

Projektseminare im Sommer 2009 ::

DIAWeb: Erstellung einer Web-Plattform für die Organisation, Verwaltung und Darstellung der Deutschen Informatik-Akademie



Übergabe der Projektdokumentation an die Deutsche Informatik-Akademie auf dem Dach der Bundeskunsthalle in Bonn (v.l.n.r.: Gabriele Trapp und Karl-Heinz Künkel von der DIA, Projektteam: Svyatoslav Hresyk, Georg Deibert, Axel Richert, Stefan Graber).

Die Deutsche Informatik-Akademie (DIA) ist seit über 20 Jahren in der Fort- und Weiterbildung im Informatik- und IT-Bereich tätig. Dabei deckt die DIA ein breites Spektrum an Themen ab, das sich von Managementaufgaben über Software-Entwicklung und eingebetteten Systemen bis hin zu juristischen Grundlagen von IT-Fragen erstreckt. Um die DIA bei der Planung und Durchführung des Seminarangebots optimal zu unterstützen, wurde das Projektseminar „DIAWeb“ im Sommersemester 2009 durchgeführt. Das neu entwickelte System bietet an IT-Weiterbildung Interessierten umfassende Informationsmöglichkeiten über die Angebote in den verschiedenen Themenfeldern. Seminare können direkt über die neue Webseite gebucht werden, wobei auch Rabatte wahrgenommen werden können. Das System unterstützt ferner eine Vielzahl von organisatorischen Abläufen und bietet spezielle Zugänge für DIA-Mitarbeiter und Dozenten sowie ein „Kundenportal“ für Seminarteilnehmer. Seit Mitte August ist die neue Web-Plattform online geschaltet und führte bereits zu positivem Feedback seitens der Dozenten und Kunden der DIA und bestätigte so den vollen Erfolg des Projektseminars.

Deutsche Informatik-Akademie: <http://www.dia-bonn.de>

Besuch aus Neuseeland

Ende August hatte die Arbeitsgruppe **Prof. Dr. Robert J. McQueen** vom ERCIS-Partner University of Waikato Management School in Hamilton, Neuseeland, zu Gast. Bob McQueen ist Professor of Electronic Commerce Technologies im Department of Management Systems in Waikato und lebt als gebürtiger Kanadier seit über 20 Jahren in Neuseeland.

Der zweiwöchige Besuch diente der Vertiefung der Beziehungen zwischen den beiden Arbeitsgruppen, aus der auch der letzte Gast aus Neuseeland, **Weiwei Li**, stammte, sowie der Auslotung zukünftiger Kooperationsmöglichkeiten. Hier ist u. a. an eine Forschungskoooperation im Rahmen der Bilateral Cooperation in Education and Research zwischen Deutschland und Neuseeland aus dem International Science and Technology (ISAT) Linkages Fund gedacht.

Skills2Web: Konzeption und Umsetzung zweier OpenSocial Gadgets für die Bewertung von Hard und Soft Skills

Persönliche Profilseiten werden in sozialen Netzwerken (SN) immer wichtiger. Diese sollen informativ und möglichst glaubwürdig sein. Daher wurden im Projektseminar „Skills2Web“ zwei Gadgets programmiert, die es den Benutzern eines SN erlauben, ihre Hard und Soft Skills strukturiert und glaubwürdig darzustellen. Das Projektseminar baute dabei auf den Ergebnissen des gleichnamigen Forschungsprojekts des ERCIS mit der Firma XING auf. Das Seminar wurde daher in Kooperation mit dem Lehrstuhl für Wirtschaftsinformatik und Informationsmanagement (Prof. Becker) und dem Psychologischen Institut IV (Prof. Holling) durchgeführt und Lehrstuhl-übergreifend betreut, auf unserer Seite durch **Till Haselmann**. Das Projektseminar fand Ende September seinen gelungenen Abschluss, die Gadgets sollen demnächst auf xing.com verfügbar sein. Siehe hierzu auch „Forschung im Fokus“.

