Editorial

Johann-Christoph Freytag · Bernhard Mitschang

Online publiziert: 24. Mai 2013
© Springer-Verlag Berlin Heidelberg 2013

1 Schwerpunktthema: RDF Data Management

In der heutigen Web-basierten und ubiquitären Datenwelt werden mehr und mehr Daten durch das W3C Resource Description Framework (RDF) modelliert und verwaltet und durch die RDF-Anfragesprache SPARQL bzw. durch das W3C SPARQL Protocol angefragt. RDF selbst ist im Allgemeinen als konzeptuelles Datenmodell für strukturierte Daten/Information bekannt, das als Standard eine Schlüsselrolle im Semantic Web spielen sollte, um Metadaten im Web zu beschreiben. Es erlaubt, Beziehungen zwischen Ressourcen als “first-class citizens” zu unterstützen, und bringt die Flexibilität für jegliche Art von Schemadefinitionen mit, selbst für den Fall, dass kein Schema existiert.

RDF erlaubt, Daten zunächst ganz ohne Schemainformation zu speichern, um dann möglicherweise zu einem späteren Zeitpunkt – falls notwendig – Schemadefinitionen hinzuzufügen. Diese Flexibilität führte dazu, dass etliche Anwendungsdomänen begannen, RDF zur Datenspeicherung und zum Datenaustausch zu nutzen, so im Bereich Life Science, im Bereich der facettenreichen Datenintegration, zum Sammeln bereichsspezifischer Daten und großer Wissenssammlungen wie beispielsweise DBpedia.

Auf Grund der wachsenden Bedeutung von RDF und seiner vielfältigen Nutzung in unterschiedlichen Bereichen

widmet sich diese Ausgabe des Datenbank-Spektrums verschiedenen Facetten von RDF. Es sollen sowohl Übersichtsarbeiten als auch neuere Entwicklungen im Bereich RDF als auch Herausforderungen, zukünftige Entwicklungen in Technologien und Anwendungen dargestellt werden. Diese Breite führte zur Auswahl von fünf Beiträgen, die in dieser Ausgabe des Datenbank-Spektrums vorgestellt werden.

Als Erstes geben Sören Auer, Jens Lehmann, Axel-Cyrille Ngonga Ngomo, Claus Stadler und Jörg Unbehauen (Universität Leipzig) in ihrem Beitrag mit dem Titel *Der Lebenszyklus von Linked Data im Web – Von der Extraktion und dem Mapping zur Verlinkung von Daten im Web* einen Einblick und Überblick über verschiedene Herausforderungen des Linked Data Management im Web. Mit der DBpedia-Wissensextraktion aus Wikipedia, dem skalierbaren Linking von Wissensbasen und dem Mapping relationaler Daten nach RDF stellen sie drei Ansätze vor, die zentrale Bausteine für das “Web der Daten” ausmachen.

Im zweiten Beitrag dieses Heftes stellt Olaf Hartig (University of Waterloo, Canada/Humboldt-Universität zu Berlin) unter dem Titel *An Overview on Execution Strategies for Linked Data Queries* verschiedene Herausforderungen vor, die bei der Ausführung von Anfragen auf RDF-Daten – bei ihm als Linked Data bezeichnet – existieren. Dabei nimmt er auch eine Klassifikation der verschiedenen Anfrageausführungsstrategien vor, wie sie aus unterschiedlichen Beiträgen der Forschungsliteratur zu entnehmen sind.

Inhaltlich verwandt zum zweiten Beitrag ist der nun folgende mit dem Titel *Effiziente Verarbeitung allgemeiner Anfragen in RDF Triple-Stores* von Andreas Brodt und Bernhard Mitschang (Universität Stuttgart), die einen Anfrageauswertungsansatz vorstellen, der die Attributwerte von Ressourcen in RDF-Datenbanken effizient ermittelt, ohne die Effizienz der komplexen Graphmustersuche zu kompromittieren. Dazu werden zunächst diejenigen Ressourcen

J.-C. Freytag (✉)
AG Datenbanken und Informationssysteme,
Humboldt-Universität zu Berlin, 12489 Berlin, Deutschland
e-mail: freytag@dbis.informatik.hu-berlin.de

B. Mitschang
Institut für Parallele und Verteilte Systeme, Universität Stuttgart,
70569 Stuttgart, Deutschland
e-mail: Bernhard.Mitschang@ipvs.uni-stuttgart.de

cen bestimmt, die der Anfrage entsprechen, ehe dann alle ihre Attributwerte gleichzeitig unter Ausnutzung zusammenhängender Speicherbereiche geladen werden. Weiterhin wird vorgestellt, wie domänenspezifische Indexe in RDF-Datenbanken integriert werden können, um auch wertbasierte Anfrageprädikate effizient ausführen zu können.

Unter dem Titel *Improving RDF Data through Association Rule Mining* stellt der Beitrag von Ziawasch Abedjan und Felix Naumann (Universität Potsdam, Hasso-Plattner-Institut) einen Ansatz zur RDF-Datenanalyse auf der Methode des Association Rule Mining vor. Die so aus den RDF-Daten analysierten Zusammenhänge werden zur Datenvervollständigung und auch zur Anfrageergänzung eingesetzt und dienen somit der Konsistenzverbesserung.

