News

Online publiziert: 27. Oktober 2021
© Gesellschaft für Informatik e.V. and Springer-Verlag GmbH Germany, part of Springer Nature 2021

1 BTW 2023

Die BTW 2021 konnte wegen der Coronasituation nicht wie geplant in Dresden ausgerichtet werden, sondern fand als Vortragsreihe während des Sommersemesters 2021 statt. Die BTW 2023, die dann bereits die 20. Auflage der Tagung sein wird, wird daher als Präsenzveranstaltung vom 6. bis 10. März 2023 an der TU Dresden stattfinden.

Weitere Informationen finden Sie zu gegebener Zeit unter https://btw2023-dresden.de.

2 Virtuelle Jahrestagung der GI-Fachgruppe Frauen und Informatik

Gudrun Schiedermeier

Die Jahrestagung 2021 der GI-Fachgruppe Frauen und Informatik fand statt am 24. April 2021 von 10:00 bis 17:00 Uhr als rein virtuelle Tagung zum Thema „Künstliche Intelligenz – vertrauenswürdig, erklärbar, fair?“. Die äußerst interessanten Vorträge von Ute Schmid „Gemeinsam klüger – Erklärbares und interaktives maschinelles Lernen“, Katharina Morik „Vertrauenswürdiges maschinelles Lernen“ und Gudrun Schiedermeier „Diskriminierende KI-Systeme“ faszinierten die über 70 Teilnehmer*innen. In einem Impulsvortrag stellte Lena Wiese einen „Beipackzettel für Machine-Learning-Modelle“ vor. Sie führte aus, dass für Personen, die von ML-basierten Entscheidungen betroffen sind, die transparente und verständliche Darstellung der Eigenschaften des zugrunde liegenden ML-Modells grundlegend für eine Einordnung solcher Entscheidungen ist. Alle Vortragenden und zusätzlich Meike Klettke diskutierten lebhaft unter der Moderation von Christine Regitz. Diese lebendige Podiumsdiskussion traf bis zum Schluss auf großes Interesse von Seiten des Publikums.

Einigkeit herrschte bei den Expertinnen, dass wir die Entwicklung und den Einsatz von KI-Systemen nicht allein den Herstellern überlassen dürfen. Essentiell für die Förderung von Vertrauen in KI-Systeme ist vielmehr, dass herstellerunabhängige Instanzen Prüfungen und Risikoabschätzungen durchführen. Diese kritische Begleitung ist wichtig, aber es bedarf auch der Durchsetzung der Regelungen. Auf der anderen Seite enthalten die Entwicklungen ein innovatives Potenzial, das wir nutzen sollten.

Großen Anfang fand die Verleihung des Preises der Fachgruppe an Jana Eisoldt für eine herausragende Bachelorarbeit und die Verleihung des Preises an Judith Hemp für eine ebenso ausgezeichnete Masterarbeit. Die Preisträgerinnen wurden aus 70 Einreichungen ausgewählt. Der Preis dient dem Ziel, Frauen mit hervorragenden Ergebnissen im Informatik-Studium sichtbar zu machen.

3 LWDA 2021

Auch im Jahr 2021 musste die Konferenz Lernen, Wissen, Daten, Analysen (LWDA) wegen der Pandemie wieder als Onlineveranstaltung stattfinden. Sie wurde von der Arbeitsgruppe von Thomas Seidl an der LMU München organisiert. Vom 1. bis 3. September 2021 fanden die parallelen Workshops der Fachgruppen Datenbanksysteme, Information Retrieval, Knowledge Discovery und Machine Learning, Business Intelligence und Analytics sowie Knowledge Management. Auch der Workshop „Grundlagen von Datenbanken“, der von Andreas Thor (HTWK Leipzig) organisiert wurde, fand in diesem Rahmen statt, da eine Präsenzveranstaltung nicht möglich war. Hier wurden elf Beiträge von Nachwuchswissenschaftlerinnen und -wissenschaftlern präsentiert.

Der Datenbankworkshop wurde vom Arbeitskreis „Data Engineering for Data Science“ gestaltet. Hier wurden zehn Beiträge vorgestellt, die für das aktuelle Themenheft „Data Engineering for Data Science“ des Datenbank-Spektrums eingereicht und angenommen wurden und in dieser oder einer der kommenden Ausgaben erscheinen werden; online auf der Homepage des Datenbank-Spektrums beim Verlag sollten bei Erscheinen des Hefts alle Beiträge abrufbar sein.

