

# Forschungsbericht 2006/2007

## Abteilung Datenbanken

Universität Leipzig, Institut für Informatik

Web: <http://dbs.uni-leipzig.de>

### Inhaltsüberblick

- Personelle Zusammensetzung
- Highlights
- Projektbeschreibungen
- Veröffentlichungen / Graduiierungsarbeiten
- Vorträge
- Mitgliedschaften in Gremien / Redaktionskollegien, Herausbergremien u. ä.

### Personelle Zusammensetzung

Univ.-Professor	Prof. Dr. Rahm, Erhard
Wiss. Mitarbeiter	Aumüller, David
Stipendiat (Graduiertenkolleg)	Golovin, Nick (bis 3/06)
Wiss. Mitarbeiter (BMBF)	Hartung, Michael
Sekretärin	Hesse, Andrea
Programmierer	Jusek, Stefan
Wiss. Mitarbeiter (DFG / IZBI)	Kirsten, Toralf
Stipendiatin (Graduiertenkolleg)	Köpcke, Hanna
Wiss. Mitarbeiter (Krebshilfe / IZBI)	Lange, Jörg (bis 9/06)
Stipendiatin (Graduiertenkolleg)	Maßmann, Sabine
Wiss. Mitarbeiter	Dr. Sosna, Dieter
Wiss. Mitarbeiter	Thor, Andreas
Stipendiatin	Weikum, Anna (bis 8/07)



Abt. Datenbanken im Okt. 2007. V.l.n.r.: T. Schnerwitzki (Student), E. Rahm, E. Tappert (Student), D. Sosna, S. Maßmann, D. Aumüller, H. Köpcke (vorn), T. Kirsten, M. Hartung, A. Thor, Q. Ao (Student)

## Highlights

Im Berichtszeitraum 2006/2007 sind folgende Ereignisse hervorzuheben:

- Im Berichtszeitraum wurden vier Dissertationen fertig gestellt bzw. verteidigt (Hai Hong Do, Toralf Kirsten, Andreas Thor, Holger Märtens)
- Im März 2006 wurde im Rahmen des MediGRID-Projekts in Kooperation mit dem OntoVerse-Projekt ein zweitägiger Workshop zum Thema „Ontologien im Grid“ durchgeführt.
- Im Jahr 2006 wurde ein BMBF-Innoregio-Projekt zur Einrichtung von E-Commerce-Plattformen für Musikinstrumente erfolgreich beendet
- Das Oberseminar der Abteilung fand im Juli 2006 sowie im Oktober 2007 bereits zum fünften bzw. sechsten Mal an der Uni-Außenstelle in Zingst/Ostsee statt.
- Die Abteilung zog im November 2006 wegen der Campus-Neubauung am Augustusplatz in ein Interimsgebäude in der Johannisgasse
- Aufgrund einer HBF-G-Bewilligung konnte die IT-Ausstattung der Abteilung 2007 durch neue Server- und Arbeitsplatzrechner modernisiert werden.
- Die Deutsche Forschungsgemeinschaft (DFG) bewilligte 2007 ein neues Projekt zu dynamischen Fusion verteilter Webdaten.



Bilder vom Workshop „Ontologien im Grid“ in Leipzig (März 2006)



Vorstellung des elektronischen Wörterbuchs Neugriechisch-Deutsch durch Dr. Sosna auf der Leipziger Buchmesse (März 2007)



Das Interimsgebäude in der Johannisgasse

## Projektbeschreibungen

### Themen

- Dynamische Integration von Webdaten
- Object Matching
- Schema- und Ontologie-Matching
- Schema- und Ontologieevolution
- MediGRID: Grid Computing für die biomedizinische Forschung - Modul Ontologie-Werkzeuge
- Semantisches kollaboratives Content Management
- E-Commerce im Musikinstrumentenbau
- Web-basiertes Lernen

