1 LWA 2013 in Bamberg

Der traditionelle Herbstworkshop der Fachgruppe Information Retrieval (IR) fand vom 7.–9. Oktober 2013 im Rahmen der Konferenz „Lernen, Wissen, Adaption“ (LWA) 2013 in Bamberg statt. Traditionell bot die LWA wieder ein Dach für die Workshops verschiedener Fachgruppen innerhalb der Gesellschaft für Informatik. Ziel des IR-Workshops war es, ein Forum zur wissenschaftlichen Diskussion und zum Austausch neuer Ideen zu schaffen. Der Workshop richtete sich daher gezielt auch an Nachwuchswissenschaftler und Teilnehmer aus der Industrie. Dementsprechend boten sich interessante Vorträge zu aktuellen Themen aus dem Information Retrieval wie Interaktives Information Retrieval, semantische Suche und Anwendungsszenarien. Gemeinsamen Vortragssitzungen mit den Fachgruppen Knowledge Discovery, Data Mining und Maschinelles Lernen (KMDL) sowie Wissensmanagement (WM) waren eine willkommene Gelegenheit, Arbeiten auch über die Grenzen der eigenen Fachgruppe hinaus zu präsentieren und zu diskutieren.

Vier Keynotes rundeten die LWA in diesem Jahr ab: „Systematic Approach to the Knowledge Extraction and Structuring for Scientific Big Data Analytics“ (Sung-Pil Cho, Yonsei University, Südkorea), „Collaborative Multi-Expert-Systems“ (Klaus-Dieter Althoff, Universität Hildesheim), „Qualitative Representations of Space and Time“ (Diedrich Wolter, Universität Bamberg), und „Smart Grid Data Analytics to Promote Energy Efficiency“ (Thorsten Staake, Universität Bamberg).

Auch im nächsten Jahr wird die Fachgruppe IR wieder mit einem Workshop an der LWA teilnehmen, die voraussichtlich im September 2014 in Aachen stattfinden wird und von Thomas Seidl organisiert wird.

2 Bericht über das Herbsttreffen der Fachgruppe Datenbanken

Am 5. und 6. Dezember 2013 veranstaltete die Fachgruppe Datenbanken erneut ein Symposium zum Thema „Big Data“. Die Veranstaltung war dabei sowohl als Nachfolger des erfolgreichen ersten Big Data Symposiums im Herbst 2012 an der TU München als auch als reguläres Herbsttreffen der Fachgruppe positioniert. Gastgeber war dieses Mal das IBM Forschungs- und Entwicklungszentrum in Böblingen. Mit über 100 Teilnehmern reicht sich das Treffen in die Liste der sehr erfolgreichen Fachgruppentreffen ein, wobei das IBM-Zentrum als Ortlichkeit und die Vor-Ort-Organisation durch Knut Stolze und Team absolut perfekt waren.

Wie üblich rankt sich das Herbsttreffen um einen sozialen Abendevent in Form einer Nachmittags- und einer am darauffolgenden Tag stattfindenden Vormittagsveranstaltung. Das Programm wurde von Wolfgang Lehner und Gerhard Weikum zusammengestellt. Das Nachmittagsprogramm wurde nach einer kurzen Begrüßung durch Udo Hertz (Leiter IBM Information Management) mit einem eingeladenen Vortrag von Prof. Christian Jensen eröffnet. Prof. Jensen arbeitet an der Aalborg University in Dänemark und ist einer der international renommiertesten Vertreter der DB-Forschungs-Community. In seinem Vortrag mit dem Titel „Topics in Geo-Spatial Data Management: Spatial Keyword Querying and Beyond“ arbeitete Christian Jensen sehr präzise die Anforderungen großer Datenmengen im Kontext einer räumlichen Anfrageverarbeitung heraus und gab ein paar Einblicke in seine aktuellen Forschungsarbeiten in diesem Umfeld.