Traditionell runden bei der LWDA Keynotes aus den verschiedenen Fachgruppen das Programm ab. In diesem Jahr

Geben Sie die gesamte Dokumenttext in einer einfacheren Sprache an.
waren die internationalen Redner Michael Leyer (Universität Rostock) mit dem Thema „How our brain reacts to and interacts with data, information and knowledge“, Mykola Pechenizkiy (TU Eindhoven) über „The origins and future of AI fairness, accountability and transparency“ und Arjen de Vries (Radboud University Nijmegen), der mit seinem Vortrag „You will want to rank your text data with a database too!“ elegant eine Verbindung von Datenbanksystemen zu Information Retrieval aufzeigte. Die beliebte Stadtführung musste leider Coronabedingt genauso ausfallen wie das gemeinsame Abendessen aller Teilnehmenden. Eine virtuelle Tour durch München, wo die Tagung normalerweise stattgefunden hätte, war ein kleiner Ersatz dafür.

Weitere Informationen: https://mcml.ai/lwda2021/

### 4 Produkt-News

Uta Störl

#### 4.1 Neo4j 4.3

Im Juni 2021 wurde Neo4j 4.3 vorgestellt. Neo4j 4.3 bringt eine Reihe von Performance-Verbesserungen: Relationship Indexes und Relationship Property Indexes erlauben die Indizierung von Beziehungen nach Typ und Attributen und ermöglichen es so, Anfragen zusätzlich zu beschleunigen. Relationship Chain Locks bieten feingranulare Sperren, so dass Knoten und Beziehungen nebeneinander geschrieben werden können – selbst bei sogenannten Superknoten mit Millionen von Beziehungen. Transaktionsdurchsatz und Datenimport profitieren davon. Das IO Scheduling priorisiert Nutzertransaktionen und Hintergrundprozesse für die Aufteilung verfügbarer IO-Ressourcen. So werden Ressourcen effizienter genutzt und ein höherer Transaktionsdurchsatz erzielt. Analytical Read Scaling bieten neue Clustering-Optionen, um reine Lesereplikate für Analyse- und Visualisierungszwecke oder Standby-Verfügbarkeit bereit zu halten.

Neo4J, https://neo4j.com/

#### 4.2 Oracle Database 21c für On-Premises

Nachdem Oracle Database 21c schon seit einiger Zeit in der Cloud auf Always Free Autonomous Database und im Oracle Database Cloud Service (DBCS) verfügbar ist, ist es nun auch für die On-premises-Plattformen soweit: Seit August steht Oracle Database 21c für Linux und Oracle Engineered System zur Verfügung.

Bei Oracle Database 21c handelt es sich um ein „Innovation Release“ mit bis zu 200 neuen Features wie z.B. native Blockchain-Tabellen, binärer JSON-Datentyp, Javascript-Ausführung in der Datenbank u. v.m. Oracle Database 21c ist die erste Oracle-Datenbankversion mit einer reinen Multitenant-CDB-Architektur (Container Database (CDB)) – eine Nicht-CDB-Architektur wird nicht mehr unterstützt. Zahlreiche Blogeinträge im Internet wie zum Beispiel der Database-Insider-Artikel „Introducing Oracle Database 21c“ https://blogs.oracle.com/database/post/introducing-oracle-database-21c beschreiben die neuen Funktionen der Version 21c. Oracle, https://www.oracle.com/

#### 4.3 IBM Db2 11.5.6

Seit kurzem ist die neue Version Db2 11.5.6 verfügbar. Neben diversen Performance- und Verfügbarkeitsverbesserungen baut sie die Multi-Model-Fähigkeiten durch die Erweiterung der Graphdatenbankunterstützung von Db2 (mit der Möglichkeit, Gremlin-Abrufe gegen Db2 auszusetzen) und durch verbesserte Spatial Analytics aus. Die Nutzung von Machine Learning durch den Optimizer wurde ausgeweitet und der Umgang mit sehr langen aktiven Transaktionen optimiert. Ein weiteres wichtiges Thema dieser Version ist der Betrieb von Db2 in einem Container bzw. in Cloud-Umgebungen. Dazu gibt es mit Click-to-Containerize ein neues Werkzeug, um bestehende Db2-Systeme automatisiert in einen Container zu überführen.

