
**Dreijahresbericht
des
Fachgebiets Wissensverarbeitung
der Universität Kassel**

2006 –2008

Prof. Dr. Gerd Stumme



ENDOWED CHAIR OF THE HERTIE FOUNDATION

Knowledge and Data Engineering

ELECTRICAL ENGINEERING & COMPUTER SCIENCE, UNIVERSITY OF KASSEL

Vorwort

Liebe Leserinnen und Leser,

im April feiert das Fachgebiet Wissensverarbeitung sein fünfjähriges Bestehen. In der Aufbauphase des Fachgebiets haben wir sehr früh und als eines der weltweit ersten Institute auf das aufkommende Web 2.0 gesetzt, wodurch wir bereits sehr früh wissenschaftliche Ergebnisse in diesem Bereich erzielen konnten, die heute regelmäßig zitiert werden. Aus der



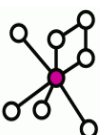
Arbeit an dieser Thematik entstand unser Flaggschiff, das kooperative Literatur-Verschlagwortungssystem „BibSonomy“. Dieses im Internet für jedermann frei benutzbare Web2.0-System hat über 5000 aktive Nutzer, und wird von mehreren deutschen Hochschul-Bibliotheken in ihren Online-Katalogsystemen als Book-marking-System eingesetzt.

Dieser Bericht fasst die Entwicklung des Fachgebiets in den letzten drei Jahren zusammen. In dieser Zeit standen zwei wichtige organisatorische Änderungen an. Zusammen mit drei anderen Informatik-Fachgebieten wechselten wir an den Fachbereich Elektrotechnik/Informatik, so dass nun die gesamte Informatik organisatorisch unter einem Dach ist; und Anfang 2008 wurde das Fachgebiet Wissensverarbeitung in das Forschungszentrum für Informationstechnik-Gestaltung (ITeG) der Universität Kassel aufgenommen. Am ITeG werden IT-Systeme unter Hinblick auf ihre gesellschaftlichen Auswirkungen in einem interdisziplinären Umfeld fortentwickelt.

Nach fünf Jahren geht die Finanzierung der Stiftungsprofessur durch die Gemeinnützige Hertiestiftung zu Ende; die Folgefinanzierung wird von der Universität Kassel übernommen. Ich danke der Hertiestiftung noch einmal herzlich dafür, dass sich durch ihre Unterstützung das Fachgebiet Wissensverarbeitung an der Universität Kassel etablieren konnte.

Ich wünsche Ihnen viel Spaß bei der Lektüre!

Prof. Dr. Gerd Stumme



Highlights

FG Wissensverarbeitung wurde in das Forschungszentrum für Informationstechnik-Gestaltung der Universität Kassel aufgenommen.

Im ITeG kooperieren Informatiker, Wirtschaftsinformatiker, Maschinenbauer und Juristen bei der gesellschaftskonformen Entwicklung von Informationssystemen.

..... Seite 2

BibSonomy erreicht hohe Nutzerzahl.

Unser kooperative Verschlagwortungssystem BibSonomy hat über 5000 aktive Nutzer und wird in mehreren nationalen und Europäischen Projekten eingesetzt.

..... Seite 4

Tagora erforscht das Entstehen von Semantik in Web2.0-Anwendungen

Physiker und Informatiker kooperieren bei der Analyse semantischer und sozialer Strukturen in kooperativen Bookmarking-Systemen. Seite 5

Nepomuk erfolgreich beendet.

Das EU-Projekt "Networked Environment for Personalized, Ontology-based Management of Unified Knowledge" hat Erweiterungen des Computer-Desktops um semantische Fähigkeiten, um die Zusammenarbeit und den Austausch von Informationen innerhalb und zwischen Arbeitsgruppen zu verbessern. Es hat eine hohe Sichtbarkeit in der wissenschaftlichen Community erreicht. Seite 5

KDubiq vergleicht soziale und Sensor-Netze

In der European Coordination Action "Knowledge Discovery in Ubiquitous Environments" koordinieren wir die Arbeitsgruppe „Ubiquitous Data – Data Collection and Interaction“. Seite 7

Dagstuhl-Seminar zu „Social Online Communities“ durchgeführt.

31 international renommierte Wissenschaftler trafen sich in Schloss Dagstuhl, um die Entwicklung der nächsten Generation von sozialer Software voranzutreiben.

..... Seite 13

Inhaltsverzeichnis

1.	Wir über uns.....	1
1.1.	Forschung im Fachgebiet Wissensverarbeitung.....	1
1.2.	Mitarbeiter	1
1.3.	Wechsel in den Fachbereich Elektrotechnik/Informatik.....	2
1.4.	Aufnahme in das Zentrum für Informationstechnik-Gestaltung.....	2
1.5.	Mitgliedschaft im Forschungszentrum L3S.....	3
1.6.	Rechner und Infrastruktur	3
2.	BibSonomy – ein Web2.0-basiertes kooperatives Lesezeichen- und Literatur-Verschlagwortungssystem	4
3.	Projekte.....	5
3.1.	TAGora – Emergent Semiotics in Social Online Communities.....	5
3.2.	NEPOMUK – The Social Semantic Desktop	5
3.3.	Microsoft Grant „Social Search“	6
3.4.	COMO - Concepts and Models.....	6
3.5.	KDUBiq - Knowledge Discovery in Ubiquitous Environments	7
3.6.	Prolearn – Professional E-Learning.....	7
3.7.	Industrieprojekt mit K+S IT-Services GmbH.....	7
3.8.	Publikations-Reporting bei SAP Research.....	7
3.9.	Publikationsmanagement bei der Fraunhofer-Gesellschaft	8
3.10.	Zukünftige Projekte.....	8
4.	Veranstaltungen.....	8
4.1.	Semantic Network Analysis, Workshop at ESWC 2006.....	9
4.2.	Web Mining Workshop at ECML/PKDD 2006.....	9
4.3.	Collaborative Knowledge Management Workshop (CoKM2007) at the 4th Knowledge Management Conference	10
4.4.	Social and Collaborative Construction of Structured Knowledge, Workshop at WWW 2007	10
4.5.	Bridging the Gap between Semantic Web and Web 2.0, Workshop at ESWC 2007.....	10
4.6.	Web Mining 2.0, Workshop at ECML/PKDD 2007.....	11
4.7.	Data Mining in Web 2.0 Environments, Workshop at IEEE International Conference on Data Mining 2007	11
4.8.	Emergent Semantics and Ontology Evolution, Workshop at ISWC 2007.....	11
4.9.	Third Conceptual Structures Tool Interoperability Workshop, Workshop at ICCS 2007.....	12
4.10.	KASW'08: Workshop on Knowledge Acquisition from the Social Web, Workshop at Triple-I 2008.....	12
4.11.	ECML/PKDD Discovery Challenge 2008	12
4.12.	Wikis, Blogs, Bookmarking Tools – Mining the Web 2.0, Workshop at ECML/PKDD 2008.....	12
4.13.	Dagstuhl Seminar "Social Web Communities"	13

4.14.	Workshop "Knowledge Discovery, Data Mining, Maschinelles Lernen 2008" der GI-Fachgruppe KDML	13
5.	Lehre	13
5.1.	Lehrveranstaltungen	14
5.2.	Projektarbeiten	15
5.3.	Abschlussarbeiten.....	16
5.4.	Konzeption und Leitung von außeruniversitären Kursen und Schulungen.....	17
5.5.	Nachwuchs-Werbung	17
6.	Forschungssemester von Prof. Gerd Stumme	17
7.	Weitere Aktivitäten	18
7.1.	Kasseler Informatik Kolloquium (KIK)	18
7.2.	Jour Fixe und Clustersprecher Informatik	19
7.3.	Akademische Selbstverwaltung	19
7.4.	Gesellschaft für Informatik	20
7.5.	Mitgliedschaft in wissenschaftlichen Beiräten	20
7.6.	Mitgliedschaft in Programmkomitees internationaler Konferenzen und Workshops	20
7.7.	Begutachtung für Forschungs-Förderinstitutionen.....	23
7.8.	Gutachtertätigkeiten für internationale Fachzeitschriften	24
8.	Publikationen und Vorträge	25
8.1.	Publikationen	25
8.2.	Vorträge	30
9.	Presseschau.....	33

1. Wir über uns

1.1. Forschung im Fachgebiet Wissensverarbeitung

Im Mittelpunkt der Forschungsaufgaben der Stiftungsprofessur für Wissensverarbeitung (Knowledge & Data Engineering) steht die Entwicklung von Methoden und Techniken, die die Verwaltung großer, teils weit verstreuter Datenmengen erleichtern. So liefert die Wissensverarbeitung das Know-how für Kooperation und Informationssuche im Internet. Hier arbeiten wir insbesondere an den Schnittstellen der Forschungsgebiete Web 2.0, Semantic Web, Knowledge Discovery und Information Retrieval.

In der Aufbauphase des Fachgebiets haben wir unseren Fokus sehr früh auf das aufkommende Web 2.0 gerichtet – zu einem Zeitpunkt, als noch nicht klar war, wie sich dieser von O'Reilly ins Gespräch gebrachte Begriff entwickeln würde. Mit unserem kooperativen Lesezeichen- und Publikations-Verschlagwortungssystem BibSonomy haben wir eine Referenzplattform entwickelt, die zu den Top 3 der Web2.0-basierten Literatur-Verschlagwortungssysteme gehört. Die Entwicklung und der Betrieb von BibSonomy bedeutet auf der einen Seite eine hohe Arbeitsauslastung mit forschungsfremden Aufgaben. Auf der anderen Seite bietet uns das System eine gute Plattform, um die bei uns entwickelten Algorithmen zu testen, und ermöglicht die Akquisition weiterer Forschungsprojekte.

Das Fachgebiet engagiert sich in nationalen und internationalen Forschungsprojekten. Die Mitarbeiter des Fachgebiets organisieren regelmäßig internationale Tagungen und Workshops rund um die Wissensverarbeitung.

1.2. Mitarbeiter

Zum Ende des Jahres 2008 waren acht Mitarbeiter im wissenschaftlichen, technischen und organisatorischen Bereich des Fachgebiets Wissensverarbeitung tätig. Zurzeit bauen wir das Fachgebiet durch Drittmittelprojekte und industriennahe Forschungsprojekte personell aus. Im wissenschaftlichen Bereich waren im Berichtszeitraum beschäftigt:

- Prof. Dr. Gerd Stumme (seit April 2004)
- Dr. Andreas Hotho (seit April 2004)
- Dr. Christoph Schmitz (UniK/L3S, April 2004 bis Oktober 2007 nach erfolgreicher Promotion)
- Dipl.-Math. Robert Jäschke (UniK/L3S, seit Juni 2005)
- Dipl.-Inform. Beate Krause (UniK/L3S, seit Oktober 2006)
- Dipl.-Inform. Miranda Grahl (Juli 2006 bis März 2008)
- Dipl.-Inform. Dominik Benz (seit Mai 2007)
- Dipl.-Inform. Folke Eisterlehner (seit November 2008)

Die Mitarbeiter werden tatkräftig durch Monika Vopicka im Sekretariat, den technischen Angestellten Sven Stefani sowie mehrere studentische Hilfskräfte unterstützt.



1.3. Wechsel in den Fachbereich Elektrotechnik/Informatik

Mit Beschluss des Präsidiums vom 21. Mai 2007 wechselten die vier am Fachbereich Mathematik/Informatik angesiedelten Informatik-Professuren an den Fachbereich Elektrotechnik/Informatik, so dass nun die gesamte Informatik organisatorisch unter einem Dach ist. Für Sponsoren und Stifter, für Drittmittelwerbungen und Rankings wird damit eine einheitliche Fachbereichszuordnung des Fachgebiets und eine klare Benennung der Fachbereiche sichergestellt, und die enge Abstimmung der Informatik-Fachgebiete in Forschung und Lehre wird erleichtert.

1.4. Aufnahme in das Zentrum für Informationstechnik-Gestaltung

Das Fachgebiet Wissensverarbeitung wurde zum 1. Januar 2008 in das Zentrum für Informationstechnik-Gestaltung (ITeG) aufgenommen. Am ITeG werden informationstechnische Anwendungen unter starker Berücksichtigung ihrer gesellschaftlichen Auswirkungen in interdisziplinären Teams entwickelt. Das ITeG dient dazu, die Kompetenzen der Universität Kassel in diesem Bereich zu bündeln.

1.5. Mitgliedschaft im Forschungszentrum L3S

Das Fachgebiet Wissensverarbeitung ist seit seiner Gründung Mitglied im Forschungszentrum L3S. Das L3S ist ein Forschungszentrum für grundlagen- und anwendungsorientierte Forschung neuer, zukunftsweisender Methoden und Technologien in den Schlüsselbereichen Wissen, Information und Lernen. Die aktuellen Schwerpunkte sind digitale Ressourcen und alle mit ihnen verbundenen technologischen Voraussetzungen wie Semantic Web und Digital Libraries, Verteilte Systeme und Netze sowie Grid Computing und auch ihre Nutzung in innovativen Lernumgebungen.

1.6. Rechner und Infrastruktur

Im Fachgebiet steht eine dem aktuellen Stand der Technik entsprechende kommunikationstechnische Infrastruktur zur Verfügung. Der Serverpark des Fachgebiets bietet die Möglichkeit, Verfahren der Wissensentdeckung und Wissensstrukturierung auf großen, realitätsnahen Datensätzen zu simulieren. Die Mitarbeiter sind je nach Bedarf mit Laptops, Workstations oder Barebones ausgestattet.

