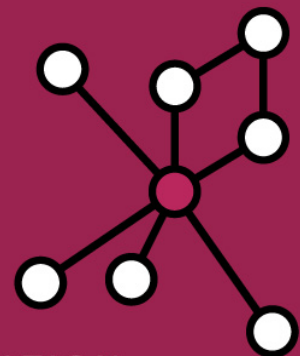

**Zweijahresbericht
des
Fachgebiets Wissensverarbeitung
der Universität Kassel**

Sommersemester 2004 – Wintersemester 2005/06

Prof. Dr. Gerd Stumme



ENDOWED CHAIR OF THE HERTIE FOUNDATION

Knowledge and Data Engineering

DEPARTMENT OF MATHEMATICS & COMPUTER SCIENCE



Vorwort

Liebe Leserinnen und Leser,

mit diesem Bericht möchte ich Ihnen einen Überblick über die Ereignisse seit dem Beginn des Fachgebiets Wissensverarbeitung geben, und Ihnen die Ergebnisse der Arbeit dieser ersten beiden Jahre vorstellen.

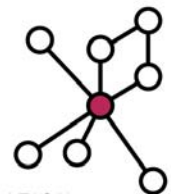
Zurückblickend kann ich sagen, dass diese Zeit sehr erfolgreich für uns war. Nach der teilweise turbulenten Aufbauphase, die wir Dank eines sehr schlagkräftigen Teams recht schnell überwandten, konnten wir uns ganz der Forschung und Lehre widmen. Hierbei ist es uns gelungen, mit einer frühzeitigen Fokussierung auf die Anwendung von Methoden der Wissensentdeckung und des Semantic Web auf das derzeit stark wachsende Forschungsgebiet Social Software ein eigenständiges Profil zu schaffen, wie Sie dem nachfolgenden Bericht entnehmen können. Mit unserem Sozialen Lesezeichen- und Literatur-Verwaltungssystem BibSonomy <<http://www.bibsonomy.org/>> können wir unsere Ergebnisse nun auch direkt einer breiten Öffentlichkeit zur Verfügung stellen.

Ich danke allen Mitarbeitern für ihr enthusiastisches Engagement und die große Leistungsbereitschaft beim Auf- und Ausbau des Fachgebiets. Besonders danke ich der Gemeinnützigen Hertiestiftung für die sehr gute Grundausstattung; und dem Stifterverband für die hilfreiche Start-Unterstützung.

Ich wünsche Ihnen viel Spaß bei der Lektüre!

A handwritten signature in black ink that reads "Gerd Stumme".

Prof. Dr. Gerd Stumme



ENDOWED CHAIR OF THE HERTIE FOUNDATION

Knowledge and Data Engineering

DEPARTMENT OF MATHEMATICS & COMPUTER SCIENCE

Highlights

L3S

Das Fachgebiet ist mit seiner Gründung dem Forschungszentrum L3S beigetreten. Das L3S arbeitet an grundlagen- und anwendungsorientierter Forschung auf dem Gebiet innovativer Informations-, Lern- und Wissenstechnologien sowie an Aus- und Weiterbildungskonzepten für Wissenschaft und Wirtschaft. Seite 3.

ICCS 2005

Das Fachgebiet veranstaltete vom 17. bis 22. Juli 2005 die 13th International Conference on Conceptual Structures an der Universität Kassel mit über 80 Teilnehmern. Seite 11.

BibSonomy

Das Fachgebiet Wissensverarbeitung bietet seit Anfang dieses Jahres einen neuen Dienst zur Verwaltung von Webseiten und Publikationen unter <http://www.bibsonomy.org/> an. Seite 4.

Hommingberger Gepardenforelle

Das Fachgebiet hat im Suchmaschinenoptimierungswettbewerb der Computerzeitschrift c't den 8. Rang von 3.030.000 bei Google erreicht. Seite 7.

Prolearn

Unsere Gruppe nimmt am European Network of Excellence PROLEARN teil, das sich mit professionellem, technologiegestützten Lernen befasst. Seite 8.

Nepomuk-Start

Als Mitglied des Forschungszentrums L3S Hannover nehmen wir am EU-Projekt "Networked Environment for Personalized, Ontology-based Management of Unified Knowledge" 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. Seite 7.

KDubiq

In der European Coordination Action "Knowledge Discovery in Ubiquitous Environments" koordinieren wir die Arbeitsgruppe „Ubiquitous Data – Data Collection and Interaction“. Seite 8.

Inhaltsverzeichnis

1	Wir über uns.....	1
1.1	Forschung im Fachgebiet Wissensverarbeitung.....	1
1.2	Mitarbeiter.....	2
1.3	Mitgliedschaft im Forschungszentrum L3S.....	3
2	Einzug und Aufbau des Fachgebiets.....	3
3	Rechner und Infrastruktur.....	4
4	BibSonomy – ein soziales Lesezeichen- und Literaturverwaltungs- system.....	4
5	Jagd auf die Hommingberger Gepardenforelle.....	7
6	Projekte.....	7
6.1	Nepomuk.....	7
6.2	KDUbiq.....	8
6.3	Prolearn.....	8
6.4	Personalized Access to Distributed Learning Resources.....	9
6.5	KDNet - Knowledge Discovery Network of Excellence.....	9
6.6	Suche und Klassifikation im Intranet der data process GmbH.....	9
6.7	COMO - COncEpts and MOdels.....	10
6.8	Zukünftige Projekte.....	10
7	Veranstaltungen.....	10
7.1	13 th International Conference on Conceptual Structures.....	11
7.2	Semantic Web Personalisation.....	12
7.3	Mining for and From the Semantic Web.....	12
7.4	Semantische Technologien für Informationsportale.....	13
7.5	European Web Mining Forum 2005.....	13
7.6	Foundational Aspects of Ontologies.....	13
7.7	Workshop on Ontologies in Peer-to-Peer Communities.....	13
7.8	Learning in Web Search.....	14
7.9	Semantic Network Analysis.....	14
8	Lehre.....	14
9	Weitere Aktivitäten.....	17
9.1	Kasseler Informatik Kolloquium (KIK).....	17
9.2	Jour Fixe Informatik.....	18
9.3	Akademische Selbstverwaltung.....	18
9.4	Gesellschaft für Informatik.....	18
9.5	Mitgliedschaft in Programmkomitees internationaler Konferenzen und Workshops.....	19
10	Publikationen.....	21
11	Presseschau.....	26

1 Wir über uns

Wissensverarbeitung – die Erforschung der Organisation und Anwendung von Wissen – gilt als eines der zukunftssträchtesten Forschungsgebiete der Informatik. Diesem hochaktuellen Thema widmet sich unser Fachgebiet Wissensverarbeitung. Dank des Engagements der Gemeinnützigen Hertie-Stiftung, die die C4-Professur mit einer Summe von rund einer Million Euro über einen Zeitraum von fünf Jahren finanziert, konnte die Wissensverarbeitung im April 2004 in den Lehr- und Forschungsbetrieb der Universität Kassel aufgenommen werden.

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. Vor dem Hintergrund stetig wachsender Informationsmengen gewinnt das Problem der Wissensverwaltung aber auch für Handel, Banken und Versicherungen immer mehr an Relevanz – nicht nur in technischer, sondern auch in personeller Hinsicht. Das Fachgebiet beschäftigt sich daher mit der organisatorischen und technischen Unterstützung von Wissensprozessen: vom Entdecken und Strukturieren des Wissens bis hin zur Kommunikation des Wissens. Durch die Arbeiten des Fachgebietes wird Wissen in Begriffssystemen strukturiert. Hierdurch wird zum einen das Wissen für automatische Verfahren nutzbar, zum anderen wird menschliches Denken und Kommunikation unterstützt. Insgesamt dient die Wissensverarbeitung also der Nutzung und dem Austausch von Wissen in Interessengruppen, wie z. B. Unternehmen.

In der Forschungslandschaft konzentriert sich das Fachgebiet auf Gebiete, die eine Semantik-basierte Unterstützung von Wissensprozessen zum Thema haben, wie Knowledge Discovery, Information Retrieval, Ontologien/Metadaten, Semantic Web, Peer to Peer. Insbesondere werden Methoden und Techniken auf den Schnittstellen dieser Forschungsgebiete entwickelt, um Synergien zu erreichen. Unser Anwendungsschwerpunkt liegt derzeit im Bereich der sozialen Software, mit dem Ziel, Kommunikation und Interaktion im Internet zu fördern.

Das Fachgebiet engagiert sich in nationalen und internationalen Forschungsprojekten, u. a. im Rahmen seiner Mitgliedschaft im Forschungszentrum L3S, Hannover, das an grundlagen- und anwendungs-orientierter Forschung auf dem Gebiet innovativer Informations-, Lern- und Wissenstechnologien sowie an Aus- und Weiterbildungskonzepten für Wissenschaft und Wirtschaft arbeitet. Die Mitarbeiter des Fachgebietes organisieren internationale Tagungen und Workshops rund um die Wissensverarbeitung.

Das Fachgebiet bietet Beratung zu allen Themen der Wissensverarbeitung an und führt Forschungs- und Entwicklungsprojekte zu Anwendungen der Wissensverarbeitung durch.

1.2 Mitarbeiter

Zum Ende des Wintersemesters 2005/06 waren fünf Mitarbeiter im wissenschaftlichen, technischen und organisatorischen Bereich des noch jungen Fachgebiets Wissensverarbeitung tätig. Zurzeit bauen wir das Fachgebiet durch Drittmittelprojekte und industrienaher Forschungsprojekte personell aus. Im wissenschaftlichen Bereich waren im Berichtszeitraum beschäftigt:

- Dr. Andreas Hotho (seit April 2004)
- Dipl.-Math. Robert Jäschke (UniK/L3S, seit Juni 2005)
- Dipl.-Inform. Christoph Schmitz (seit April 2004)
- Prof. Dr. Gerd Stumme (seit April 2004)
- Julien Tane, D.E.A. (L3S, April 2004 – Februar 2005)

Die Mitarbeiter werden tatkräftig durch Fr. Finis im Sekretariat, einen technischen Angestellten und durch mehrere studentische Hilfskräfte unterstützt.



1.3 Mitgliedschaft 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. Die Projekte des L3S sind den Bereichen Technology Enhanced Learning, Semantic Web and Digital Libraries sowie Distributed Systems and Networks zugeordnet, wobei die Aufgaben des L3S Forschung, Beratung und Technologietransfer gleichmaßen umfassen.

Das L3S hat sich national und international, z.B. als Koordinator des Networks of Excellence PROLEARN im Rahmen des 6. Forschungsrahmenprogramms der EU sowie als Core-Partner der Netzwerke KnowledgeWeb und REWERSE im Bereich Semantic Web etabliert und ist aktiv an der Ausrichtung der bedeutendsten Konferenzen in diesen Gebieten beteiligt. Das L3S ist in zehn Forschungsprojekten der EU im Rahmen des 6. EU/IST Forschungsrahmenprogramms involviert und koordiniert zwei davon.