IBM, https://www.ibm.com/

#### 4.4 SAP: HANA Cloud 2108

SAP hat mit dem Juli-Release der HANA Cloud weite
tere Ergänzungen zur Verfügung gestellt: Die Cloud-Datenbank erlaubt nun Near-Zero-Downtime Upgrades und bietet High Availability über synchrone Replikation innerhalb einer Verfügbarkeitszone mit nun maximal 6 TB großen Instanzen bei 99,95% Verfügbarkeit.

Im Bereich Security kann nun für den SAP HANA Cloud Data Lake ein externer Identity-Provider über JWT oder LDAP genutzt werden. Dazu bietet sich für Kunden jetzt die Möglichkeit, Encryption Keys selber zu verwalten. Im Bereich Multi-Model-Prozessierung gibt es Erweiterungen für den JSON Document Store, in der Graph Engine sowie weitere Predictive- und Cluster-Algorithmen im Bereich Machine Learning.

SAP, https://www.sap.com/

#### 4.5 MongoDB 5.0

Die neue Version von MongoDB enthält u. a. eine optimierte Unterstützung für Zeitreihenanwendungen und bessere Unter
stüzung beim Resharding. MongoDB bietet nun einen
nativen Zeitreihendatentyp mit dafür optimierten Kolle
kationen, der sowohl die Speicherung als auch Abfragen und weitere Analysen (Echtzeitanalyse, Visualisierung etc.) un-
Die Funktion Live-Resharding flexibilisiert das Skalieren der Datenbank. Bei Bedarf kann der Sharding Key, welcher die Verteilung der Dokumente einer Kollektion auf die unterschiedlichen Shards im Cluster festlegt, geändert werden. Das Resharding erfolgt dann live und ohne Unterbrechung der Datenbank.

MongoDB, Inc., https://www.mongodb.com/

4.6 Oracle: MySQL Autopilot Funktionen für den MySQL Heatwave Cloud Service

MySQL und der korrespondierende Cloud-Service MySQL Database Service (MDS) sind von Haus aus für OLTP optimiert. Bei OLAP-Abfragen werden die MySQL-Daten in einem ETL-Prozess transformiert und diese in einer OLAP-optimierten Datenbank ausgeführt.

Hier setzt der Cloud Service „MySQL HeatWave für real-time Analytics“ an, der nun um die Funktion MySQL Autopilot erweitert wurde. Der MySQL HeatWave Service besteht aus zwei Komponenten: Dem MySQL Leading Node und dem MySQL HeatWave Backend. Der Leading Node ist eine erweiterte MDS-Instanz, auf welche Applikationen über die MySQL-Konnektoren zugreifen. Da der MySQL-Parser unverändert ist, unterstützt HeatWave alle Applikationen, die auch MySQL 8 unterstützen. Der Leading Node ist für alle OLTP-Abfragen und die Speicherung der Daten zuständig. Über eine neue Storage RAPID Engine (Rapid Analytics Processing in DRAM) werden diese Daten auf das HeatWave Backend System in Echtzeit synchronisiert. Die Daten werden im HeatWave Backend im Arbeitsspeicher optimiert im Spaltenformat gespeichert und das HeatWave Backend bei Bedarf skaliiert. Mit dem neuen MySQL Autopilot erfolgen diese Schritte weitestgehend automatisiert und werden durch Machine-Learning-Algorithmen optimiert. Der Prozess des ETL entfällt komplett.

Anfragen, die über den Leading Node gestellt werden, werden im MySQL-Optimizer bewertet und wenn möglich nicht auf dem lokalen Leading Node ausgeführt, sondern an das HeatWave Backend weitergeleitet, verarbeitet und zurückgegeben. Der MySQL Leading Node antwortet dann auf die Anfragen der Applikationen. Einfache Abfragen (z.B. Primary Key Lookups) werden vom MySQL-Optimizer klassisch „lokal“ bearbeitet. Somit kann ein MySQL HeatWave Service sowohl für OLTP als auch für OLAP-Betrieb effizient verwendet werden und ermöglicht Anwendern gemischte Arbeitslasten oder Echtzeit-Analysen auf Basis von MySQL.

Oracle stellt verschiedene Benchmarks (TPC-H, CHbenCHmark) auf der Website zu MySQL HeatWave vor. Je nach verwendetem Benchmark und der verwendeten Abfrage ergeben sich nach diesen Angaben Leistungssteigerungen von über 1000x.

