

Editorial

Ralf Schenkel · Christa Womser-Hacker

Online publiziert: 25. Januar 2014
© Springer-Verlag Berlin Heidelberg 2014

1 Schwerpunktthema: Information Retrieval

Die Menge verfügbarer Information ist in den letzten Jahrzehnten dramatisch gewachsen, so dass es ohne Unterstützung durch Suchmaschinen und ähnliche Werkzeuge nahezu unmöglich wäre, relevante Information in der unüberschaubaren Menge des Rauschens zu identifizieren. Jeder kennt dieses Problem aus eigener Erfahrung bei der Suche nach speziellen Informationen im Web, wo neben der reinen textuellen Relevanz auch andere Faktoren wie Qualität der Quelle, Aktualität und Ortsbezüge eine große Rolle spielen. Suchprobleme auf großen Datenmengen treten aber nicht nur im Web auf, sondern sind inhärent in vielen anderen Anwendungsbereichen, wie zum Beispiel Patentrecherche, Suche nach möglichen Diagnosen in der Medizin oder Suche nach historischen Belegen für Aussagen und Ereignisse.

In der Information-Retrieval-Community sind viele grundlegende Verfahren entwickelt worden, ohne die Suche in der heutigen Qualität unmöglich wäre. Allerdings sind längst noch nicht alle Probleme zufriedenstellend gelöst, und insbesondere der Schritt von der anfrage- zur aufgabenbasierten Suche und die Unterstützung des Benutzers bei komplexen Suchen sind immer noch eine große Herausforderung für die Forschung.

R. Schenkel (✉)
Lehrstuhl für Informatik mit Schwerpunkt
Informationsmanagement, Universität Passau, 94032 Passau,
Deutschland
e-mail: ralf.schenkel@uni-passau.de

C. Womser-Hacker
Institut für Informationswissenschaft und Sprachtechnologie,
Universität Hildesheim, 31141 Hildesheim, Deutschland
e-mail: womser@uni-hildesheim.de

Diese Ausgabe des Datenbank-Spektrums stellt einige aktuelle Ergebnisse insbesondere aus dem Bereich der Interaktion von IR-Systemen mit dem Benutzer vor.

Der erste Beitrag *Limitations of automatic Patent IR – A plea for more interactivity* von Julia Jürgens und Christa Womser-Hacker (Universität Hildesheim) befasst sich mit der Recherche in großen Patentsammlungen. Er gibt zunächst einen Einblick in die Besonderheiten, die Patentrecherche von der Suche in anderen Dokumentarten unterscheiden, und stellt einen typischen Suchprozess vor. Nach einem umfassenden Überblick über diverse Lösungsansätze, die in der Literatur diskutiert wurden, werden einige interaktive Ansätze zur Verbesserung der Suchprozesse vorgeschlagen.

Im zweiten Beitrag *Strategic Search Support on Macro and Micro Level* stellen Andreas Tacke und Sascha Kriewel (Universität Duisburg-Essen) ein System zur Suche in medizinischen Dokumenten vor, das im Kontext des Kreshmoi-Projektes entstanden ist. Das System richtet sich gleichermaßen an Experten wie an Laienbenutzer und bietet daher eine Auswahl an speziellen Tools zur Unterstützung des Suchenden. Der Beitrag berichtet außerdem über die Ergebnisse einer vorläufigen Benutzerstudie, die die Effektivität ausgewählter Unterstützungswerkzeuge zeigt.

Der dritte Beitrag *Collaborative Information Seeking in Professional Work-Settings: A Study of Equipment Utilization* von Thilo Böhm, Claus-Peter Klas und Matthias Hemmje (Fernuniversität Hagen) betrachtet das Problem der kollaborativen Suche nach Information. Er stellt Ergebnisse einer Studie vor, die das kollaborative Suchverhalten dreier Arbeitsgruppen aus Wissenschaft und Industrie untersucht, und leitet daraus Anforderungen an die Entwicklung solcher kollaborativer Suchsysteme ab.