Neun Linux-Server und ein externes RAID-System mit etwa 8 TB Plattenplatz werden für Standarddienste wie Mail, Web-Server, Content-Management, Wikis und Datenbanken gemeinsam mit dem Fachgebiet Software Engineering betrieben. Die Server werden darüber hinaus für Experimente mit neuen Algorithmen auf großen Datensätzen eingesetzt.

Der Serverpark wird ergänzt durch eine Softwarewerkstatt, die derzeit aus neun Rechnern besteht, die jeweils mit genügend Speicher und schneller Netzwerkanbindung ausgestattet sind, um als Diskless Clients für Forschungsprojekte dienen zu können. Die Rechner können für umfangreiche Berechnungen als Rechen-Cluster benutzt werden. Sie bekommen ihr Betriebssystem über das Netzwerk von einem zentralen Server. Dieses Verfahren wird ebenfalls von den Fachgebieten Software Engineering und Datenbanken eingesetzt, wodurch der Konfigurationsaufwand erheblich verringert wurde, da nur noch ein Masterimage aktualisiert werden muss. Die Benutzerverwaltung für die Fachgebiete Wissensverarbeitung und Software Engineering ist integriert, d.h. die Studierenden und Mitarbeiter können mit einheitlichen Zugangsdaten an jedem Rechner der beiden Softwarewerkstätten sowie an allen dafür vorgesehenen Servern arbeiten.

Die Rechner sind per Gigabit-Ethernet untereinander und an das Hochschulrechenzentrum angebunden. An einer zentralen Firewall laufen unterschiedliche Netzwerke für Mitarbeiter, Außenanbindung und Softwarewerkstatt zusammen. Neben der üblichen Infrastruktur wie Drucker, Scanner etc. steht außerdem ein H.323-Videokonferenzsystem mit zwei Großbildschirmen zur Verfügung.

Das Fachgebiet Wissensverarbeitung betreut den Webauftritt <http://www.cs.uni-kassel.de/> der Kasseler Informatik.

2. BibSonomy – ein Web2.0-basiertes kooperatives Lesezeichen- und Literatur-Verschlagwortungssystem

Jeder Leser kennt die lähmende Frage: Wo habe ich das gelesen, wie hieß gleich der Autor und wo ist die Quelle nun wieder zu finden? Was für den Hausgebrauch eine temporäre Irritation darstellt, die nach gründlicher Durchsicht des Bücherregals in der Regel beseitigt wird, ist für wissenschaftliche Arbeit „lebenswichtig“. Das Auffinden und Nutzen von einschlägiger Literatur ist dabei nicht nur für den einzeln arbeitenden Wissenschaftler von hoher Bedeutung, sondern hat in der Zusammenarbeit von Forschergruppen einen noch höheren Stellenwert.

The screenshot shows the BibSonomy website interface. At the top, there is a search bar with a dropdown menu set to 'search: all' and a text input field for a full-text search. To the right of the search bar are links for 'login', 'register', 'help', 'blog', and 'about', along with a German flag icon. Below the search bar, there are input fields for 'username:' and 'password:'. The main content area is divided into two columns. The left column contains a list of bookmarks and publications, each with a title, a brief description, and a timestamp. The right column features a 'filter:' section with a dropdown menu and a list of 'busy tags' such as 'analysis', 'data', 'information', and 'learning'. The interface is clean and organized, with a blue header and a white main content area.

Literaturrecherche, die koordiniert und arbeitsteilig vonstatten geht, bringt einen enormen Zeitvorteil. Das Fachgebiet Wissensverarbeitung hat dafür mit Bibsonomy ein webgestütztes System entwickelt, das es erlaubt, Literatur und Web-Lesezeichen zu verwalten. Was immer vom Leser für aufhebens- und wieder leenswert angesehen wird, gibt dieser selbst in das System ein und macht es so sich selbst – und anderen (wieder) zugänglich, sei es nur für den Nutzer selbst, für eine geschützte Gruppe oder für die ganze Welt. Bibsonomy lässt frei wählbare Schlagwörter – so genannte „Tags“ – zu, weil jeder Forscher typischerweise sein eigenes Verwaltungs- und Ablageschema hat.

Das Problem solcher selbst verwalteter Systeme liegt in der Effizienz der Suche. Bibliotheken greifen deswegen bei der Verschlagwortung auf standardisierte Schlagwortsysteme und geschultes Fachpersonal zurück. Die Verschlagwortung mit Tags durch bibliothekarische Laien und in hoch spezialisierten Fachgebieten, die sich den üblichen Standards entziehen, schafft Probleme wie bspw. Synonyme und doppelt vorhandene Literatureinträge. Daher entwickelt das Fachgebiet Wissensverarbeitung Algorithmen zur Suche und Strukturierung von Wissen in solchen Web2.0-Systemen.

Die über 5000 aktiven Nutzer von BibSonomy profitieren von den vielfältigen Export-Formaten der Literaturlisten, wie etwa BibTeX, Endnote, RTF und HTML, um die Pflege ihrer eigenen Publikationen effizienter zu gestalten. Mit BibSonomy lassen sich die Publikationslisten für den eigenen Lebenslauf sowie die diversen Web-Präsentationen und Jahresberichte zentral verwalten, eine Mehrfacheingabe der Daten entfällt. Bibsonomy ist zu erreichen unter <http://www.bibsonomy.org>, die Benutzung ist kostenlos. Interessierten Instituten wird die Einrichtung einer Gruppe zur institutsinternen Literaturverwaltung angeboten.

3. Projekte

3.1. TAGora – Emergent Semiotics in Social Online Communities



EU STREP, 06/2006 – 05/2009

Im Mittelpunkt dieses EU-Projektes steht die Untersuchung von Web2.0-Applikationen, welche es den Benutzern ermöglichen, benutzerspezifische Inhalte selber zu gestalten. Durch die Abbildung sozialer Strukturen im Internet entstehen neuartige Benutzerstrukturen, welche sich von bisher untersuchten Datenmodellen unterscheiden. Aspekte sind hierbei die Entstehung von semiotischen Beziehungen und deren Entwicklung über die Zeit. Für die Erforschung und Entwicklung von Ansätzen und Lösungen von Modellen und Analysemethoden erfolgt die Zusammenarbeit im interdisziplinären Kontext. Forschungspartner hierbei sind: University of Roma (La Sapienza), Sony CSL, University of Koblenz-Landau und University of Southampton.

Weitere Informationen:

<http://www.tagora-project.eu/>

3.2. NEPOMUK – The Social Semantic Desktop



EU Integrated Project, 6. RP, 01/2006 – 12/2008

Als Mitglied des Forschungszentrums L3S Hannover nehmen wir am EU-Projekt NEPOMUK teil. Ziel des Projektes ist die Erweiterung des Computer-Desktops um semantische Fähigkeiten, um die Zusammenarbeit und den Austausch von Informationen innerhalb und zwischen Arbeitsgruppen

zu verbessern. Der "Social Semantic Desktop" vereint Fähigkeiten des Semantic Web mit denen der Social Network Analysis (SNA).

In dem Projekt hat sich das Fachgebiet Wissensverarbeitung insbesondere mit der Entdeckung und Strukturierung von Communities befasst und dabei Verfahren der sozialen Netzwerkanalyse untersucht und weiterentwickelt. So wurden Beziehungen zwischen Benutzern, Ressourcen und Metadaten zur Erkennung von Benutzern mit ähnlichen Interessen genutzt, um den Austausch von Wissen auch zwischen einander unbekanntem Nutzern zu fördern. Die während des Projekts entwickelte Implementierung ist frei verfügbar und erreichte eine hohe Sichtbarkeit, u.a. auf heise online und im Technology Review (siehe Presseschau).

Weitere Informationen:

<http://nepomuk.semanticdesktop.org>

3.3. Microsoft Grant „Social Search“



Folksonomien ermöglichen eine individuelle Kategorisierung des eigenen Wissens durch die Zuordnung von Schlüsselbegriffen (Tags) zu relevanten Ressourcen (z.B. Urls) und zugleich eine kollektive Nutzung der gesammelten Daten. Trotz der richtunggebenden Funktion von Tags, die zur Navigation und Orientierung in der Datenmenge genutzt werden können, bietet eine gezielte Suche wie sie von den herkömmlichen Suchmaschinen für das World Wide Web bekannt ist, eine sinnvolle Ergänzung/Erweiterung für die Informationsfindung. Aus diesem Grund untersuchte das Projekt, das durch eine Forschungsprämie von Microsoft Research unterstützt wurde, die Anwendung und Fortentwicklung auf Linkpopularität basierender Verfahren auf die tripartite Struktur der Folksonomie.

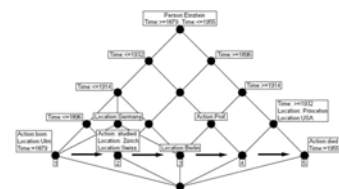
Weitere Informationen:

<http://www.microsoft.com/presspass/features/2006/may06/05-31LiveLabs.msp>

3.4. COMO - Concepts and Models

DFG, 12/2005 – 11/2007

Das DFG-Projekt COMO ist eine deutsch-russische Kooperation zur Untersuchung begrifflicher und modelltheoretischer Strukturen in der Wissensverarbeitung. Dabei wurden Ansätze zu einer begrifflichen Restrukturierung der Logik im Hinblick auf praktische Verwertbarkeit mit modelltheoretischen Bestrebungen verknüpft werden, um offene Probleme im Bereich der pragmatischen Verwendung aktueller mathematisch-logischer Theorien lösen zu können.



3.5. KDUBiq - Knowledge Discovery in Ubiquitous Environments

EU Coordination Action, 6. RP, 12/2005 – 05/2008

KDUBiq war eine EU-geförderte Coordination Action für ubiquitäre Wissensverarbeitung. Dr. Andreas Hotho leitete die Arbeitsgruppe „Ubiquitous Data Types“, in der die Ähnlichkeiten von Daten, die von Sensor-Netzen erzeugt werden, und von Daten, die aus sozialer Interaktion im Web 2.0 entstehen, im Hinblick auf die Wissensentdeckung untersucht wurde.



Weitere Informationen:

<http://www.kdubiq.org/kdubiq/control/index>“.

3.6. Prolearn – Professional E-Learning

EU Network of Excellence, 6. RP, 01/2004 – 12/2007

Das Network of Excellence PROLEARN wurde im Programm IST (Information Society Technology) der Europäischen Kommission gefördert und befasste sich mit professionellem, technologiegestütztem Lernen.



Das Ziel dieses Netzwerks war es, die wichtigsten Forschungsgruppen, Organisationen und Industriepartner im Bereich des Lernens und der Weiterbildung zusammenzubringen, um so die bestehende Lücke zwischen Forschung und Praxis in diesem Bereich zu verringern.

PROLEARN zielte darauf ab, die Forschungsergebnisse in diesem Bereich auf europäischer Ebene besser zu integrieren. Dazu wurden Kooperationen zwischen verschiedenen Beteiligten in Forschung und Anwendung initiiert und vertieft. Das Fachgebiet Wissensverarbeitung engagierte sich in dem Arbeitspaket Social Software und stellte mit BibSonomy eine Plattform für die Sammlung und Verwaltung von projektrelevanten Publikationen zur Verfügung.

3.7. Industrieprojekt mit K+S IT-Services GmbH



Untersuchung und Anwendung von Suchmaschinen bei K+S AG.

3.8. Publikations-Reporting bei SAP Research



SAP Research setzt seit 2008 das vom Fachgebiet Wissensverarbeitung betriebene kooperative Publikationsverwaltungssystem BibSonomy für das interne Publikations-Reporting ein.

3.9. Publikationsmanagement bei der Fraunhofer-Gesellschaft



Fraunhofer Gesellschaft

Das kooperative Publikationsverwaltungssystem BibSonomy des Fachgebiets Wissensverarbeitung wird vom Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme für das interne Publikations-Reporting eingesetzt.

3.10. Zukünftige Projekte

Das Fachgebiet startet momentan in mehreren gerade genehmigten bzw. beantragten Drittmittelprojekten die Kooperation mit inneruniversitären Partnern. Im DFG-Projekt „Info 2.0 – Informationelle Selbstbestimmung im Web 2.0“ bearbeiten wir zusammen mit der Projektgruppe verfassungsverträgliche Technikgestaltung der Universität Kassel die Frage, wie sich die neuen technischen Möglichkeiten des Web 2.0 auf die informationelle Selbstbestimmung der Nutzer auswirken und wie dieser Prozess gesellschaftsverträglich gestaltet werden kann. Im DFG-Projekt „PUMA – Akademisches Publikationsmanagement“ werden wir in Zusammenarbeit mit der Bibliothek die Verknüpfung von BibSonomy mit dem Bibliotheks-OPAC, der Open-Access-Plattform KOBRA und dem Forschungsbericht der Universität angehen.

Darüber hinaus sind mehrere Anträge bei der EU, DFG, BMBF und dem Land Hessen eingereicht bzw. in Vorbereitung.

