

Salzburg, 29. September 2011

**Einladung zu den öffentlichen Hearings
im Rahmen der Besetzung der Universitätsprofessur für „Informationssysteme“**

Die pünktlich beginnenden Vorträge finden an allen Terminen im **Hörsaal T01 des Fachbereichs Computerwissenschaften der PLUS, Jakob-Haringer-Str. 2, A-5020 Salzburg** statt und sind öffentlich zugänglich.

Interessentinnen und Interessenten sind herzlich willkommen.

Univ.-Prof. Dr. Marian Vajtersic
Vorsitzender der Berufungskommission

Programm

Donnerstag, 13. Oktober 2011 um 9:00 Uhr

Dr. Nikolaus Augsten
Freie Universität Bozen, Fakultät für Informatik

„Effiziente Ähnlichkeitssuche in hierarchischen Daten mit pq-Grams“

Häufig müssen Daten von verschiedenen Quellen abgeglichen werden. Dabei reicht es meist nicht, sich auf exakte Übereinstimmungen zu beschränken, sondern auch ähnliche Einträge müssen in Betracht gezogen werden. Dies ist besonders schwierig, wenn die Daten eine Baumstruktur aufweisen, da nicht nur einfache Datenwerte, sondern auch deren hierarchische Beziehungen bewertet werden müssen. In diesem Vortrag werden pq-Grams als neue Methode zur Berechnung der Ähnlichkeit zwischen Bäumen eingeführt. Die pq-Gram Distanz ist eine gute Näherung und eine untere Schranke für die sogenannte Edit-Distanz, die als Referenzmaß für die Ähnlichkeit zwischen Bäumen gilt. Die pq-Gram Distanz lässt sich effizient mithilfe eines Datenbanksystems berechnen oder als eigener Operator ins System einbauen. Die empirische Evaluierung mit synthetischen Daten und mit Daten aus realen Anwendungen bestätigt die analytischen Ergebnisse und zeigt die praktische Relevanz der Lösung.

Donnerstag, 13. Oktober 2011 um 11:00 Uhr

Prof. Dr. Mario Döllner
Universität Passau, Lehrstuhl für verteilte Informationssysteme

„Verteilte Multimedia Suche“

Der Vortrag ist thematisch in zwei Bereiche gegliedert und beleuchtet aktuelle Ergebnisse meiner Forschungsgruppe an der Universität Passau im Rahmen der verteilten Multimedia Suche. Der erste Teil des Vortrages gibt einen Überblick über meine Arbeiten in den Teilbereichen Multimedia Annotation, Multimedia Datenbanken, Multimedia Middleware und Mobile Multimedia Applikationen. In diesem Zusammenhang werden meine Beiträge in der Standardisierung, Kooperationen mit anderen Forschungseinrichtungen und die Themen der Drittmittelprojekte in das Gesamtbild eingeordnet. Der zweite Teil des Vortrages konzentriert sich auf zwei ausgewählte Forschungsergebnisse und stellt diese im Detail vor. Für jedes dieser Forschungsprojekte werden die Problemstellung, die Methodik der gewählten Ansätze und erzielte Ergebnisse präsentiert.

