



Klemens Böhm ist seit 2004 Inhaber des Lehrstuhls für Datenbanken und Informationssysteme am KIT. Davor war er gut zwei Jahre Professor für Angewandte Informatik/ Data and Knowledge Engineering an der Otto-von-Guericke-Universität Magdeburg. Von 1998 bis 2002 war er Oberassistent an der ETH Zürich in der Datenbankgruppe, von 1993 bis 1998 wissenschaftlicher Mitarbeiter der GMD – Forschungszentrum Informationstechnik GmbH am Darmstädter Institut für Integrierte Publikations- und Informationssysteme (IPSI). Klemens Böhm promovierte 1997 an der Technischen Hochschule Darmstadt mit einer Arbeit über die Verwaltung semistrukturierter Daten mit Datenbanksystemen. Er studierte Informatik mit Nebenfach Betriebswirtschaftslehre in Frankfurt, Darmstadt und Lissabon (Diplom von der TH Darmstadt 1993).

Zu den derzeitigen Forschungsschwerpunkten des Lehrstuhls gehören die Entwicklung von Analysetechniken für große Datenbestände, die Verwaltung wissenschaftlicher Daten, technische Fragen des Datenschutzes sowie Unterstützung für Geschäftsprozesse. Der Lehrstuhl arbeitet viel mit anderen Lehrstühlen, seien sie Informatik, seien sie aus anderen Disziplinen, und mit Anwendern zusammen, seien sie Wissenschaftler, seien sie aus der Industrie. Klemens Böhm ist Sprecher des 2015 bewilligten DFG-Graduiertenkollegs „Energiezustandsdaten – Informatik-Methoden zur Erfassung, Analyse und Nutzung“, in dem KIT-Forscherinnen und -Forscher aus unterschiedlichen Disziplinen datengestützt an der Entwicklung effizienter, nachhaltiger, robuster und benutzerfreundlicher Energiesysteme arbeiten.

ÜBERBLICK UND ALLGEMEINES

Zu den derzeitigen Forschungsschwerpunkten der Professur gehören die Entwicklung von Analysetechniken für große Datenbestände, die Verwaltung wissenschaftlicher Daten, technische Fragen des Datenschutzes sowie Unterstützung für Geschäftsprozesse.

Ein wichtiges Forschungsziel ist das Erkennen von Auffälligkeiten, insbesondere in Datenbeständen hoher Dimensionalität (also beispielsweise Datenobjekte mit sehr vielen Attributen) oder in Zeitreihen bzw. Datenströmen. Dazu gehört auch, wie sich durch möglichst geringfügiges Verfälschen der Daten bestimmte Phänomene wirksam verstecken lassen. Daraus ergeben sich vielfältige Möglichkeiten für Anwender, beispielsweise bessere Verfahren für Predictive Maintenance. Zu den weiteren Anwendungsfällen, die die Arbeitsgruppe derzeit bearbeiten, gehört die Analyse von SQL-Query Logs, um zu höherer Nutzerzufriedenheit mit Datenbanken und zu besserem Leistungsverhalten zu kommen.

Bei der **Verwaltung wissenschaftlicher Daten** interessiert uns insbesondere die Effizienz und Skalierbarkeit gängiger Analysetechniken. Beispielsweise in sehr großen Mengen von Zeitreihen das Auffinden von Paaren, deren zeitliche Verläufe maximal ähnlich zueinander sind. Hierzu gehört auch die sichere Speicherung der Daten, ohne dass sich die Auswertung wichtiger Anfragen wesentlich verlangsamt.

Für die **Unterstützung von Geschäftsprozessen** beschäftigen wir uns mit der Modellierung von Prozessen unter besonderer Berücksichtigung von Randbedingungen. Hierzu zählen Verfahren zur Analyse solcher Modellierungen sowohl bezüglich struktureller als auch verhaltens- und datenbezogener Eigenschaften sowie zur Generierung von Prozessmodellen, ausgehend von deskriptiv vorgegebenen Eigenschaften der Abläufe und von Optimierungszielen.

Diese Forschungsthemen passen gut zur thematischen Ausrichtung in der Lehre, insbesondere zu Datenbanktechnologie und zur Datenanalyse, aber auch zu Workflow-Management und Systemen zur Privatheitsthematik, seien es Vorlesungen, Seminare oder Praktika.

ERGEBNISSE UND ERFOLGE

Das DFG-Graduiertenkolleg „Energiezustandsdaten – Informatik-Methoden zur Erfassung, Analyse und Nutzung“, dessen Sprecher Herr Professor Böhm ist, hat neben sehr guten wissenschaftlichen Erfolgen auch durch eine Reihe von Aktivitäten auf sich aufmerksam gemacht, z. B. durch die erstmalige Publikation von Energieverbrauchsdaten einer industriellen Anlage oder durch die Durchführung eines Workshops zum Thema „Data Science“, der sich an KIT-Wissenschaftlerinnen und -Wissenschaftler anderer Disziplinen gerichtet hat und mit großem Interesse angenommen wurde.

Ein EU-Projekt im Bereich Predictive Maintenance mit Beteiligung der Professur wurde erfolgreich abgeschlossen. Beim Data Mining Cup, einem sehr renommierten, weltweiten Wettbewerb im Bereich „Datenanalyse“ mit üblicherweise weit mehr als 200 teilnehmenden Teams, hat ein Team von KIT-Studierenden im Rahmen eines einschlägigen Praktikums der Professur den 11. Platz errungen.

AUSGEWÄHLTE PUBLIKATIONEN

N. Arzamasova, M. Schäler, K. Böhm: Cleaning Antipatterns in an SQL Query Log. IEEE Transactions on Knowledge and Data Engineering (TKDE). Vol. 30, Number 3, 2018.

M. T. Cazzolato, A. J. M. Traina, K. Böhm: Efficient and Reliable Estimation of Cell Positions. 27th International Conference on Information and Knowledge Management (CIKM). Turin, Italien, 2018. Annahmequote Langbeiträge: 17%.

Á. Elekes, A. Englhardt, M. Schäler, K. Böhm: Resources to Examine the Quality of Word Embedding Models Trained on n-Gram Data. The SIGNLL Conference on Computational Natural Language Learning (CoNLL). Brüssel, Belgien, 2018. Annahmequote: 20,65%.

S. Bischof, H. Trittenbach, M. Vollmer, D. Werle, T. Blank, K. Böhm: HIPE – An Energy-Status-Data Set from Industrial Production. International Workshop on Energy Data and Analytics (EDA), zusammen mit ACM e-Energy. Karlsruhe, 2018.

M. Vollmer, I. Rutter, K. Böhm: On Complexity and Efficiency of Mutual Information Estimation on Static and Dynamic Data. 21st International Conference on Extending Database Technology (EDBT). Wien, Österreich, 2018. Annahmequote Langbeiträge: 24,8%.

MITARBEITERINNEN UND MITARBEITER

Verwaltungspersonal

Barbara Breitenstein
Bettina Wagner

Wissenschaftliches Personal

Vadim Arzamasov
Natalia Arzamasova
Jakob Bach
Dr. Aboubakr Achraf El Ghazi
Ábel Elekes
Adrian Englhardt
Edouard Fouché
Jutta Mülle
Elaheh Ordoni
Dr. Martin Schäler
Georg Steinbuß
Gabriela Suntaxi
Saeed Taghizadeh
Christine Tex
Holger Trittenbach
Michael Vollmer
Jens Willkomm

Technisches Personal

Christian Möck
Herma Teune