Column Stores: Bewertung von Column Stores im Kontext eines Data Marts im Finanzdienstleistungsbereich

Die in diesem Projektseminar bearbeitete Aufgabe bestand darin, einen bisher relational implementierten Data Mart des Praxispartners zeb/ mittels Column Stores zu realisieren und seine Leistung zu untersuchen. Column Stores, die bereits Gegenstand des Newsletters Nr. 2 (Oktober 2006) waren, speichern im Unterschied zu „Row Stores“ Daten nicht zeilen-, sondern spaltenweise und eignen sich dadurch sehr gut für Data Warehouse- und vergleichbare Anwendungen. Im Projektseminar wurden die Column Stores MonetDB, Sybase IQ und Vertica hinsichtlich ihrer Stärken und Schwächen analysiert und in zwei Benchmarks sowohl untereinander als auch gegenüber der bisherigen relationalen Implementierung hinsichtlich ihrer Geschwindigkeit bei Ladeprozessen und der Auswertung von Berichten bewertet. Im Ergebnis zeigte sich, dass Column Stores bereits mit relativ wenig Aufwand ein beeindruckendes Leistungsverhalten aufweisen, das mit der relationalen, aufwändig optimierten Alternative nur in wenigen Teilbereichen erzielt werden konnte.



Aktuelles Schlagwort :: Data Center

Wenn man der Berichterstattung in der Fachpresse glauben darf, kommt seit einiger Zeit das zentralisierte Rechenzentrum im Stil der 1970er und 1980er Jahre zurück, denn man hört und liest in letzter Zeit viel von „Data Centers“. Diese weisen allerdings zwei gravierende Unterschiede zu der Art Rechenzentrum auf, die man aus den Anfängen der Informatik und der Zeit vor der Erfindung des Personal Computers kennt: Erstens ist bei heutigen Data Centers,

Kommt das klassische Rechenzentrum der 70er Jahre zurück?

Berechnungen abwickeln lässt, genau steht. Dies hat seinen Grund hauptsächlich darin, dass durch die flächendeckende Verfügbarkeit von Breitbandnetzen die Platzierung eines Data Centers dort erfolgen kann, wo insbesondere Baugrund und Stromversorgung preiswert sind. Zweitens spricht man nicht mehr von „Rechenzentrum“ („Computing Center“), sondern man spricht bewusst von einem „Data Center“, eigentlich also einem Datenzentrum. Wir erleben damit die Konsequenz dessen, was **Tim O'Reilly** vor wenigen Jahren als ein Charakteristikum des Web 2.0 bezeichnet hat: „Data is the next ‚Intel Inside‘“. Dass dies immer noch gilt, hat er in seinem jüngsten Paper¹ klargestellt: „Collective intelligence applications depend on managing, understanding, and responding to massive amounts of user-generated data in real time. The „subsystems“ of the emerging internet operating system are increasingly data subsystems: location, identity (of people, products, and places), and the skeins of meaning that tie them together.“

Ein Data Center hat de facto viel mit einem klassischen Rechenzentrum gemeinsam, denn es umfasst neben vielen Rechnern (nicht selten Cluster bestehend aus 1.800 bis 2.000 Servern) auch umfangreiche Kommunikations- und Speichersysteme, ferner ausfallsichere Stromversorgung, Klimatisierung oder auch Brandschutzeinrichtungen. In ihrer heutigen Form entstanden sind sie um die Jahrtausendwende im Rahmen der ersten Dot-Com-Blase, als das Internet mit seinen Diensten und Angeboten stark anwuchs. Firmen

Data Center in ihrer heutigen Form entstanden im Rahmen der ersten Dot-Com-Blase

wie Google, Yahoo! oder Amazon begannen mit dem Bau großer Datenzentren (allein Google betreibt inzwischen mehr als 35 weltweit), um die tägliche Masse an Suchanfragen oder Kauftransaktionen effizient bewältigen zu können. Allerdings sind die dabei entstandenen Infrastrukturen so weit angewachsen, dass sie der Allgemeinheit kostenlos oder kostenpflichtig zur Verfügung gestellt werden können (z. B. in Form von Google Apps oder Amazon Web Services), was Teil des „Cloud Computing“ ist.