4. Veranstaltungen

Mitglieder des Fachgebiets Wissensverarbeitung haben in den letzten drei Jahren zusammen mit Kollegen anderer Universitäten die folgenden wissenschaftlichen Workshops auf internationalen Tagungen als Program Co-Chairs organisiert:

- Co-Chair of the ESWC Workshop on Semantic Network Analysis, Budva, Montenegro, 12. Juni 2006 (Stumme)
- Web Mining Workshop at ECML/PKDD 2006, Berlin, Deutschland, 18. September 2006 (Hotho)
- Collaborative Knowledge Management Workshop (CoKM2007) at the 4th Knowledge Management Conference, Potsdam, Deutschland, 28.-30. März 2007 (Hotho)
- Member of the organising team of the WWW07 Workshop on Social and Collaborative Construction of Structured Knowledge, Banff, Kanada, 8. Mai 2007 (Stumme)

- Bridging the Gap between Semantic Web and Web 2.0, Workshop at ESWC 2007, Innsbruck, Österreich, 7. Juni 2007 (Hotho)
- Web Mining 2.0, Warsaw, Poland, Workshop at ECML/PKDD 2007, 21. September 2007 (Stumme)
- Data Mining in Web 2.0 Environments, Workshop at IEEE International Conference on Data Mining 2007, Omaha, U.S.A., 28. Oktober 2007 (Hotho)
- Emergent Semantics and Ontology Evolution, Busan, Korea, Workshop at ISWC 2007, 13. November 2007 (Hotho)
- Third Conceptual Structures Tool Interoperability Workshop (CS-TIW 2008), Workshop at ICCS 2008, Toulouse, Frankreich, 7. Juli 2008 (Jäschke)
- KASW'08: Workshop on Knowledge Acquisition from the Social Web, Graz, Österreich, 3. September 2008 (Hotho)
- ECML PKDD Discovery Challenge 2008, Antwerpen, Belgien, 15. September 2008 (Hotho, Benz, Jäschke, Krause)
- Wikis, Blogs, Bookmarking Tools - Mining the Web 2.0 - Workshop at ECML PKDD 2008, Antwerpen, Belgien, 15. September 2008 (Hotho)
- Dagstuhl Seminar "Social Web Communities", Dagstuhl, Deutschland, 21.-26. September 2008 (Stumme)
- Workshop "Knowledge Discovery, Data Mining, Maschinelles Lernen 2008" der Fachgruppe KDML, Würzburg, Deutschland, 6.-8. Oktober 2008 (Jäschke)

Die Tagungen und Workshops werden im Folgenden kurz beschrieben, weitere Details sind unter <http://www.kde.cs.uni-kassel.de/events> zu finden.

4.1. Semantic Network Analysis, Workshop at ESWC 2006

Mit dem schnellen Wachstum von Web2.0-Anwendungen richtete sich der Fokus vieler Wissenschaftler aus unterschiedlichen Disziplinen (wieder) auf soziale Netzwerke. Zwei Disziplinen, die durch diesen Fokus gegenseitig ins Blickfeld rückten, sind das Semantic Web und die Soziale Netzwerkanalyse (SNA). Semantic-Web- und Web2.0-Anwendungen sind neue Anwendungsgebiete für die SNA, stellen aber auch neue Herausforderungen, sei es aufgrund der Größe der betrachteten Netzwerke oder aufgrund ihrer komplexeren Struktur. Beispielsweise sind die meisten SNA-Methoden nicht darauf ausgelegt, zwischen mehreren Relationstypen – wie sie auf natürliche Weise im Semantic Web vorkommen – zu unterscheiden. Ziel des Workshops war es, Forscher aus beiden Disziplinen erstmals zusammenzubringen und ihnen die Möglichkeit zu bieten, Schnittstellen zu diskutieren.

4.2. Web Mining Workshop at ECML/PKDD 2006

Der im Rahmen der Tagung ECML/PKDD 2006 in Berlin organisierte Workshop "Web Mining (WebMine)" diente als Plattform für Forscher, die sich mit der Ent-

wicklung von Methoden und Algorithmen im Bereich des Web Mining beschäftigen. Dabei werden Data-Mining-Verfahren, z.B. aus dem Bereich des Text Mining, auf die Gegebenheiten des World Wide Web angepasst. Anwendungsgebiete wie z.B. große Web-Verkaufsplattformen erzeugen neue Web-Mining-Herausforderungen durch die Forderung nach besseren Verfahren zur Personalisierung. Um diesen Anforderungen zu begegnen, werden immer häufiger Methoden des Semantic Web in das Data Mining integriert, um die reichhaltigen Informationen besser verarbeiten zu können. Arbeiten aus den genannten Themengebieten wurden vorgestellt und diskutiert.

4.3. Collaborative Knowledge Management Workshop (CoKM2007) at the 4th Knowledge Management Conference

In Zusammenarbeit mit Forschern aus dem Bereich der Sozialen Netzwerkanalyse, der Psychologie und dem Data Mining wurde der Workshop Collaborative Knowledge Management (CoKM2007) auf der "4. Konferenz Professionelles Wissensmanagement - Erfahrungen und Visionen" in Potsdam organisiert. Web2.0-Systeme bieten auf der einen Seite einen sehr einfachen Zugang, haben aber auf der anderen Seite den Nachteil, dass die gespeicherten Daten nur sehr schwach strukturiert sind. Ziel des Workshops war es daher, die aus dem Bereich der Wissensrepräsentation bekannten Methoden zur Strukturierung von Wissen auf die Probleme in Web2.0-Systemen zu übertragen.

4.4. Social and Collaborative Construction of Structured Knowledge, Workshop at WWW 2007

In diesem Workshop wurde diskutiert, welche Möglichkeiten heutige Ontologie-Systeme bieten, um in einer Gruppe von Leuten gemeinsam Wissen zu akquirieren, zu formalisieren und zu strukturieren. Verschiedene Systeme, die an unterschiedlichen Stellen des Ontologie-Lebenszyklus eingesetzt werden, wurden vorgestellt und miteinander verglichen. Der Workshop wurde durch einen eingeladenen Workshop von Jamie Taylor (Metaweb) abgerundet.

4.5. Bridging the Gap between Semantic Web and Web 2.0, Workshop at ESWC 2007

Mit dem Aufkommen des Web 2.0 und der damit verbundenen Menge an benutzererzeugten Daten stand sehr schnell die Frage im Raum, wie man diese große Menge an schwach strukturierten Daten für das Semantic Web nutzbar machen kann. Mit diesem Thema beschäftigte sich der Internationale Workshop "Bridging the Gap between Semantic Web and Web 2.0" der auf der 4th European Semantic Web Konferenz am 7. Juni 2007 in Innsbruck, Österreich veranstaltet wurde. Der sehr gut besuchte Workshop deckte das ganze Spektrum an Web2.0-Anwendungen ab. Er führte zu sehr interessanten Diskussionen über die Analyse von Web2.0-Daten, sowie über Kombinationsmöglichkeiten von Semantic Web und Web 2.0.

4.6. Web Mining 2.0, Workshop at ECML/PKDD 2007

Methoden der Sozialen Netzwerk-Analyse wurden populär für die Analyse von Web2.0-Communities und andere Paradigmen des Social Computing. Während viele Plattformen eine einfache Unterstützung zum Management von persönlichen sozialen Netzwerken bieten, gibt es nur wenige Ansätze, die soziale Relationen modellieren und diese dann für Dienste einsetzen, die gleichzeitig soziale und Kontext-Aspekte berücksichtigen. Ziel des Workshops war es, Modellierungen zu diskutieren, die den Kontext einbeziehen und für Algorithmisierungen genutzt werden können. Einsatzmöglichkeiten gibt es in einem breiten Spektrum von mobilen Diensten, bspw. in den Feldern kontextuell-soziale Awareness, kontextuell-soziale Empfehlungen, sowie kontextsensitive Autorisierung.

4.7. Data Mining in Web 2.0 Environments, Workshop at IEEE International Conference on Data Mining 2007

Um neue Data-Mining-Verfahren ging es auf dem Internationalen Workshop über "Data Mining in Web 2.0 Environments", der in Kooperation mit der IEEE International Conference on Data Mining (ICDM 2007) veranstaltet wurde. Der Workshop richtete sich sowohl an Forscher als auch an Praktiker, die im Bereich Data und Web Mining arbeiten. Bei der Anwendung von Data-Mining-Methoden auf die im Web 2.0 vorkommenden unstrukturierten Daten trifft man auf völlig neue Probleme für das Data Mining. Zum Beispiel findet man viele lose gekoppelte Textstücke, die von Blogs, aber auch von Bookmarking-Systemen, stammen und die zum Teil widersprüchlich sein können. Andere Herausforderungen liegen in der schieren Masse der Daten oder rühren von Echtzeitanforderung an die zu entwickelten Algorithmen her. Lösungen wurden auf dem Workshop vorgestellt und diskutiert. Die Veranstaltung wurde durch einen eingeladenen Vortrag von Lyle Ungar zum Thema Informationsextraktion abgerundet.

4.8. Emergent Semantics and Ontology Evolution, Workshop at ISWC 2007

Der internationale Workshop über "Emergent Semantics and Ontology Evolution" wurde im Rahmen der ISWC 2007 veranstaltet. Ein zentrales Thema der Veranstaltung war die Frage: Wie kann man mit Hilfe von Daten aus dem Web 2.0 den Wissensakquisitions-Flaschenhals bei der Erstellung von Ontologien umgehen. Um dieses Ziel zu erreichen, müssen Methoden entwickelt werden, die die Ambiguität des unkontrolliert entstandenen Vokabulars sowie die flache Struktur an die Bedürfnisse einer Ontologie anpassen. Die präsentierten Arbeiten stellten hierfür Lösungen vor. Die beiden eingeladenen Vorträge von Karl Aberer und Paul Buitelaar rundeten das Programm ab.

4.9. Third Conceptual Structures Tool Interoperability Workshop, Workshop at ICCS 2007

Gemeinsam mit Wissenschaftlern aus Southampton und Karlsruhe organisierten wir 2008 den dritten Conceptual Structures Tool Interoperability Workshop (CS-TIW) in Toulouse, welcher im Rahmen der Tagung International Conference on Conceptual Structures (ICCS) stattfand. Auf dem Workshop präsentierten Forscher aus aller Welt ihre Ergebnisse zum Thema Conceptual Structures Tools and the Web und diskutierten bei der anschließenden Podiumsdiskussion rege über verschiedene Möglichkeiten, die Interoperabilität zwischen existierenden begrifflichen Programmen zu verbessern.

4.10. KASW'08: Workshop on Knowledge Acquisition from the Social Web, Workshop at Triple-I 2008

Im Rahmen der Triple-I-Konferenz in Graz wurde der "Workshop on Knowledge Acquisition from the Social Web" organisiert. Er fand am 3. September 2008 statt und hatte zum Ziel, Forscher aus verschiedenen Forschungsgebieten, die Interesse an der Extraktion von Wissen aus dem "Sozialen Web" haben und Methoden in diesem Bereich entwickeln, zusammenzubringen. Neben der Analyse von Web2.0-Daten wurde auch intensiv über die Kombination von Semantic Web und Web 2.0 diskutiert.

4.11. ECML/PKDD Discovery Challenge 2008

Im Rahmen der großen europäischen Konferenz zu maschinellem Lernen und Wissensentdeckung in Datenbanken (ECML/PKDD 2008) organisierten wir die sogenannte Discovery Challenge. Sie ermöglichte Wissenschaftlern über die Grenzen Europas hinweg ihre Lernverfahren an zwei vorgegebenen Problemen, basierend auf einem Datensatz unseres sozialen Verschlagwortungssystems BibSonomy, zu testen. Eine der beiden Aufgaben verfolgte das Ziel, Spammer von guten Benutzern automatisch unterscheiden zu lernen. In der zweiten Aufgabe sollten die Teilnehmer ein Empfehlungssystem für Schlagwörter entwickeln, welches möglichst gut die vom Benutzer gewählten Schlagwörter vorhersagt. Mehr als 150 Forscher aus aller Welt griffen auf den Datensatz zu, um ihn zu untersuchen und ihre Verfahren zu testen; 18 von ihnen reichten ihre Ergebnisse ein.

4.12. Wikis, Blogs, Bookmarking Tools – Mining the Web 2.0, Workshop at ECML/PKDD 2008

Der Workshop "Wikis, Blogs, Bookmarking Tools - Mining the Web 2.0 Workshop" wurde im Rahmen der ECML/PKDD 2008 am 15. September 2008 in Antwerpen, Belgien veranstaltet. Das Web 2.0 stellt gerade für Data-Mining-Forscher eine neue und sehr reichhaltige Quelle an Daten dar, die in verschiedenster Art und Weise analysiert und ausgewertet werden können. Ein zentrales Anliegen des Workshops war, den aktuellen Stand der Forschung zu erfassen und aktuelle

Trends im Bereich "Web 2.0 Mining" zu diskutieren. Dabei stand die Anwendung von neuen Data-Mining-Methoden und -Algorithmen im Vordergrund. Der Workshop wurde zusammen mit der ECML/PKDD Discovery Challenge abgehalten.

4.13. Dagstuhl Seminar "Social Web Communities"

Web2.0-Anwendungen wie Blogs, Wikis und kooperative Verschlagwortungssysteme haben in sehr kurzer Zeit viele Benutzer anziehen können. Ein wesentlicher Grund hierfür ist die einfache Bedienung der Systeme, die keine Informatikexpertise voraussetzt. Seit einiger Zeit arbeiten Wissenschaftler daran, diese "Graswurzel-Wissensrepräsentationen" mit semantisch reichhaltigeren Darstellungsformen zu verbinden, ohne die Anwender damit zu überfordern. In dem aus dem EU-Projekt "TAGora – Emergent Semiotics in Social Online Communities" hervorgegangenen Dagstuhl-Seminar trafen sich 31 Wissenschaftler der Disziplinen Semantic Web, maschinelles Lernen, Information Retrieval, Soziale Netzwerkanalyse, Sprachverarbeitung, Bibliothekswissenschaften und Hypermedia-Systeme, um die Entwicklung der nächsten Generation von sozialer Software gemeinsam voranzutreiben.