Donnerstag, 13. Oktober 2011 um 14:00 Uhr

Assoc. Prof. Dr. Sebastian Maneth
NICTA & University of New South Wales Sydney

„Neue Informationssysteme durch NoSQL Datenbanken“

Daten moderner Anwendungen sind oft "flüchtig", das heißt abhängig vom augenblicklichen Zustand des Webs, oder leicht veränderbar durch primitive single-user Operationen. Transaktionen und die ACID Eigenschaften sind in solchen Szenarien nur bedingt wichtig und auf den Einsatz teurer Datenbankschiffe wie Oracle oder DB2 kann daher verzichtet werden. Stattdessen können neue "NoSQL" Datenbanken, wie z.B. key-value stores (Bigtable, Hadoop) oder Dokument und Graphdatenbanken (MarkLogic Server), eingesetzt werden. Daten werden hier schemalos gespeichert und mit Hilfe einfacher Anfragen analysiert. Entsprechende Anfrageauswerter arbeiten extrem effizient, skalieren horizontal und erreichen Geschwindigkeiten unvergleichbar zu konventionellen Datenbanken. Wir behaupten dass eine Großzahl interessanter Anfragen in Web Portalen nicht deswegen nicht unterstützt wird weil die Benutzerinteraktion zu kompliziert wäre, sondern schlicht weil derartige Anfragen bisher noch nicht effizient unterstützt werden konnten. Mittels moderner NoSQL Datenbanken ändert sich diese Situation. Zukünftige Informationssysteme unterstützen komplexe Anfragen und ermöglichen dadurch völlig neue Sichten auf die bestehenden Daten. Wir diskutieren dies anhand von Beispielen. Aus technischer Sicht sind effektive Indexe auf den neuartigen Daten eine der wichtigen Errungenschaften von NoSQL Datenbanken. Wir erläutern am Beispiel von XML die Index-Erstellung von Strukturdaten mittels Wortbuch-basierter Komprimierungsverfahren.

Donnerstag, 13. Oktober 2011 um 16:00 Uhr

Jun.-Prof.in Dr.in Daniela Nicklas
Universität Oldenburg, Department für Informatik

„Datenmanagement für hochdynamische Miniwelten“

Die Aufgabe eines Informationssystems ist es, einen bestimmten, für Anwendungen relevanten Ausschnitt der Realität - die sogenannte Miniwelt - möglichst korrekt abzubilden. Werden Miniwelten betrachtet, deren Zustand sich hochfrequent ändert (wie z.B. dynamische Umgebungsmodelle für mobile Systeme oder Überwachungssysteme für komplexe Industrieanlagen), stoßen traditionelle Verarbeitungskonzepte an ihre Grenzen.

Dieser Vortrag geht auf einige Herausforderungen im Datenmanagement hochdynamischer Miniwelten ein, wie bi-temporale Datenmodelle und deren effiziente Implementierung, und zeigt Lösungsmöglichkeiten auf, die im Kontext von Logistik- und Verkehrsanwendungen entwickelt wurden. Es wird hierzu ein datenstrombasiertes Verarbeitungskonzept vorgestellt. Im Ausblick wird auf die Rolle von Historien und statischer Anteile solcher Miniwelten eingegangen und ein über Datenbank- und Datenstrommanagementsysteme vereinheitlichtes Verarbeitungskonzept diskutiert.

Freitag, 14. Oktober 2011 um 8:00 Uhr

Ao.Univ.-Prof. Dr. Werner Retschitzegger
Universität Linz, Information Systems Group

**„Surviving the Heterogeneity Jungle with Composite Mapping Operators –
From Wishful Thinking to Viable Engineering”**

Overcoming diverse kinds of heterogeneities in scenarios like information integration, schema evolution or data model translation has been identified as one of the “Achilles heels in computing” and is the only topic raised in all seven DB research assessment reports since 1988. Especially the resolution of *structural heterogeneities*, being a result of applying different modelling constructs for the same semantic concept is a challenging, recurring problem.

This talk gives insight into our approach to systematically resolve recurring structural heterogeneities at a conceptual level by means of reusable *mapping operators* expressed in terms of UML component diagrams. Based on that, *executable transformations* on basis of coloured Petri Nets are automatically derived in order to integrate, evolve or translate according schema instances. A reflection on the achievements and future efforts concludes the talk.

Freitag, 14. Oktober 2011 um 10:00 Uhr

PD Dr.-Ing. Ralf Schenkel
Universität des Saarlandes Saarbrücken, Exzellenzcluster Multimodal Computing and Interaction

„Managing Large-Scale Semantic Information“

Mehr als zehn Jahre nach der Ausrufung des "Semantic Web" spielen semantische Daten heute eine wichtige Rolle in vielen Anwendungen. Ein wesentlicher Grund für diesen Erfolg ist die breite Verfügbarkeit enorm großer Faktensammlungen wie Linked Open Data, aber auch das mit automatischen Methoden extrahierte Wissen in Ontologien wie DBpedia und YAGO.

