

Norbert Ritter, Andreas Henrich,
Wolfgang Lehner, Andreas Thor, Steffen Friedrich,
Wolfram Wingerath (Hrsg.)

**Datenbanksysteme für
Business, Technologie und Web
(BTW 2015)**

Workshopband

**02. – 03.03.2015
in Hamburg, Germany**

Gesellschaft für Informatik e.V. (GI)

Lecture Notes in Informatics (LNI) - Proceedings

Series of the Gesellschaft für Informatik (GI)

Volume P-242

ISBN 978-3-88579-636-7

ISSN 1617-5468

Volume Editors

Norbert Ritter

Universität Hamburg
Fachbereich Informatik
Datenbanken und Informationssysteme
22527 Hamburg, Germany
E-Mail: ritter@informatik.uni-hamburg.de

Andreas Henrich

Otto-Friedrich-Universität Bamberg
Fakultät Wirtschaftsinformatik und Angewandte Informatik
Lehrstuhl für Medieninformatik
96047 Bamberg, Germany
E-Mail: andreas.henrich@uni-bamberg.de

Wolfgang Lehner

Technische Universität Dresden
Fakultät Informatik
Institut für Systemarchitektur
01062 Dresden, Germany
Email: wolfgang.lehner@tu-dresden.de

Andreas Thor

Deutsche Telekom Hochschule für Telekommunikation Leipzig
Gustav-Freytag-Str. 43-45
04277 Leipzig, Germany
E-Mail: thor@hft-leipzig.de

Steffen Friedrich

Universität Hamburg
Fachbereich Informatik
Datenbanken und Informationssysteme
22527 Hamburg, Germany
E-Mail: friedrich@informatik.uni-hamburg.de

Wolfram Wingerath

Universität Hamburg
Fachbereich Informatik
Datenbanken und Informationssysteme
22527 Hamburg, Germany
E-Mail: wingerath@informatik.uni-hamburg.de

Series Editorial Board

Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Austria
(Chairman, mayr@ifit.uni-klu.ac.at)

Dieter Fellner, Technische Universität Darmstadt, Germany

Ulrich Flegel, Hochschule für Technik, Stuttgart, Germany

Ulrich Frank, Universität Duisburg-Essen, Germany

Johann-Christoph Freytag, Humboldt-Universität zu Berlin, Germany

Michael Goedicke, Universität Duisburg-Essen, Germany

Ralf Hofestädt, Universität Bielefeld, Germany

Michael Koch, Universität der Bundeswehr München, Germany

Axel Lehmann, Universität der Bundeswehr München, Germany

Peter Sanders, Karlsruher Institut für Technologie (KIT), Germany

Sigrid Schubert, Universität Siegen, Germany

Ingo Timm, Universität Trier, Germany

Karin Vosseberg, Hochschule Bremerhaven, Germany

Maria Wimmer, Universität Koblenz-Landau, Germany

Dissertations

Steffen Hölldobler, Technische Universität Dresden, Germany

Seminars

Reinhard Wilhelm, Universität des Saarlandes, Germany

Thematics

Andreas Oberweis, Karlsruher Institut für Technologie (KIT), Germany

© Gesellschaft für Informatik, Bonn 2015

printed by Köllen Druck+Verlag GmbH, Bonn

Vorwort

In den letzten Jahren hat es auf dem Gebiet des Datenmanagements große Veränderungen gegeben. Dabei muss sich die Datenbankforschungsgemeinschaft insbesondere den Herausforderungen von „Big Data“ stellen, welche die Analyse von riesigen Datenmengen unterschiedlicher Struktur mit kurzen Antwortzeiten im Fokus haben. Neben klassisch strukturierten Daten müssen moderne Datenbanksysteme und Anwendungen semistrukturierte, textuelle und andere multimodale Daten sowie Datenströme in völlig neuen Größenordnungen verwalten. Gleichzeitig müssen die Verarbeitungssysteme die Korrektheit und Konsistenz der Daten sicherstellen.