### Dynamische Integration von Webdaten

E. Rahm, A. Thor, D. Aumüller

Für Webdaten müssen Datenintegrationsaufgaben z. T. in Echtzeit erfolgen, wie es in aktuellen interaktiven Webapplikation, sogenannten Mashups, gefordert ist. Solche Anwendungen vereinen dynamisch Inhalte verschiedener Datenquellen und Web-Services. Vorhandene Systeme unterstützen derzeit nur relativ einfache Integrationsansätze. Unser Framework, das auf unserem Peer-to-Peer-Datenintegrationsansatz iFuice aufbaut, nutzt Mappings zwischen verschiedenen Datenquellen, die z.B. in Form von Hyperlinks zur Verfügung stehen. Eine effektive Bündelung von Informationen verschiedener Quellen ist durch die Anwendung mächtiger Operatoren möglich, die nicht nur skriptgetrieben sondern auch interaktiv eingesetzt werden können. Für ein angenehmes Nutzererlebnis implementierten wir einen sequentiellen Verfeinerungsprozess mit dem erste Ergebnisse einer Datenintegration schnell visuell verfügbar sind und automatisch bzw. auf Abruf Schritt für Schritt im Hintergrund mittels komplexerer Strategien verfeinert und dem Nutzer angezeigt werden. Eine Reihe von Skriptoperatoren steht zur Verfügung um die Ausführung der verschiedenen Mappings und die Datenfusion zu steuern. Basierend auf der generischen Architektur werden mehrere domänenspezifische Integrationsysteme entwickelt, z.B. ein Online Citation Service zur Bestimmung von Zitierungszahlen und Einflussfaktoren von Journalen, Konferenzen und Autoren – basierend auf einer Menge einschlägiger bibliografischer Datenquellen.

### Object Matching

E. Rahm, H. Köpcke, A. Thor

Aufbauend auf der Datenintegrationsplattform iFuice wurden im Berichtszeitraum Studien zum Object Matching durchgeführt. Object Matching oder Objektkonsolidierung ist ein entscheidender Aspekt der Datenintegration. Es bezeichnet das Problem Instanzen in verschiedenen Datenquellen zu finden, die dasselbe Objekt der realen Welt beschreiben. Dazu wurde mit MOMA ein flexibles Framework zur Objektkonsolidierung auf Basis instanzbasierter Mappings entwickelt. Es erlaubt die Konstruktion sogenannter Match-Workflows, welche die Ausführung und Kombination verschiedener Matcher ermöglicht. Dadurch ermöglicht MOMA sowohl die Verwendung der Ähnlichkeit zwischen Attributen als auch die Verwendung kontextueller Information, d.h. Beziehungen zwischen Objekten, für das Matching. Dazu bietet MOMA verschiedene Operatoren zur Mapping-Kombination. Mit Hilfe realer Daten wurden verschiedene Match-Strategien entwickelt und evaluiert. Ein weiterer methodischer Arbeitsschwerpunkt lag in der Untersuchung von Möglichkeiten des Self-Tunings, um den erforderlichen Aufwand bei der Bestimmung geeigneter Match-Strategien zu reduzieren. Verschiedene trainingsbasierte maschinelle Lernverfahren wurden untersucht, um Matcher im Rahmen einer Match-Strategie möglichst automatisiert auszuwählen, zu kombinieren und zu konfigurieren. Für die erforderliche Auswahl von Trainingsdaten wurden mehrere Methoden realisiert, die eine automatisierte Auswahl ermöglichen. Die Self-Tuning Ansätze wurden anhand verschiedener Match-Aufgaben evaluiert.