Im vierten Beitrag *Applying Geographic Information Retrieval – An Experience Report on Developing Local Search*

for a Developing Country berichtet Dirk Ahlers (NTNU Trondheim, Norwegen) über Erfahrungen bei der Entwicklung einer Suchmaschine für Webseiten aus Honduras, die insbesondere räumliche Suche unterstützt. Er zeigt dabei, welche besonderen Probleme in einem Land wie Honduras auftreten und wie sie im konkreten Projektkontext gelöst wurden.

Die vier Schwerpunktbeiträge werden in der Rubrik „Kurz erklärt“ ergänzt durch einen Beitrag *DB2 10.5 mit BLU Acceleration: Optimierte für komplexe analytische Anfragen* von Manfred Päßler und Matthias Nicola, IBM.

Weiterhin finden Sie unter der Rubrik „Datenbankgruppen vorgestellt“ einen Beitrag von Ingo Frommholz, Udo Kruschwitz und John Tait, die die *Information Retrieval Special Group* der BCS vorstellen, das britische Pendant zur Fachgruppe Information Retrieval der Gesellschaft für Informatik.

In diesem Heft erreicht die Rubrik „Dissertationen“ einen neuen Rekord mit 15 Kurzfassungen von Dissertationen aus der deutschen DBIS-Community.

In der Rubrik „Community“ beschreibt Jens Dittrich, Universität des Saarlandes, in einem Beitrag *Die Umgedrehte Vorlesung – Chancen für die Informatiklehre* seine Erfahrungen mit einer Umgedrehten Datenbankvorlesung, wo sich die Studierenden den Vorlesungsstoff vorab mit Hilfe von Lehrvideos aneignen sollen, damit er in der eigentlichen Vorlesung vertieft werden kann. Wie immer enthält diese Rubrik auch einen Beitrag *News* mit aktuellen Informationen.

2 Künftige Schwerpunktthemen

2.1 Datenmanagement in der Cloud

Anwendungen, Institutionen und Unternehmen müssen mit immer größeren Datenmengen hantieren. Sowohl wissenschaftliche Experimente (z.B. Genomsequenzierung, Teilchenbeschleuniger) als auch die Aufzeichnung menschlicher Aktivitäten (z.B. soziale Netzwerke, Online-Shopping) erzeugen Datenmengen, deren Speicherung und Verarbeitung oft nur noch unter Ausnutzung von Cloud-Infrastrukturen möglich ist. Die Kopplung tausender Rechner ermöglicht dabei eine ökonomische Verarbeitung großer Datenmengen, da die erforderlichen Ressourcen (u.a. Rechenkapazität, Speicherplatz) flexibel „on-demand“ an den Bedarf von Nutzern und Anwendungen angepasst werden können. Gleichzeitig ergeben sich jedoch eine Vielzahl wissenschaftlicher Fragestellungen hinsichtlich der Speicherung, Verarbeitung und Analyse großer Datenmengen. Dieses Themenheft des Datenbank-Spektrums soll einen Überblick über aktuelle Entwicklungen und Herausforderungen im Bereich Cloud-Datenmanagement geben und

mögliche Richtungen für Forschung, Lehre und Entwicklung aufzeigen.

Mögliche Themen (Auswahl):

- Anfrageverarbeitung und -optimierung
- Datenanalyse
- Datenschutz und Datensicherheit
- Benchmarks
- Ressourcen- und Workload-Management
- Geschäftsmodelle

Beitragsformat: 8–10 Seiten, zweispaltig

Gastherausgeber:

Andreas Thor, Hochschule für Telekommunikation Leipzig
thor@hft-leipzig.de

Stefanie Scherzinger, Hochschule Regensburg
stefanie.scherzinger@hs-regensburg.de

Günther Specht, Universität Innsbruck
guenther.specht@uibk.ac.at

Einreichung der Beiträge bis zum 1. Februar 2014

3 Data Management on New Hardware

Continued progress and evolution of computer hardware and infrastructure provide new optimization opportunities for data management concerning performance, scalability, availability, energy efficiency, and more. The traditional DB focus on I/O optimization is not enough to fully utilize this potential. New aspects include processors (multi-core CPUs, GPUs, FGPA, etc.), multi-level memory hierarchies, storage systems (storage-class memories such as flash (SSDs) and phase-change memory), or low-power hardware. They impose new challenges and bottlenecks to DB performance, but, in turn, provide opportunities to improve and accelerate data-intensive systems for differing kinds of workloads (e.g., OLTP, DW/OLAP, ETL, Streaming/Realtime, or XML/RDF processing). An important goal is the automatic and transparent adaptation to the new hardware characteristics, thereby achieving maximum gain (e.g., performance, energy efficiency, etc.) for the applications.