Die jüngsten Fortschritte bei Hardware und Rechnerarchitektur ermöglichen neuartige Datenmanagementtechniken, die von neuen Index- und Anfrageverarbeitungsparadigmen (In-Memory, SIMD, Multicore) bis zu neuartigen Speichertechniken (Flash, Remote Memory) reichen. Diese Entwicklungen spiegeln sich in aktuell relevanten Themen wie Informationsextraktion, Informationsintegration, Data Analytics, Web Data Management, Service-Oriented Architectures, Cloud Computing oder Virtualisierung wider.

Wie auf jeder BTW-Konferenz gruppieren sich um die Tagung eine Reihe von Workshops, die spezielle Themen in kleinen Gruppen aufgreifen und diskutieren. Im Rahmen der BTW 2015 finden folgende Workshops statt:

- Databases in Biometrics, Forensics and Security Applications: DBforBFS
- Data Streams and Event Processing: DSEP
- Data Management for Science: DMS

Dabei fasst der letztgenannte Workshop DMS als Joint Workshop die beiden Initiativen Big Data in Science (BigDS) und Data Management for Life Sciences (DMforLS) zusammen.

Mit seinen Schwerpunkten reflektiert das Workshopprogramm aktuelle Forschungsgebiete von hoher praktischer Relevanz. Zusätzlich präsentieren Studenten im Rahmen des Studierendenprogramms die Ergebnisse ihrer aktuellen Abschlussarbeiten im Bereich Datenmanagement. Für jeden Geschmack sollte sich somit ein Betätigungsfeld finden lassen!

Die Materialien zur BTW 2015 werden auch über die Tagung hinaus unter <http://www.btw-2015.de> zur Verfügung stehen.

Die Organisation einer so großen Tagung wie der BTW mit ihren angeschlossenen Veranstaltungen ist nicht ohne zahlreiche Partner und Unterstützer möglich. Sie sind auf den folgenden Seiten aufgeführt. Ihnen gilt unser besonderer Dank ebenso wie den Sponsoren der Tagung und der GI-Geschäftsstelle.

Hamburg, Bamberg, Dresden, Leipzig, im Januar 2015

Norbert Ritter, Tagungsleitung und Vorsitzender des Organisationskomitees

Andreas Henrich und Wolfgang Lehner, Leitung Workshopkomitee

Andreas Thor, Leitung Studierendenprogramm

Wolfram Wingerath, Steffen Friedrich, Tagungsband und Organisationskomitee

Tagungsleitung

Norbert Ritter, Universität Hamburg

Organisationskomitee

Norbert Ritter

Anne Hansen-Awizen

Steffen Friedrich

Wolfram Wingerath

Felix Gessert

Fabian Panse

Volker Nötzold

Studierendenprogramm

Andreas Thor, HfT Leipzig

Koordination Workshops

Andreas Henrich, Univ. Bamberg

Wolfgang Lehner, TU Dresden

Tutorienprogramm

Norbert Ritter, Univ. Hamburg

Thomas Seidl, RWTH Aachen

Andreas Henrich, Univ. Bamberg

Wolfgang Lehner, TU Dresden

Second Workshop on Databases in Biometrics, Forensics and Security Applications (DBforBFS)

Vorsitz: Jana Dittmann, Univ. Magdeburg; Veit Köppen, Univ. Magdeburg;

Gunter Saake, Univ. Magdeburg; Claus Vielhauer, FH Brandenburg

Ruediger Grimm, Univ. Koblenz

Dominic Heutelbeck, FTK

Stefan Katzenbeisser, TU Darmstadt

Claus-Peter Klas, GESIS

Günther Pernul, Univ. Regensburg

Ingo Schmitt, BTU Cottbus

Claus Vielhauer, FH Brandenburg

Sviatoslav Voloshynovskiy, UNIGE, CH

Edgar R. Weippl, SBA Research, Austria

Data Streams and Event Processing (DSEP)

Vorsitz: Marco Grawunder, Univ. Oldenburg, Daniela Nicklas Univ. Bamberg

Andreas Behrend, Univ. Bonn

Klemens Boehm, KIT

Peter Fischer, Univ. Freiburg

Dieter Gawlick, Oracle

Boris Koldehofe, TU Darmstadt

Wolfgang Lehner, TU Dresden

Richard Lenz, Univ. Erlangen-Nürnberg

Klaus Meyer-Wegener, Univ. Erlangen-Nürnberg

Gero Mühl, Univ. Rostock

Kai-Uwe Sattler, TU Ilmenau

Thorsten Schöler, HS Augsburg

Joint Workshop on Data Management for Science (DMS)