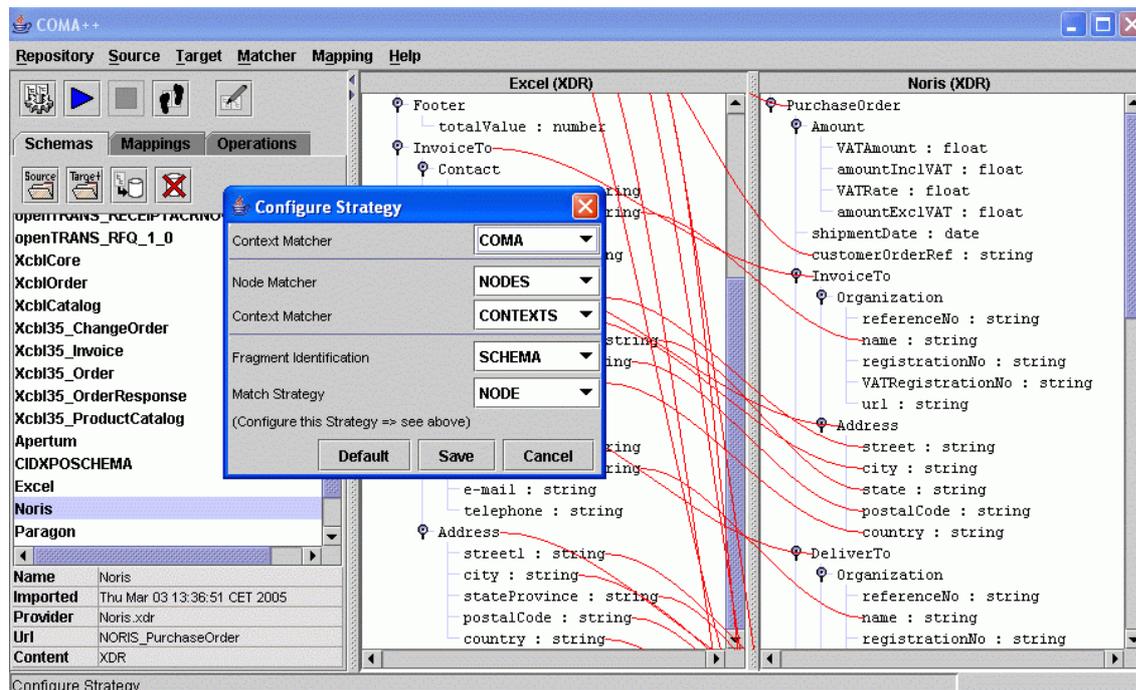
### Schema- und Ontologie-Matching

E. Rahm, T. Kirsten, S. Maßmann, A. Thor

Der früher erstellte Prototyp COMA++ zum Schema- und Ontologie-Matching benutzt hauptsächlich Schemainformationen. Zur Ausnutzung von Instanzdaten wurden nun neben einem erweiterten Import Match-Algorithmen umgesetzt, die aus den Instanzen bestimmte Constraints und den Inhalt an sich verwenden. Dabei erzielte der Einsatz eines Propagierungsansatzes auf die Instanz-Matchergebnisse eine Verbesserung, wobei eine Evaluierung der geeignetsten Konfiguration noch ansteht. Mit der Teilnahme an dem OAEI-Contest 2006 wurden erste Ergebnisse für das so erweiterte System gewonnen. Weitere Untersuchungen zum instanzbasierten Abgleich von Ontologien erfolgten für hierarchische Produktkataloge im eBusiness sowie für umfangreiche Ontologien der Bioinformatik. Dabei werden neben der Verwendung von Metadaten (z.B. Produktnamen) insbesondere die zuvor erzielten Ergebnisse einer Objektkonsolidierung zur Bestimmung korrespondierende Produktkategorien genutzt. Durch die Identifikation gleicher Produkte wurden Korrespondenzen der assoziierten

Kategorien bestimmt. Dabei wurden neben 1:1-Korrespondenzen auch Korrespondenzen zwischen Mengen von Kategorien ermittelt. Die entwickelten Verfahren wurden an Hand realer Daten evaluiert.

Die Abbildung zeigt die Oberfläche von COMA++.



## Schema- und Ontologieevolution

E. Rahm, M. Hartung, T. Kirsten, A. Weikum

Durch die große Verbreitung von schema-basierten XML-Dokumenten, Web-Services, Ontologien und Datenbanken stellt sich zunehmend die Problematik der Schemaevolution, d.h. Änderungen in solchen Schemas zu unterstützen. Die Realisierung neuer bzw. geänderter Anforderungen oder das Beheben von Designfehlern sind mögliche Gründe für eine Evolution von Schemas, welche traditionellerweise nur mit hohen manuellen Aufwand durchführbar sind. Neben der Evolution von Schemas in Datenbanken bzw. semistrukturierten Dokumenten, spielt ebenfalls die Evolution in Ontologien mehr und mehr eine entscheidende Rolle. Innerhalb des Berichtszeitraums wurde im speziellen die Evolution von XML-Schema betrachtet. Evolutionsoperatoren für XML-Schema wurden ermittelt, hauptsächlich Operatoren zum Verschieben von Schemaelementen, und bzgl. ihrer Umsetzbarkeit für die Migration von Instanzdaten untersucht. Weiterhin wurde eine Evolutionsunterstützung in das System für den Aufbau und das Management der D-Grid Ontologie (DGO) integriert. Außerdem wurde eine statistische Analyse der Evolution aktuell in Entwicklung befindlicher biomedizinischer Ontologien und zugehöriger Ontologie-Matchings begonnen.