4.14. Workshop "Knowledge Discovery, Data Mining, Maschinelles Lernen 2008" der GI-Fachgruppe KDML

Im Rahmen der Workshopreihe „Lernen, Wissen, Adaption (LWA)“ fand vom 6. bis 8. Oktober der Workshop Knowledge Discovery, Data Mining, Maschinelles Lernen (KDML) der gleichnamigen Fachgruppe der Gesellschaft für Informatik in Würzburg statt. Die Workshopreihe dient dem Austausch aktueller Forschungsfragen und Ergebnisse insbesondere der deutschsprachigen Wissenschaftler und versteht sich als Forum, bei dem etablierte und neu auf einem Gebiet arbeitende Wissenschaftler miteinander diskutieren können. Der KDML-Workshop als Teil der LWA wurde gemeinsam mit Kollegen aus Darmstadt und Würzburg organisiert.

5. Lehre

Das Fachgebiet Wissensverarbeitung bietet im Grundstudium die Vorlesung Datenbanken I an. Im Hauptstudium werden vertiefende Vorlesungen zu den Themen Wissensentdeckung (Knowledge Discovery), Internet-Suchmaschinen (Information Retrieval), Semantic Web und Formale Begriffsanalyse angeboten. Das Angebot wird ergänzt durch regelmäßige Seminare und Projekte in diesen Themengebieten sowie durch die Betreuung von Diplomanden, Doktoranden und Habilitanden.

5.1. Lehrveranstaltungen

Folgende Lehrveranstaltungen wurden im Berichtszeitraum angeboten:

Wintersemester 2005/06

- Vorlesung Semantic Web
- Übung Semantic Web
- Vorlesung Internet-Suchmaschinen
- Übung Internet-Suchmaschinen
- Projekt

Sommersemester 2006

- Vorlesung Datenbanken
- Übung Datenbanken
- Vorlesung Knowledge Discovery
- Übung Knowledge Discovery
- Seminar Online Communities
- Projekt Data Mining Cup
- Projekt Semantic Web

Wintersemester 2006/07

- Vorlesung Künstliche Intelligenz
- Übung Künstliche Intelligenz
- Vorlesung Internet-Suchmaschinen
- Übung Internet-Suchmaschinen
- Projekt Web 2.0

Sommersemester 2007

- Vorlesung Datenbanken
- Übung Datenbanken
- Vorlesung Knowledge Discovery
- Übung Knowledge Discovery
- Seminar Clustering
- Projekt Data Mining Cup
- Projekt Web 2.0

Wintersemester 2007/08

- Vorlesung Künstliche Intelligenz
- Übung Künstliche Intelligenz
- Vorlesung Internet-Suchmaschinen
- Übung Internet-Suchmaschinen
- Projekte "Semantic Web" und "Wissensverarbeitung"
- Seminar Web 2.0

Sommersemester 2008 (Forschungssemester von Prof. Stumme)

- Vorlesung NLP - Analyse des Wissensrohstoffs Text
- Übung NLP
- Seminar Einführung in das Schreiben wissenschaftlicher Texte (im Bereich Informatik)
- Projekt Data Mining Cup
- Projekte "Semantic Web" und "Wissensverarbeitung"

Wintersemester 2008/09

- Vorlesung Künstliche Intelligenz
- Übung Künstliche Intelligenz
- Vorlesung Knowledge Discovery
- Übung Knowledge Discovery
- Projekte "Semantic Web" und "Wissensverarbeitung"
- Seminar Web 2.0

5.2. Projektarbeiten

In den Projekten Semantic Web (Bachelor) und Wissensverarbeitung (Master) wurden im Berichtszeitraum die folgenden Arbeiten durchgeführt.

Diplom I/Bachelor

- Thomas Steuber: Ontology Learning in BibSonomy, WS 05/06
- Alexander Podlich: Suchmaschinen, Vergleich von TF/IDF und Pagerank, WS 05/06
- Alex Armonis und Samuel Greef: Rankingverfahren von Suchmaschinen, WS 05/06
- Lado Kumsiashvili: Suchmaschinen-Crawler, WS 05/06

- Christian Schenk: Tagging von Files, WS 05/06
- Serak Rezane: BibSonomy – Ein Folksonomy System, WS 05/06
- Markus Matz: Multilinguales Klassifizieren, SS 06
- Jens Illig: Entwurf und Integration eines Item-Based Collaborative Filtering Tag Recommender Systems in das BibSonomy-Projekt, WS 05/06
- Johannes Zschoche: Graph-Visualisierung von Peer-to-Peer-Systemen, SS 07
- Christian Voigtmann, Stefan Achler, Martin Göb, Data Mining Cup 2007, SS 07
- Stefan Stützer: Spam-Entdeckung in BibSonomy, WS 07/08
- Julia Trabert: Erweiterung von BibSonomy um AJAX-Elemente, WS 07/08
- Andreas Schmidt: Lernen von Tag-Recommender, WS 07/08

Master

- Design einer API für BibSonomy, Manuel Bork, SS 06
- Jörn Dreyer, From Natural Text to Object Diagrams - NT2OD, SS 07
- Andreas Koch: Framework for System Tags, SS 08
- Michael Wagner: Erweiterung der Gruppenfunktionalität in BibSonomy, SS 08
- Stefan Lind: Firefox-3.0-Erweiterung für BibSonomy, SS 08
- Thomas Steuber: Performance-Tests von RDF-Datenbanken, SS 08

5.3. Abschlussarbeiten

Promotionen

- Dr. Christoph Schmitz: Self-Organized Collaborative Knowledge Management, Juni 2007

Diplom I

- Serak Rezane: Folksonomies in Intranets, Juni 2006
- Markus Matz: Geotagging mit mobilen Endgeräten, Juli 2006
- Thomas Steuber: Automatische Extraktion von bibliographischen Informationen, Februar 2007
- Jens Illig: Machine Learnability Analysis of Textclassifications in a Social Bookmarking Folksonomy, September 2008

5.4. Konzeption und Leitung von außeruniversitären Kursen und Schulungen

- Text Clustering and Semantic Web Mining, ICCL Summer School 2006, Dresden, Deutschland, 4.-5. Juli 2006 (Dr. Andreas Hotho)
- Ubiquitous Data Mining for Web 2.0, Tutorial at the 2nd European Summer School on Knowledge Discovery for Ubiquitous Computing, Porto, Portugal, 2.-9. März 2008 (Dr. Andreas Hotho)

5.5. Nachwuchs-Werbung

Der Fachbereich Elektrotechnik/Informatik informiert regelmäßig am Tag der Technik Schülerinnen und Schüler sowie deren Lehrkräfte über das Studienangebot und den Ingenieurberuf. Seit dem Sommersemester 2007 organisiert das Fachgebiet in diesem Rahmen ein Experiment zu Suchverfahren im Internet. Im Sommer 2007 boten wir dieses Experiment außerdem im Rahmen der Sommeruniversität an, die das Ziel hat, begabte Schülerinnen und Schüler für ein Frühstudium zu interessieren.

6. Forschungssemester von Prof. Gerd Stumme

Im Rahmen des Forschungssemesters hat Prof. Stumme an der Vrije Universiteit Brussel die inhaltliche Arbeit an dem EU-Projekt "TAGora - Emergent Semiotics in Social Online Communities" vorangetrieben und die bereits bestehenden Kontakte an der Vrije Universiteit Brussel und der Katholieke Universiteit Leuven vertieft sowie neue Kontakte an der Katholieke Universiteit Leuven und an der Université Catholique de Louvain-la-Neuve geknüpft. Während des Forschungssemesters wurden die folgenden Themen zu kooperativen Verschlagwortungssystemen bearbeitet: die Entdeckung von Spam, der Vergleich des Verhaltens der Benutzer von Suchmaschinen und von kooperativen Verschlagwortungssystemen, die Entwicklung eines Empfehlungssystems bei der Schlagwortzuordnung in kooperativen Verschlagwortungssystemen, sowie die Analyse der Semantik von Ähnlichkeitsmaßen für Schlagwörter. Das Dagstuhl-Seminar im Leibniz-Zentrum für Informatik "Schloss Dagstuhl" über "Social Online Communities", bei dem 30 eingeladene, international renommierte Teilnehmer über die Integration verschiedener Forschungsrichtungen (u.a. Web 2.0, Semantic Web, Wissensentdeckung, Information Retrieval, Bibliothekswesen) diskutierten, fand ebenfalls im Rahmen des Forschungssemesters statt.

Der Lehrbetrieb lief während des Forschungssemesters weiter. Die Grundvorlesung Datenbanken wurde von Herrn Kollegen Wegner übernommen. Dr. Andreas Hotho bot in diesem Semester erstmals die neue Vorlesung „NLP – Wissensrohstoff Text“ im Master Informatik an.

7. Weitere Aktivitäten

7.1. Kasseler Informatik Kolloquium (KIK)

Im Kasseler Informatik-Kolloquium (KIK) präsentieren Forscherinnen und Forscher aktuelle Ergebnisse zu Grundlagen und Anwendungen der Informatik. statt. Das KIK wird von allen Informatik-Kollegen bestritten, und vom FG Wissensverarbeitung koordiniert. Unser besonderer Dank gilt der Gesellschaft zur Förderung des kaufm. und techn. Führungsnachwuchses Kassel e. V. (GFF), die diese Aktivität durch Spenden fördert. Die aktuellen KIK-Termine sind unter <http://www.cs.uni-kassel.de> zu finden.

Termin	Vortragender	Titel
27.11.2008	Dr. Christoph Schmitz	Kleine Nadeln, großer Heuhaufen – Data Mining im industriellen Maßstab
06.11.2008	Prof. Dr. Dr.h.c. Volker Claus	40 Jahre Informatik – wie wir wurden, was wir sind
26.06.2008	Prof. Dr. Martin Riedmiller	Roboterfußball und mehr - Konzepte und Anwendungen maschineller Lernverfahren
29.05.2008	Dr. Andreas Hotho	Data Mining in kooperativen Verschlagwortungssystemen
24.04.2008	Dr. Markus Brede	Grenzen mathematischer Software am Beispiel der Computeralgebrasysteme Maple und Mathematica
31.01.2008	Dipl.-Ing. Klaus Bosch	Validieren und Testen von sicherheitskritischer Software unter Berücksichtigung von COTS, Applikationen, ASIC-Entwicklungsprozessen
20.12.2007	Prof. Dr. Barbara Chapman	Challenges for Application Development in the Multi-Core Era
29.11.2007	Prof. Dr. Ralph Bergmann	Ambient Intelligence zur Entscheidungsunterstützung bei Feuerwehreinsätzen
01.11.2007	Dr. Gerrit Hornung	Verdeckte Online-Durchsuchungen - technisch machbar und rechtlich zulässig?
28.06.2007	Prof. Dr. Otto Spaniol	The Excellence Cluster UMIC (Ultra High-Speed Mobile Information and Communication)
26.04.2007	Prof. Dr.-Ing. Georg Frey	Analyse vernetzter Systeme mittels Probabilistic Model Checking
08.02.2007	Doktoranden und Habilitanden der Kasseler Informatik	Wissen für alle - dank Informatik
06.02.2007	Prof. Klaus Wucherer	Stand und Weiterentwicklung der Automa-

		tisierungstechnik
01.02.2007	Prof. Dr. Claudia Eckert	Ubiquitäre Sicherheit im Zeitalter des Ubiquitous Computing -- Neue Herausforderungen für die IT-Sicherheit
11.01.2007	Prof. Dr.-Ing. Birgit Vogel-Heuser	Antrittsvorlesung: Systemarchitektur und Entwicklungsmethoden für verteilte, verlässliche eingebettete Systeme in der Domäne Automatisierungstechnik
08.12.2006	Marit Hansen	Identitätsmanagement und Datenschutz: Ansätze im Projekt PRIME - Privacy and Identity Management for Europe
08.12.2006	Prof. Dr. Kai Rannenberg	Identity changes in a changing world: Some results from the EU Network of Excellence "Future of Identity in the Information Society" (FIDIS)
30.11.2006	Dr. Bettina Hoser	Netzwerke als grundlegendes Modell interaktiver Prozesse: Analyse und Interpretation
26.10.2006	Prof. Dr. Luttenberger	WiSeBEES: A sparse and mobile Wireless Sensor Network
29.06.2006	Prof. Dr. Dieter Hogrefe	Das Internet vom "Best-Effort"-Netz zum sicheren Kommunikationsmedium
01.06.2006	Prof. Dr. Andreas Nürnberger	Multimedia Retrieval: Quo Vadis?
26.01.2006	Prof. Dr. Rudi Studer	Semantic Web: Methoden, Werkzeuge und Anwendungen

7.2. Jour Fixe und Clustersprecher Informatik

Die zwölf Informatiker des Fachbereichs 16 treffen sich regelmäßig etwa einmal monatlich im Semester, um alle Fragen, die die Organisation von Forschung und Lehre der Informatik an der Universität Kassel betreffen, zu besprechen. Das FG Wissensverarbeitung koordiniert den Jour Fixe. Zum WS 2008/09 hat Prof. Gerd Stumme darüber hinaus die Rolle des Clustersprechers Informatik übernommen.