Das Fachgebiet Wissensverarbeitung ist dem Forschungszentrum L3S am 22. März 2004 beigetreten. Die Projekte PADLR, Prolearn und Nepomuk (s.u.) werden über das L3S verwaltet.

2 Einzug und Aufbau des Fachgebiets

Dem Fachgebiet wurden Räume im Erdgeschoss der Wilhelmshöher Allee 73 in direkter Nachbarschaft zur Standort-Bibliothek zugewiesen. Diese mussten vor Bezug umgebaut und von Grund auf renoviert werden.



Aufgrund der baulichen Situation (große Raumtiefen, großes Fensterraster, bereits bestehende Innenwände für Laboranforderungen) schlug die Bauabteilung abweichend vom Fachgebietsleiter eine Raumgestaltung vor, die über den Normflächenwerten liegt, um die Umbaukosten geringer zu halten. Dies wird bei den anstehenden Diskussionen um das Raumhandelskonzept der UniK zu berücksichtigen sein.

Beim Umbau waren Trennwände und Türen einzusetzen, Bodenbeläge zu erneuern, die Schließanlage auszutauschen, eine Klimaanlage in den Serverraum einzubauen, neue Telefonverbindungen sowie das Glasfasernetz zu schalten, etc. Auch das gesamte Mobiliar wurde neu bestellt. Der knappe Zeitplan machte es erforderlich, die Planung der Arbeiten sehr aufmerksam zu überwachen und die Verwaltungsvorgänge nach Möglichkeit zu beschleunigen. Der Bezug lief dann dank der

guten Unterstützung aller Beteiligten weitgehend reibungslos, wenn wir auch die ersten Wochen noch auf provisorischen Möbeln und aus den Umzugskartons arbeiten mussten. Insgesamt ist durch das Engagement aller ein sehr angenehmes Arbeitsumfeld entstanden.

Mit der Antrittsvorlesung von Prof. Dr. Gerd Stumme über „Begriffliches Wissen und seine Verarbeitung“ und einer anschließenden Einweihung der Räume des Fachgebiets stellte sich das FG Wissensverarbeitung im Oktober 2004 der Öffentlichkeit vor.

3 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 Rechnerausstattung des Fachgebiets umfasst fünf zentrale Server sowie Arbeitsplatzrechner und eine Softwarewerkstatt. Neben Standarddiensten wie Mail, Web-Server, Content-Management, Wikis und Datenbanken stellen die Linux-Server über ein externes RAID-System etwa 2 TB Plattenplatz für Experimente und Datensätze zur Verfügung. Die Mitarbeiter sind je nach Bedarf mit Laptops, Workstations oder Barebones ausgestattet.

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. Derzeit arbeiten wir an der Einbindung weiterer Fachgebiete, um z. B. die Softwarewerkstätten zu vernetzen, die Verwaltung von Benutzerkennungen zu vereinheitlichen und Serverdienste zu konsolidieren.

Die Softwarewerkstatt besteht derzeit aus acht Rechnern, 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 Rechencluster benutzt werden. 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.

4 BibSonomy – ein soziales Lesezeichen- und Literaturverwaltungssystem

Seit Anfang 2006 bieten wir unter <http://www.bibsonomy.org/> einen neuen Dienst zur Verwaltung von Webseiten und Publikationen an. Dieses für jeden frei be-

nutzbare System erlaubt es, Lesezeichen (Favoriten, Bookmarks) für Webseiten zentral auf dem BibSonomy-Server abzuspeichern und zu verschlagworten. Die vom Nutzer frei wählbaren Schlagworte, im Englischen Tag genannt, erlauben es dem Nutzer, seine Lesezeichensammlung zu strukturieren und mit Hilfe der Tags einen Eintrag später einfach wieder zu finden.

The screenshot shows the BibSonomy website interface in a Mozilla Firefox browser. The search bar contains the word 'web'. Below the search bar, there are navigation links for 'tags', 'groups', 'popular', 'myBibSonomy', 'post bookmark', and 'post bibtex'. The main content area is divided into two columns: 'bookmarks' on the left and 'publications' on the right. The 'bookmarks' column lists several items, including 'www.ajaxwrite.com', 'Radio Melodic', 'Hard Rockin 80s.com', 'REST web services', 'PIMO-a PIM Ontology for the Semantic Desktop', and 'SPARQL: Web 2.0 Meet the Semantic Web - O'Reilly XML Blog'. The 'publications' column lists 'The Semantic Web', 'Semantic Web Mining - State of the Art and Future Directions', 'Semantic Network Analysis of Ontologies', and 'Efficient Phrase-Based Document Indexing for Web Document Clustering'. On the right side, there is a 'related tags' section with a list of tags such as 'semantic', 'Semantic', 'tools', 'web2.0', 'software', 'folksonomy', 'ajax', '2.0', 'mining', 'search', 'rdf', 'internet', 'design', 'social', 'programming', 'ontology', 'myown', 'services', 'ontology', 'javascript', 'technology', 'opensource', 'free', 'online', 'development', 'seminar2006', 'cool', 'knowledge', 'xml', 'wiki', 'list', 'acquisition', 'socialsoftware', 'google', 'collaboration', 'html', 'api', 'population', 'Sam', 'Integration', 'Information', 'Chapman', 'Armadillo', 'AKT', 'webdesign', 'tagging', 'reference', 'css', 'network', and 'learning'. The browser's address bar shows 'http://www.bibsonomy.org/tag/web'.

Darüber hinaus bietet das System jedem Benutzer die Möglichkeit, in den Lesezeichen anderer Benutzer mit verwandten Interessen zu suchen, und interessante neue Webseiten zu entdecken. Diese soziale Komponente erzeugt also personalisierte Empfehlungen, die globale Suchmaschinen wie Google nicht leisten können, da sie die Interessen des Anfragenden nicht kennen. Durch die zentrale Speicherung hat der Benutzer außerdem jederzeit von jedem Rechner Zugriff auf seine Lesezeichen.

Ein zentrales Thema bei der wissenschaftlichen Arbeit ist die Literaturrecherche. Für die systematische Katalogisierung der gefundenen Publikationen gibt es jedoch wenig systematische Unterstützung; jeder Forscher entwickelt typischerweise sein eigenes Verwaltungs- und Ablageschema. Da auch das Fachgebiet Wissensverarbeitung eine professionelle Literaturverwaltung benötigte, wurde BibSonomy in dieser Richtung erweitert.

Neben den üblichen Publikationsdaten wie Autor, Titel, Verlag, Jahr etc. kann man – wie bei den Lesezeichen – frei wählbare Tags zur Beschreibung der Doku-

mente hinzufügen. Ähnlich gehen Bibliotheken vor, wenn Sie Bücher in ihren Bestand aufnehmen und diese zum Zwecke des Wiederfindens verschlagworten. Während in einer Bibliothek jedoch die Schlagwörter aus einem fest vorgegebenen Katalog ausgewählt werden, kann in BibSonomy jeder Nutzer seine Tags frei wählen. Die Bibliothekslösung hat den Vorteil, dass die Literatursuche nicht durch mehrdeutige Schlagwörter erschwert wird. Bibliothekare werden allerdings für die systematische Verschlagwortung geschult, was für einen breiteren Benutzerkreis nicht vorausgesetzt werden kann. Die größere Flexibilität von Lesezeichen-Systemen bei der Eingabe wird durch Schwierigkeiten bei der Suche erkauft. Das Fachgebiet Wissensverarbeitung forscht daher an Methoden, die die Suche unterstützen, indem etwa Synonyme und doppelt vorhandene Literatureinträge erkannt werden.

BibSonomy verwendet das so genannte BibTeX-Format zur Speicherung von Publikationsdaten. Seine Publikationsverwaltung ist somit direkt in das moderne Satzsystem LaTeX integriert, mit dem Forscher – insbesondere in den Naturwissenschaften – ihre wissenschaftlichen Veröffentlichungen gleich druckfertig gestalten. Das System erzeugt automatisch Literaturlisten in weiteren Formaten und nimmt so den Forschern weitere unnötige Mehrarbeit ab, die ihnen bisher vielfach noch durch die Verwendung unterschiedlicher Systeme aufgezwungen wurde. Bisher musste häufig eine Sekretärin die Veröffentlichungsliste eines Instituts in verschiedenen Formaten pflegen, um sie an die Webseite des Instituts, den Jahresbericht und den diversen Berichten zu Forschungsprojekten anzupassen. Nun kann die Liste einmal zentral verwaltet und dann automatisch in die gewünschten Formate exportiert werden. Der Export in das – vor allem für Wordanwender relevante – Endnote-Formats für Literaturdaten wird ebenfalls unterstützt.

BibSonomy ist eines von mehreren Lesezeichen-Systemen, die im Zuge des Web 2.0 – der nächsten Generation des World Wide Webs – entstanden sind. Andere Lesezeichen-Systeme dienen der Verwaltung von Photos und Musik – und selbst dem Austausch von guten Vorsätzen! BibSonomy ist derzeit das einzige System, das die Verwaltung von Lesezeichen und Publikationen verbindet. Neben Lesezeichen-Systemen sind weitere typische Web-2.0-Anwendungen Wikis (Webseiten, die von jedermann verändert werden können, wie die Online-Enzyklopädie Wikipedia) und Blogs (Webtagebücher).

Neben der Idee, der Öffentlichkeit und Wissenschaftlern einen modernen Dienst zur Verwaltung und zum Austausch von Wissen zu bieten, ist die wesentliche Motivation des Fachgebiets Wissensverarbeitung zum Betreiben des Lesezeichen-Systems BibSonomy, dass es praktisch relevante Fragen zur Verarbeitung und Pflege von Wissen im Internet aufwirft, die Anstoß zu Forschungsprojekten bieten: Aufgrund der Tatsache, dass es in Lesezeichen-Systemen keine Vorschriften zur Verschlagwortung gibt (wie etwa in Bibliotheken), ist das verwendete Vokabular hier sehr viel heterogener. Ein wichtiges Forschungsthema ist daher, wie untrainierte Benutzer so unterstützt werden können, dass sie nicht durch die ver-

wendeten Verfahren irritiert werden. Das Fachgebiet entwickelt Suchverfahren und Empfehlungssysteme, die den Benutzern interessante Lesezeichen und Publikationen vorschlagen. Diese Ansätze werten das Netz aus, das die Benutzer durch die Vergabe der Schlagworte zwischen den Inhalten spannen.

Das FG Wissensverarbeitung entwickelt die Methoden und Verfahren zum Einsatz in sozialen Lesezeichen-Systemen derzeit in den zwei Europäischen Forschungsprojekten Nepomuk und Tagora (s.u.) bearbeitet, an denen das Fachgebiet Wissensverarbeitung beteiligt ist. Die hier entwickelten Verfahren werden sukzessive in das BibSonomy-System eingebaut.