## MediGRID: Grid Computing für die biomedizinische Forschung - Modul Ontologie-Werkzeuge

E. Rahm, M. Hartung

Das BMBF-Verbundprojekt MediGRID strebt die Entwicklung einer Grid-Middleware-Integrationsplattform und darauf aufsetzende eScience-Dienste für die biomedizinische Wissenschaft an. Für das Projekt wurde eine modulare Aufgabenverteilung gewählt: In den vier methodischen Projekt-Modulen (Middleware, Ontologie-Werkzeuge, Ressourcenfusion und eScience) entwickeln die entsprechenden Konsortialpartner schrittweise eine Grid-Infrastruktur. Sie berücksichtigen dabei insbesondere die Anforderungen der Grid-Nutzer aus dem biomedizinischen Umfeld, welche exemplarisch in den drei anwendungsorientierten Projektmodulen (Biomedizinische Informatik, Bildverarbeitung, Klinische Forschung) erarbeitet werden. Die erfolgreiche Nutzung von Diensten, Daten und physischen Ressourcen im Grid erfordert neben einer technischen auch eine semantische Interoperabilität zwischen den beteiligten Organisationen und Knoten. Ontologien stellen hierzu ein wesentliches Mittel dar, da sie helfen, unterschiedliche Ressourcen in semantisch eindeutiger Weise zu beschreiben und nutzbar zu machen. Das mit dem IZBI (Interdisziplinäres Zentrum für Bioinformatik) und IMISE (Prof. Löffler) kooperierende Teilmodul Ontologiewerkzeuge, eines von sieben Modulen des MediGRID-Konsortiums, strebt die Bereitstellung von Werkzeugen und Grid-Diensten für die Nutzbarmachung bereits existierender Ontologien für MediGRID-Anwendungen an. Im Berichtszeitraum wurde die Realisierung einer Grid-Middleware für die Integration unterschiedlicher biomedizinischer Ontologien in das MediGRID abgeschlossen.

Derzeit laufende Grid-Applikationen in MediGRID nutzen den realisierten Ontologieservice zum Zugriff auf Ontologieinformationen, welche in den Anwendungsszenarien benötigt werden. Weitere Ontologien können je nach Bedarf in der Community bzw. den Applikationen über die entwickelte Middleware in das Grid eingebunden werden. Ferner wurde eine kollaborative, wiki-artige Umgebung für den Aufbau und das Management einer Ontologie für das gesamte D-Grid (D-Grid Ontology DGO) entwickelt und implementiert. Neben den kollaborativen Fragestellungen und der Nutzerfreundlichkeit des Systems wurde insbesondere die Problemstellung der Schema- bzw. Ontologieevolution thematisiert (siehe Abschnitt Schema- und Ontologieevolution). Das System ist derzeit unter <http://buell.izbi.uni-leipzig.de/dgo> erreichbar.

## Semantische Annotierung von Webdaten

E. Rahm, D. Aumüller

Im Web stehen semistrukturierte Informationen zur Verfügung, die nur mittels Extraktion zur Weiterverarbeitung, Analyse und Darstellung sinnvoll genutzt werden können. Mittels Abgleich (Matching) zu speziell für diesen Zweck erstellten Wörterbüchern und z.B. Einträgen der Wikipedia-Enzyklopädie sollen Daten aus dem Web strukturiert werden. Als Beispielanwendung wird ein Werkzeug zur Extraktion von Detailinformationen aus semistrukturierten Metadaten wissenschaftlicher Publikationen entwickelt. Zur sinnvollen Darstellung der so erhaltenen Daten findet u.a. unser semantisches, kollaboratives Content Management System Caravela Verwendung. Mit diesem Framework wird das Ziel verfolgt, Informationen aus verschiedenen Datenquellen zusammenzuführen und zu annotieren. Kommentare, Querverweise und Kategorisierung anhand verschiedener Ontologien erfolgen kollaborativ, um den Wissensaustausch zu fördern. Die anhand eines im Vorfeld entwickelten semantischen Wiki-Systems gesammelten Erfahrungen bzgl. Einbringung von Metadaten, kontextabhängiger Navigation und Suche sowie semantischer Anfragemöglichkeiten fließen in das System mit ein.