Daran anschließend konnte der Leiter der Datenbankgruppe der Oracle Labs, Eric Sedlar, begrüßt werden. Eric Sedlar, aus Redwood Shores (CA) angereist, zeigte am Beispiel von zwei Projekten die Aktivitäten der Oracle Labs
im Umfeld von Big Data: In einem ersten Projekt wird die Optimierung des „time-to-query“ adressiert. Wesentlich in agilen Big-Data-Anwendungen ist es, die Zeit, Daten in ein System zu laden und physisch zu optimieren, so minimal wie möglich zu halten. Oracle Labs arbeitet hierbei an hardware-unterstützten Ladeprozessoren, die Daten „in Echtzeit“ in das Datenbanksystem einbringen können. Das zweite von Eric Sedlar skizzierte Projekt gab einen Einblick in eine auf der Java VM basierenden neuen Laufzeit für das Datenbanksystem. Die Idee, Anwendungsprogramme sehr eng mit Datenbankoperatoren zu vermischen, scheint in diesem Projekt tatsächlich einen großen Schritt vorangegangen zu sein. Im Anschluss an Oracle gab Jan Teichmann einen Einblick in die Big-Data-Strategie der SAP, insbesondere im Kontext des hauptspeicherbasierten Datenbanksystems HANA. Die Take-Away Message des Beitrags bestand darin, dass (a) Hardware-Entwicklungen es ermöglichen, auch sehr große Datenbestände im Hauptspeicher halten zu können, dass (b) es fundamental wichtig ist, sowohl transaktionale als auch analytische Zugriffe im gleichen System und sogar auf den gleichen Datenbeständen durchführen zu können, und dass (c) SAP für sich „Big“ nicht nur für „Data“, sondern „Big“ auch mit Blick auf eine große Anzahl an Nutzern und eine große Anzahl an Anwendungen definiert und Lösungen auch jenseits der reinen Datenbanktechnik auf Ebene der Geschäftsanwendungen anbietet.

Nach einer bewusst ausführlichen Kaffeepause wurde die Vortragsreihe wiederum von einem Mitglied der internationalen Datenbank-Community begonnen. Prof. Anastasia Ailamaki (EPFL) referierte zum Thema „Exploring Big Brain Data“. In der ihr typischen sehr motivierenden und engagierten Art begeisterte Prof. Ailamaki mit einer Reise in die Thematik der data-driven science und gab – neben einführenden und grundlegenden Themen des Information Retrieval und des Information Foraging – interessante Einblicke in die damit verbundenen Datenbankfragenstellungen. Als Beispiele nannte Anastasia Ailamaki „brain indexing“ und „structural analysis“. Indizierung von Gehirnstrukturen ist dahingehend komplex, als dass klassische räumliche Indexstrukturen diese speziellen biologischen Strukturen nur sehr schlecht repräsentieren können. Mit ihrem speziellen Ansatz FLAT zeigte sie eine innovative Art die Beschränkungen zu umgehen. Als weitere Herausforderung adressierte Prof. Ailamaki die Eigenschaft, sich an Gehirnstrukturen bei einer Navigation in der Datenanalyse zu orientieren. Um hier interaktiv arbeiten zu können, stellte sie mit dem Projekt SCOUT eine inhaltbezogene Pre-Fetching-Strategie vor – Experimente auf Basis echter Datenbestände belegen einen signifikanten Speedup. Als Vertreter der Gruppe von Big-Data-Anwendungen stellte in einem letzten Vortrag an diesem Nachmittag Justin Coffey von der Firma Criteo ein Big-Data-Anwendungsszenario vor. Criteo analysiert Web-Logs und leitet daraus personalisierte Werbestrategien ab. Im Rahmen des Vortrags erläuterte Justin Coffey die weltweite Datensammel- und Datenkonsolidierungsinfrastruktur und zeigte eindrucksvoll, dass zwar Big Data bereits möglich ist, aber auch noch sehr viel (auch von Seiten der Forschung) zu tun ist, um Big Data-Anwendungen im großen Stil zu ermöglichen.

Den Abend rundete ein ausgezeichnetes Dinner im Mercure Bristol Hotel ab – in diesem Zusammenhang ein großes Dankeschön für das Sponsoring durch IBM.