Über den Entwurf und Betrieb von Data Centers ist in den letzten Jahren viel geforscht worden, wobei Aspekte wie Zuverlässigkeit, Verfügbarkeit, Ausfallsicherheit und Sicherheit höchste Priorität eingeräumt wird. Standards wie der TIA-942 der amerikanischen Telecommunications Industry Association beschäftigen sich primär mit dem Layout, der Verkabelung und dem technischen Entwurf

eines Data Centers. Redundanz ist ein zentraler Aspekt, denn nach Google-Erfahrungen sind gerade im ersten Betriebsjahr eines Data Centers Ausfälle von Servern und Festplatten häufig, und ebenso kommen Stromausfälle, Überhitzung und Verkabelungsfehler vor. Dem lässt sich in weiten Teilen durch die Bereitstellung neuartiger Datenstrukturen und Algorithmen, also letztlich durch Software begegnen. Im Fall Google gehören hierzu das Google File System (GFS), welches Daten über viele Server verteilt speichert und auf fast allen Rechnern läuft, BigTable, ein verteiltes Speichersystem zur Verwaltung strukturierter Daten, und MapReduce zur Ausführung paralleler Berechnungen über große Datenmengen auf vielen Rechnern. Dabei werden Eingabedaten auf Prozesse verteilt, die vorgegebene Map-Funktionen berechnen; auf die entstehenden Zwischenergebnisse werden Reduce-Prozesse angewendet, welche parallel die Ausgabedaten erzeugen. Beispiele für eine Anwendung dieses Paradigmas sind die verteilte Suche und Filterung von Zeichenketten aus Dateien oder die verteilte Häufigkeitsanalyse. Derartige Funktionalität ermöglicht es Google, Suchanfragen so schnell zu beantworten, wie wir das heute gewohnt sind.

Die Verfügbarkeit riesiger Rechenkapazität und gigantischer Datenspeicher gibt also Anlass zu neuen Fragestellungen. Als weiteres Beispiel seien Multi-Tenant-Datenbanken genannt, hierbei wird in einem Data Center ein Datenbanksystem betrieben, in welchem sich in einer oder mehreren Datenbank-Instanzen Daten verschiedener Anwender befinden können. Im einfachsten Fall verwenden alle Anwender das gleiche Schema; Unterschiede etwa in den verwendeten Attributen einer Tabelle werden über Erweiterungstabellen abgefangen, die formal zu einer Schemaerweiterung führen.

<http://www.datacenterknowledge.com>

<http://msdn.microsoft.com/en-us/library/aa479086.aspx>

¹ <http://www.web2summit.com/web2009/public/schedule/detail/10194>

Literatur:

S. Aulbach et al.: Multi-tenant databases for software as a service: schema-mapping techniques; Proc. ACM SIGMOD International Conference on Management of Data 2008, 1195-1206

F. Chang et al.: Bigtable: A Distributed Storage System for Structured Data; Proc.7th Symposium on Operating System Design and Implementation 2006, 205-218

J. Dean, S. Ghemawat: MapReduce: Simplified Data Processing on Large Clusters; Proc.6th Symposium on Operating System Design and Implementation 2004, 137-150

Impressum

Herausgeber:

Prof. Dr. Gottfried Vossen

Lehrstuhl für Informatik

Universität Münster

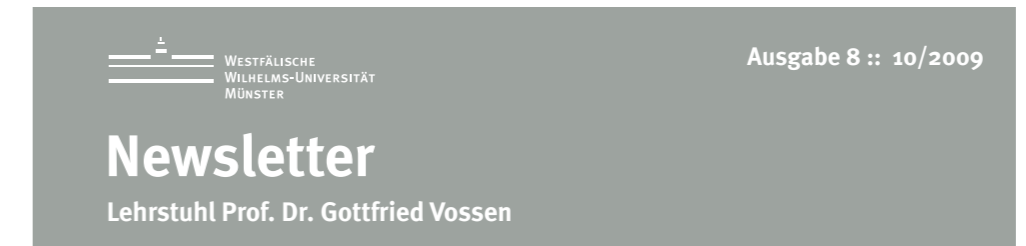
Leonardo-Campus 3 | 48149 Münster

fon +49 251 83 38150 | fax +49 251 83 38159

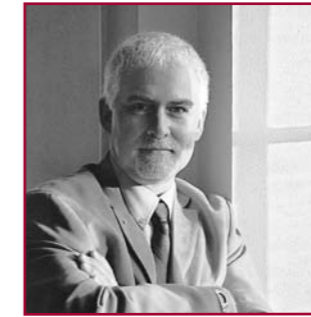
dbis-group.uni-muenster.de

Fotonachweis:

Titelseite: Gabriele Peters



dbis-group.uni-muenster.de



Liebe Leserinnen und Leser,

vor Ihnen liegt die achte Ausgabe unseres DBIS Newsletters, in welchem wir Ihnen wieder Aktuelles rund um die Arbeit meiner Gruppe zusammengestellt haben. Wir präsentieren unseren Newsletter heute in neuem Design und hoffen, dass Sie dieses so informativ und übersichtlich finden wie das vorherige. Wir berichten über Tagungsbesuche in unterschiedlicher Funktion sowie Messebeteiligungen weltweit und neue Publikationen, das ERCIS Launchpad 2009, Dissertationen und Besucher. Im abgelaufenen Semester haben wir drei Projektseminare zu ganz unterschiedlichen Themen erfolgreich beendet, wobei Praxisbezug in allen Fällen großgeschrieben wurde; vielleicht weckt das ja Ihr Interesse an einer zukünftigen Zusammenarbeit.