## E-Commerce im Musikinstrumentenbau

E. Rahm, S. Maßmann

Die Nutzung des E-Commerce für den Vertrieb von Musikinstrumenten wird für KMU zunehmend bedeutsamer. Allerdings fehlt es in der Regel noch an Wissen und Unterstützung für den konsequenten und professionellen Einsatz in der Branche. Im Rahmen des Projektes wird sowohl der State-of-the-Art für die Webpräsenzen von Musikinstrumentenbauern als auch von Online Shops der Händler ausgewertet. Daraus resultieren Empfehlungen für den Inhalt und zu unterstützende Features für erfolversprechende E-Commerce-Lösungen. In Zusammenarbeit mit einer Internetagentur und Referenzunternehmen in der vogtländischen Musikinstrumentenregion wurde begonnen, zwei Plattformen mit Open Source Software zu realisieren und bestehende DV-Systeme einzubinden. Eine Plattform wird für den B2C- und die andere für den B2B-Bereich mit mehreren Unternehmen konzipiert. Sie werden erprobt und evaluiert und stehen nach Ende des Projektes zur weiteren Nutzung den Musikinstrumentenbauern zur Verfügung. Das Projekt ist Teil des BMBF InnoRegio-Vorhabens „Musicon Valley“ (<http://www.musiconvalley.de/>).

## Web-basiertes Lernen

E. Rahm, D. Sosna

Die im Einsatz befindliche E-Learning-Plattform LOTS (Leipzig Online-Test-System, <http://lots.uni-leipzig.de/>) zur Erstellung, Bearbeitung und automatischen Auswertung von Übungsaufgaben wurde weiter entwickelt. Im Berichtszeitraum wurden eine Reihe von Detailverbesserungen der Implementation sowie im Erscheinungsbild realisiert. Die Integration eines Tutorials zu XQuery ist in Vorbereitung. Es wurde ein Zustand erreicht und ein entsprechender rechtlicher Rahmen geschaffen, die es ermöglichen, LOTS als Open-Source-Projekt anderen interessierten Nutzern zur Verfügung zu stellen (<http://lots.sourceforge.net/>). Im Berichtszeitraum wurde ferner die Realisierung eines internet-gestützten elektronischen Wörterbuchs Neugriechisch-Deutsch in Zusammenarbeit mit dem Institut für Klassische Philologie fortgesetzt. Der Prototyp mit derzeit ca. 9000 Einträgen ist unter der URL <http://teiresias.uni-leipzig.de/> zu erreichen. Neben der üblichen Web-Navigation in den Suchergebnissen ist insbesondere die phonetische Suche hervorzuheben, die den speziellen Belangen des Neugriechischen Rechnung trägt. Davon ausgehend wurde ein Algorithmus erarbeitet, wie eine sprachspezifische phonetische Suche für andere indoeuropäische Sprachen entwickelt werden kann.



## Veröffentlichungen / Graduiierungsarbeiten

### Bücher

- Do, H.-H.: Schema Matching and Mapping-based Data Integration. Verlag Dr. Müller (VDM), ISBN 3-86550-997-5, 2006
- Jarke, M.; Seidl, T.; Quix, C.; Kensche, D.; Conrad, S.; **Rahm, E.**; Klamma, R.; Kosch, H.; Granitzer, M.; Apel, S.; Rosenmüller, M.; Saake, G.; Spinczyk, O. (Eds.): Datenbanksysteme in Business, Technologie und Web (BTW 2007), Workshop Proceedings ISBN 3-86130-929-7, Verlagshaus Mainz, Aachen 2007.



### Zeitschriften

- Do, H.-H.; **Rahm, E.**: Matching Large Schemas: Approaches and Evaluation. Information Systems, 32(6), 2007
- Köpcke, H.**; **Rahm, E.**: Analyse von Zitierungshäufigkeiten für die Datenbankkonferenz BTW. Datenbank-Spektrum, Heft 20, Feb. 2007
- Prufer, K.; Muetzel, B.; Do, H.-H.; Weiss, G.; Khaitovich, P.; **Rahm, E.**; Pääbo, S.; Lachmann, M.; Enard, W.: FUNC: a package for detecting significant associations between gene sets and ontological annotations. BMC Bioinformatics 2007, 8:41
- Rahm, E.**: Model Management. Datenbank-Spektrum, Heft 23, Nov. 2007
- Rahm, E.**; Bernstein, P. A.: An Online Bibliography on Schema Evolution. ACM Sigmod Record 35(4), Dec. 2006
- Rahm, E.**; **Kirsten, T.**; **Lange, J.**: The GeWare data warehouse platform for the analysis of molecular-biological and clinical data. Journal of Integrative Bioinformatics, 4(1):47, 2007
- Sax U.; Weisbecker, A.; Falkner, J.; Viezens, F.; Mohammed, Y.; **Hartung, M.**; , Bart, J.; Krefting, D.; Knoch, T.; Semler, S.C.: Auf dem Weg zur individualisierten Medizin – GRID-basierte Services für die Elektronische Patientenakte der Zukunft. E-HEALTH-COM, Jul-Aug, 2007