Der zweite Tag wurde eröffnet mit einem Vortrag von Felix Naumann vom Hasso-Platten-Institut in Potsdam, der neue Ergebnisse aus dem Bereich Big Data Profiling vorstellte, insbesondere neue effiziente Algorithmen für das Finden von eindeutigen Spaltenkombinationen in großen Tabellen. Jens Seifert (IBM) gab anschließend einen Überblick über die Architektur von DB2 BLU, der Hauptspeicher-Datenbank von IBM, und ging besonders auf Komprimierung von Daten und die transparente Einbindung von Hauptspeicherdaten in die Ausführung von Anfragen ein. Marc Spaniol (Max-Planck-Institut für Informatik Saarbrücken) ging in seinem anschließenden Vortrag auf Herausforderungen und mögliche Lösungsansätze für Temporal Web Analytics ein. Den Abschluss der Vorträge bildete der Vortrag von Harald Schöning (Software AG), der Einblicke in Big Data bei der Software AG gab.

Die interessanten Vorträge wurden abgerundet durch ein Panel über den Einfluss von Big Data auf die Informatikausbildung, das von Bernhard Mitschang (Uni Stuttgart) moderiert wurde. Die Panelisten Johann Christoph Freytag (Humboldt-Universität zu Berlin), Harald Schöning (Software AG), Gerhard Weikum (Max-Planck-Institut für Informatik) und Thomas Ruf (GfK) beleuchteten dabei sowohl Aspekte aus der Sicht von Wissenschaft und Forschung als auch aus Anwendungssicht.

An dieser Stelle also nochmals vielen Dank für die Teilnahme, an alle Sprecher und insbesondere auch Dank an IBM für die umfangreiche Unterstützung.

3 Autumn School 2014 for Information Retrieval and Information Foraging (Herbstschule)

Die Fachgruppe Information Retrieval veranstaltet vom 21.–26. September 2014 erneut eine Herbstschule, die diesmal erstmals komplett auf Englisch durchgeführt wird. Im bewährten Ambiente von Schloss Dagstuhl, dem Ausbildungsbetrieb und Begegnungszentrum der Gesellschaft für Informatik, werden Kenntnisse über die Problembereiche und Erfolgsfaktoren im Information Retrieval vermittelt. Dazu werden neben einführenden und grundlegenden Themen des Information Retrievals und des Information Foraging insbesondere aktuelle Forschungsfragen angesprochen. Die folgenden Themen sind vorgesehen und werden von namhaften Experten vorgestellt:
The Autumn School 2014 for Information Retrieval and Information Foraging richtet sich an Anwender, Entwickler sowie Studierende aus Wissenschaft und Praxis. Die Autumn School steht in der Tradition der Herbstschule IR und der Summer School Information Foraging, welche von der Radboud University in Nijmegen veranstaltet wurde. Weitere Informationen und Möglichkeit zur Anmeldung finden Sie unter http://tinyurl.com/asirf2014. Bei Anmeldung bis zum 15. Juli 2014 gelten ermäßigte Teilnahmebeiträge.

5 Produkt-News

Uta Störl

5.1 Oracle Database 11g – Final Release freigegeben

Nachdem im Sommer 2013 bereits das neueste Release der Oracle Datenbank (12.1) freigegeben wurde (siehe Datenbank-Spektrum 03/2013), erhielt Ende Oktober 2013 das sogenannte final release der Datenbank mit der Versionsnummer 11.2.0.4 die Freigabe. Das Release enthält neben den üblichen Fehlerkorrekturen einige neue Features, die aus dem 12er Release zurückübernommen (backported) wurden. Im Security Bereich etwa wurde das Feature Data Redaction übernommen, mit dem es möglich ist, sicherheitskritische Daten bei jeder Ausgabe zu maskieren, ohne diese Daten im Original zu verändern. Data Redaction wird verwendet, um zum Beispiel Teile von Kreditkartennummern auf Zahlungsbelegen durch das Zeichen ‘*’ oder ein anderes beliebiges Zeichen zu ersetzen. Der Standardsupport für die Version 11.2.0.4 endet im Januar 2015.