Ein zentrales **Thema** in dieser Ausgabe ist das **Thema Data Center**, das im Zusammenhang

ERCIS Launchpad 2009; Data Center und SkillRank; Exponatec Cologne

mit **Cloud Computing** in letzter Zeit von unterschiedlichen Seiten behandelt wird und das von seinen großen Treibern auch auf neue Forschungsfragen hin untersucht wird. Man kommt bei diesem Thema nicht an Google vorbei, denn Google ist nicht nur Betreiber zahlreicher derartiger Zentren weltweit, sondern entwickelt auch die meiste dort verwendete Software selbst. Ebenfalls auf Google geht der Begriff des **PageRank** zurück, dessen Idee wir in diesem Newsletter auf einen völlig anderen Anwendungsbereich, die Bewertung von Fertigkeiten (Skills), übertragen; damit schließt sich der Kreis zu einem unserer Projektseminare.

Unser MoVIS-System für Museen, über das wir in diesem Newsletter schon des Öfteren berichtet haben, stellen wir in diesem Jahr auf der **Exponatec Cologne** vom 17.–20. November 2009 in Halle 11.1 am Stand G 55 aus; wir freuen uns natürlich auf jeden Besucher.

Weitere und stets aktuelle Informationen erhalten Sie wie immer über unsere Webseite unter dbis-group.uni-muenster.de (bitte beachten Sie die neue URL). Ich wünsche Ihnen viel Spaß bei der Lektüre.

Münster, im Oktober 2009

Ihr
G. Vossen
Prof. Dr. Gottfried Vossen



Ankündigungen

Ideen- und Gründerwettbewerb :: ERCIS Launch Pad 2009

In diesem Jahr findet zum zweiten Mal der erfolgreiche Ideen- und Gründerwettbewerb des European Research Center for Information Systems (ERCIS) in Münster statt. Der diesjährige IT-Gründertag steht ganz unter dem Motto „Internet Business“ und möchte alle Gründungsinteressierten ansprechen, die innovative Ideen für Applikationen oder Dienstangebote haben, welche sich im Internet realisieren lassen. Der Gründertag findet am 28. Oktober 2009 ab 10 Uhr im Hörsaal Leo 18 auf dem Leonardo-Campus statt und ist öffentlich. Für alle Teilnehmer winken attraktive Preise mit einem Gesamtwert von über 9.500 €, die von unseren Sponsoren großzügig bereitgestellt werden. Alle Besucher erhalten eine 6-monatige kostenlose Premium-Mitgliedschaft von der Projektbörse projektwerk. Außerdem bietet sich hier eine hervorragende Plattform, um mit aktuellen und ehemaligen IT-Gründern in Kontakt zu kommen. Unterstützung findet das Launch Pad erneut vom Technologiehof Münster, vertreten durch Gabriele Felder, Leiterin des Bereichs „Existenzgründung und Unternehmenswachstum“. Weitere Informationen, die Regeln und Teilnahmebedingungen sowie alle wichtigen Termine finden Sie unter: <http://ercis-launchpad.de>

Exponatec Cologne

MoVIS auf der Internationalen Fachmesse für Museen, Konservierung und Kulturerbe in Köln, 17.–20.11.2009, Halle 11.1, Stand G55. www.exponatec.de

Was macht eigentlich ... Dr. Daniel Dahl?