Oracle, https://www.oracle.com

4.7 Neo4j: Graph Data Science 1.6

Seit Juni 2021 steht Neo4j Graph Data Science in Version 1.6 zur Verfügung. Die neue Version der Graph-Analysekomponente bietet unter anderem neue Skalierungs- und Normalisierungsfunktionen sowie die Möglichkeit der direkten Filterung von In-Memory-Graphen. Auch Graph-Embedding-Algorithmen werden dadurch verbessert.

Neo4j, https://neo4j.com/

4.8 Neo4j: GraphQL Library 2.0.0

Seit August 2021 steht die Neo4j GraphQL Library 2.0.0 zur Verfügung – die erweiterbare, Low-Code- und Open-Source-Bibliothek zum Erstellen von Anwendungen mit der Neo4j-Graphdatenbank. Die neue Version verbessert unter anderem die Unterstützung für Relationship-Properties und bietet Cursor-basierte Pagination (wichtig für Infinite Scrolling) sowie eine sehr einfach zu benutzende Unterstützung für Count-Abfragen.

Neo4j, https://neo4j.com/product/graphql-library/

4.9 Server-seitiges JavaScript in Oracle 21c und Application Express 20.2

Eine wesentliche Neuerung mit der Einführung der Oracle Datenbank Version 21c ist die Möglichkeit, neben PL/SQL und Java auch JavaScript-Code in der Oracle-Datenbank auszuführen. Umgesetzt wird dies durch die Oracle Database Multilingual Engine (MLE), welche auf einer Integration der universellen GraalVM basiert. Hierzu gibt es ein neues PL/SQL Package DBMS_MLE, über das der JavaScript-Code zur Ausführung übergeben wird und Daten zwischen den PL/SQL- und JavaScript-Umgebungen ausgetauscht werden können. JavaScript-Datentypen werden dabei automatisch auf Datenbank-Datentypen abgebildet und umgekehrt. Damit ist auch ein gemischter Betrieb von JavaScript und PL/SQL möglich.

Mit Oracles Low-Code-Entwicklungsumgebung Application Express (Apex) in der Version 20.2 wird diese Funktionalität direkt unterstützt. Während die Client-seitige Entwicklung mit Apex schon immer mit JavaScript erfolgte, können Entwickler jetzt wahlweise auch Server-seitig mit Apex entwickeln.

Oracle, https://www.oracle.com/

4.10 IBM: Data Virtualization as a Service

Datenvirtualisierung steht jetzt auch als Managed Service in der Public Cloud zur Verfügung. Damit können Daten aus den unterschiedlichsten Datenquellen – seien es relationale Datenbanken wie Db2, Oracle oder PostgreSQL oder
NoSQL Systeme wie MongoDB – in einem einheitlichen Format und mit einheitlicher Governance präsentiert werden, ohne dass sie wie bei einem traditionellen Data Warehouse oder Data Lake zunächst in ein zentrales System kopiert werden müssen. Vorteile der Datenvirtualisierung gegenüber diesem traditionellen Ansatz sind nicht nur das Einsparen des zusätzlichen Speicherplatzes, sondern auch die Sichtbarkeit von Änderungen in den Quellsystemen in Echtzeit.

IBM, https://www.ibm.com/

4.11 IBM: Data Fabric und Cloud Pak for Data Version 4

*Data Fabric* ist ein Architekturmuster, das verteilte Daten dynamisch zusammenbringt, katalogisiert, kategorisiert und abfragbar macht. Es erleichtert die Nutzung von Daten als Unternehmensasset und erlaubt, verschiedene Arten von Daten im Self-Service zu kombinieren und effizient zu verwalten. Governance und Security werden dabei durchgängig über alle Datenquellen hinweg integriert. Data Fabric stellt daher vertrauenswürdige Daten bereit, eine Basis für die Transparenz der KI-Initiativen. IBM Cloud Pak for Data Version 4 erleichtert mit seinen Komponenten AutoSQL, AutoCatalog, AutoPrivacy und AutoAI den Aufbau eines Data Fabric in einer hybriden Multi-Cloud-Umgebung.

IBM, https://www.ibm.com/

Dank an Jörg Latza (SAP), Arne Brüning, Carsten Thalheim, Ulrike Schwinn (alle Oracle), Hannes Voigt (Neo4j) sowie Wilfried Hoge und Andreas Weininger (beide IBM) für ihren fachlichen Input.