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

5.2 PostgreSQL 9.3 veröffentlicht

Version 9.3 des freien Datenbankmanagementsystems PostgreSQL wurde veröffentlicht. Als wichtigste Neuerung wird von den Entwicklern die Möglichkeit des bidirektionalen...
Datenaustauschs mit anderen Systemen basierend auf einem Foreign Data Wrapper (FDW) angeben. Weitere Veränderungen sind u.a. zusätzliche JSON-Funktionalitäten, Updatable VIEWs und die Parallelisierung der Dump-Funktionalität zur Verbesserung der Backup-Geschwindigkeit.

PostgreSQL, http://www.postgresql.org/

5.3 Couchbase lite für Mobilgeräte

Couchbase hat eine mobile Version seines dokumentenorientierten NoSQL-Datenbankmanagementsystems vorgestellt: Couchbase lite. Couchbase lite steht derzeit als Beta Version 2 zur Verfügung und unterstützt iOS und Android. Für die Programmierung werden derzeit REST- und Objective C-APIs unterstützt – ein Java-API ist angekündigt.

Couchbase, http://www.couchbase.com/mobile

5.4 Oracle: Exadata Database Machine X4-2 und Oracle Database 12c

Seit Dezember 2013 ist die fünfte Generation der Exadata Database Machine offiziell verfügbar. Mit der neuen Maschinengeneration ist es nun möglich, nicht nur Oracle Database 11g Release 2, sondern auch die im letztem Sommer vorgestellte Oracle Database 12c (siehe Datenbank-Spektrum 03/2013) mit dem vollen Funktionumfang zu nutzen. Der Kunde kann wie bisher zwischen dem Modell X4-2 als Eighth-, Quarter-, Half- oder Full Rack und dem Modell X3-8 als Full Rack mit jeweils unterschiedlicher Festplattenkonfiguration wählen.

In der Exadata Database Machine X4-2 kommen nun aktuelle Standard Sun Server X4-2 mit neuem 12-Core Intel Xeon E5-2697 v2 Prozessor (2,7GHz) als Datenbankserver zum Einsatz. Deren interne Festplatten wurden von 300 GB auf 600 GB Größe verdoppelt, der standardmäßig zur Verfügung stehende Hauptspeicher von 256 GB kann mit einem optionalen Memory Expansion Kit auf 512 GB vergrößert werden. Für die interne Vernetzung der Server in der Exadata Database Machine über InfiniBand werden nun PCIe 3.0 Ports genutzt, wobei durch den Betrieb beider IB Ports im Active/Active Modus noch mehr Bandbreite zur Verfügung steht. Mit den vorgestellten Hardwareerweiterungen eignet sich die Exadata Database Machine noch besser für die Kon solidierung mehrerer Oracle Datenbanken.

Die Datenbanksysteme im Modell X3-8 sind unverändert geblieben, jedoch wird auch in diesem Modell jetzt auf Sun Server X4-2L als zentraler Speicher-Baustein aller Engineered Systems mit Exadata Storage Technologie gesetzt. Mit neuen, schnelleren Six-Core Intel Xeon E5-2630 v2 Prozessoren (2,6 GHz) ausgestattet, verfügt jeder Exadata Storage Server nun über 96 GB Hauptspeicher. Der zusätzliche Hauptspeicher wird benötigt, um den auf vier Sun Flash Accelerator F80 PCIe Karten befindlichen Flash Speicher von insgesamt 3,2 TB zu adressieren. Als Festplatten kann der Kunde zwischen 1,2 TB 10k RPM High Performance oder 4 TB 7,2 RPM High Capacity Festplatten wählen. Damit stehen in einer Exadata Database Machine Full Rack bis zu 672 TB reiner Festplattenplatz, 44,8 TB reiner Flash Speicher und bis zu 4 TB Hauptspeicher zur Verfügung.

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

Danksagung Dank an Heinz-Wilhelm Fabry und Frank Schneede (beide Oracle) für ihren fachlichen Input.