DANIEL DAHL promovierte im Juni 2008 mit dem Thema „Integration von Lernobjekten auf Metadatenebene – Ein Web 2.0-geprägter Ansatz“ am Lehrstuhl für Informatik. Im Anschluss nahm er seine bis heute andauernde Tätigkeit im Web 2.0 Competence Center bei Bayer Business Services in Leverkusen auf. Bayer Business Services bietet dem gesamten Bayer-Konzern neben Diensten in den Bereichen Employee Services oder Training die Beratung in und Durchführung von IT-Projekten an. Im Rahmen seiner Arbeit im Web 2.0 Competence Center leitet Daniel Dahl national wie international ausgerichtete Kundenprojekte, die sich mit der Einführung und Etablierung von Web 2.0-Konzepten wie Wikis, Blogs, Mashups oder Social Networking beschäftigen. Daniel Dahl ist auch nach seiner Zeit als Mitarbeiter in Prof. Vossens Team im wissenschaftlichen Umfeld tätig; so wird eine Publikation unter dem Titel „From Theory to Practice – A Reality Check on Enterprise Mashups“ im Herbst 2009 im Rahmen einer Web 2.0-/Internet-Tagung veröffentlicht.

Neue Publikationen der DBIS-Group

Tagungsbände

:: T. Morzy, G. Vossen (eds.): *Advances in Databases and Information Systems – ADBIS 2009*; Proc. 13th East European Conference, Riga, Latvia, September 2009, Lecture Notes in Computer Science 5739, Springer-Verlag, Berlin, 2009
:: G. Vossen, D.D.E. Long, J. Xu Yu (eds.): *Web Information Systems Engineering – WISE 2009*; Proc. 10th International Conference, Poznan, Poland, October 2009, Lecture Notes in Computer Science 5802, Springer-Verlag, Berlin, 2009

Data Warehousing

:: J.-N. Mazón, J. Lechtenböcker, J. Trujillo: *A survey on summarizability issues in multi-dimensional modelling*; erscheint in *Data & Knowledge Engineering*, 2009

E-Learning

:: D. Dahl, G. Vossen: *Towards Learning Object Metadata Quality Control – The Next Steps*; Proc. EDEN (European Distance and E-Learning Network) Annual Conference 2009, Gdansk, Poland (on CD-ROM, ISBN 978-963-87914-3-6)

RFID

:: J. Schwieren, G. Vossen: *A Design and Development Methodology for Mobile RFID Applications based on the ID-Services Middleware Architecture*; Proc. 10th International Conference on Mobile Data Management (MDM) 2009, Taipei, Taiwan, 260-266
:: J. Schwieren, G. Vossen: *ID-Services: An RFID Middleware Architecture for Mobile Applications*; Information Systems Frontiers, Special Issue on Advances in RFID Technology, 2009

Aktuelle Diplomarbeiten

MICHAEL BUNGE hat in seiner Diplomarbeit „Anreizsysteme im Wissensmanagement am Beispiel der Lernplattform Learnr“ untersucht, welche Arten von Anreizsystemen für eine Steigerung der Beteiligung im Rahmen der Lernplattform Learnr geeignet sind. Ausgehend von psychologischen Grundlagen der Motivation sowie der Analyse existierender Anreizsysteme in virtuellen Communities entwickelt der Autor ein Konzept zur Realisierung eines Anreizsystems für Learnr, das Bewertungsmechanismen beinhaltet, den sozialen Vergleich ermöglicht und persönliche Ansprachen integriert.

Medizininformatik

:: M. Dugas, B. Breil, V. Thiemann, J. Lechtenböcker, G. Vossen: *Single Source Information Systems to Connect Patient Care and Clinical Research*; erscheint in Proc. 22nd International Conference of the European Federation for Medical Informatics: Medical Informatics in a United and Healthy Europe (MIE) 2009, Sarajevo, Bosnia and Herzegovina
:: M. Dugas, B. Breil, V. Thiemann, J. Lechtenböcker, G. Vossen: *Single Source Informationssysteme – Nutzung von Routinedaten für die klinische Forschung*; erscheint in Proc. 54. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS) 2009, Essen, Germany

Web 2.0

:: G. Vossen: *Web 2.0: A Buzzword, a Serious Development, Just Fun, or What?* Proc. International Conference on e-Business (ICE-B) 2009, Milan, Italy, IS-33-IS-40

Service-Orientierung

:: G. Thies, G. Vossen: *Governance in Web-Oriented Architectures*; wird erscheinen in Proc. IEEE Asia-Pacific Services Computing Conference, Dez. 2009, Singapur

Springer Encyclopedia of Database Systems 2009 Entries

:: G. Vossen: *ACID Properties*
:: G. Vossen: *Concurrency Control – Traditional Approaches*
:: G. Vossen: *Transaction*
:: G. Vossen: *Transaction Management*
:: G. Vossen: *Transaction Models – the Read/Write Approach*
:: J. Lechtenböcker: *2-PC Protocol* (www.springer.com; ISBN 978-0-387-49616-0)