7.3. Akademische Selbstverwaltung

Prof. Dr. Gerd Stumme

- Cluster-Sprecher Informatik (seit 10/08)
- Mitglied in der Berufungskommission Theoretische Informatik (2008/9)
- Mitglied im Prüfungsausschuss Bachelor Informatik, seit 10/08 kommissarischer Vorsitzender
- Mitglied im Prüfungsausschuss Diplom Informatik
- Mitglied im Prüfungsausschuss Master Informatik

Dr. Andreas Hotho

- Mitglied in der Berufungskommission Theoretische Informatik (2008/9)
- Mitglied im Prüfungsausschuss Bachelor Informatik (seit April 2005)
- Mitglied im Prüfungsausschuss Master Informatik (seit August 2006)

Dominik Benz

- Stellvertretendes Mitglied im Prüfungsausschuss Bachelor Informatik

7.4. Gesellschaft für Informatik

Prof. Dr. Gerd Stumme

- Mitglied im Sprechergrremium des GI-Arbeitskreises „Grundlagen von Informationssystemen“

Dr. Andreas Hotho

- Mitglied der Fachgruppenleitung der GI-Fachgruppe „Knowledge Discovery, Data Mining und Maschinelles Lernen“
- Mitglied im Sprechergrremium des GI-Arbeitskreises „Knowledge Discovery“

7.5. Mitgliedschaft in wissenschaftlichen Beiräten

Prof. Dr. Gerd Stumme

- Advisory Board des Europäischen Projekts "WeKnowIt – Emerging, Collective Intelligence for Personal, Organisational and Social Use"

7.6. Mitgliedschaft in Programmkomitees internationaler Konferenzen und Workshops

Prof. Dr. Gerd Stumme

- 4th International Conference on Formal Concept Analysis, Dresden, 13.-17. Feb. 2006
- 3rd European Semantic Web Conference, (ESWC 2006), Budva, Montenegro, 11.-14. Juni 2006
- Editorial Board of the 14th Intl. Conference on Conceptual Structures (ICCS 2007), Aalborg, Dänemark, 16.-21. Juli 2006
- IFIP International Conference on Artificial Intelligence Applications and Innovations, Santiago, Chile, 20.-22. Aug. 2006
- 8th International Conference on Data Warehousing and its Applications (DaWaK 2006), Krakau, Polen, 4.-8. Sept. 2006
- 2nd VLDB Workshop on Ontology Based Techniques for Databases and Information Systems, Seoul, Korea, 11. Sept. 2006

- 12th International Conference on Artificial Intelligence: Methodology, Systems, Applications (AIMSA 2006), Varna, Bulgarien, 13.-15. Sept. 2006
- 17th European Conference on Machine Learning/10th European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2006), Berlin, 18.-22. Sept. 2006
- 15th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2006), Prag, Tschechische Republik, 2.-6. Okt. 2006
- Bildung von Sozialen Netzwerken in Anwendungen der Sozialen Software, Workshop at Informatik 2006, Dresden, 6. Okt. 2006
- 1st International Workshop on Applications of Semantic Technologies, Workshop at Informatik 2006, Dresden, 6. Okt. 2006
- 5th International Semantic Web Conference, Athens, USA, 5.-9. Nov. 2006
- 33rd International Conference on Current Trends in Theory and Practice of Computer Science (SOFWEM 2007), Harrachov, Tschechische Republik, 20.-26. Jan. 2007
- 5th Intl. Conference on Formal Concept Analysis, Clermont-Ferrand, Frankreich, 12.-16. Februar 2007
- Social Network Analysis Workshop at ICFCFA 2007
- Workshop on Collaborative Knowledge Management, Potsdam, 28. März 2007
- 16th Intl. World Wide Web Conference, Banff, Kanada, 8.-12. Mai 2007
- 1st IEEE Intl. Conference on Research Challenges in Information Science, Ouarzazate, Marokko, 23.-26. Juli 2007
- 5th Intl. Conf. on Concept Lattice Applications 2007, Montpellier, Frankreich, 24.-26. Oktober 2007
- Editorial Board Member of the 15th Intl. conf. on Conceptual Structures, Sheffield, Großbritannien, 22.-27. Juli 2007
- 18th European Conference on Machine Learning and the 11th European Conference on Principles and Practice of Knowledge Discovery, Warschau, Polen, 17.-21. September 2008
- 6th International Semantic Web Conference, Busan, Korea, 11.-15. November 2007
- Editorial Board Member of the 6th Intl. Conf. on Formal Concept Analysis, Montreal, Kanada, 25.-28. Februar 2008
- 17th Intl. World Wide Web Conference, Beijing, China, 21.-25. April 2008
- 5th Europ. Semantic Web Conference, Teneriffa, Spanien, 1.-6. Juni 2008
- Hypertext 2008, Pittsburg, U.S.A., 19.-21. Juni 2008

- 18th Europ. Conference on Artificial Intelligence, Patras, Griechenland, 21.-25. Juli 2008
- 9th Intl. Conf. on Web Information Systems Engineering, Auckland, Neuseeland, 1.-3. September 2008
- 3rd Conceptual Structures Tool Interoperability Workshop at ICFCA 2008, Toulouse, Frankreich, 7. Juli 2008
- Editorial Board Member of the 16th Intl. Conf. on Conceptual Structures, Toulouse, Frankreich, 7.-11- Juli 2008
- 6th International Conference on Knowledge Engineering and Knowledge Management, Catania, Italien, 29. Sept. – 3. Okt. 2008
- Workshop “Wikis, Blogs, Bookmarking Tools - Mining the Web 2.0” at ECML/PKDD 2008, Antwerpen, Belgien, 15 September 2008
- Vice Chair of the 7th Intl. Semantic Web Conference, Karlsruhe, 26.-30. Oktober 2008
- 2008 IEEE/WIC/ACM Intl. Conf. on Web Intelligence, Sydney, Australia, 9.-12. Dezember 2008

Dr. Andreas Hotho

- Workshop on Web Mining 2006 (WebMine), Workshop at the 17th European Conference on Machine Learning and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases, Berlin, Germany, 18. September 2006
- Workshop der Fachgruppe Knowledge Discovery, Data Mining und Maschinelles Lernen und des Arbeitskreises Knowledge Discovery (KDML 2006), im Rahmen der LWA 2006, Hildesheim, 9.-11. Oktober 2006
- Collaborative Knowledge Management (CoKM2007), Workshop der 4. Konferenz Professionelles Wissensmanagement, Erfahrungen und Visionen, Potsdam, Deutschland, 28. März 2007
- Bridging the Gap between Semantic Web and Web 2.0, Workshop der 4th European Semantic Web Conference 2007, Innsbruck, Österreich, 7. Juni 2007
- International Workshop on Data Mining in Web 2.0 Environments, Workshop at the IEEE International Conference on Data Mining 2007, Omaha, U.S.A., 28. Oktober 2007
- International Workshop on Emergent Semantics and Ontology Evolution, Workshop at the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference, Busan, Korea, 12. November 2007
- International Workshop on Knowledge Acquisition from the Social Web (KASW'08), Workshop at the Triple-I Conference, Graz, Österreich, 3. September 2008

- ECML PKDD Discovery Challenge 2008 (RSDC'08), Workshop at the 19th European Conference on Machine Learning and 12th Conference on Principles and Practice of Knowledge Discovery in Databases, Antwerpen, Belgien, 15. September 2008
- Wikis, Blogs, Bookmarking Tools - Mining the Web 2.0 Workshop (WBBTMine'08), Workshop at the 19th European Conference on Machine Learning and 12th Conference on Principles and Practice of Knowledge Discovery in Databases, Antwerpen, Belgien, 15. September 2008

Dominik Benz

- Workshop on „Integrating Taxonomies and Folksonomies for Enhanced Knowledge Navigation“ (ITFEKN 2008) at the 7th international Conference on Practical Aspects of Knowledge Management (PAKM2008), Yokohama, Japan, November 2008
- “Social Linking” Track of the ACM Hypertext Conference 2008, Pittsburgh, USA

Robert Jäschke

- Social Semantic Web: Where Web 2.0 Meets Web 3.0, Workshop at AAAI-SSS-09, Stanford, U.S.A., 23.-25. März, 2009
- ECAI 2008 Workshop on Mining Social Data, Patras, Griechenland, 21. Juli 2008

7.7. Begutachtung für Forschungs-Förderinstitutionen

Prof. Gerd Stumme

Neben regelmäßigen Antragsbegutachtungen für die Deutsche Forschungsgemeinschaft (DFG) war Prof. Dr. Gerd Stumme als Gutachter für die folgenden Institutionen tätig:

- Belgian Science Policy (BELSPO), September 2006
- UK Economic and Social Research Council (ESRC), November 2006
- Chilean Research Fund Council (FONDECYT), Oktober 2007
- Science Foundation Ireland (SFI), März 2008
- Fonds zur Förderung wissenschaftlicher Forschung (FWF) Österreich, Juni 2008
- Bayrische Forschungstiftung, August 2008
- Schweizerischer Nationalfonds (SNF) zur Förderung der wissenschaftlichen Forschung, September 2008

Dr. Andreas Hotho

- Czech Science Foundation (GACR), Juni 2007

- Luxembourg National Research Fund (FNR), August 2008

7.8. Gutachtertätigkeiten für internationale Fachzeitschriften

Prof. Dr. Gerd Stumme

- ACM Transactions on Knowledge Discovery from Data, Januar 2007
- Intl. Journal on Approximate Reasoning, Januar 2007
- ACM Transactions on the Web, Juni 2007
- Intl. Journal of Metadata, Semantics and Ontologies, 2007
- Handbook on Ontologies, November 2007
- Order, Januar 2008
- J. Data Mining and Knowledge Discovery, Februar 2008
- J. General Systems, März 2008
- Knowledge and Information Systems J., April 2008

Dr. Andreas Hotho

- Knowledge and Information Systems (KAIS): An International Journal, Juni 2006
- ACM Transactions On Internet Technology, Januar 2006
- IEEE Transactions on Knowledge and Data Engineering (TKDE), Februar 2006
- KES Journal, Februar 2006
- International Journal of Electronic Commerce, März 2006
- IEEE/ACM Transactions on Computational Biology and Bioinformatics, April 2006
- Knowledge and Information Systems (KAIS): An International Journal, April 2006
- Data and Knowledge Engineering (DKE), Mai 2006
- Data and Knowledge Engineering (DKE), September 2006
- IEEE Intelligent Systems, November 2006
- IEEE Transactions on Systems, Man, and Cybernetics, April 2007
- Advances in Data Analysis and Classification (ADAC), Mai 2007
- International Journal of Data & Knowledge Engineering, September 2007
- International Journal of Human Computer Studies, Oktober 2007
- Journal of Web Semantics, November 2007
- Journal of Classification, März 2008

- Knowledge and Information Systems (KAIS): An International Journal, Mai 2008
- Computers and Mathematics with Applications, Juni 2008
Beate Krause
- Data and Knowledge Engineering, Dezember 2008
Dominik Benz
- Journal of Information Science, August 2008

8. Publikationen und Vorträge

8.1. Publikationen

2006

- M. Ackermann, B. Berendt, M. Grobelnik, A. Hotho, D. Mladenic, G. Semeraro, M. Spiliopoulou, G. Stumme, V. Svatek, and M. van Someren (eds.): Semantics, Web and Mining. Springer, Heidelberg, 2006.
- H. Alani, B. Hoser, C. Schmitz, G. Stumme (eds.), Proceedings of the 2nd Workshop on Semantic Network Analysis. 2006. Workshop at the 3rd European Semantic Web Conference, Budva, Montenegro, June 11-14, 2006.
- D. Benz, K. H. L. Tso, L. Schmidt-Thieme: Automatic Bookmark Classification - A Collaborative Approach. Proceedings of the 2nd Workshop in Innovations in Web Infrastructure (IWI2) at WWW2006, Edinburgh, Scotland, 2006.
- B. Berendt, A. Hotho, D. Mladenic, G. Semeraro (eds.): Workshop on Web Mining 2006 (WebMine), 2006.
- S. Bloehdorn, A. Hotho: Boosting for Text Classification with Semantic Features. In Advances in Web Mining and Web Usage Analysis, Springer, LNCS 3932, 149-166, Heidelberg, 2006.
- S. Bloehdorn, W. Buntine, A. Hotho: Introduction to the Special Issue "Learning in Web Search". Informatica, 30(2):141-141, 2006.
- S. Bloehdorn, P. Cimiano, A. Hotho: Learning Ontologies to Improve Text Clustering and Classification. From Data and Information Analysis to Knowledge Engineering, 334-341, 2006.
- P. Haase, M. Ehrig, A. Hotho, B. Schnizler: Personalized Information Access in a Bibliographic Peer-to-Peer System. In S. Staab, H. Stuckenschmidt (eds.):

Peer-to-Peer and Semantic Web, Decentralized Management and Exchange of Knowledge and Information, Springer, 143-158, 2006.