A special issue of Datenbank-Spektrum wants to publish research contributions providing an overview over ongoing work in the data management area. Submissions covering topics from the following non-exclusive list are encouraged:

- Database algorithms (e.g., caching) and data structures (e.g., indexes) on modern hardware
- Hardware or memory support for query processing
- Cost models and query optimization for novel hierarchical memory systems
- Database architectures on multi-threaded and chip multi-processors
- Databases and transactional memory systems

- Database systems supporting energy efficiency and energy proportionality
- New benchmarks for or performance analysis of database workloads on modern hardware

Paper format: 8–10 pages, double column

Notice of intent for a contribution: April 15th, 2014

Guest editors:

Theo Härder, University of Kaiserslautern,

haerder@cs.uni-kl.de

Jens Teubner, TU Dortmund,

jens.teubner@cs.tu-dortmund.de

Deadline for submissions: June 1st, 2014

4 Informationsmanagement für Digital Humanities

In den Geisteswissenschaften fallen in immer größerer Menge digitale Forschungsdaten an. Dabei ergeben sich durch die spezifischen Rahmenbedingungen zahlreiche Herausforderungen für Datenbanken und IR-Systeme: Die Daten und Dokumente sind heterogen in Sprache, Struktur und Qualität. Es gibt zwar eine Vielzahl von Standards und Methoden, eine übergreifende Sicht existiert aber kaum. Relevante Kollektionen mit elektronischen Texten, Metadaten, Bildern und anderen multimedialen Ressourcen liegen in verschiedenen Disziplinen und Institutionen vor und bilden eine hochgradig verteilte und heterogene Informationslandschaft, deren Verarbeitung oft im Rahmen spezifischer, geisteswissenschaftlicher Forschungsfragen erfolgt. Von besonderer Bedeutung sind die Erschließung, Veröffentlichung und Verwaltung digitaler Ressourcen im Rahmen spezifischer Anwendungen z. B. in der Archäologie, den Geschichts-, Sprach- oder Religionswissenschaften, aber insbesondere auch im Kontext interdisziplinärer Forschung. Im Themenheft sollen einführende und überblicks-

artige Artikel sowie aktuelle Forschungsergebnisse zu ausgewählten Themen ein breites Bild zum aktuellen Stand des Informationsmanagements für Digital Humanities geben.

Mögliche Themen aus diesem Bereich könnten z. B. sein:

- Integrierte Analyse, Verarbeitung und Visualisierung verteilter bzw. heterogener Kollektionen
- Nutzung, Entwicklung und Auswertung von Vokabularen, Thesauri und Ontologien
- Langzeitarchivierung und Datenprovenienz
- Katalogisierung, Annotation und Dokumentation von Ressourcen (Data Curation)
- Erkennung, Analyse und Visualisierung kollektionsinterner oder -übergreifender Zusammenhänge z. B. durch Analyse von Ort und Zeit, Themen, Named Entities
- Aspekte der Usability im Umgang mit verteilten und heterogenen Ressourcen
- Anwendungen zum Datenmanagement, zur Suche und zur Analyse in speziellen Anwendungsfeldern aus den Geisteswissenschaften
- *Big Data*-Technologien für die Digital Humanities
- Forschungsinfrastrukturen für die Digital Humanities

Beitragsformat: 8–10 Seiten, zweispaltig

Ankündigung einer Beitragseinreichung bis zum 1. August 2014

Gastherausgeber:

Andreas Henrich, Otto-Friedrich-Universität Bamberg

andreas.henrich@uni-bamberg.de

Gerhard Heyer, Universität Leipzig

hey@informatik.uni-leipzig.de

Christoph Schlieder, Otto-Friedrich-Universität Bamberg

christoph.schlieder@uni-bamberg.de

Einreichung der Beiträge bis zum 1. Oktober 2014