Dem Thema „Aggregation korrelierender Maße im Data-Ware-house-Umfeld“ hat sich **FLORIAN BLENKLE** gewidmet. Am Beispiel der Kennzahl „Value at Risk“ (VaR) zeigt er, dass zur adäquaten Behandlung von Korrelationen, die bspw. zwischen unterschiedlichen Portfolios einer Bank auftreten, umfangreiche Berechnungen erforderlich sind, die bei Datenbeständen realistischer Größe nicht innerhalb vertretbarer Antwortzeiten durchgeführt werden können. Daher implementiert, untersucht und bewertet er verschiedene Materialisierungsstrategien zur Vorberechnung verdichteter Informationen und stellt eine im vorliegenden Anwendungskontext geeignete Lösung vor.

Forschung im Fokus :: SkillRank

Im Internet finden sich immer mehr soziale Netzwerke und generell Plattformen, auf denen Benutzer ihre Persönlichkeit in Form einer Profilseite präsentieren können. Im Rahmen seriöser Netzwerke zum Pflegen geschäftlicher Kontakte, wie z. B. LinkedIn oder XING, ist es für die Benutzer häufig wichtig, ihre Profile glaubwürdig darzustellen. Diesem bisher ungelösten Problem nähern wir uns mit unserer aktuellen Forschung am sogenannten SkillRank. Das momentan erforschte Verfahren nennen wir SkillRank, weil es schlussendlich die Personen nach ihrer tatsächlichen Fachkompetenz sortieren soll. Hier haben wir jedoch ausschließlich den Themenkomplex der Glaubwürdigkeitsbeurteilung herausgegriffen.

Glaubwürdigkeit als Katalysator für Geschäftskontakte

Bewertungen und Zeugnisse sind im geschäftlichen Umfeld äußerst wichtig, wenn es darum geht, neue Kontakte zu knüpfen. Zwischen unbekanntenen Personen schaffen diese Unterlagen eine Vertrauensbasis für einen Erstkontakt. Insbesondere Unternehmen, die auf der Suche nach neuen Mitarbeitern oder Freiberuflern sind, müssen regelmäßig Fähigkeiten auf Basis von Bewerbungsschreiben und Ähnlichem abschätzen.

In der heutigen Zeit verlagert sich der Aufbau neuer Geschäftsbeziehungen zunehmend ins Internet, da Plattformen wie LinkedIn und XING – sozusagen die ernst gemeinten Varianten des berühmt-berüchtigten StudiVZ – eine ideale Basis schaffen, um mit geeigneten Personen auf der ganzen Welt in Kontakt zu treten. Diese Plattformen erlauben es den Benutzern, detaillierte Profilseiten anzulegen, auf denen sie ihre Fachkompetenzen und Berufserfahrung präsentieren können. Bisher sind diese Profilseiten jedoch nur semi-strukturiert und enthalten notgedrungen viele Freitextfelder. Das liegt daran, dass die manuelle Verwaltung der zahlreichen möglicherweise relevanten Fachkompetenzen i. d. R. bei Weitem die Kapazitäten der Betreiberunternehmen übersteigt.

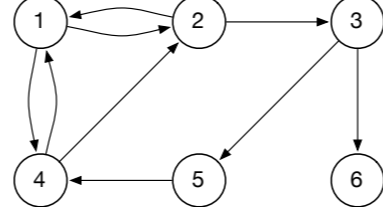
In Bezug auf die Fachkompetenz kann das Blaue vom Himmel versprochen werden

Leider führt dies auch dazu, dass grundsätzlich in Bezug auf die Fachkompetenz das Blaue vom Himmel versprochen werden kann. Insofern ist es für Suchende wie auch Anbieter von großer Bedeutung, eine „objektive“ Glaubwürdigkeit ausweisen zu können, die das Profil zumindest grundsätzlich untermauert.