5 Jagd auf die Hommingberger Gepardenforelle



Wo findet man am schnellsten Informationen über die Hommingberger Gepardenforelle?

Diese Frage stellte die Computerzeitschrift c't

und rief zu einem Wettbewerb auf: *Welche Webseite wird bei der Eingabe von "Hommingberger Gepardenforelle" in Google an den Stichtagen 15. Mai und 15. Dezember am höchsten gerankt?*

Am 15. April, einen Tag vor der Veröffentlichung, verzeichnete Google keinen Treffer, denn es gibt weder den Ort Hommingberg noch eine Gepardenforelle. Drei Tage später waren es bereits 104.000 Treffer. Der Wettbewerb hat viele Teilnehmer angelockt, unter ihnen viele Werbefirmen, die sich darauf spezialisiert haben, ihre Kunden in Suchmaschinen möglichst weit oben zu platzieren.

Im Rahmen des Seminars Internet-Suchmaschinen. beteiligte sich das Fachgebiet an dem Wettbewerb, und kam in der Endwertung bei Google auf den 8. Rang von insgesamt 3.030.000 Treffern. Die Studierenden erfuhren so anschaulich, welche Relevanz die von uns vermittelte Theorie in der Praxis hat. Mittlerweile hat die Fortsetzung dieser Aktivität in Verbindung mit BibSonomy zu dem Microsoft Live Labs Award „Accelerating Search in Academic Research“ geführt. Details zu dem Forellen-Wettbewerb sind unter <http://www.kde.cs.uni-kassel.de/lehre/ss2005/googlespam> zu finden.

6 Projekte

6.1 Nepomuk

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

Als Partner des Forschungszentrums L3S Hannover

nehmen wir am European Integrated Project 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. Die Vision ist der "Social Semantic Desktop", der die Fähigkeiten des Semantic Web mit denen der Social Network Analysis vereint.



Innerhalb eines Konsortiums aus Forschern, Industrie und einer wachsenden Community befasst sich das Fachgebiet Wissensverarbeitung insbesondere mit der Entdeckung und Strukturierung von Communities und untersucht dabei Verfahren aus dem Gebiet der Social Network Analysis. So werden Beziehungen zwischen Benutzern, Ressourcen und Metadaten zur Erkennung von Benutzern mit ähnlichen Interessen genutzt. Dies soll den Austausch von Wissen auch zwischen einander unbekanntem Nutzern fördern.

6.2 KDUBiq

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



KDUBiq (Knowledge Discovery in Ubiquitous Environments) ist eine europäische Coordination Action, die Forscher zusammenbringt, die sich mit der Wissensentdeckung in ubiquitären Systemen beschäftigen. KDUBiq unterstützt und koordiniert Forschungsaktivitäten in diesem Bereich. Die Hauptaufgabe des Projektes liegt darin, in einer Roadmap Forschungsfragen an der Schnittstelle von hochverteilten (und mobilen) Systemen und fortgeschrittenen Wissensentdeckungssystemen zu identifizieren und einer breiten Öffentlichkeit vorzustellen. Dr. Andreas Hotho koordiniert in dem Projekt die Arbeitsgruppe „Ubiquitous Data – Data Collection and Interaction“.

6.3 Prolearn

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



Unsere Gruppe nimmt teil am Network of Excellence PROLEARN, das im Programm IST (Information Society Technology) der Europäischen Kommission gefördert wird und sich mit professionellem, technologiegestütztem Lernen befasst. Das Ziel dieses Netzwerks ist 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 zielt darauf ab, die Forschungsergebnisse in diesem Bereich auf europäischer Ebene besser zu integrieren; dazu werden Kooperationen zwischen verschiedenen Beteiligten in Forschung und Anwendung initiiert und vertieft. Das Fachgebiet Wissensverarbeitung engagiert sich in dem Arbeitspaket Social Software und stellt mit BibSonomy eine Plattform für die Sammlung und Verwaltung von projektrelevanten Publikationen zur Verfügung.

6.4 Personalized Access to Distributed Learning Resources

BMBF & MWK Niedersachsen, 04/2001 – 02/2005



Das Fachgebiet Wissensverarbeitung der Universität Kassel bearbeitete im Rahmen seiner Mitgliedschaft im Forschungszentrum L3S ein Modul des Projektes „Personalized Access to Distributed Learning Resources (PADLR)“, den Courseware Watchdog. Dieser dient dazu, Lehrmaterialien im WWW oder im Peer-to-Peer-Netz Edutella zu finden und dem Benutzer zu präsentieren.

Mit Hilfe eines ontologiebasierten fokussierten Webcrawlers und durch Integration in Edutella kann Material gesammelt werden. Subjektives Clustering erweitert bekannte Algorithmen mit ontologiebasiertem Hintergrundwissen und erlaubt so die Beschreibung von Präferenzen und Erzeugung subjektiver Sichten. Eine Visualisierung basierend auf formaler Begriffsanalyse bietet intelligentes Browsing. Strategien zur Ontologie-Evolution erlauben es, Veränderungen im Interessengebiet des Lernenden in der Ontologie zu reflektieren.

6.5 KNet - Knowledge Discovery Network of Excellence

EU Network of Excellence, 5. RP, 02/2002 – 12/2004



KNet ist ein offenes Netzwerk von Teilnehmern aus Wissenschaft, Industrie und Verwaltung. Der Hauptzweck dieses internationalen Projekts ist die Integration von Problemen aus dem Geschäftsalltag in Forschungsdiskussionen und die Zusammenarbeit im Hinblick auf die Zukunft von Knowledge Discovery und Data Mining.

6.6 Suche und Klassifikation im Intranet der data process GmbH

data process GmbH, Beginn 10/2004



Im Rahmen einer Forschungskoooperation mit der Firma data process GmbH werden verschiedene Verfahren zum Klassifizieren und Suchen von Textdokumenten (auch mehrsprachig) untersucht. Dabei wurden empirische Studien mit unterschiedlichen Datensätzen zum Vergleich des aktuellen Standes der Forschung mit der SAP-Suchmaschine TREX durchgeführt. Weiterhin wurde ein Begriffssystem zur Strukturierung der Intranetdokumente automatisch extrahiert und mit dem bestehenden System verglichen und evaluiert. Dabei wurden Verfahren des Ontologielernens eingesetzt und weiterentwickelt, sowie auf ihre praktische Anwendbarkeit hin evaluiert. Das SAP-TREX-System konnte erfolgreich bei data process in Betrieb genommen werden.

Um den Austausch von Wissen innerhalb der K+S Gruppe weiter zu fördern, wurde in einer Bachelor-Arbeit die Webapplikation BibSonomy in das Intranet der K+S Gruppe integriert, sowie um eine Such- und Rankingfunktion erweitert. Dadurch ist es möglich, den Mitarbeitern sowohl eine zentrale Speicherung ihrer Bookmarks anbieten zu können, als auch mit Hilfe der Rankingfunktion die Suche nach wichtigen Dokumenten innerhalb des Unternehmens zu unterstützen. Eine Integration dieser Suche mit der SAP-Suchmaschine TREX ist geplant.

6.7 COMO - COncEpts and MOdels

DFG, 12/2005 – 11/2007

Das Projekt COMO ist ein deutsch-russisches Austauschprogramm zur Untersuchung begrifflicher und modelltheoretischer Strukturen in der Wissensverarbeitung. Dabei sollen 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.

Von besonderer Bedeutung sind dabei einerseits die praktischen Kommunikationsprobleme, die mit der Verwendung unterschiedlicher Semantiken (Ontologien) zusammenhängen. Andererseits sollen mit den in der begrifflichen Systemtheorie erarbeiteten allgemeinen Begriffsbildungen Granularitätsprobleme bei Systembeschreibungen in Verbindung mit modalen und temporalen Logiken untersucht werden.

6.8 Zukünftige Projekte

Während des Berichtszeitraums wurde das EU-Projekt „TAGora – Semiotic dynamics in on-line social communities“ akquiriert, dass zum 1. Juni 2006 an den Start geht. Inhalt des Projektes ist die Analyse des Wachstums von sozialen Netzen in sozialen Lesezeichensystemen.

Ausserdem haben wir mit dem Vorschlag, soziale Lesezeichensysteme zur Verbesserung von Internetsuchmaschinen einzusetzen, ein Microsoft-Forschungsstipendium erhalten, das in der zweiten Hälfte des Jahres 2006 startet. Hierüber werden wir im nächsten Jahresbericht ausführlicher berichten. Weitere Projektanträge sind derzeit in der Begutachtung.

7 Veranstaltungen

Das Fachgebiet Wissensverarbeitung hat im Juli 2005 die 13th International Conference on Conceptual Structures an der Universität Kassel ausgerichtet. Darüber hinaus haben Mitglieder des Fachgebiets in den letzten zwei Jahren zusammen mit Kollgen anderer Universitäten die folgenden wissenschaftlichen Workshops auf internationalen Tagungen als Program Co-Chairs organisiert:

- Semantic Web Personalization, Workshop at the 19th National Conference on Artificial Intelligence (AAAI 2004), San Jose, California, U. S. A, 25.-26. Juli 2004 (Hotho)
- Mining for and from the Semantic Web, Workshop at the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22. August 2004 (Hotho)
- Semantische Technologien für Informationsportale, Ulm, Deutschland im Rahmen der 34. Jahrestagung der Gesellschaft für Informatik, 23. September 2004 (Stumme)
- Workshop on Ontologies in Peer-to-Peer Communities, Workshop at ESWC 2005, Heraklion, Greece, 30. Mai 2005 (Schmitz)
- Learning in Web Search, Workshop at ICML 2005, Bonn, Deutschland, 7. August 2005 (Hotho)
- Foundational Aspects of Ontologies, Workshop at KI 2005, Koblenz, Deutschland, 11. September 2005 (Stumme)
- European Web Mining Forum 2005, Workshop at ECML/PKDD 2005, Porto, Portugal, 3. Oktober 2005 (Hotho, Stumme)
- Semantic Network Analysis. Workshop at ISWC 2005, Galway, Ireland, 7. November 2005 (Schmitz, Stumme)

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

7.1 13th International Conference on Conceptual Structures

Wie Wissensverarbeitung durch Computer unterstützt werden kann, war Gegenstand der 13th International Conference on Conceptual Structures, zu der wir vom 17. bis 22. Juli ca. 80 Teilnehmer an der Universität Kassel begrüßen konnten. Die Tagung ist Teil einer jährlichen Tagungsreihe, die seit 1993 abwechselnd in Europa, Australien und Nordamerika stattfindet. Der Schwerpunkt der Tagungsreihe liegt auf zwei Ansätzen: den Begrifflichen Graphen, die eine Formalisierung der mathematischen Logik zum Ziel haben, die sich (im Gegensatz etwa zur Prädikatenlogik) so dicht wie möglich am menschlichen Denken und an der Sprache orientieren soll; und der Formalen Begriffsanalyse, einer in Deutschland entstandenen Methode zum Erstellen von Begriffshierarchien basierend auf Datensammlungen.