Workshop on Big Data in Science (BigDS)

Vorsitz: Birgitta König-Ries, Univ. Jena; Erhard Rahm, Univ. Leipzig;
Bernhard Seeger, Univ. Marburg

Alsayed Algergawy, Univ. Jena
Peter Baumann, Jacobs Univ.
Matthias Bräger, CERN
Thomas Brinkhoff, FH Oldenburg
Michael Diepenbroeck, AWI
Christoph Freytag, HU Berlin
Michael Gertz, Univ. Heidelberg
Frank-Oliver Glöckner, MPI-MM
Anton Güntsch, BGBM Berlin-Dahlem
Thomas Heinis, IC, London
Thomas Hickler, Senckenberg

Jens Kattge, MPI für Biogeochemie
Alfons Kemper, TU München
Meike Klettke, Univ. Rostock
Alex Markowetz, Univ. Bonn
Thomas Nauss, Univ. Marburg
Jens Nieschulze, Univ. Göttingen
Kai-Uwe Sattler, TU Ilmenau
Stefanie Scherzinger, OTH Regensburg
Myro Spiliopoulou, Univ. Magdeburg
Uta Störl, Hochschule Darmstadt

Workshop on Data Management for Life Sciences (DMforLS)

Vorsitz: Sebastian Dorok, Bayer Pharma AG; Matthias Lange, IPK Gatersleben;
Gunter Saake, Univ. Magdeburg

Sebastian Breß, TU Dortmund
Sebastian Dorok, Bayer Pharma AG
Mourad Elloumi, UTM Tunisia
Ralf Hofestädt, Univ. Bielefeld
Andreas Keller, Saarland Univ.
Jacob Köhler, DOW AgroSciences
Horstfried Läßle, Bayer HealthCare

Matthias Lange, IPK Gatersleben
Ulf Leser, HU Berlin
Wolfgang Müller, HITS GmbH
Erhard Rahm, Univ. Leipzig
Gunter Saake, Univ. Magdeburg
Uwe Scholz, IPK Gatersleben
Can Türker, ETH Zürich

Inhaltsverzeichnis

Workshopprogramm

Second Workshop on Databases in Biometrics, Forensics and Security Applications (DBforBFS)

Jana Dittmann, Veit Köppen, Gunter Saake, Claus Vielhauer

Second Workshop on Databases in Biometrics, Forensics and Security

Applications (DBforBFS)..... 19

Veit Köppen, Mario Hildebrandt, Martin Schäler

On Performance Optimization Potentials Regarding Data Classification in Forensics.....21

Maik Schott, Claus Vielhauer, Christian Krätzer

Using Different Encryption Schemes for Secure Deletion While Supporting Queries.....37

Data Streams and Event Processing (DSEP)

Marco Grawunder, Daniela Nicklas

Data Streams and Event Processing (DSEP).....49

Timo Michelsen, Michael Brand, H.-Jürgen Appelrath

Modulares Verteilungskonzept für Datenstrommanagementsysteme.....51

Niko Pollner, Christian Steudtner, Klaus Meyer-Wegener

Placement-Safe Operator-Graph Changes in Distributed Heterogeneous

Data Stream Systems.....61

Michael Brand, Tobias Brandt, Carsten Cordes, Marc Wilken, Timo Michelsen

Herakles: A System for Sensor-Based Live Sport Analysis using Private

Peer-to-Peer Networks.....71

Christian Kuka, Daniela Nicklas

Bestimmung von Datenunsicherheit in einem probabilistischen

Datenstrommanagementsystem.....81

Cornelius A. Ludmann, Marco Grawunder, Timo Michelsen,

H.-Jürgen Appelrath

Kontinuierliche Evaluation von kollaborativen Recommender-Systeme in

Datenstrommanagementsystemen.....91

Sebastian Herbst, Johannes Tenschert, Klaus Meyer-Wegener

Using Data-Stream and Complex-Event Processing to Identify Activities of Bats.....93