### Proceedings

- Aumueller, D.**; **Rahm, E.**: Caravela: Semantic Content Management with Automatic Information Integration and Categorization. Proc. European Semantic Web Conference (ESWC), LNCS 4519, 2007
- Böhme, T.; **Rahm, E.**: Parameterized XPath Views. Proc. British National Conference on Databases (BNCOD), LNCS 4587, 2007
- Böhme, T.; **Rahm, E.**; **Sosna, D.**: LOTS - Online-Training an der Universität Leipzig. In: Proceedings of the Workshop on e-Learning 2006, HTWK Leipzig, 10.-11. Juli 2006, ISSN 1610-1014
- Drumm, C.; Schmitt, M.; Do, H.-H.; **Rahm, E.**: QuickMig - Automatic Schema Matching for Data Migration Projects. Proc. ACM Conference on Information and Knowledge Management (CIKM), 2007
- Engmann, D.; **Massmann, S.**: Instance Matching with COMA++. Proc. BTW Workshop „Model Management und Metadaten-Verwaltung“, 2007.
- Hartung, M.**; **Rahm, E.**: A Grid Middleware for Ontology Access. Proc. German e-Science Conference, 2007
- Hartung, M.**: Automatisierte Umsetzung von komplexen XML-Schemaänderungen. Proc. BTW Workshop „Model Management und Metadaten-Verwaltung“, 2007
- Helmchen, Chr.; **Sosna, D.**: Online-Wörterbuch – eine Motivation für e-learning? Proc. Workshop on e-Learning, 2006
- Kirsten, T.**; **Lange, J.**; **Rahm, E.**: An integrated platform for analyzing molecular-biological data within clinical studies EDBT-Workshop Information Integration in Healthcare Applications, Springer LNCS 4254, 2006
- Kirsten, T.**; **Rahm, E.**: BioFuice: Mapping-based data integration in bioinformatics. Proc. of 3rd Int. Workshop on Data Integration in the Life Sciences (DILS), Springer LNCS 4075, 2006
- Kirsten, T.**; **Thor, A.**; **Rahm, E.**: Instance-based matching of large life science ontologies. Proc. Workshop on Data Integration in the Life Sciences (DILS), LNCS 4544, 2007
- Lange, J.**, **Kirsten, T.**, **Rahm, E.**: An integrated analysis platform for experimental and clinical data in modern cancer research studies. Proc. GMDS 2006
- Maßmann, S.**; Engmann, D.; **Rahm, E.**: COMA++: Results for the Ontology Alignment Contest OAEI 2006. Proceedings of the 1st International Workshop on Ontology Matching (OM) Collocated with the 5th International Semantic Web Conference (ISWC) Athens, Georgia, USA, November 5, 2006

- **Rahm, E.; Thor, A.; Aumueller, D.:** Dynamic Fusion of Web Data. Proc. XML Database Symposium (Xsym), LNCS 4704, 2007
- Sax, U.; Weisbecker, A.; Falkner, J.; Viezens, F.; Mohammed, Y.; **Hartung, M.;** Bart, J.; Krefting, D.; Knoch, T.A.; Semler, S.C.: Auf dem Weg zur individualisierten Medizin - Grid-basierte Services für die EPA der Zukunft. Proc. Telemed 2007 - Electronic Health Record und Gesundheitsportale, Berlin, 2007
- **Thor, A.; Aumueller, D.; Rahm, E.:** Data Integration Support for Mashups. Proc. Workshop on Information Integration on the Web (IIWeb), 2007
- **Thor, A.; Kirsten, T.; Rahm, E.:** Instance-based matching of hierarchical ontologies. Proc. Datenbanksysteme in Business, Technologie und Web (BTW), 2007.
- **Thor, A.; Rahm, E.:** MOMA - A Mapping-based Object Matching System. Proc. Conference on Innovative Data Systems Research (CIDR), 2007