Neun international bekannte eingeladene Vortragende brachten neue Impulse in die Diskussion:

- Rose Dieng, INRIA, Sophia Antipolis, France: *Conceptual Graphs for Semantic Web Applications*

- Norbert Fuchs, University of Zürich, Switzerland: *Knowledge Representation and Reasoning in (Controlled) Natural Language*
- Joseph Goguen, UCSD, San Diego, USA: *What is a Concept?*
- Ian Horrocks, Manchester, UK: *Applications of Description Logics: State of the Art and Research Challenges*
- Eduard Hovy, ISI, Los Angeles, USA: *Methodologies for Reliable Construction of Ontological Knowledge*
- Yannis Kalfoglou, University of Southampton, UK: *Using Formal Concept Analysis and Information Flow for modelling and sharing common semantics: lessons learnt and emerging issues*
- Aldo de Moor, STAR LAB, Brussels, Belgium: *Patterns for the Pragmatic Web*
- Ambjörn Naeve, KTH, Stockholm, Sweden: *Conzilla: a Conceptual Interface to the Semantic Web*
- George Paliouras, NCSR, Athens, Greece: *On the Need to Bootstrap Ontology Learning with Extraction Grammar Learning*

Die Tagung wurde von der Deutschen Forschungsgemeinschaft (DFG), der Gesellschaft zur Förderung des technischen und kaufmännischen Führungsnachwuchses in Nordhessen – Informatik (GFF-I) und dem Ernst-Schröder-Zentrum für Begriffliche Wissensverarbeitung (ESZ) finanziell unterstützt. Die Vorträge der Tagung wurden in einem Band der Lecture Notes in Artificial Intelligence (Springer) und einem Band der kassel university press dokumentiert.

7.2 Semantic Web Personalisation

Der Workshop wurde im Rahmen der 19th National Conference on Artificial Intelligence (AAAI 2004) in San Jose, California, USA am July 25-26, 2004 veranstaltet. Die Personalisierung von Webseiten erlaubt die zielgerichtete Präsentation von Information, die von Interesse für den Anwender sind. Dabei sollen diese möglichst aus dem bisherigem Nutzerverhalten bestimmt werden. Im Vordergrund des Workshops standen Einreichungen, die zeigen, wie man mit intelligenten Techniken bestehendes Wissen in Form von existierenden, extrahierten oder auch gelernten Ontologien ausnutzen kann, um die Ergebnisse der Web-Personalisierung zu verbessern. Ausgewählte Beiträge des Workshops wurden im „ACM TOIT Special Issue on Intelligent Techniques in Web Personalization“ veröffentlicht.

7.3 Mining for and from the Semantic Web

Ziel dieses Workshops war es, die Verbindung zwischen Forschern und Anwendern der beiden Forschungsbereiche Semantic Web und Data Mining/Maschinelles Lernen herzustellen und den fachlichen Austausch zu vertiefen. Dafür bot die Intl. Conference on Knowledge Discovery and Data Mining in Seattle, WA, USA einen idealen Rahmen. Auf dem Workshop, der am 22. August 2004 stattfand, wurden sowohl Methoden, die den Aufbau des „Semantic Web“ unterstützen, wie z.B. das Lernen von Ontologien, als auch die Analyse von semanti-

schen Daten diskutiert. Der Workshop wurde in enger Kooperation mit dem „Workshop on Web Mining and Web Usage Analysis (WebKDD)“ organisiert. Ausgewählte Beiträge wurden in einem Band der *Lecture Notes in Artificial Intelligence* (Springer) publiziert.

7.4 Semantische Technologien für Informationsportale

Im Rahmen der 34. Jahrestagung der Gesellschaft für Informatik veranstalteten wir zusammen mit Kollegen aus Karlsruhe am 23. September 2004 in Ulm diesen Workshop, auf dem Forscher und Praktiker aus den Bereichen Informationsportale, Semantische Technologien und Maschinelles Lernen zusammenkamen, um neue Methoden und Technologien für Informationsportale vorzustellen. Informationsportalen stellt sich zudem das Problem der Integration von Daten aus heterogenen Quellen in besonderem Maße, insbesondere wenn die Nutzer aktiv an der Informationserfassung mitwirken. Semantische Technologien bieten hierfür die Basis, Methoden des Maschinellen Lernens müssen für eine automatische oder computergestützte Strukturierung der Informationsbestände herangezogen werden. Die Beiträge wurden in einem Tagungsband der Reihe *Lecture Notes in Informatics* veröffentlicht.

7.5 European Web Mining Forum 2005

Das European Web Mining Forum wurde als Workshop der 16th European Conference on Machine Learning/9th European Conference on Principles and Practice on Knowledge Discovery in Databases am 3. Oktober 2005 in Porto, Portugal, abgehalten. Das Forum beschäftigte sich mit der Entdeckung von versteckten Mustern in der Struktur und der Benutzung des World Wide Webs. Ein besonderer Schwerpunkt war die Analyse von sozialen Netzen. Ausgewählte Ergebnisse des Workshops wurden in einem LNAI-Band des Springer-Verlags publiziert.

7.6 Foundational Aspects of Ontologies

Am 11. September 2005 wurde auf diesem Workshop der 28. Deutschen Tagung zur Künstlichen Intelligenz in Koblenz über philosophische und technische Grundlagen von Ontologien diskutiert. Zusammen mit Kollegen aus Karlsruhe und Dresden veranstalteten wir ein Vortragsprogramm, das Forscher aus Philosophie, Algebra, Logik und Kategorientheorie zusammenbrachte.

7.7 Workshop on Ontologies in Peer-to-Peer Communities

Dieser Workshop fand als Teil der 2nd European Conference on the Semantic Web am 30. Mai 2005 in Heraklion, Griechenland, statt. Ein Schwerpunkt des Workshops war die Selbstorganisation von Peers in Peer-to-Peer-Netzen analog zu menschlichen Gemeinschaften auf Basis von Semantic-Web-Technologie.

7.8 Learning in Web Search

Heutige Suchmaschinen sind in der Lage, riesige Datenmengen zu verarbeiten, wobei jedoch nur recht einfache Techniken bei der Analyse der Webdokumente zum Einsatz kommen. Im Workshop „Learning in Web Search“, der am 7. August 2005 in Bonn auf der International Conference on Machine Learning (ICML) stattfand, wurden neue Methoden und Techniken des maschinellen Lernens vorgestellt und diskutiert, die mehr und gezieltere Informationen aus Webdokumenten extrahieren. Dabei reichte das Spektrum der diskutierten Themen von Methoden zur begriffsbasierten Textklassifikation bis zu personalisierten Metasuchmaschinen. Ausgewählte Beiträge wurden im Special Issue „Learning in Web Search“ des Informatica-Journals publiziert.

7.9 Semantic Network Analysis

Dieser zusammen mit Kollegen aus Karlsruhe und Southampton auf der 4th International Conference on the Semantic Web durchgeführte Workshop brachte Forscher aus den Bereichen Semantic Web und Social Network Analysis zusammen. Wir diskutierten zusammen mit den Autoren und einem breiteren Publikum die Impulse, die die Forschungsrichtungen Communities of Practice, Wissensmanagement, Web Communities, and Peer-to-Peer durch das Zusammenwachsen der oben genannten Bereiche Semantic Web und Social Network Analysis derzeit erhalten.

Ein eingeladener Vortrag von Stefan Decker zeigte die Perspektiven für das persönliche Wissensmanagement auf dem eigenen PC auf, wenn dieser mit den Rechnern von Kollegen semantisch vernetzt wird.

8 Lehre

Das Fachgebiet Wissensverarbeitung bietet im Grundstudium die Vorlesung Datenbanken I an. Im Hauptstudium werden vertiefende Vorlesungen zu den Themen Wissensentdeckung (Knowledge Discovery), 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. Folgende Lehrveranstaltungen wurden angeboten:

Sommersemester 2004

- Vorlesung Datenbanken
- Übung Datenbanken
- Vorlesung Knowledge Discovery
- Übung Knowledge Discovery
- Seminar Semantic Web and Web Mining

- Projektseminar Data Mining mit Hintergrundwissen

Wintersemester 2004 - 2005

- Vorlesung Grundlagen der Wissensverarbeitung
- Übung Grundlagen der Wissensverarbeitung
- Vorlesung Objektorientierte und Deduktive Datenbanken
- Übung Objektorientierte und Deduktive Datenbanken
- Vorlesung Information Retrieval
- Übung Information Retrieval
- Projektseminar Semantic Web

Sommersemester 2005

- Vorlesung Datenbanken
- Übung Datenbanken
- Vorlesung Knowledge Discovery
- Übung Knowledge Discovery
- Vorlesung Formale Begriffsanalyse
- Übung Formale Begriffsanalyse
- Seminar Internet-Suchmaschinen (siehe auch den Bericht zur Hommingberger Gepardenforelle)
- Projektseminar Semantic Web

Wintersemester 2005 - 2006

- Vorlesung Semantic Web
- Übung Semantic Web
- Vorlesung Internet-Suchmaschinen
- Übung Internet-Suchmaschinen
- Projektseminar Semantic Web

Die folgenden ausgewählten Themen zeigen die Spannweite der in den Projektseminaren behandelten Themen auf:

- **Multilinguales Klassifizieren** (Dr. Andreas Hotho)
Das automatische Klassifizieren von Dokumenten in eine vorgegebene Menge von Klassen ist mittlerweile eine Standardanwendung im Bereich des Text Mining. Allerdings bedarf es einer größeren Zahl an Beispieldokumenten, die gerade beim multilingualen Klassifizieren auch in jeder der Sprachen vorliegen müssen. Im Projekt wurde das bestehende System "ConceptBooster" um die

Möglichkeit, die zwei weiteren Sprachen Französisch und Spanisch vorzuarbeiten, ergänzt. Um nicht immer Dokumente, die in allen Sprachen klassifiziert sind, zum Lernen des Modells zu benötigen, wurde anhand des FAO-Korpus die Möglichkeit der sprachübergreifenden Klassifikation auf Basis einer multilingualen Ontologie untersucht und evaluiert.