Googles PageRank als Grundlage

Ein **ähnliches** Problem wurde bereits 1995 von Bomze und Gutjahr untersucht und in **ähnlicher** Form von Google als PageRank aufgegriffen und populär gemacht. Das Problem, das der PageRank löst, ist die Bewertung, wie wichtig eine Webseite ist. Bis dato wurden Suchanfragen im Internet auf Basis von inhaltlichen Kriterien ausgewertet und nach Relevanz sortiert. Der PageRank machte sich erstmals die Link-Struktur des World Wide Web zunutze. Dazu bedienten sich die Erfinder einer rekursiven Definition: Eine Webseite ist wichtig, wenn sie von vielen wichtigen Webseiten verlinkt wird. Obwohl diese Definition auf den ersten Blick nicht besonders hilfreich erscheint, lässt sie sich erstaun-

lich einfach operationalisieren. Dazu muss man sich das Web als gerichteten Graphen vorstellen, bei dem Webseiten als Knoten und Links zwischen den Webseiten als Kanten aufgefasst werden. So entstehen Graphen wie der sehr einfache Graph in der nebenstehenden Abbildung.



Beispielgraph mit sechs Webseiten/Personen und ihren Links/Bestätigungen

Links. Die Wahrscheinlichkeit, dass ein Besucher einem dieser Links folgt, ist jeweils 0,5.

Mathematisch versierte Leser werden schnell erkennen, dass es sich bei der beschriebenen Situation im Prinzip um eine **Markov-Kette** handelt. Berechnet man nun den Steady State der Kette, so erhält man die relative Verteilung der Zeit, die ein Surfer auf den einzelnen Seiten verbringt. Je höher dieser Anteil ist, desto wichtiger muss diese sein. Demnach ist die Steady-State-Verteilung der Markov-Kette eine Angabe zur Wichtigkeit der einzelnen Seiten. Die Berechnung einer solchen Verteilung ist ausreichend erforscht. In unserem Beispiel ergibt sich die Rangfolge 1, 2 > 4 > 3 > 5, 6, was auch der Intuition entspricht, dass die „eng verflochtenen“ Seiten 1, 2 und 4 am wichtigsten sind.

Weiter gedacht: SkillRank

Wir haben diesen Gedanken aufgegriffen und weitergesponnen. Statt Aussagen über die Wichtigkeit einer Webseite zu treffen, möchten wir die Glaubwürdigkeit einer Person beurteilen. Dazu haben wir die PageRank-These umformuliert: Eine Person ist glaubwürdig, wenn sie von vielen glaubwürdigen Personen als glaubwürdig bestätigt wird. Das Problem ist damit mathematisch gesehen sehr ähnlich. Wir können die Personen als Knoten und die Bestätigungen zwischen ihnen als Kanten auffassen und erhalten somit einen struktur-identischen Graphen zum PageRank. Auch hier zeigt die Intuition, dass die drei Personen 1, 2 und 4 offensichtlich die am besten bestätigten sind und daher die höchste Glaubwürdigkeit in der Gruppe erhalten sollten. Einige der Annahmen, die zur effizienten Berechnung des PageRank getroffen werden, sind jedoch für die Berechnung des SkillRank nicht mehr gerechtfertigt, was wir demnächst in einem Konferenzbeitrag näher beschreiben werden.

Quellennachweise:

:: Bomze, I. and Gutjahr, W. (1995). Estimating qualifications in a self-evaluating group. *Quality and Quantity*, 29(3):241-250
:: Langville, A. N. and Meyer, C. D. (2006). *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, Princeton, NJ, USA
:: Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). *The PageRank Citation Ranking: Bringing order to the Web*. Technical Report 1999-66, Stanford InfoLab

Aktuelle Dissertation ::

A Framework for the Consistent Usage of Tag-based Mashups



In seiner Dissertation beschreibt **STEPHAN HAGEMANN** ein System zur nahtlosen Integration von Mashups in Webseiten während des Web-Surfens. Aus der Konzeption und seinem Prototypen leitet er weitreichende Verwendungsmöglichkeiten ab und entwickelt Vorschläge für den Transfer in andere Kontexte.

Die Zahl der Web-Plattformen, mit denen Nutzer Inhalte er- und einstellen können, steigt rasant. Dieses Ökosystem im Web lebt von den Beziehungen zwischen den Nutzern, den Fähigkeiten der Suchmaschinen und den Möglichkeiten der Nutzer, Inhalte zu strukturieren. Eine Möglichkeit, Inhalte zu strukturieren, ist Tagging: freie Annotation mittels Schlagwörtern. Sogenannte Tags stellen persönliche Metadaten mit sozialen Auswirkungen dar, da sie typischerweise individuell erstellt, aber global genutzt werden können. Ihre Unkontrolliertheit macht den Inhalt von Tags schwierig systematisch nutzbar. Es zeigen sich jedoch Konventionen, die innerhalb von bestimmten Gruppen oder Diensten genau definieren, welche Bedeutung Tags haben