- B. Hoser, A. Hotho, R. Jäschke, C. Schmitz, G. Stumme: Semantic Network Analysis of Ontologies. In Y. Sure, J. Domingue (eds.): *The Semantic Web: Research and Applications*, Springer, LNAI 4011, 411-426, Heidelberg, 2006.
- A. Hotho, R. Jäschke, C. Schmitz, G. Stumme: BibSonomy: A Social Bookmark and Publication Sharing System. In A. de Moor, S. Polovina, H. Delugach (eds.): *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, Aalborg University Press, 87-102, Aalborg, Denmark, 2006.
- A. Hotho, R. Jäschke, C. Schmitz, G. Stumme: Information Retrieval in Folksonomies: Search and Ranking. In Y. Sure, J. Domingue (eds.): *The Semantic Web: Research and Applications*, Springer, LNAI 4011, pages 411-426, Heidelberg, 2006.
- A. Hotho, R. Jäschke, C. Schmitz, G. Stumme; Trend Detection in Folksonomies. In Y. S. Avrithis, Y. Kompatsiaris, S. Staab, N. E. O'Connor (eds.): *Proc. First International Conference on Semantics and Digital Media Technology (SAMT)*, LNCS 4306, 56-70, Springer, Heidelberg, 2006.
- A. Hotho, R. Jäschke, C. Schmitz, G. Stumme: FolkRank: A Ranking Algorithm for Folksonomies, *Proc. FGIR, Hildesheimer Informatikberichte, Universität Hildesheim*, 111-114, 2006.
- A. Hotho, R. Jäschke, C. Schmitz, G. Stumme: Emergent Semantics in BibSonomy. In C. Hochberger, R. Liskowsky (eds.): *Informatik 2006 – Informatik für Menschen, Bd. 2*, P-94, Köllen Druck+Verlag, Bonn, 2006.
- R. Jäschke, A. Hotho, C. Schmitz, B. Ganter, G. Stumme: TRIAS - An Algorithm for Mining Iceberg Tri-Lattices . *ICDM '06: Proceedings of the Sixth International Conference on Data Mining*, 907-911, IEEE Computer Society, Washington, DC, USA, 2006.
- R. Jäschke, A. Hotho, C. Schmitz, G. Stumme: Wege zur Entdeckung von Communities in Folksonomies. In S. Braß, A. Hinneburg (eds.): *Proc. 18. Workshop Grundlagen von Datenbanken*, 80-84. Martin-Luther-Universität, Halle-Wittenberg, 2006.
- C. Schmitz, A. Hotho, R. Jäschke, G. Stumme: Content Aggregation on Knowledge Bases using Graph Clustering. In *Proceedings of the 3rd European Se-*

mantic Web Conference, LNCS 4011, pages 530-544. Springer, Heidelberg, 2006.

- C. Schmitz, A. Hotho, R. Jäschke, G. Stumme: Mining Association Rules in Folksonomies. In V. Batagelj, H.-H. Bock, A. Ferligoj, A. Ziberna (eds.): Data Science and Classification (Proc. IFCS 2006 Conference), Studies in Classification, Data Analysis, and Knowledge Organization, 261-270, Springer, Heidelberg, 2006.
- C. Schmitz, A. Hotho, R. Jäschke, G. Stumme: Kollaboratives Wissensmanagement. In T. Pellegrini, A. Blumauer (eds.): Semantic Web – Wege zur vernetzten Wissensgesellschaft, Springer, 273-290, Berlin, 2006.
- C. Schmitz, A. Löser: How to model Semantic Peer-to-Peer Overlays? Proc. P2PIR Workshop, Informatik 2006, Dresden, 2006.
- G. Stumme, A. Hotho, B. Berendt: Semantic Web Mining - State of the Art and Future Directions. Journal of Web Semantics, Elsevier, 4(2):124-143, 2006.

2007

- D. Benz A. Hotho. Position Paper: Ontology Learning from Folksonomies. In A. Hinneburg (ed.). LWA 2007: Lernen - Wissen - Adaption, Halle, September 2007, Workshop Proceedings (LWA), 109-112. Martin-Luther-University Halle-Wittenberg, 2007.
- D. Benz, K. H. L. Tso, L. Schmidt-Thieme: Supporting Collaborative Hierarchical Classification: Bookmarks as an Example. Special Issue of the Computer Networks journal on Innovations in Web Communications Infrastructure, (51)16:4574-4585, 2007.
- B. Berendt, A. Hotho, D. Mladenic, G. Semeraro, editors: From Web to Social Web: Discovering and Deploying User and Content Profiles, Springer, LNCS 4736, 2007.
- Y. Cao, K. Ehms, S. Fiedler, M. Hofer, A.-K. Kaiamo, B. Kieslinger, R. Klamma, B. Krause, M. Kravcik, T. Ryyppö, M. Spaniol, G. Stumme, F. Wild: Case study on social software use in distributed working environments, 2007.
- C. Cattuto, C. Schmitz, A. Baldassarri, V. D. P. Servedio, V. Loreto, A. Hotho, M. Grahl, G. Stumme: Network Properties of Folksonomies. AI Communications, 20(4):245-262, 2007.

- L. Chen, P. Cudré-Mauroux, P. Haase, A. Hotho, E. Ong, editors: Proceedings of the First International Workshop on Emergent Semantics and Ontology Evolution, ESOE 2007, co-located with ISWC 2007 + ASWC 2007, November 12th, 2007, 292 CEUR Workshop Proceedings. 2007.
- M. Grahl, A. Hotho, G. Stumme: Conceptual Clustering of Social Bookmarking Sites. In 7th International Conference on Knowledge Management (I-KNOW '07), 356-364. Know-Center, Graz, Austria, 2007.
- M. Grahl, A. Hotho, G. Stumme. Conceptual Clustering of Social Bookmark Sites. In A. Hinneburg, editor, Workshop Proceedings of Lernen - Wissensentdeckung - Adaptivität (LWA 2007), 50-54. Martin-Luther-Universität Halle-Wittenberg, 2007.
- A. Hotho, G. Stumme: Mining the World Wide Web – Methods, Applications, and Perspectives. *Künstliche Intelligenz*, 3:5-8, 2007.
- R. Jäschke, M. Grahl, A. Hotho, B. Krause, C. Schmitz, G. Stumme: Organizing publications and bookmarks in BibSonomy. In H. Alani, N. Noy, G. Stumme, P. Mika, Y. Sure, D. Vrandečić (eds.), Workshop on Social and Collaborative Construction of Structured Knowledge at WWW 2007, Banff, Canada, 2007.
- R. Jäschke, A. Hotho, C. Schmitz, G. Stumme: Analysis of the Publication Sharing Behaviour in BibSonomy. In Proceedings of the 15th International Conference on Conceptual Structures, LNCS 4604, Springer, Heidelberg, 2007.
- R. Jäschke, L. B. Marinho, A. Hotho, L. Schmidt-Thieme, G. Stumme: Tag Recommendations in Folksonomies. In J. N. Kok, J. Koronacki, R. L. de Mantaras, S. Matwin, D. Mladenic, A. Skowron, editors, Knowledge Discovery in Databases: PKDD 2007, 11th European Conference on Principles and Practice of Knowledge Discovery in Databases, Proceedings, LNCS 4702, 506-514, Springer, Heidelberg, 2007.
- R. Jäschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, G. Stumme. Tag Recommendations in Folksonomies. In A. Hinneburg (ed.): Workshop Proceedings of Lernen - Wissensentdeckung - Adaptivität (LWA 2007), 13-20. Martin-Luther-Universität Halle-Wittenberg, 2007.
- C. Schmitz. Self-Organized Collaborative Knowledge Management . Kassel University Press, 2007.
- C. Schmitz, M. Grahl, A. Hotho, G. Stumme, C. Catutto, A. Baldassarri, V. Loreto, V. D. P. Servedio: Network Properties of Folksonomies. In: Tagging and

Metadata for Social Information Organization, Proc. WWW2007 Workshop. Banff, Kanada, 2007.

- J. Völker, D. Vrandečić, Y. Sure, A. Hotho: Learning Disjointness. In E. Franconi, M. Kifer, W. May (eds.): Proceedings of the European 19 Semantic Web Conference, ESWC2007, LNCS 4519, 175-189, Springer, Heidelberg, 2007.

2008

- H. Alani, S. Staab, G. Stumme (eds.): Proceedings of the Dagstuhl Seminar on Social Web Communities. Schloss Dagstuhl, September 2008.
- K. Bade, D. Benz: Evaluation Strategies for Learning Algorithms of Hierarchical Structures, in Proceedings of the 32nd Annual Conference of the German Classification Society - Advances in Data Analysis, Data Handling and Business Intelligence (GfKI 2008), 2008. In Press.
- D. Benz, M. Grobelnik, A. Hotho, R. Jäschke, D. Mladenic, V. D. P. Servedio, S. Sizov, M. Szomszor: Analyzing Tag Semantics Across Collaborative Tagging Systems, in H. Alani, S. Staab, G. Stumme (eds.): Proceedings of the Dagstuhl Seminar on Social Web Communities, 2008.
- B. Berendt, N. Glance, A. Hotho, editors. Wikis, Blogs, Bookmarking Tools - Mining the Web 2.0 Workshop. Workshop at 18th Europ. Conf. on Machine Learning (ECML'08) / 11th Europ. Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD'08), 2008.
- C. Cattuto, D. Benz, A. Hotho, G. Stumme: Semantic Analysis of Tag Similarity Measures in Collaborative Tagging Systems, in Proceedings of the 3rd Workshop on Ontology Learning and Population (OLP3), Patras, Greece, 2008.
- C. Cattuto, D. Benz, A. Hotho, G. Stumme: Semantic Grounding of Tag Relatedness in Social Bookmarking Systems, The Semantic Web - ISWC 2008, LNAI 5318, 615-631, Springer, Heidelberg, 2008.
- A. Hotho: Social Bookmarking. In A. Back, N. Gronau, K. Tochtermann (eds.): Web 2.0 in der Unternehmenspraxis: Grundlagen, Fallstudien und Trends zum Einsatz von Social Software, pages 26-38. Oldenbourg Verlag, München, 2008.
- A. Hotho, D. Benz, R. Jäschke, B. Krause (eds.): ECML PKDD Discovery Challenge 2008 (RSDC'08), Workshop at the 18th Europ. Conf. on Machine Learning (ECML'08) / 11th Europ. Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD'08), 2008.

- A. Hotho, R. Jäschke, D. Benz, M. Grahl, B. Krause, C. Schmitz, G. Stumme: Social Bookmarking am Beispiel BibSonomy, in A. Blumauer, T. Pellegrini (eds.): Social Semantic Web, Springer, Berlin, Heidelberg, pp. 363–391, 2009.
- R. Jäschke, A. Hotho, C. Schmitz, B. Ganter, G. Stumme; Discovering Shared Conceptualizations in Folksonomies. Web Semantics: Science, Services and Agents on the World Wide Web, 6(1):38-53, 2008.
- R. Jäschke, B. Krause, A. Hotho, G. Stumme: Logsonomy A Search Engine Folksonomy. In Proceedings of the Second International Conference on Weblogs and Social Media (ICWSM 2008), AAAI Press, 2008.
- R. Jäschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, G. Stumme: Tag Recommendations in Social Bookmarking Systems, AI Communications, 21(4): 231-247, 2008.
- B. Krause, A. Hotho, G. Stumme: A Comparison of Social Bookmarking with Traditional Search, in Advances in Information Retrieval, 30th European Conference on IR Research, ECIR 2008, LNAI 4956, 101-113, Springer, Heidelberg, 2008.
- B. Krause, R. Jäschke, A. Hotho, G. Stumme: Logsonomy – social information retrieval with logdata, in HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia, ACM, New York, NY, USA, 157-166, 2008.
- B. Krause, C. Schmitz, A. Hotho, G. Stumme: The Anti-Social Tagger - Detecting Spam in Social Bookmarking Systems, in Proc. of the Fourth International Workshop on Adversarial Information Retrieval on the Web, 2008.
- J. Völker, D. Vrandečić, Y. Sure, A. Hotho; AEON - An Approach to the Automatic Evaluation of Ontologies. J. of Applied Ontology 3(1-2), 41-62, 2008.

8.2. Vorträge

Zusätzlich zu den Präsentationen der obigen Tagungsbeiträge wurden folgende Vorträge gehalten:

- Robert Jäschke: Information Retrieval in Folksonomies: Search and Ranking, L3S Seminar, Hannover, 20. Januar 2006.
- Gerd Stumme: Folksonomies und Soziale Netze. Informatik-Kolloquium, TU Clausthal-Zellerfeld, 25. Januar 2006.

- Andreas Hotho: Kollaboratives Wissensmanagement mit Social Resource Sharing Systems, Workshop: Social Software in der Wertschöpfung, MFG Stiftung Baden-Württemberg, Stuttgart, 18. Juli 2006.
- Andreas Hotho: Bibsonomy - Lesezeichen und Literatur Verwalten in sozialen Systemen, XML-Tage in Berlin 2006, 26. September 2006.
- Andreas Hotho: Using Background Knowledge to Improve Text Clustering, Info-Lunch, Forschungszentrum L3S, Universität Hannover, 1. Nov. 2006.
- Andreas Hotho: BibSonomy - Ein soziales System zum Verwalten von Lesezeichen und Literatur, LearnTec 2007, Karlsruhe, 14. Februar 2007.
- Beate Krause: BibSonomy - A Social Bookmarking and Publication Management System; Wirtschaftsinformatik, Humboldt Universität zu Berlin, Februar 2007
- Beate Krause: Collaborative Publication Management with BibSonomy, Thematic Workshop on Social Software Connecting People, Wien, März 2007
- Andreas Hotho: BibSonomy – Research around a Folksonomy System, Forschungskolloquium Data and Knowledge Engineering, Universität Magdeburg, 31. Mai 2007.
- Andreas Hotho: Data Mining in Social Bookmarking Tools, Kolloquiumsvortrag des Knowledge Management Department und der Kompetenz-Center CCSW and C4, DFKI Kaiserslautern, 29. Juni 2007.
- Andreas Hotho: Data Mining in Social Bookmarking Tools, CS Seminar of the SFU Computer Science Department, Simon Fraser University, Vancouver, 19. Juli 2007.
- Andreas Hotho: Analysis of Folksonomies, Kolloquium der Knowledge Discovery and Machine Learning Group, Fraunhofer IAIS, Sankt Augustin, 10. Dezember 2007.
- Gerd Stumme: Information Retrieval in Folksonomies, Informatik-Kolloquium, Universität Magdeburg, 10. Januar 2008
- Andreas Hotho: Analytische Methoden zur Nutzerunterstützung in Tagging-Systemen, Workshop: Social Tagging in der Wissensorganisation Perspektiven und Potenziale, Institut für Wissensmedien in Tübingen, 21./22. Februar 2008.