- **BibSonomy - Ein Folksonomy-System** (Dr. Andreas Hotho)
Die Konzeption und Implementierung eines sozialen webbasierten Bookmark- und Publikationsverwaltungssystems stand im Zentrum dieser Arbeit. Ziel war es, ein effizientes und voll funktionsfähiges Folksonomy-System in einer Tomcat/MySQL-Umgebung in Anlehnung an ähnliche Systeme wie z.B. Del.icio.us zu erstellen.
- **Ontology Learning in BibSonomy** (Dr. Andreas Hotho)
Die Erweiterung der Webapplikation BibSonomy um zwei Funktionalitäten stellen den Kern dieser Projektarbeit dar. Auf der einen Seite wurde eine Tag-Relationenverwaltung hinzugefügt, die es erlaubt, Beziehungen zwischen Tags auszudrücken. Als zweite Funktionalität wurde ein sogenanntes Fragensystem konzipiert und implementiert, welches dazu dient, dem Benutzer von BibSonomy Inkonsistenzen innerhalb seiner Tags, aber auch in seiner Tag-Hierarchie aufzuzeigen und bei der Beseitigung zu unterstützen.
- **Tagging von Dateien** (Dr. Andreas Hotho)
Im Zuge dieser Projektarbeit sollte das bestehende Bibsonomy-System um eine WebDAV-Schnittstelle erweitert werden. Über diese Schnittstelle können dann die im Bibsonomy-System gespeicherten BibTEX-Einträge heruntergeladen sowie allgemeine Dateien hoch- und heruntergeladen werden. Dies erlaubt die Nutzung von sozialen Bookmarkssystemen auch zur Speicherung von Dokumenten und damit zum Dokumentenmanagement.
- **Entwurf und Integration eines Item-Based Collaborative Filtering Tag Recommender Systems in Bibsonomy** (Dr. Andreas Hotho)
Die Projektarbeit beschäftigt sich mit der Erweiterung des sozialen Bookmarksystems Bibsonomy um automatisch erstellte Tagvorschläge. Kern der Arbeit ist das Ermitteln von individuell passenden Tagging-Vorschlägen für neue URL- und BibTeX-Einträge aus einem bisherigen Datenbestand von Taggings, um so den Benutzer beim Taggen der Ressourcen zu unterstützen und die Konvergenz des genutzten Vokabulars zu beschleunigen.
- **Rankingverfahren von Suchmaschinen** (Dr. Andreas Hotho)
Die Analyse von Einflußfaktoren auf die Ergebnisse aktueller Suchmaschinen stellt einen wesentlichen Schritt zum Verständnis aktueller Suchmaschinen dar und erlaubt deren Weiterentwicklung. Während bei der ersten Arbeit die manuelle Analyse von Einflußfaktoren im Vordergrund stand, beschäftigt sich die zweite Arbeit mit der Berechnung des Rankings für das Suchwort "Hommingberger Gepardenforelle" auf einem sehr großen Dokumentenkorpus.

- **Peer-Clustering mit Schwarmverhalten** (Christoph Schmitz)
In P2P-Netzen ist es günstig, wenn ähnliche Inhalte oder ähnliche Peers geclustert sind, d.h. topologisch nah beieinander liegen. Dieses Projekt untersucht, wie Inhalte in P2P-Netzen mit Hilfe von simuliertem Schwarmverhalten geclustert werden können. Dabei wird das Verhalten von natürlichen Schwärmen, etwa von Vögeln, Insekten oder Fischen, nachgebildet.
- **Implementierung und Erprobung eines Webcrawlers** (Christoph Schmitz)
In Zusammenarbeit mit anderen Projekten zum Thema Suchmaschinenoptimierung sollen große Mengen von Webseiten gesammelt werden, die sich mit einem bestimmten Thema befassen. Aufgabe ist es, einen bestehenden Crawler zu aktualisieren, einzelne Komponenten (z.B. Datenbankbindung, HTML-Parser) auszutauschen und einen verteilten Crawler zu implementieren. Der Crawler wurde auf einer großen Datenmenge erprobt, und lieferte Daten für andere Projekte.

9 Weitere Aktivitäten

9.1 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
26.01.2006	Prof. Dr. Rudi Studer	Semantic Web: Methoden, Werkzeuge und Anwendungen
15.12.2005	Prof. Dr. Josef Börcsök	Antrittsvorlesung: Grundlagen zum Entwurf von Rechnerarchitekturen für Systeme mit funktionaler Sicherheit
24.11.2005	Prof. Dr. Lothar Litz	Ambient Intelligence und die Anforderungen an Networked Control Systems
14.07.2005	John F. Sowa	Knowledge Signatures for Cognitive Modeling
30.06.2005	Prof. Dr. Clemens Jochum	IT vom Discounter - wie weit geht die Kommoditisierung?
28.04.2005	Prof. Dr. Wilhelm Schäfer	Ein UML-basierter Ansatz für die Modellierung, Verifikation und Implementierung der Informationsverarbeitung in mechatronischen Systemen

10.02.2005	Dr. Gerhard Bauch	Codierungs- und Übertragungsverfahren für den Mobilfunk
27.01.2005	Prof. Dr. Claudia Leopold	Antrittsvorlesung: Effizient und einfach parallel programmieren
12.01.2005	Prof. Dr. Kurt Geihs	Antrittsvorlesung: Antrittsvorlesung: Von Middleware zu Modelware
25.11.2004	Prof. Dr. Gunter Saake	Adaptive Informationssysteme
28.10.2004	Prof. Dr. Gerd Stumme	Antrittsvorlesung: Begriffliches Wissen und seine Verarbeitung
15.07.2004	Prof. Dr. Albert Zündorf	Antrittsvorlesung: Das UML-Projekt Fujaba: Vorgehensmodell, Sprache, Werkzeug
24.06.2004	Prof. Dr. Martina Zitterbart	Zero-Touch Networks: Chance oder Chaos
27.05.2004	Prof. Dr. Wolfgang Nejdil	Effizienter und sicherer Zugriff auf verteilte Daten im Semantic Web

9.2 Jour Fixe Informatik

Die zwölf Informatiker der Fachbereiche 16 und 17 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.

9.3 Akademische Selbstverwaltung

Prof. Dr. Gerd Stumme

- Mitglied im Diplom-Prüfungsausschuss Informatik
- Mitglied im Bachelor-Prüfungsausschuss Informatik
- Mitglied im Master-Prüfungsausschuss Informatik
- Entwurf der Master-Prüfungsordnung Informatik
- Stellvertretendes Mitglied im Fachbereichsrat Mathematik/Informatik
- Zweitmitglied im Fachbereich Elektrotechnik/Informatik

Dr. Andreas Hotho

- Mitglied im Bachelor-Prüfungsausschuss Informatik
- Mitglied im Master-Prüfungsausschuss Informatik

9.4 Gesellschaft für Informatik

Prof. Dr. Gerd Stumme

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

Dr. Andreas Hotho

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

9.5 Mitgliedschaft in Programmkomitees internationaler Konferenzen und Workshops

Prof. Dr. Gerd Stumme

- 12th International Conference on Conceptual Structures, Huntsville, Alabama, USA, July 19-23, 2004
- Member of the Editorial Board of the International Journal of Data Warehousing and Mining
- 1st Asia-Pacific Conference on Conceptual Modelling (APCCM '04), Dunedin, Neuseeland, 18.-22. Januar 2004
- 2nd International Conference on Formal Concept Analysis (ICFCA '04), Sydney, Australien, 23. -26. Februar 2004
- AAAI Workshop on Semantic Web Personalization at the 19th National Conference on Artificial Intelligence (AAAI 2004), San Jose, California, U. S. A, 25. -26. Juli 2004
- First IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI 2004), Toulouse, Frankreich, 22.-27. August 2004
- 6th International Conference on Data Warehousing and Knowledge Discovery, Zaragosa, Spanien, 1. -3. September 2004
- 15th European Conference on Machine Learning (ECML), Pisa, Italien, 20.-24. Sept. 2004
- 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), Pisa, Italien, 20. -24. Sept. 2004
- Workshop Statistical approaches to Web Mining of ECML/PKDD 2004, Pisa, Italien, 20. Sept. 2004
- Workshop Knowledge Discovery and Ontologies of the ECML/PKDD 2004, Pisa, Italien, 24. Sept. 2004
- 2004 IEEE/WIC/ACM International Conference on Web Intelligence (WI'04), Peking, China, 20. -24. Sept. 2004
- Workshop of Ontologies and their Applications, Sao Luis, Brasilien, 28. Sept. 2004
- Fachgruppentreffen Maschinelles Lernen, Wissensentdeckung, Data Mining, Berlin, 4.-6. Okt. 2004

- IADIS International Conference WWW/Internet, Madrid, Spanien, 6.-9. Okt. 2004
- 3rd International Semantic Web Conference (ISWC2004), Hiroshima, Japan, 7. -11. Nov. 2004
- 3rd Intl. Conference on Formal Concept Analysis, Lens, Frankreich, 14.-18. Febr. 2005
- IADIS Virtual Multi Conference on Computer Science and Information Systems, 11.-29. April 2005
- 14th International Conference on the World Wide Web (WWW 2005), Chiba, Japan, 10.-14. Mai 2005
- 7th International Conference on Data Warehousing and its Applications (DaWaK 2005), Kopenhagen, Dänemark, 22.-26. Aug. 2005
- VLDB Workshop on Ontology Based Techniques for Databases and Information Systems, Trondheim, Norwegen, 2.-3. Sept. 2005
- International Workshop on Web Semantics, Kopenhagen, Dänemark, 22.-26. Aug. 2005
- 3rd International Conference on Concept Lattices and Their Applications, Olomouc, Czech Republic, 7.-9. Sept. 2005
- 2005 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Compiègne, Frankreich, 19.-22. Sept. 2005
- 16th European Conference on Machine Learning/9th European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2005), Porto, Portugal, 3.-7. Okt. 2005
- 2nd International Workshop on Knowledge Discovery and Ontologies, Porto, Portugal, 7. Okt. 2005
- 4th International Semantic Web Conference, Galway, Irland, 6.-10. Nov. 2005
- 4th International Conference on Formal Concept Analysis, Dresden, 13.-17. Feb. 2006

Dr. Andreas Hotho

- Workshop on the Application of Language and Semantic Technologies to support Knowledge Management Processes, 8th October 2004
- 2004 IEEE/WIC/ACM International Conference on Web Intelligence (WI'04), Peking, China, 20. -24. Sept. 2004
- Workshop on "Statistical Approaches to Web Mining" SAWM'04, 20 September 2004, Pisa, Italy
- Knowledge Discovery and Ontologies (KDO-2004), Pisa, Italy, September 24, 2004