Der Vortrag stellt wichtige Herausforderungen vor, die diese neuartige Form von Daten an Informationssysteme stellt, und zeigt einige Lösungen auf.

Dazu gehören klassische Datenbanktechniken wie die effiziente Ausführung von Anfragen in verteilten Wissensbasen, Methoden aus dem Information Retrieval, um gute Ergebnisse zu "unscharfen" Anfragen zu finden, sowie Interfaces, die Benutzern das Arbeiten mit semantischen Arbeiten erleichtern.

Freitag, 14. Oktober 2011 um 12:30 Uhr

PD Dr. Matthias Schubert
Universität München, Institut für Informatik

„Lokalisierungsanfragen in Computer-Tomographie Datenbanken“

Bildgebende Verfahren wie Computertomographien sind wichtige Werkzeuge der modernen Medizin, die für die Diagnose verschiedenster Krankheiten zunehmende Bedeutung finden. Der Zugriff auf aktuelle medizinische Bilddatenbanken ermöglicht es den Medizinern einen bestimmten CT-Scan durch Spezifizierung des Patientennamen, des Scan-Zeitpunkts oder der Scan-ID zu laden. Weitergehende Anfrageoptionen, wie zum Beispiel die Einschränkung der Anfrage auf eine bestimmte Körperregion oder die Navigation innerhalb eines Scans, müssen jedoch manuell durch den Benutzer vorgenommen werden.

Im Rahmen dieses Vortrags werden Zugriffsverfahren vorgestellt, die die automatische Navigation innerhalb eines Scans nach Bildähnlichkeit und semantischen Konzepten erlauben. Die vorgestellten Techniken basieren dabei auf einem neuen Positionierungssystem, das die Position eines Bildes innerhalb des menschlichen Körpers unabhängig von den Aufnahme- oder Patienteneigenschaften beschreibt. Zur Abbildung der Bildinformation auf eine Position kommen dabei neue Bilddeskriptoren und Methoden der Ähnlichkeitssuche zum Einsatz.

Freitag, 14. Oktober 2011 um 14:30 Uhr

Dr. Jens Teubner
ETH Zürich, Systems Group

„Datenbank-Algorithmen für kommende Rechner-Generationen“

Rasante Entwicklungen in der Hardwaretechnologie sorgen für immer stärkeren Druck auf die Softwareseite, die Fortschritte durch geeignete Algorithmen auch effizient auszunutzen. Es zeigt sich, dass dies gerade in der Datenbankwelt nur unzureichend gelingt. Nachhaltige Erfolge erfordern ein radikales Umdenken beim Design von Datenbankalgorithmen.

Im Vortrag demonstriere ich wie ein solches Umdenken aussehen kann. Dazu stelle ich zwei Arbeiten aus meinem "Avalanche"-Projekt an der ETH Zürich vor, die ein breites Anwendungsspektrum einerseits sowie verschiedene Ausprägungen moderner Rechner-Architekturen andererseits abdecken. Zunächst werde ich zeigen, wie die massive Hardware-Parallelität von FPGAs zur Lösung des "frequent item"-Problems (ein Basisproblem des Data Mining) ausgenutzt werden kann, obwohl das Problem bislang notorisch schwer zu parallelisieren galt. Im zweiten Teil werde ich einen Algorithmus zur Joinverarbeitung vorstellen, der sich auf quasi beliebig vielen CPUs eines Multi-Core-Systems parallelisieren lässt. Beide Beispiele folgen ähnlichen Designkonzepten und erreichen deutliche Performanzsteigerungen, teilweise um Größenordnungen und mehr.