Abgeschlossen wird die Reihe der Schwerpunktbeiträge im vorliegende Heft mit einem internationalen Beitrag über Benchmark-Aktivitäten im Bereich des Linked Data. Peter Boncz (CWI, Amsterdam), Irini Fundulaki (FORTH, Griechenland), Andrey Gubichev (Technische Universität München), Josep Larriba-Pey (UPC, Spanien) und Thomas Neumann (Technische Universität München) stellen den *Linked Data Benchmark Council (LDBC)* vor, der sich sowohl aus akademischen als auch industriellen Mitgliedern zusammensetzt. Ähnlich wie der *Transaction Processing Performance Council (TPC)* soll im Rahmen eines Europäischen Projektes ein herstellerunabhängiger Benchmark spezifiziert werden. Der vorliegende Beitrag beschreibt den existierenden Stand des Benchmarks auf RDF- und Graph-Datenbanken. Gleichzeitig wird mit diesem Beitrag eine Einladung an alle Leser ausgesprochen, sich an der Entwicklung des Linked Data Benchmark zu beteiligen.

Die fünf Schwerpunktbeiträge werden ergänzt durch einen Fachbeitrag *Was lange währt, wird endlich gut: Temporale Daten im SQL-Standard* von Dušan Petković (Fachhochschule Rosenheim). Dieser Aufsatz setzt sich mit den Erweiterungen des SQL-Standards für die Verwaltung temporaler Daten auseinander. Neben den so genannten “application-time period tables” werden systemversionierte Tabellen eingeführt und deren mögliche Nutzung in der praktischen Anwendungen diskutiert. Außerdem nimmt der Autor zu dieser erstmaligen Aufnahme temporaler Daten in den SQL:2011-Standard kritisch Stellung und erörtert fehlende Spezifikationsmöglichkeiten weiterer wünschenswerter Eigenschaften des temporalen Datenmodells, die in nachfolgenden Erweiterungen des Standards aufgenommen werden sollten.

Weiterhin finden Sie unter der Rubrik “Datenbankgruppen vorgestellt” einen Beitrag von Erhard Rahm zu *Der Lehrstuhl Datenbanken an der Universität Leipzig*, der nach einer kurzen Einleitung zur Entwicklung des Lehrstuhls einen Überblick zu den aktuellen Forschungsthemen und zum Leipziger Lehrprofil gibt. Die Rubrik “Dissertationen” ist in diesem Heft mit 8 Kurzfassungen von Dissertationen erfreulicherweise wiederum recht umfangreich.

In der Rubrik “Community” gibt es mehrere Beiträge. Dirk Reinel berichtet von der *Herbstschule “Information Retrieval” 2012* im Leibniz-Zentrum für Informatik auf Schloss Dagstuhl. Die BTW 2013 an der Otto-von-Guericke-Universität in Magdeburg hatte mit über 250 Teilnehmern eine erfreuliche Resonanz. Der Beitrag *BTW 2013 – Zwischen wissenschaftlicher Geschichte und moderner Herausforderung* von Veit Köppen, Martin Schäler, Alexander Grebhahn und Gunter Saake gibt einen detaillierten Einblick in die verschiedenen Veranstaltungen unserer größten Fachtagung. Wie immer enthält diese Rubrik auch einen Beitrag *News* mit aktuellen Informationen.

2 Künftige Schwerpunktthemen

2.1 Best Workshop Papers of BTW 2013

This special issue of the “Datenbank-Spektrum” is dedicated to the Best Papers of the Workshops running at the BTW 2013 at the TU Magdeburg. The selected Workshop contributions should be extended to match the format of regular DASP papers.

Paper format: 8–10 pages, double column

Selection of the Best Papers by the Workshop chairs and the guest editor: April 15th, 2013

Guest editor:

Theo Härder, University of Kaiserslautern,
haerder@cs.uni-kl.de

Deadline for submissions: June 10th, 2013

2.2 Information Retrieval

The amount of available information has increased dramatically in the last decades. At the same time, the way in which this information is presented has changed rapidly: Multimedia data such as audio, images, and video complements or even replaces textual information, user-generated content from blogs or social networks replaces static Web sites, and highly dynamic content such as tweets is published in real-time. Information Retrieval methods allow to quickly find relevant pieces of information for a possibly complex information need from this huge pile of data.

This special issue of the Datenbank-Spektrum aims to provide an overview of recent developments, challenges, and future directions in the field of Information Retrieval technologies and applications.

Topics of interest include (but are not limited to)

- Crawling, Indexing, Query Processing
- Information Extraction and Mining
- Interactive Information Retrieval
- Personalized and Context-Aware Retrieval
- Structured and Semantic Search
- Evaluation and Benchmarking
- Archiving and Time-Aware Retrieval Models
- Enterprise Search
- Realtime Search: Streams, Tweets, Social Networks
- Multimedia Retrieval

Paper format: 8–10 pages, double column

Notice of intent for a contribution: July 15th, 2013

Guest editors:

Ralf Schenkel, Max-Planck-Institut für Informatik,
schenkel@mpi-inf.mpg.de

Christa Womser-Hacker, Universität Hildesheim,
womser@uni-hildesheim.de

Deadline for submissions: October 1st, 2013

2.3 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 ermög-

licht 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
 Ankündigung einer Beitragseinreichung bis zum 15. Dezember 2013

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