- Artificial Intelligence: Methodology, Systems, and Applications, 11th International Conference, AIMS 2004, Varna, Bulgaria, September 2-4, 2004
- Advances in Web Intelligence, Second International Atlantic Web Intelligence Conference, AWIC 2004, Cancun, Mexico, May 16-19, 2004.
- Data Warehousing and Knowledge Discovery, 6th International Conference, DaWaK 2004, Zaragoza, Spain, September 1-3, 2004.
- The Fifth International Conference on Knowledge Based Computer Systems (KBCS-2004), Hyderabad, India, 20-22 December 2004.
- Workshop on Web Mining and Web Usage Analysis (WebKDD), August 22, 2004, Seattle, Washington.
- Advances in Web Intelligence, Second International Atlantic Web Intelligence Conference, AWIC 2005, 6-9 June, 2005; Lodz, Poland, 2005
- The First International Conference on Advanced Data Mining and Applications, 22-24 July, Wuhan, China, 2005
- PIA 2005 – Workshop on New Technologies for Personalized Information Access, 24-25 July, 2005, Edinburgh, Scotland, UK
- ICML 2005 Workshop: 4th International Workshop on Chance Discovery: From Data Interaction to Scenario Creation, 7-11 August, Bonn, Germany, 2005
- Data Warehousing and Knowledge Discovery, 6th International Conference, DaWaK 2005, 22 - 26 August, Copenhagen, Denmark, 2005
- The 16th European Conference on Machine Learning (ECML) and the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD) will be co-located in Porto, Portugal, 3-7 October, 2005.
- 7. Workshop des GI-Arbeitskreis Knowledge Discovery (AK KD), Saarbrücken, 10-12 October, 2005
- 3rd Intl. Wordnet Conference, January, 22–26 January, Jeju Island, Korea, 2006
- Student Research Workshop at EACL 2006, 3-7 April, 2006, Trento, Italy, 2006

10 Publikationen

2004

- B. Berendt, Andreas Hotho, Gerd Stumme. Usage Mining for and on the Semantic Web. In H. Kargupta, A. Joshi, K. Sivakumar, Y. Yesha, editors, *Data Mining: Next Generation Challenges and Future Directions*, 461-481, AAAI Press, Boston, 2004.
- Bettina Berendt, Andreas Hotho, Dunja Mladenic, Maarten van Someren, Myra Spiliopoulou, Gerd Stumme, editors, *Web Mining: From Web to Semantic Web*. LNAI 3209, Springer, Heidelberg, 2004.

- Bettina Berendt, Andreas Hotho, Dunja Mladenic, Maarten van Someren, Myra Spiliopoulou, Gerd Stumme. A Roadmap for Web Mining: From Web to Semantic Web. In: Web Mining: From Web to Semantic Web, LNAI 3209, 1-22, Springer, Heidelberg, 2004.
- Stephan Bloehdorn, Andreas Hotho. Text Classification by Boosting Weak Learners based on Terms and Concepts. Proceedings of the Fourth IEEE Intl. Conference on Data Mining, 331-334, IEEE Computer Society Press, 2004.
- Stephan Bloehdorn, Andreas Hotho. Boosting for Text Classification with Semantic Features. Proceedings of the MSW 2004 workshop at the 10th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 70-87, 2004.
- Philipp Cimiano, Andreas Hotho, Gerd Stumme, Julien Tane. Conceptual Knowledge Processing with Formal Concept Analysis and Ontologies. In Peter Eklund, ed., Concept Lattices, LNAI 2961, 189-207, Springer, Heidelberg, 2004.
- Philipp Cimiano, Andreas Hotho, Steffen Staab. Comparing Conceptual, Partitional and Agglomerative Clustering for Learning Taxonomies from Text. Proceedings of the European Conference on Artificial Intelligence (ECAI'04), 2004.
- Philipp Cimiano, Andreas Hotho, Steffen Staab. Clustering Ontologies from Text. Proc. of the Conference on Lexical Resources and Evaluation, 2004.
- Marc Ehrig, Jens Hartmann, Christoph Schmitz. Ontologie-basiertes Web Mining. Workshop "Semantische Technologien für Informationsportale" (GI-Jahrestagung 2004), 2004.
- Peter Haase, Marc Ehrig, Andreas Hotho, Björn Schnizler. Personalized information access in a bibliographic peer-to-peer system. Proceedings of the AAAI Workshop on Semantic Web Personalization, 2004, 1-12. AAAI Press,
- Andreas Hotho. Clustern mit Hintergrundwissen. Disk 286. Akademische Verlagsgesellschaft Aka GmbH, Berlin, 2004.
- Andreas Hotho, York Sure, Lise Getoor, editors, International Workshop on Mining for and from the Semantic Web (MSW2004). Located at the 10th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining KDD 2004, 22nd August 2004, Seattle, WA, USA.
- Andreas Hotho, York Sure, Lise Getoor. A workshop report: mining for and from the Semantic Web at KDD 2004. SIGKDD Explorations (6)2:142-143, 2004.

- A. Maedche, K. Sattler, G. Stumme, editors, DBFusion 2002 Information Integration and Mining in Databases and on the Web. CEUR Workshop Proceedings, (124) CEUR-WS, Aachen, 2004.
- Bamshad Mobasher, Sarabjot Singh Anand, Bettina Berendt, Andreas Hotho, editors, Proc. Semantic Web Personalization. September 2004.
- Wolfgang Nejdl, Martin Wolpers, Wolf Siberski, Christoph Schmitz, Mario Schlosser, Ingo Brunkhorst, Alexander Löser. Super-Peer-Based Routing Strategies for RDF-Based Peer-to-Peer Networks. Journal of Web Semantics, (Special issue WWW 2003) 2004.
- Lars Schmidt-Thieme, Gerd Stumme, Ute Rusnak, Andreas Eberhart, editor(s), Semantische Technologien für Informationsportale. LNI 51, 158-212, Gesellschaft für Informatik, Bonn, 2004.
- Christoph Schmitz, Steffen Staab, Christoph Tempich. Socialisation in Peer-to-Peer Knowledge Management. Proc. International Conference on Knowledge Management (I-Know 2004), Graz, Austria, 2004.
- G. Stumme. Iceberg Query Lattices for Datalog. In Karl Erich Wolff, Heather D. Pfeiffer, Harry S. Delugach, editors, Conceptual Structures at Work: 12th International Conference on Conceptual Structures (ICCS 2004), LNAI 3127, 109-125, Springer, Heidelberg, 2004.
- Julien Tane, Christoph Schmitz, Gerd Stumme. Semantic resource management for the web: an e-learning application. Proc. 13th International World Wide Web Conference (Alternate Track Papers & Posters), 1-10, 2004.

2005

- Bettina Berendt, Andreas Hotho, Gerd Stumme. Semantic Web Mining and the Representation, Analysis, and Evolution of Web Space. In Vojtech Svatek, Václav Snasel, editors, Proc. of the 1st Intl. Workshop on Representation and Analysis of Web Space, 1-16, Technical University of Ostrava, 2005.
- Stephan Bloehdorn, Philipp Cimiano, Andreas Hotho, Steffen Staab. An Ontology-based Framework for Text Mining. LDV Forum - GLDV Journal for Computational Linguistics and Language Technology, (20)1:87-112, 2005.
- P. Cimiano, A. Hotho, S. Staab. Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis. Journal of Artificial Intelligence Research (JAIR), (24):305-339, AAAI Press, 2005.

- F. Dau, M.-L. Mugnier, G. Stumme, eds., *Conceptual Structures: Common Semantics for Sharing Knowledge*, 13th International Conference on Conceptual Structures, Proceedings. ICCS 3596, Springer, 2005.
- Frithjof Dau, Marie-Laure Mugnier, Gerd Stumme, editors, *Contributions to ICCS 2005*. Contributions to ICCS 2005, kassel university press, Kassel, 2005.
- B. Ganter, G. Stumme, R. Wille, editors, *Formal Concept Analysis – Foundations and Applications*. LNAI 3626, Springer, Heidelberg, 2005.
- Peter Haase, Andreas Hotho, Lars Schmidt-Thieme, York Sure. *Collaborative and Usage-Driven Evolution of Personal Ontologies*. In Asunción Gómez-Pérez, Jérôme Euzenat, editors, *Proc. 2nd European Semantic Web Conference*, LNAI 3532, 486-499, Springer, 2005.
- Peter Haase, Christoph Schmitz, York Sure, editors, *Ontologies in Peer-to-Peer Communities*. Heraklion, Greece, 2005.
- Pascal Hitzler, Carsten Lutz, Gerd Stumme, editor(s), *Foundational Aspects of Ontologies*. Fachberichte Informatik der Universität Koblenz-Landau 9-2005.
- Andreas Hotho. *Text Clustern mit Hintergrundwissen (Dissertationsbeschreibung)*. *Künstliche Intelligenz* (1):62-64, Böttcher IT Verlag, 2005.
- Andreas Hotho, Andreas Nürnberger, Gerhard Paaß. *A Brief Survey of Text Mining*. *LDV Forum - GLDV Journal for Computational Linguistics and Language Technology* (20)1:19-62, 2005.
- Lotfi Lakhal, Gerd Stumme. *Efficient Mining of Association Rules Based on Formal Concept Analysis*. *Formal Concept Analysis*, 180-195, 2005.
- Christoph Schmitz. *Towards Self-Organizing Communities in Peer-to-Peer Knowledge Management*. *Proc. ESWC 2005 Workshop on Ontologies in Peer-to-Peer Communities*, Heraklion, Greece, 2005.
- Christoph Schmitz. *Towards Content Aggregation on Knowledge Bases through Graph Clustering*. In Stefan Braß, Christian Goldberg, editors, *Proc. 17. GI-Workshop "Grundlagen von Datenbanken"*, Wörlitz, 2005.
- Gerd Stumme. *Ontology Merging with Formal Concept Analysis*. In Yannis Kalfoglou, W. Marco Schorlemmer, Amit P. Sheth, Steffen Staab, Michael Uschold, editors, *Semantic Interoperability and Integration*, IBFI 04391, Schloss Dagstuhl, Germany, 2005.

- Gerd Stumme, Bettina Hoser, Christoph Schmitz, Harith Alani, editors, Semantic Network Analysis. CEUR Proceedings, Aachen, 2005.

2006

- Peter Haase, Marc Ehrig, Andreas Hotho, Björn Schnizler. Personalized Information Access in a Bibliographic Peer-to-Peer System. In S. Staab, H. Stuckenschmidt, editors, Peer-to-Peer and SemanticWeb, Decentralized Management and Exchange of Knowledge and Information, 143-158, Springer, 2006.
- B. Hoser, A. Hotho, R. Jäschke, C. Schmitz, G. Stumme. Semantic Network Analysis of Ontologies. In Y. Sure, J. Domingue, editors, The Semantic Web: Research and Applications, LNAI 4011, 411-426, Springer, Heidelberg, 2006.
- Andreas Hotho, Robert Jäschke, Christoph Schmitz, Gerd Stumme. BibSonomy: A Social Bookmark and Publication Sharing System. Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures, 2006.
- Andreas Hotho, Robert Jäschke, Christoph Schmitz, Gerd Stumme. Information Retrieval in Folksonomies: Search and Ranking. Proceedings of the 3rd European Semantic Web Conference, 411-426, Budva, Montenegro, 2006.
- Christoph Schmitz, Andreas Hotho, Robert Jäschke, Gerd Stumme. Content Aggregation on Knowledge Bases using Graph Clustering. Proceedings of the 3rd European Semantic Web Conference, Budva, Montenegro, 2006.
- C. Schmitz, A. Hotho, R. Jäschke, G. Stumme. Kollaboratives Wissensmanagement. In T. Pellegrini, A. Blumauer, editors, Semantic Web – Wege zur vernetzten Wissensgesellschaft, 273-290, Springer, 2006.
- G. Stumme, A. Hotho, B. Berendt. Semantic Web Mining - State of the Art and Future Directions. Journal of Web Semantics (4)2:124-143, Elsevier, 2006.