Konferenzberichte ::

ADBIS :: Advances in Databases and Information Systems 2009



Anfang September 2009 hat Prof. Vossen an der 13th East-European Conference on Advances in Databases and Information Systems (ADBIS) teilgenommen, die von der Riga Technical University in Riga, Lettland, veranstaltet wurde. Diese Konferenz hat es sich zur Aufgabe gemacht, Forscher vor allem aus Mittel- und Osteuropa mit Kollegen aus aller Welt zusammenzubringen. Prof. Vossen hat in diesem Jahr zusammen mit **PROF. MORZY** von der Technischen Universität Poznan in Polen das Programmkomitee geleitet und die vorgestellten Arbeiten ausgewählt. Themen waren zwar überwiegend „klassische“ Datenbankthemen wie Query Languages, Query Processing und Optimization, Indexing and Caching, aber bereits hier wurden viele neue Aspekte diskutiert, ebenso in Bereichen wie Data Integration oder Business Processes. Die Tagung wurde von einem Doctoral Consortium begleitet, das Doktoranden die Möglichkeit gab, vor einem internationalen Publikum

sollen. Ein paralleler Trend ist die Verfügbarmachung von Dienstfunktionalität über Web-Programmierschnittstellen. Eine Form, diese zu nutzen, ist die Schaffung von Mashups, also von Web-Anwendungen, die Daten und Funktionen einer oder mehrerer Quellen in einer Oberfläche integrieren. In der Kombination können diese Trends, Tagging und Mashups, eine einfache Möglichkeit darstellen, Funktionalität auf der Basis von Tags zur Verfügung zu stellen. Dazu wird die Flexibilität der Tags genutzt, um in ihnen notwendige Daten zur Erstellung von Mashups zu kodieren. Der von Stephan Hagemann entwickelte Prototyp setzt diese Idee von Tag-basierten Mashups so um, dass die Mashups während des Surfens Client-seitig direkt in Webseiten „injiziert“ werden. So können diese auf Basis der Inhalte der originalen Webseite deren Inhalt um relevante Funktionalitäten und Informationen erweitern. Stephan Hagemann beschreibt Nutzungsszenarien, die die Idee von Tag-basierten Mashups nutzen, um weitgehende Neuerungen in den Anwendungsmöglichkeiten von Webseiten zu ermöglichen. Als theoretische Erweiterung schlägt er die Nutzung des gleichen Ansatzes in anderen Kontexten vor und diskutiert Mindestanforderungen für die Umsetzung solcher Systeme. Schließlich entwickelt er eine Transformation seines Prototypen in den relationalen Kontext, deren elegante Umsetzung auf mögliche Erweiterungen der Idee von Tag-basierten Mashups hinweist. <http://activetags.de/>

ihre Arbeiten vorzustellen. Am letzten Tag der Konferenz hat Prof. Vossen eine Podiumsdiskussion zum Thema *Academic and Industrial Interactions: Models and Experiences* geleitet (siehe Bild). Während die osteuropäischen Panelisten sich ausgesprochen frustriert über ihre Erfahrungen mit Industriekooperationen äußerten, konnte Prof. Vossen das Erfolgsmodell ERCIS erläutern.

MDM :: Mobile Data Management 2009

Prof. Vossen hat im Mai 2009 an der International Conference on Mobile Data Management 2009 in Taipei, Taiwan, teilgenommen. Dieser Konferenzbesuch war für die Arbeitsgruppe doppelt von Bedeutung: Einerseits wurde die zusammen mit **JOACHIM SCHWIERN** verfasste Arbeit „A Design and Development Methodology for Mobile RFID Applications based on the ID-Services Middleware Architecture“ präsentiert, in welcher gezeigt wird, wie sich auf der Basis der im Rahmen des MoVIS-Projektes entwickelten ID-Services Middleware konkrete RFID-



Applikationen entwickeln lassen. Andererseits wurde eine derartige Applikation, das *Sentry Patrol Control System (SPCS)*, im Rahmen einer an die Konferenz angegliederten Ausstellung einem interessierten Fachpublikum demonstriert. Das Bild zeigt den SPCS-Stand mit seinem exemplarischen RFID-basierten Wächterkontrollrundgang.