedicated to the area of “argumentative intelligence” concerned with the question of how to design intelligent systems that reach a deep understanding of states of affairs and domains in order to be able to analyze and generate arguments to support users in the activity of rationalizing as a basis for informed decision making. We will invite submissions from different research fields including semantics, natural language processing, information retrieval, knowledge representation and artificial intelligence in general to contribute to this new field of research. The special issue will feature key results and advancements achieved as a result of the DFG-funded priority program “Robust Argumentation Machines”.

Expected size of the paper: 8–10 pages, double-column (cf. the author guidelines at www.springer.com/13222). Contributions either in German or in English are welcome.

Deadline for submissions: Feb. 1st, 2020

Issue delivery: DASP-2-2020 (July 2020)

3.3 Data Management for Future Hardware

This special issue of the „Datenbank-Spektrum“ is dedicated to the research achieved by the DFG Priority Programme „Scalable Data Management on Future Hardware“. We invite submissions on original research as well as overview articles addressing the challenges and opportunities of modern and future hardware for data management such as many-core processors, co-processing units, new memory and network technologies.

Paper format: 8–10 pages, double-column (cf. the author guidelines at www.springer.com/13222).

Deadline for submissions: June 1st, 2020

Issue delivery: DASP-3-2020 (November 2020)

Guest editors:

Kai-Uwe Sattler, TU Ilmenau
 kus@tu-ilmenau.de
 Alfons Kemper, TU München
 alfons.kemper@in.tum.de