Peter M. Fischer, Io Taxidou

Streaming Analysis of Information Diffusion.....95

Henrik Surm, Daniela Nicklas <i>Towards a Framework for Sensor-based Research and Development Platform for Critical, Socio-technical Systems</i>	97
Felix Beier, Kai-Uwe Sattler, Christoph Dinh, Daniel Baumgarten <i>Dataflow Programming for Big Engineering Data</i>	101
Joint Workshop on Data Management for Science (DMS)	
Sebastian Dorok, Birgitta König-Ries, Matthias Lange, Erhard Rahm, Gunter Saake, Bernhard Seeger <i>Joint Workshop on Data Management for Science (DMS)</i>	105
Alexandr Uciteli, Toralf Kirsten <i>Ontology-based Retrieval of Scientific Data in LIFE</i>	109
Christian Colmsee, Jinbo Chen, Kerstin Schneider, Uwe Scholz, Matthias Lange <i>Improving Search Results in Life Science by Recommendations based on Semantic Information</i>	115
Marc Schäfer, Johannes Schildgen, Stefan Deßloch <i>Sampling with Incremental MapReduce</i>	121
Andreas Heuer <i>METIS in PARADISE Provenance Management bei der Auswertung von Sensordatenmengen für die Entwicklung von Assistenzsystemen</i>	131
Martin Scharm, Dagmar Waltemath <i>Extracting reproducible simulation studies from model repositories using the CombineArchive Toolkit</i>	137
Robin Cijvat, Stefan Manegold, Martin Kersten, Gunnar W. Klau, Alexander Schönhuth, Tobias Marschall, Ying Zhang <i>Genome sequence analysis with MonetDB: a case study on Ebola virus diversity</i>	143
Ahmet Bulut <i>RightInsight: Open Source Architecture for Data Science</i>	151
Christian Authmann, Christian Beilschmidt, Johannes Dröner, Michael Mattig, Bernhard Seeger <i>Rethinking Spatial Processing in Data-Intensive Science</i>	161
Studierendenprogramm	
Marc Büngener <i>CBIR gestütztes Gemälde-Browsing</i>	173
David Englmeier, Nina Hubig, Sebastian Goebel, Christian Böhm <i>Musical Similarity Analysis based on Chroma Features and Text Retrieval Methods</i>	183

Alexander Askinadze <i>Vergleich von Distanzen und Kernel für Klassifikatoren zur Optimierung der Annotation von Bildern</i>	193
Matthias Liebeck <i>Aspekte einer automatischen Meinungsbildungsanalyse von Online-Diskussionen</i>	203
Martin Winter, Sebastian Goebel, Nina Hubig, Christopher Pleines, Christian Böhm <i>Development and Evaluation of a Facebook-based Product Advisor for Online Dating Sites</i>	213
Daniel Töws, Marwan Hassani, Christian Beecks, Thomas Seidl <i>Optimizing Sequential Pattern Mining Within Multiple Streams</i>	223
Marcus Pinnecke <i>Konzept und prototypische Implementierung eines föderativen Complex Event Processing Systeme mit Operatorverteilung</i>	233
Monika Walter, Axel Hahn <i>Unterstützung von datengetriebenen Prozessschritten in Simulationsstudien durch Verwendung multidimensionaler Datenmodelle</i>	243
Niklas Wilcke <i>DduP – Towards a Deduplication Framework utilising Apache Spark</i>	253

Tutorienprogramm

Christian Beecks, Merih Uysal, Thomas Seidl <i>Distance-based Multimedia Indexing</i>	265
Kai-Uwe Sattler, Jens Teubner, Felix Beier, Sebastian Breß <i>Many-Core-Architekturen zur Datenbankbeschleunigung</i>	269
Felix Gessert, Norbert Ritter <i>Skalierbare NoSQL- und Cloud-Datenbanken in Forschung und Praxis</i>	271
Jens Albrecht, Uta Störl <i>Big-Data-Anwendungsentwicklung mit SQL und NoSQL</i>	275