- Prof. Dr. Gerd Stumme: Data Mining in Folksonomies, Université Catholique de Louvain-la-Neuve, Belgien, 25. April 2008
- Andreas Hotho: Web 2.0 Data for Ubiquitous Knowledge Discovery, Final Symposium KDUBiq, Bonn, 13./14. Mai 2008.
- Andreas Hotho: Extracting Semantics from Folksonomies, International Workshop on Knowledge Acquisition from the Social Web, KASW'08, Graz, Austria, 3. September 2008.
- Andreas Hotho: Data Mining in kooperativen Verschlagwortungssystemen, KIK – Kasseler Informatik-Kolloquium, Universität Kassel, 29. Mai 2008.
- Andreas Hotho: Analysis of Emerging Structures in Tagging Systems, Forschungskolloquium der TU Berlin, DAI-Labor, Technische Universität Berlin, 22. August 2008.
- Gerd Stumme: Information Retrieval and Data Mining in Social Bookmarking Systems, Vrije Universiteit Brussel, Belgien, 17. September 2008
- Gerd Stumme: Improving the User Experience in the Web (2.0) with Knowledge Discovery, Information Retrieval and Social Network Analysis, Dagstuhl, 22. September 2008
- Andreas Hotho: Semantics in Social Tagging Systems, User Generated Metadata, Workshop at the International Conference on Dublin Core and Metadata Applications, Berlin, 26. September 2008.
- Gerd Stumme: Information Retrieval and Data Mining in Folksonomies, Kolloquium, Katholieke Universiteit Leuven, Belgien, 3. Oktober 2008
- Beate Krause: Recommendations in Emergent Networks, Microsoft, Redmond, U.S.A., 8. Oktober 2008
- Dominik Benz: Aspects of Social Content Retrieval in Collaborative Tagging Systems, Workshop of the PetaMedia Network of Excellence, Berlin, 21. Oktober 2008
- Gerd Stumme: Information Retrieval in Folksonomies, Kolloquium, Forschungszentrum L3S, Hannover, 10. Dezember 2008

9. Presseschau

heise online, 12.01.2009

<http://www.heise.de/newsticker/Semantic-Desktop-fuer-KDE--/meldung/121519>



Semantic Desktop für KDE

Menschen gruppieren Informationen normalerweise nach Themen und merken sich dann Beziehungen zwischen Objekten – etwa den Zusammenhang zwischen einer Person und der Firma, für die diese arbeitet. Computern dieses Konzept ebenfalls beizubringen, ist seit Jahren Gegenstand intensiver wissenschaftlicher Arbeit. In letzter Zeit konzentriert sich die Forschung dabei vor allem auf die Idee des "Semantic Web", der semantischen Aufbereitung von Internet-Angeboten, um beispielsweise Suchmaschinen zu verbessern. Ein europäisches Forschungsvorhaben versucht nun, diese Technik auf den PC-Schreibtisch zu holen: Das so genannte Nepomuk-Projekt arbeitet am "Semantic Desktop".

Die Koordination hat das deutsche Forschungszentrum für künstliche Intelligenz (DFKI) übernommen. Die Projektteilnehmer arbeiten bereits seit drei Jahren an einer Software, die sinnvolle Verbindungen zwischen Dateien auf einem einzelnen Rechner aufspüren soll. Die dabei entstandene Technik liegt inzwischen für verschiedene Plattformen vor und ist seit Kurzem standardmäßig als Komponente in der Linux-Benutzeroberfläche KDE enthalten, berichtet das Technologiema­gazin *Technology Review* in seiner Online-Ausgabe.

Die Semantic-Desktop-Idee ist nicht neu. Die Nonprofit-Organisationen Open Source Applications Foundation sowie SRI arbeiteten bereits an ähnlichen Projekten. Viel kam dabei bislang allerdings nicht heraus. Das Problem: Es ist nach wie vor sehr schwer, gute semantische Informationen automatisch zu generieren. Damit die Technik funktioniert, müssen Nutzer Dateien und Dokumente häufig selbst entsprechend markieren. Fehlt es jedoch an praktischen Anwendungen für dieses "Tagging", sind Nutzer auch nur schlecht zu motivieren, die notwendigen Metadaten zu erstellen.

Nepomuk unterscheidet sich von früheren Ansätzen durch eine praxisnähere Vision, meint Ansgar Bernardi, stellvertretender Leiter des Forschungsbereiches Wissensmanagement am DFKI. Die Software ergänzt viele semantische Informationen bereits selbst und ermutigt den Nutzer, weitere Inhalte beizutragen, weil ihm das hilft, Dateien später wieder aufzufinden. Außerdem enthält das Werkzeug eine einfache Methode, bereits angelegte Informationen mit anderen Nutzern auszutauschen.

Die Software generiert semantische Informationen mit Hilfe eines Crawler-Programms, das den Rechner durchkämmt und dabei so viele Dateien wie möglich mit Zusatzinformationen versieht. Diese auch von Suchmaschinen bekannten Werkzeuge gehen beispielsweise das Adressbuch eines Nutzers durch und suchen nach Dateien, die mit den darin gefundenen Personen zusammenhängen könnten. Nepomuk kann dann eine Datei, die von einer bestimmten Person abgeschickt wurde, mit einer Firma korrelieren, für die sie arbeitet. Neben der KDE-Version sind grundlegende Technologien aus Nepomuk auch für Windows und Mac OS X verfügbar.

Ben Schwan

Semantischer Sinn für den Schreibtisch

<http://www.heise.de/tr/Semantischer-Sinn-fuer-den-Schreibtisch--/artikel/121040>

Von [Erica Naone](#)

Menschen gruppieren Informationen normalerweise nach Themen und merken sich dann Beziehungen zwischen Objekten – etwa den Zusammenhang zwischen einer Person und der Firma, für die sie arbeitet. Computern dieses Konzept ebenfalls beizubringen ist seit Jahren Gegenstand intensiver wissenschaftlicher Arbeit. In letzter Zeit konzentriert sich die Forschung dabei vor allem auf die Idee des "Semantic Web", der semantischen Aufbereitung von Internet-Angeboten, um beispielsweise Suchmaschinen zu verbessern. Ein europäisches Forschungsvorhaben versucht nun, diese Technologie auf den PC-Schreibtisch zu holen: Das so genannte Nepomuk-Projekt arbeitet am "Semantic Desktop".



Die Koordination hat das deutsche Forschungszentrum für künstliche Intelligenz (DFKI) übernommen. Die Projektteilnehmer arbeiten bereits seit drei Jahren an einer Software, die sinnvolle Verbindungen zwischen Dateien auf einem einzelnen Rechner aufspüren soll. Die dabei entstandene Technologie liegt inzwischen für verschiedene Plattformen vor und ist seit kurzem standardmäßig als Komponente in der Linux-Benutzeroberfläche KDE enthalten.

Die Semantic Desktop-Idee ist nicht neu. Die Nonprofit-Organisationen Open Source Applications Foundation sowie SRI arbeiteten bereits an ähnlichen Projekten. Viel kam dabei bislang allerdings nicht heraus. Das Problem: Es ist nach wie vor sehr schwer, gute semantische Informationen automatisch zu generieren. Damit die Technik funktioniert, müssen Nutzer Dateien und Dokumente häufig selbst entsprechend markieren. Fehlt es jedoch an praktischen Anwendungen für dieses "Tagging", sind Nutzer auch nur schlecht zu motivieren, die notwendigen Metadaten zu erstellen.

Nepomuk unterscheidet sich von früheren Ansätzen durch eine praxisnähere Vision, meint Ansgar Bernardi, stellvertretender Leiter des Forschungsbereiches Wissensmanagement am DFKI. Die Software ergänzt viele semantische Informationen bereits selbst und ermutigt den Nutzer, weitere Inhalte beizutragen, weil ihm das hilft, Dateien später wieder aufzufinden. Außerdem enthält das Werkzeug eine einfache Methode, bereits angelegte Informationen mit anderen Nutzern auszutauschen.

Die Software generiert semantische Informationen mit Hilfe eines Crawler-Programms, das den Rechner durchkämmt und dabei so viele Dateien wie möglich mit Zusatzinformationen versieht. Diese auch von Suchmaschinen bekannten Werkzeuge gehen beispielsweise das Adressbuch eines Nutzers durch und suchen nach Dateien, die mit den darin gefundenen Personen zusammenhängen könnten. Nepomuk kann dann eine Datei, die von einer bestimmten Person abgeschickt wurde, mit einer Firma korrelieren, für die sie arbeitet. Das mache aus Nepomuk ein besonders praktisches lokales Suchwerkzeug, meint Bernardi.

Die meisten Betriebssysteme lassen Nutzer ihren Rechner nur nach einfachen Schlüsselbegriffen durchsuchen. Nepomuk kann hingegen deutlich nützlichere Informationen aufdecken, in dem es sich auf die Verbindungen zwischen Daten konzentriert. Die Software lokalisiert also auch relevante Dateien, die die Schlüsselbegriffe, die für die Suche verwendet wurden, nicht enthalten. Eingebaut ist außerdem ein Peer-to-Peer-Datenaustauschsystem zum leichten Teilen von Dateien samt ihrer semantischen Anhängsel zwischen mehreren Nutzern.

"Das könnte der semantische Desktop sein, der tatsächlich überlebt", meint Nova Spivack, Chef und Gründer von Radar Networks, der mit Twine einen eigenen semantischen Lesezeichen- und Social-Networking-Dienst betreibt. Er sieht viel Potenzial in dem, was das Projekt bislang geleistet hat. "Darauf kann man aufbauen."

Andere Projekte, semantische Technologien auf den Desktop zu holen, hätten bislang beim Endanwender keinen Erfolg gehabt. "Nepomuk ist für echte Nutzer und Endwickler gedacht." Besonders der direkte Einbau in KDE sei wichtig, glaubt Spivack. "Die Oberfläche ist weit verbreitet und lässt sich von Programmierern leicht erweitern und modifizieren."

Obwohl die offizielle Förderung für Nepomuk Ende Dezember auslief, erwartet Bernardi weitere Entwicklungsschritte – dann als Open-Source-Software-Projekt. Eine Spinoff-Firma ist ebenfalls angedacht. Außerdem wurde mit der Open Semantic Collaboration Architecture Foundation eine Plattform geschaffen, die die zukünftige Arbeit an Nepomuk-Technologien koordinieren soll.

Die Software ist nicht nur für KDE erhältlich. Nutzer können die grundlegenden Werkzeuge auch für Windows, Macintosh und andere Linux-Oberflächen kostenlos herunterladen. Nepomuk lässt sich außerdem auch nur für bestimmte Daten einsetzen – beispielsweise nur für Web-Seiten, die man im Browser Firefox betrachtet hat.

Erica Naone

Bücher jetzt einfacher verwalten

Uni Kassel führt das
System BibSonomy ein

KASSEL. Das Fachgebiet Wissensverarbeitung der Uni Kassel bietet einen neuen Dienst zur Verwaltung für Webseiten und Publikationen an. Das Verwaltungssystem ist frei zugänglich und heißt BibSonomy. Die Nutzer können sich eine Lesezeichensammlung anlegen und so bestimmte Einträge in Büchern und auf Internetseiten schneller wiederfinden.

Das System der Uni Kassel soll die Literaturrecherche bei wissenschaftlichen Arbeiten erleichtern. So wird automatisch eine Literaturliste erstellt. Das erspart das langwierige Zusammensuchen der genutzten Publikationen. Neben dem Anlegen von virtuellen Lesezeichen in den eigenen Interessengebieten können die Nutzer auch die Lesezeichen von anderen Nutzern durchstöbern und so beispielsweise neue Internetseiten entdecken. (jwe)

Kontakt: Prof. Gerd Stumme, Telefon 0 56 1/8 04 62 52. Das frei nutzbare BibSonomy-System gibt es im Internet unter: <http://www.bibsonomy.org>

The Wild vs The Orderly: Folksonomies and Semantics (TRIPLE-I 2008)



September 04, 2008 By: Jana Herwig

Category: Collective Intelligence, Search Engines, Social Software, Vocabularies & Languages

This second day of TRIPLE-I 2008 was my personal folksonomy day, even though the theme was already set yesterday, with Andreas Hotho's invited talk about "Extracting Semantics from Folksonomies" which was the opening lecture of the workshop "Knowledge acquisition from the Social Web."

Andreas Hotho is directing the Bibsonomy project at Kassel University's Knowledge and Data Engineering research group; Bibsonomy is a social bookmark and publication sharing system catering especially for researchers who, next to bookmarking also wish to manage publications. Next to other interesting things, Bibsonomy supports the import of bookmarks from del.icio.us, Firefox bookmarks and local BibTeX files. Being a project led by a university's computer science department, Bibsonomy is at the same time the result, the object and a stimulus for research in the area of tagging and folksonomies. Andreas describes this double appeal of folksonomies to both ordinary people and researchers in a 12 seconds vlog post:



Andreas Hotho's statement about folksonomies and research (see www.bibsonomy.org) on 12seconds.tv

One of the outcomes of the research into folksonomies is FolkRank, a search algorithm that exploits the structure of folksonomies; the name reveals that it was inspired by PageRank, but as the graph of folksonomy structures does not correspond to the web graph, some adaptations had to be made. The specifics of these adaptations can be found in an online article by Andreas and his colleagues: "FolkRank: A Ranking Algorithm for Folksonomies" (PDF, 268 KB).