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

- Gerd Stumme: Ontology Merging with Formal Concept Analysis. Dagstuhl Seminar on Semantic Interoperability and Integration. 20. September 2004
- Gerd Stumme: Begriffliches Wissen und seine Verarbeitung. Antrittsvorlesung, Universität Kassel. 28. Oktober 2004

- Gerd Stumme: Begriffliches Wissen in Peer-to-Peer-Netzen. Hertie-Stiftung, Frankfurt, 20. Januar 2005
- Gerd Stumme: Begriffliche Wissensverarbeitung als Angewandte Informatik. ErnstSchröderKolloquium, Darmstadt, 8. Juli 2005
- Andreas Hotho, Steffen Staab: Machine Learning and the Semantic Web. Tutorial at ICML 2005, Bonn 7. August 2005
- Gerd Stumme: Conceptual Knowledge Processing. Invited Talk at the VLDB Workshop on Ontology-Based Techniques for Databases and Information Systems. Oslo, 2. September 2005
- Gerd Stumme: Allgegenwärtige Wissensverarbeitung. Haydauer Hochschulgespräche. 29. September 2005
- Christoph Schmitz: Ontologien als Basis für P2P-Wissensmanagement. Doctoral Mentoring Programme der Third German Conference on Multi-Agent System Technologies (MATES 05), Koblenz, September 11-13, 2005
- Gerd Stumme: Jagd auf die Hommingberger Gepardenforelle – Suchmaschinen und Wissensverarbeitung im Internet. MINT Schülermesse, Kassel, 30. Nov. 2005
- Robert Jäschke: Information Retrieval in Folksonomies: Search and Ranking, L3S Seminar Talk, Hannover, 20. Januar 2006
- Gerd Stumme: Folksonomies und Soziale Netze. Informatik-Kolloquium, TU Clausthal-Zellerfeld, 25. Januar 2006

11 Presseschau

Das Fachgebiet Wissensverarbeitung hat seine Forschungsaktivitäten mit den Pressemitteilungen einer größeren Öffentlichkeit bekannt gemacht. Dies führte zu einigen Mitteilungen in regionalen und überregionalen Zeitungen.

Eine Karriere im Eiltempo

Neuer Informatik-Professor an der Universität Kassel – Förderung durch Hertie-Stiftung

Von Ralph-Michael Krum

KASSEL. Ein Auge ist himmelblau, das andere haselnussbraun. Doch diese nette Laune der Natur ist nicht das einzig Ungewöhnliche an Gerd Stumme. Stumme, Jahrgang 1967, studierte in Darmstadt und Bordeaux Mathematik. Nach seiner Promotion lehrte er als Privatdozent, Gast- und Vertretungsprofessor an den Hochschulen von Karlsruhe, Clermont-Ferrand und Magdeburg. Promotion mit 30, Habilitation mit 35. Nun ist der Mann mit dem offenen, freundlichen Gesicht und der zurückhaltenden Art neuer Professor für Informatik an der Kasseler Uni. Eine Karriere in Rekordzeit.

Die Wissensverarbeitung ist sein Fachgebiet. „Mir geht es darum, das Wissen aus Rechnern verfügbar und kom-

Wissen verfügbar und kommunizierbar machen

munizierbar zu machen.“ Soll heißen: Stumme arbeitet daran, Daten aus großen Datenbanken zu extrahieren, zu sortieren und diese durch einen Zusammenschluss gleichberechtigter Rechner vielen Menschen zur Verfügung zu stellen.

Dem normalen Anwender dürfte dies nicht unbekannt

vorkommen. Musikbörsen, illegale wie legale, arbeiten nach diesem Prinzip. Wer am Netz ist, kann sich Daten aus anderen vernetzten Rechnern herunterladen. Ein möglicher Effekt: Dozenten an anderen Universitäten und Fakultäten könnten so vom Wissen ihrer Kollegen profitieren und gleichzeitig die eigenen Erkenntnisse anderen zur Verfügung stellen.

Was Stumme von anderen unterscheidet, ist sein interdisziplinärer Ansatz. „Mein Ziel ist es, die Menschen zu unterstützen und ihnen lästige und zeitraubende Arbeiten abzunehmen, aber auch gleichzeitig die Grenzen dieser Entwicklung zu erkennen und zu definieren.“

Um dieses Ziel zu erreichen, arbeitet Stumme in engem Kontakt auch mit Professor Wille von der Uni Darmstadt zusammen, der auch die psychologischen Faktoren dieser Entwicklung durchdenkt. Möglich wurde Stummes Professur in

Kassel durch das Engagement der Gemeinnützigen Hertie-Stiftung, die die C4-Professur mit exakt 971000 Euro finanziert.

Allein im vergangenen Jahr flossen rund fünf Millionen Hertie-Euro hessenweit in wissenschaftliche Projekte. Das Besondere: Auch die allererste Förderung in Hessen vergab die Stiftung nach Kassel. Mit dieser knappen Million wird die Professur, die zudem eineinhalb Stellen für wissenschaftliche Mitarbeiter sowie

zweieinhalb Verwaltungsstellen umfasst, für fünf Jahre finanziert.

Danach übernehmen das Land Hessen und die Universität Kassel die Finanzierung. Dass Stumme ein ganz Schneller ist, macht deutlich, dass ein Mitarbeiter Stummes noch während seiner Ernennungs-Pressekonferenz dabei war, die nötigen Rechner zu verkabeln und für die Praxis fit zu machen.

Wenn das kein gutes Zeichen ist ...



Alle tragen ihn mit: Gerd Stumme wird von Uni-Präsident Prof. Dr. Rolf-Dieter Postlep, Dr. Roland Kaehlbrandt, Geschäftsführer der Hertie-Stiftung, Prof. Dr. Lutz Wegner vom Studiengang Informatik sowie Dekan Prof. Dr. Hans-Georg Rück schon jetzt auf Händen getragen. FOTO: KRUM

Stiftungsprofessor

Gerd Stumme ordnet das Wissen

KASSEL · 4. APRIL · RAP · Wie die wachsenden Mengen an Wissen geordnet und nutzbar gemacht werden können, das ist das Fachgebiet von Gerd Stumme, der jetzt eine Stiftungsprofessur für Wissensverarbeitung im Studiengang Informatik der Kasseler Universität übernommen hat. Der 37-Jährige hat in Darmstadt und Bordeaux Mathematik studiert und als Privatdozent sowie Gast- und Vertretungsprofessor an verschiedenen Hochschulen gelehrt. Ermöglicht wird die C4-Professur durch die Hertiestiftung, sie stellt für die ersten fünf Jahre eine Million Euro zur Verfügung, danach werden Uni und Land die Professur zahlen.

Die Wissensverarbeitung erforscht, wie Datenmengen verwaltet werden können. So sind Experten an der Entwicklung von Internet-Suchmaschinen beteiligt. Die neue Kasseler Professur wird sich auch der Frage widmen, wie Computer und Internet im Bildungswesen eingesetzt werden können. Der Studiengang Informatik ist 2001 von den Fachbereichen Elektrotechnik und Mathematik eingerichtet worden. Neben acht neuen und umgewidmeten Professuren sind drei Stiftungsprofessuren geschaffen worden: Wissensverarbeitung, Praktische Informatik/Software-Engineering und für Betriebssysteme.

KASSEL: INFORMATIK-KOLLOQUIUM · Aktuelle Ergebnisse zu Grundlagen und Anwendungen der Informatik präsentieren Wissenschaftlerinnen und Wissenschaftler jeweils am letzten Donnerstag im Monat bei einem Informatik-Kolloquium an der Kasseler Universität. Die Veranstaltungsreihe beginnt am Donnerstag, 27. Mai, mit einem Vortrag von Professor Wolfgang Nejdil (Titel: „Effizienter und sicherer Zugriff auf verteilte Daten im Semantic Web“). Die Veranstaltungen starten jeweils um 17 Uhr im Raum 0315 in der Wilhelmshöher Allee 73 in Kassel.

Neues Fachgebiet für die Informatik

Universität Kassel erhält Stiftungsprofessur für Wissensverarbeitung

Nur zufriedene Gesichter gab es beim Pressetermin am 2. April, zu dem die Universität Kassel und die Gemeinnützige Hertie-Stiftung ins Gästehaus der Universität gebeten hatten. Anlass war der Dienstantritt von Prof. Dr. Gerd Stumme auf der Stiftungsprofessur „Wissensverarbeitung“ am Fachbereich Mathematik/Informatik der Universität. Wissensverarbeitung – die Erforschung der Organisation und Anwendung von Wissen – gilt als eines der zukunftsträchtigsten Fachgebiete der Informatik. Dank des Engagements der Hertie-Stiftung, die die C4-Professur mit einer Summe von rund einer Million Euro über einen Zeitraum von fünf Jahren finanziert, konnte dieses Fachgebiet nun in den Lehr- und Forschungsbetrieb der Kasseler Hochschule aufgenommen werden.

Uni-Präsident Prof. Dr. Rolf-Dieter Postlep dankte der Hertie-Stiftung für ihr Engagement. „Die Integration der Wissensverarbeitung in Forschung und Lehre ist ein weiterer Meilenstein bei der Entwicklung der Informatik an der Universität Kassel“, sagte Postlep. Die fachliche Ausrichtung der Hertie-Professur habe sich bereits

als Treffer erwiesen, denn für die Zeit zwischen 2005 und 2008 wird Wissensverarbeitung einer von drei Förderungsschwerpunkten des Bundesministeriums für Bildung und Forschung (BMBF) sein. Das hatte ein Vertreter des BMBF bereits im Herbst auf dem Fakultätentag Informatik angekündigt.

„Wir unterstützen die Stiftungsprofessur, weil sie mit ihrem Schwerpunkt – der Anwendung neuer Technologien im Bildungswesen – an der Schnittstelle von Technik und Bildung angesiedelt ist“, sagte Dr. Roland Kaehlbrandt, Geschäftsführer der Gemeinnützigen Hertie-Stiftung, der größten privaten Stiftung in Hessen.