### Technische Berichte und Poster

- **Hartung, M.;** Herre, H.; Loebe, F.; **Rahm, E.:** D-Grid Ontology - Semantic description of the D-Grid initiative. Poster for the D-Grid All Hands Meeting (AHM), Göttingen, 2007
- Sax U., Mohammed Y., Viezens F., Lingner Th., Morgenstern B., **Hartung M.;** Rienhoff O.: MediGRID - Medical Grid Computing. Poster für die ISC2006, Dresden.

### Dissertationen, Graduiierungsarbeiten

- **Do, H.:** Schema Matching and Mapping-based Data Integration. Dissertation, Univ. Leipzig, Jan. 2006
- **Kirsten, T.:** Data-Warehouse- und Mapping-basierte Datenintegrationsplattformen in der Bioinformatik. Dissertation, Univ. Leipzig, 2007
- **Thor, A.:** Automatische Mapping-Verarbeitung auf Webdaten. Dissertation, Univ. Leipzig, 2007 (eingereicht)
- **Märtens, H.:** Beiträge zur dynamischen Lastbalancierung in parallelen Datenbanksystemen. Dissertation, Univ. Leipzig, 2007 (eingereicht)
- Engmann, D.: Konzeption und Implementierung eines instanzbasierten Matching-Verfahrens. Diplomarbeit, Univ. Leipzig, 2007
- Gebhardt, S.: Extraktion lokaler Daten für iFuice. Diplomarbeit, Univ. Leipzig, 2007
- Heller, N.: Integration bibliographischer Daten mit iFuice. Diplomarbeit, Univ. Leipzig, 2007
- Helmchen, Chr.: Teiresias - Datenbankbasiertes Online-Wörterbuch Neugriechisch-Deutsch. Diplomarbeit, Univ. Leipzig, 2007
- Pickert, T.: Generischer Datenimport für ein Web-Content-Management System. Bachelorarbeit, Univ. Leipzig, 2007
- Scherf, M.: Extraktion semistrukturierter Daten für iFuice. Diplomarbeit, Univ. Leipzig, 2007
- Schliebe, T.: Erkennung von Nachbarschaftsbeziehungen räumlicher Strukturen am Beispiel von DNA - Daten, Univ. Leipzig, 2007
- Schmidt, O.: Implementierung der XPath-Anfragesprache für XML-Daten in RDBMS unter Ausnutzung des Nummerierungsschemas DLN. Diplomarbeit, Univ. Leipzig, 2007
- Schwipps, M.: Entwicklung eines komponentenbasierten Applikationsframeworks für iFuice. Diplomarbeit, Univ. Leipzig, 2007
- Stoll, S.: Terminologieorientiertes Website-Matching mit COMA++. Diplomarbeit, Univ. Leipzig, 2007



Nach der erfolgreichen Verteidigung: H.-H. Do (Januar 2006) und T. Kirsten (Dez. 2007) mit Doktorhut

## Vorträge

- Böhme, T.: LOTS - Online-Training an der Universität Leipzig. Workshop on e-Learning 2006, HTWK Leipzig, Juli 2006
- Böhme, T.: Parameterized XPath Views. British National Conference on Databases, Glasgow, Juli 2007
  
- Hartung, M.: A Grid Middleware for Ontology Access. German e-Science Conference, Baden-Baden, Mai 2007
- Hartung, M.: Automatisierte Umsetzung von komplexen XML-Schemaänderungen. BTW Workshop "Model Management und Metadaten-Verwaltung", Aachen, März 2007
- Hartung, M.: Integration und Nutzung von Ontologien in MediGRID. D-Grid Metadaten Workshop, Göttingen, März 2007
- Hartung, M.: Ontological Applications at MediGRID. eHealth Conference, Berlin, April 2007
- Hartung, M.: Ontologie- und Datenbankzugriff im Grid-Umfeld. Workshop Ontologien im Grid. Leipzig, März 2006
- Hartung, M.: Portlet-basierter Zugang zu Ontologien im Grid. Anwenderworkshop zur Anwendungs- und Portletprogrammierung. Göttingen, Juli 2006
  