Andreas Hotho's talk more specifically addressed the search for methods to identify tags which describe the same concept (or a more specific / a more general concept respectively) within a folksonomy. He suggested two approaches:

1. Applying measures directly to folksonomy statistics, allowing to describe tags as a vector; e.g. co-occurrence frequency and FolkRank could serve as a similarity measure (with these two having a tendency towards high-frequency tags) or a cosine method (which is more likely to produce "siblings")
2. Looking up tags in an external thesaurus/vocabulary (for instance achieving semantic grounding by mapping a tag and its most similar tags with Wordnet Synsets)

Future areas of interest within folksonomy research Andreas proposed were trend detection, tag recommendation, detecting spam (a major challenge!), logsonomies (i.e. the structure of search engine query log files) and learning synsets, hierarchies, and structures of folksonomies. Andreas Hotho can be contacted via his homepage, if you have any further questions regarding Bibsonomy, FolkRank or this present piece of research.

<http://blog.semantic-web.at/2008/09/04/the-wild-vs-the-orderly-folksonomies-and-semantics-triple-i-2008/>

e-teaching.org: e-teaching im Gespräch mit Robert Jaeschke 28.8.2007



The image shows a screenshot of the e-teaching.org website. At the top, there is a banner with the logo and several small images of people in a classroom or meeting. Below the banner is a navigation menu with links for Login, FAQ, and Glossar. On the left side, there is a vertical menu with various categories, each with a colored square icon. The 'News & Trends' section is highlighted in green, and 'NotizBlog' is selected. The main content area shows the breadcrumb 'e-teaching.org » News & Trends » NotizBlog' and the title 'Alle Blogbeiträge'. Below this, it says 'Anzahl Einträge: 751'. The main article is titled 'e-teaching im Gespräch mit Robert Jaeschke' and includes a small profile picture of Robert Jaeschke, a progress bar, and the text: 'e-teaching.org im Gespräch mit Robert Jaeschke, einer der Entwickler des Social-Bookmarking-Dienstes "Bibsonomy". Er schildert, wie Bibsonomy für Forschung und Lehre eingesetzt werden kann.' At the bottom of the article, it says 'Kategorien: Podcasts' and 'Gepostet von: fhillebrecht am 28.8.2007 um 14:58'. There are also links for 'Kommentar schreiben' and '0 Kommentare, Permalink'.

http://www.e-teaching.org/news/eteaching_blog/et_showEntries?permlink=1188305918

Pushing the Boundaries of Search, 31. Mai 2006: Microsoft Live Labs Award für BibSonomy
<http://www.microsoft.com/presspass/features/2006/may06/05-31LiveLabs.msp>

Pushing the Boundaries of Search

Microsoft

PressPass - Information for Journalists

The 12 winners in Microsoft Live Labs “Accelerating Search in Academic Research” are part of a quest to identify bold and innovative approaches to information retrieval, data mining, machine learning and human/computer interactions.

REDMOND, Wash., May 31, 2006 – According to the “Live Labs Manifesto” of Gary William Flake Ph.D., the Internet operates in a manner fundamentally unlike anything that has ever preceded it. It is a world where “something small and intangible – a better algorithm – can massively increase global utility and welfare.”

Microsoft Live Labs is a partnership between MSN and Microsoft Research that takes a holistic approach toward applied research for Internet-enabled products and services. The partnership brings together people with a variety of skills and perspectives to foster research programs, incubate entirely new inventions, and improve and accelerate new Web-based technologies.

This week Flake – head of Live Labs and a Microsoft technical fellow – and his team are announcing 12 winners of a new Live Labs request for proposals (RFP) entitled “Accelerating Search in Academic Research.” The RFP aims to identify bold, innovative and new approaches to information retrieval, data mining, machine learning and human/computer interactions, with the ultimate goal of creating new technologies that can drastically change the way we interact with the Web and its vast array of resources.

Recipients of the Live Labs grants are posing some of the most compelling questions in search technology today. Even if the user gets relevant results, can he or she trust that information? How can a search tool get the best data from the Web? What’s happening on that part of the Web that’s “below the surface,” that’s not being crawled today? How can user behavior help predict economic or social changes?

Although winners will receive cash grants of between US\$35,000 and \$50,000, according to Evelyne Viegas, Ph.D., program manager for External Research & Programs at Microsoft Research, most applicants say what drew them to the RFP was not the cash, but the promise of a wealth of real-world user query and click-through data from MSN. Awardees will gain access to extensive data logs from MSN to aid in their research, as well as an increased quota of queries to the MSN software development kit that enables programmatic access to real-world search results. (The data being used by the awardees contains no personally identifiable user information.)

“We’ve been hearing a lot from academia that they need this kind of data,” Viegas says. “That’s what’s really unique about this RFP. We’re giving them access to more than 15 million real-user queries, with click-through information. The academic community is hungry for this kind of resource, and we’re excited to see what they can do with it.”

Real-World Data

Researcher Amélie Marian, a grant recipient from Rutgers University, explains the value of MSN’s data for the academic community.

“One of the main difficulties in data management research, and specifically in Web data management, is to find large volumes of real data that can be used in experiments and evaluations,” she says. “So I was very interested in the opportunity to access the MSN query logs and excerpts.”

Marian will use the data provided by Live Labs to help understand how users search and access information, in an effort to create more reliable, trustworthy search results. Her project may be the first to consider corroborating evidence as exhibited by the presence of the same information in multiple Web pages as a factor in ranking search results.

“Web resources often have unreliable data because of erroneous, biased, misleading or outdated information,” Marian says. “My project aims at providing an interface that does the extra work of checking different Web sources to corroborate query answers, in order to save the users from the hassle of having to go through multiple Web sites to compare information.”

Search as a Social Phenomenon

Beyond ranking the relevance and veracity of results, the depth and breadth of those MSN data sets can provide researchers with a range of new insights into the way people use the Web, and into the Web itself, says Viegas. So, not surprisingly, researchers around the world were jumping at the opportunity. Live Labs received 182 proposals from institutions in 36 countries, representing the gamut of disciplines from computer science to sociology.

“We had economists, educators and sociologists responding to our search RFP, so that was really exciting to see,” she says. “It’s not just search as a technology, but also as a social phenomenon. With this RFP, we’ve been able to gather experts in academia who can help us examine search beyond relevance, and look into this Internet cultural shift where people are the makers of information as well as consumers of information. The hope is that, by providing researchers with large-scale data sets, they can start getting some answers and direction on solving many of the questions we have about search.”

According to grant recipient Zoubin Ghahramani, who conducts research through the University of Cambridge, Carnegie Mellon University and University College London, many of those answers may come from an approach that combines the rigors of technology and science with the inherent unpredictability of human behavior.

Ghahramani’s research is focused on using Bayesian statistical techniques to create more intelligent machines that can “learn” to deal with the uncertainties of search posed by users. His group is focused on several problems, including how to identify users who anticipate what other people will search for, personalizing search by combining information across similar users, and predicting what people will be searching for.

“Search is an inference problem,” he says. “What is the probability that the user is interested in this Web page given that he or she typed this query? Once we start exploring this database of search query logs using machine learning methods, I imagine we will find many interesting and practical challenges.”

Live Labs Grant Recipients

The following is a list of grant awardees, along with the title of their proposal and a brief abstract:

[...]

Gerd Stumme, University of Kassel

“Social Search: Bringing the Social Component to the Web.”

“Social bookmark tools like del.icio.us are rapidly emerging on the Web. Unlike link-based search approaches à la PageRank, these systems provide personal recommendations based on input from similar users. We will extend link-based search with social search, in order to provide enhanced functionality and multiple search paradigms for the Web.”

[...]

publik Nr. 6 vom 11.7.2006



Soziale Komponente für das Web

Kasseler Informatiker entwickeln komfortable Literatur- und Lesezeichenverwaltung

Jeder Leser kennt die lähmende Frage: Wo habe ich das gelesen, wie hieß gleich der Autor und wo ist die Quelle nun wieder zu finden? Was für den Hausgebrauch eine temporäre Irritation darstellt, die nach gründlicher Durchsicht des Bücherregals in der Regel besiegt wird, ist für wissenschaftliche Arbeit „lebenswichtig“. Das Auffinden und Nutzen von einschlägiger Literatur ist dabei nicht nur für den einzeln arbeitenden Wissenschaftler von hoher Bedeutung, sondern hat in der Zusammenarbeit von Forschergruppen einen noch höheren Stellenwert. Literaturrecherche, die koordiniert und arbeitsteilig vonstatten geht, bringt einen enormen Zeitvorteil. Informatiker an der Universität Kassel haben dafür mit www.bibsonomy.org ein webgestütztes System entwickelt, das es erlaubt, Literatur ebenso wie Lesezeichen des Webs zu verwalten. Was immer vom Leser für aufhebens- und wieder lesenswert angesehen wird, gibt dieser selbst in das System ein und macht es so sich selbst – und anderen (wieder) zugänglich, sei es nur für den Nutzer selbst, eine geschützte Gruppe oder die ganze Welt.

Das Problem selbst verwalteter Systeme liegt in der Verschlagwortung und der Effizienz der Suche. Bibliotheken greifen hierfür auf standardisierte Schlagwortsysteme

und geschultes Fachpersonal zurück. Die Verschlagwortung durch bibliothekarische Laien und in hochspezialisierten Fachgebieten, die sich den üblichen Standards entziehen, macht eine erfolgreiche Suche häufig zu einem Glücksspiel, weil jeder Forscher typischerweise sein eigenes Verwaltungs- und Ablageschema hat. Bibsonomy lässt solche frei wählbaren Schlagworte (in der gängigen EDV-Sprache: „Tags“) zu – und schafft sich damit Probleme bei der Suche. „Wir forschen daher an Methoden, die die Suche unterstützen, indem etwa Synonyme und doppelt vorhandene Literatureinträge erkannt werden“, sagt Professor Dr. Gerd Stumme, Leiter des Fachgebiets Wissensverarbeitung. Ziel sei, Algorithmen (mathematische Rechenvorschriften) zur Suche und Strukturierung von Wissen zu entwickeln.

Forschungsauftrag von Microsoft

Dass die Kasseler Informatiker um Professor Stumme in der Lage sind, komplexe Suchvorgänge zu steuern, haben sie jüngst im Wettbewerb um die „Homingberger Gepardenfordele“ bewiesen. Es galt, diesen real nicht existierenden Begriff den gängigen Suchmaschinen so schmackhaft zu machen, dass er durch Querverweise und andere Referenzen möglichst häufig registriert wurde. Die Kasseler sind da-

bei weit oben gelandet. Google und Co. sind unter anderem deshalb zu überlisten, weil sie die Interessen der Anfragenden nicht kennen. Die aber sind in einem sozialen System, das gemeinsame Lese- und Forschungsinteressen hat, nicht nur bei der Schlagwortvergabe von hoher Bedeutung. Das hat auch der Software-Riese Microsoft erkannt und einen seiner 12 weltweit vergebenen „Grants“ zur Verbesserung von Internet-Suchmaschinen jetzt für das Projekt „Social Search: Bringing the Social Component to the Web“ an das Kasseler Fachgebiet vergeben.

Für die jetzt schon registrierten 500 Nutzer von Bibsonomy wird die kostenlose Nutzung durch die Weiterentwicklung noch attraktiver werden. Auf einen wichtigen Nebenaspekt, der Wissenschaftlern und ihren Sekretariaten immer wieder ungeliebte Arbeit macht, weist Stumme mit einer gewissen Befriedigung noch besonders hin: Das Programm speichert die Publikationsdaten im so genannten BibTeX-Format. „Damit können die im Wissenschaftsgeschäft immer wieder geforderten Literaturlisten einmal zentral verwaltet und dann automatisch in die verschiedenen gewünschten Formate exportiert werden, u.a. Endnote, adäquat formatiertes HTML, XML und als RSS-Feed“, sagt Stumme.

Jens Brömer

Bibsonomy ist zu erreichen unter www.bibsonomy.org, die Benutzung ist kostenlos. Interessierten Instituten wird die Einrichtung einer Gruppe zur institutsinternen Literaturverwaltung angeboten. Das Fachgebiet Wissensverarbeitung unter www.kde.cs.uni-kassel.de/



Uni-Bibliothek

www.bibsonomy.org

Der Begriff „Lesezeichen“ – die Älteren von Ihnen werden sich erinnern – stammt ja ursprünglich aus dem Umgang mit Büchern. Der an der Uni Kassel entwickelte englischsprachige Dienst **BibSonomy** verwaltet Web-Lesezeichen ebenso wie Hinweise auf gedruckte Literatur. Zugeschnitten ist BibSonomy auf die Bedürfnisse des Wissenschaftlers, der sich gleichermaßen souverän aus Online- wie aus Offline-Quellen bedient.



Erfasst werden die Bücher in dem verbreiteten BibTeX-Format, das bibliographische Daten und meist eine Zusammenfassung enthält; diese Publikationsdaten lassen sich in verschiedenen Dateiformaten ausgeben, etwa als HTML, XML oder als RSS-Feed. BibSonomy kombiniert typische Web-2.0-Techniken wie Tagging mit einer hierarchischen Struktur. Nutzer können sich zu Arbeitsgruppen zusammenschließen, um ihre Wissensquellen untereinander bequem auszutauschen. (heb)