Im Mittelpunkt der Forschungsaufgaben der Stiftungsprofessur für Wissensverarbeitung steht die Entwicklung von Methoden und Techniken, die die Verwaltung großer, teils weit verstreuter Datenmengen erleichtern. So liefert die Wissensverarbeitung das Know-how für die Suchmaschinen im Internet. Vor dem Hintergrund stetig wachsender Informationsmengen gewinnt das Problem der Datenverwaltung aber auch für Handel, Banken und Versicherungen immer mehr an

Relevanz – nicht nur in technischer, sondern auch in personeller Hinsicht. Deshalb ist es eine weitere Aufgabe des Hertie-Professors, dringend benötigte Fachleute für die Wirtschaft auszubilden. Ferner widmet sich die Professur der Frage, wie Technologien wie Computer und Internet im Bildungswesen eingesetzt werden können.

Vorne in der Forschung

„Ich freue mich natürlich sehr über das mit der Berufung verbundene Vertrauen. Zugleich bin ich zuversichtlich, mit meiner Entscheidung für Kassel und sein junges Kollegium in der Informatik die richtige Wahl getroffen zu haben“, führte Prof. Stumme aus. „In engem Kontakt mit dem Stifter und auf der Basis bereits eingeworbener Drittmittel und vieler Kontakte werden wir die Universität Kassel mit Beiträgen zur Wissensverarbeitung – etwa dem Semantischen Web – und Entwicklungsarbeiten zum E-learning ganz vorne in der Forschungslandschaft platzieren.“

Bei so viel gleichgestimmter Energie, wie an diesem Tag erkennbar, kein unrealistisches Vorhaben – ein Glücksfall eben. *Lutz Wegner/p*



Auf Händen getragen: Der neue Professor für Wissensverarbeitung, Dr. Gerd Stumme, unterstützt von Uni-Präsident Postlep, Dr. Roland Kaehlbrandt, Hertie-Stiftung, sowie Prof. Dr. Lutz Wegner und Prof. Dr. Hans-Georg Rück, beide FB Mathematik/Informatik (von links).

Foto: publik

Frecher Fischzug mit falschen Forellen

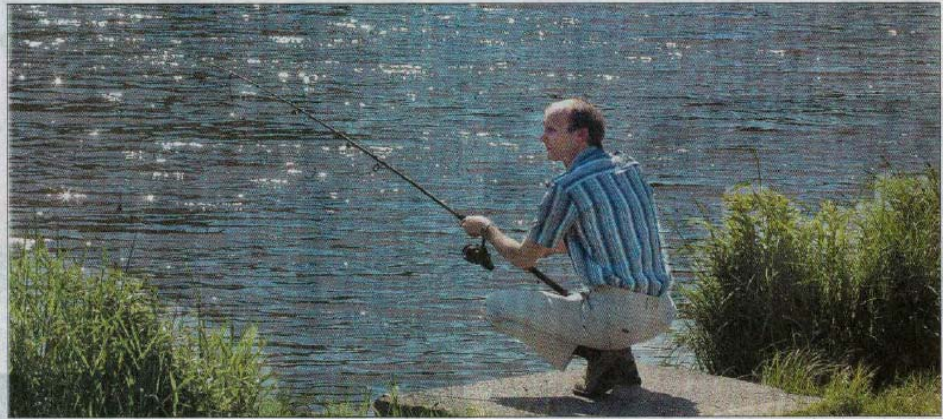
Suchmaschinen-Wettbewerb im Internet: Universität Kassel wetteifert um bestmögliche Platzierung

Von Alexander Röder

KASSEL. Kasseler Professoren werfen ihre Angel aus: Denn seit einigen Wochen sorgt die *Hommingberger Gepardenforelle* für große Aufregung. Millionenfach ging sie zuletzt weltweit ins Netz - ins virtuelle Netz, dem Internet. Und dabei ist dieser dicke Brocken schlichtweg eine Kopfgeld, ein Hirngespinnst.

Die Computerzeitschrift *c't* rief im April einen Wettbewerb aus, die unsinnige Wort-hülse *Hommingberger Gepardenforelle* mit Leben zu füllen. Ziel ist es, die eigene Internetseite in einer Suchmaschine bestmöglich, sprich: an erster Stelle zu platzieren. Nicht einfach, bedenkt man, dass ein Suchbegriff nicht selten tausende Treffer auflistet. Klar, dass auch die Uni Kassel an diesem ungewöhnlichen Fischfang teilnimmt.

Der Wettbewerb kam dabei genau zum richtigen Zeitpunkt. Denn das an der Uni Kassel um Prof. Dr. Gerd Stumme neu initiierte Fachgebiet Wissensverarbeitung veranstaltet derzeit ein Seminar zum Thema Suchmaschinen. „Wir können so unseren Studierenden deutlich machen, welche Relevanz die von uns vermittelte Theorie in der Praxis hat“, sagt Informatiker Christoph Schmitz. In der



Ob sie anbeißt? So wie hier mit einer Angelrute am Fuldaufer lauert der Kasseler Forscher Dr. Andreas Hotho geduldig im Internet, um die fiktive *Gepardenforelle* den Datenmeeren zu entreißen. FOTO: HERZOG

Praxis wird eine Internetseite in einer Suchmaschine als relevant, also als bedeutsam, eingestuft, je mehr Seiten durch Links auf sie verweisen, erklärt Dr. Andreas Hotho. Er ist wie Schmitz Wissenschaftler im Fachgebiet Wissensverarbeitung. Dabei haben sich Werbefirmen als besonders findig erwiesen, ihre Kunden in den Suchmaschinen weit oben zu platzieren. Beliebter

Trick: Link-Farmen, ein unauflösbarer Kreis aufeinander verweisender Seiten. Kein Wunder also, dass jene Werbefirmen auch im Wettbewerb um die *Hommingberger Gepardenforelle* gut vertreten sind. Aber auch die Uni Kassel scheint ein überaus engmaschig geknüpftes Netz ausgeworfen zu haben. Am 15. Mai, dem ersten Wertungsstichtag, lag die Seite auf Rang 8 - bei

mittlerweile mehr als einer Million Teilnehmern.

seien Studierende aufgefordert worden, sich am Wettbewerb zu beteiligen. Um die

**UNIKASSEL
VERSITÄT**

„Wissenschaftler sind im Internet sehr gut miteinander vernetzt“, begründet Hotho diesen Zwischenstand. Zudem

halten zu können, ruft Hotho zu Linkspenden auf.
► www.kde.cs.uni-kassel.de/lehre/ss2005/googlespam

public (Uni Kassel)

24. 1. 2006

Erfolgreiche Forellenjagd

Wettbewerb um Suchmaschinen-Ranking

Beim Rennen um die beste Suchmaschinenplatzierung mit dem Begriff „Hommingberger Gepardenforelle“ hat das Fachgebiet Wissensverarbeitung der Universität Kassel mit einer Gruppe Studierender einen sehr guten achten Platz belegt. Die Suchmaschinen Google und Yahoo führten einen Link zur Universität Kassel jeweils an achter Position ihrer Suchergebnisse auf, wenn nach „Hommingberger Gepardenforelle“ gesucht wurde. So jedenfalls zum letzten Stichtag des von der Computerzeitschrift *c't* ausgerufenen Wettbewerbs am 15. Dezember. Google verzeichnete insgesamt 3 030 000 und Yahoo 878 000 Teilnehmer beim Wettbewerb.

Weder den Ort Hommingberg noch eine Gepardenforelle gibt es tatsächlich. Beim Wettbewerb sollten die Teilnehmer am fiktiven Beispiel Fertigkeiten im Programmieren und Organisieren unter Beweis stellen, um den Auswahlalgorithmen der Suchmaschinen wichtig zu erscheinen. „Dies wird für Betreiber von Internetseiten im Konkurrenzkampf um die Aufmerksamkeit des Internetnutzers immer wichtiger. Ein beliebter

Trick sind Linkfarmen, d.h. große Ansammlungen von scheinbar unabhängigen Websites, die aufeinander verweisen und damit gegenseitig die Relevanz erhöhen“, erklärt Prof. Dr. Gerd Stumme, Hertie-Stiftungsprofessur für Wissensverarbeitung an der Universität Kassel.

Das Fachgebiet Wissensverarbeitung veranstaltet ein Seminar zum Thema Suchmaschinen. „Der Wettbewerb der *c't* kam für uns genau zum richtigen Zeitpunkt. Wir können so unseren Studierenden deutlich machen, welche Relevanz die von uns vermittelte Theorie in der Praxis hat“, stellt Dipl.-Inform. Christoph Schmitz fest. Die Studierenden lernen, welche Datenstrukturen und Algorithmen in Suchmaschinen angewandt werden. Um festzustellen, welche Seiten überhaupt angezeigt werden, wird zunächst ein Index erstellt – wie das Schlagwortverzeichnis eines Buches, nur mit dem Unterschied, dass der Index für das ganze World-Wide-Web (WWW) Millionen von Einträgen hat. Im Seminar werden hierzu effiziente Datenstrukturen diskutiert.

mwü/au



Für regelmäßig aktualisierte Inhalte haben auch die Teilnehmer des Seminars „Internet-Suchmaschinen“ des Fachgebietes Wissensverarbeitung an der Universität Kassel auf ihrer Seminar-Homepage gesorgt. Da es in dem Seminar ohnehin um die Analyse von Spam-Methoden gehen sollte, wurde die Wettbewerbsteilnahme kurzerhand zur praktischen Übung.

Die Seminar-Website zeigt aber auch anschaulich, dass von Google geschätzte, aktuelle Inhalte für den Surfer ohne jeden Wert sein können. Die Studenten und wissenschaftlichen Mitarbeiter haben das Programm SCLgen – An Automatic CS Paper Generator angeworfen [7] und sich automatisch ein pseudowissenschaftliches Nonsense-Paper generieren lassen. Das englischsprachige Dokument haben sie durch Babelfish ins Deutsche übersetzen lassen und anschließend so viele Substantive ersetzt, bis das Suchwort häufig genug vorkam. Ehrlicherweise enthält die URL ihrer Seite den Begriff googlespam.

Haupterfolgsrezept der Seminareilnehmer – neben der von Google vielen Uni-Websites zugebilligtem inhaltlich hohen Wertschätzung – ist die massive Verlinkung der Seite. Die Studenten haben ihre Kontakte spielen lassen und Webmaster anderer Institute um Linksetzungen gebeten. Auch die Uni-Pressestelle hat mitgespielt, eine Pressemitteilung veröffentlicht und für einen Link von der Universitäts-Homepage gesorgt. Selbst in der Lokalpresse wurde um Link-Spenden gebeten. Der Ertrag: Immerhin drei Backlinks von Sites mit Page Rank 6 und 18 5er-Backlinks und ein achter Platz bei Google.

c't 2006, Heft 1

Publik
Nr. 6
11.07.06

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 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. 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 aufhebungs- 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 Rechenverfahren) 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 „Hommingberger Gepardenforelle“ 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 instituts-internen Literaturverwaltung angeboten.

Das Fachgebiet Wissensverarbeitung unter www.kde.cs.uni-kassel.de/