- Kirsten, T.: An integrated platform for analyzing molecular-biological data within clinical studies. EDBT-Workshop Information Integration in Healthcare Applications, München, März, 2006
- Kirsten, T.: BioFuice: Mapping-based data integration in bioinformatics Workshop on Data Integration in the Life Sciences (DILS), Hinxton, Juli, 2006
- Kirsten, T.: Data-Warehouse- und Mapping-basierte Datenintegrationsplattformen in der Bioinformatik. Forschungskolloquium DBS der HU Berlin, Berlin, März, 2007
- Kirsten, T.: Instance-based matching of hierarchical ontologies. Datenbanksysteme in Business, Technologie und Web (BTW), Aachen, März 2007
- Kirsten, T.: Instance-based matching of large life science ontologies. Workshop on Data Integration in the Life Sciences (DILS), Philadelphia, Juni, 2007
- Kirsten, T.: Mapping-based Data Integration in Bioinformatics: The BioFuice System. NGFN-2 Database Meeting, Heidelberg, Mai 2006
  
- Maßmann, S.: Instance Matching with COMA++. BTW Workshop "Model Management und Metadaten-Verwaltung", Aachen, März 2007
- Maßmann, S.: Schema and Ontology Matching with COMA++. Forschungsseminar, HPI Potsdam, Juni 2007
  
- Rahm, E.: Datenintegration in den Lebenswissenschaften. Invited Talk, BMBF-Workshop „Ontologien im Grid“, Univ. Leipzig, März 2006
- Rahm, E.: Object Matching in Peer-to-Peer Data Integration. Invited Talk at 2nd Int. Workshop on Exchange and Integration of Data (EID), Brixen, Juni 2006
- Rahm, E.: Fusion verteilter Webdaten. Informatik-Kolloquium, Univ. Marburg, Juni 2006
- Rahm, E.: A Data Warehouse Platform for the Analysis of Molecular-biological and Clinical Data. Invited Talk at Symposium on Data Warehouse Technologies in Bioinformatics (DWTB), Wittenberg, Dez. 2006
- Rahm, E.: Architekturunterstützung sehr großer Datenbanken. Keynote ULDB-Konferenz (Ultra Large Data Bases), Frankfurt, Juli 2007
- Rahm, E.: Data Integration in Bioinformatics and Life Sciences. EDBT Summer School, Bozen, Sep. 2007
- Rahm, E.: Dynamic Fusion of Web Data: Beyond Mashups. Keynote XSym, Wien, Sep. 2007
- Rahm, E.: Next-Generation Schema Matching. Panel Presentation, XSym, Wien, Sep. 2007
  
- Sosna, D.: Online Wörterbuch - Eine Motivation für e-Learning? Workshop on e-Learning 2007, HTWK Leipzig, Juli 2007
  
- Thor, A.: MOMA - A Mapping-based Object Matching System. Conference on Innovative Data Systems Research (CIDR), Asilomar, Jan. 2007
- Thor, A.: Data Integration Support for Mashups. Workshop on Information Integration on the Web (IIWeb), Vancouver, Juli 2007

## **Mitgliedschaften in Gremien/Redaktionskollegien, Herausbergremien u.ä.**

### **Rahm, E.:**

- Vorstandsmitglied IZBI (Interdisz. Zentrum für Bioinformatik, Leipzig)
- Vorsitzender Prüfungsausschuß Informatik
- Stv. Sprecher des Fachbereichs "Datenbanken und Informationssysteme" der Gesellschaft für Informatik
- Sprecher des Arbeitskreises "Web und Datenbanken" der Gesellschaft für Informatik
- Advisory Board Europar Conference
- Mitherausgeber VLDB Journal
- Mitherausgeber der Zeitschrift "Datenbank-Spektrum"
- Mitherausgeber der Zeitschrift "Parallel Computing"
- Mitherausgeber der Buchreihe DISDB (Dissertationen zu Datenbanken) des infix-Verlages
- Programmkomitee verschiedener Konferenzen (u.a. SIGMOD2006, VLDB 2007, VLDB 2008, ICDE 2008 [Vice-PC-Chair], EDBT 2006, BTW 2007, DILS 2006, DILS 2008)

### **Kirsten, T.:**

- Programmkomitee der Konferenzen Intelligent Systems for Molecular Biology (ISMB) und European Conference on Computational Biology (ECCB) 2007