

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
Wiss. Mitarbeiter (DFG)	Endrullis, Stefan (seit 9/08)
Wiss. Mitarbeiter (DFG)	Groß, Anika
Wiss. Mitarbeiter (BMBF)	Hartung, Michael
Sekretärin	Hesse, Andrea
Programmierer	Jusek, Stefan
Wiss. Mitarbeiter (BMBF)	Müller, Robert (12/08-4/09)
Wiss. Mitarbeiter (IZBI/IMISE)	Dr. Kirsten, Toralf
Wiss. Mitarbeiter (DFG)	Köpcke, Hanna
Wiss. Mitarbeiter	Kolb, Lars (seit 10/09)
Wiss. Mitarbeiter (BMBF) / Stipendiat	Maßmann, Sabine
Externer Doktorand	Peukert, Eric (seit 5/09)
Wiss. Mitarbeiter	Dr. Sosna, Dieter
Wiss. Mitarbeiter	Dr. Thor, Andreas



Abt. Datenbanken im Juni 2009. V.l.n.r. (vorne): D. Aumüller, A. Groß, I. Landsmann (Studentin), H. Köpcke, S. Raunich (Gastwissenschaftler), L. Huang (Student), S. Horn (Student), D. Sosna, S. Maßmann; V.l.n.r. (hinten): M. Hartung, A. Thor, E. Rahm, C. Wartner (Student), S. Endrullis, T. Kirsten, L. Kolb

Highlights

Im Berichtszeitraum 2008/2009 sind folgende Ereignisse hervorzuheben:

- Im Berichtszeitraum wurden drei Dissertationen verteidigt bzw. fertiggestellt (Andreas Thor, Holger Märtens, Nick Golovin)
- Die Deutsche Forschungsgemeinschaft (DFG) bewilligte 2008 ein neues Projekt zur dynamischen Fusion verteilter Webdaten.
- In 2009 wurde ein weiteres DFG-Projekt zum Thema „Evolution von Ontologien und Mappings“ bewilligt.
- Im Jahr 2008 wurde ein Antrag zur semantischen Integration von Webdaten im Rahmen der ersten Phase der BMBF-ForMat (Forschung für den Markt im Team) Initiative bewilligt. Im April 2009 wurde hierzu ein Industrie-Workshop mit lokalen IT-Unternehmen erfolgreich durchgeführt.
- Im Anschluß an die erste Phase des Format-Projekts wurde 2009 der Folgeantrag im Umfang von ca. 1,5 Millionen Euro für die zweite Phase bewilligt. In der zweijährigen Projekt-Laufzeit sollen ab Januar 2010 Forschungsprototypen der Abteilung Datenbanken im Rahmen eines neuen Innovationslabors zur Webdatenintegration (WDI-Lab) für den Praxiseinsatz weiterentwickelt werden.
- Im Jahr 2009 wurde das MediGRID-Projekt (Modul Ontologiewerkzeuge) zur Einrichtung einer deutschlandweiten Grid-Infrastruktur für die Medizin und Lebenswissenschaften erfolgreich beendet.
- Mit SAP Research, Dresden, wurde eine Kooperation im Bereich Schema Matching begonnen.
- David Aumüllers Demonstration zum pdfmeat-Tool wurde mit dem „Best Demo Award“ auf der größten deutschsprachigen Datenbankkonferenz (BTW 2009) ausgezeichnet.
- Das Oberseminar der Abteilung fand im Juli 2008 sowie im Juni 2009 bereits zum siebten bzw. achten Mal an der Uni-Außenstelle in Zingst/Ostsee statt.
- In 2009 weilten zwei internationale Gäste im Rahmen mehrmonatiger Forschungsaufenthalte an der Abteilung: Salvatore Raunich (Univ. of Roma Tre/Basilicata) sowie Jorge Martinez (Univ. of Malaga; gefördert durch ein DAAD-Stipendium)
- Die DFG bewilligte Ende 2009 ein Stipendium für Dr. Andreas Thor für einen Forschungsaufenthalt an der University of Maryland, College Park, USA.



Bilder der Demonstration des pdfmeat-Tools sowie der Übergabe des Best Demo Awards während der BTW (März 2009)



Bilder vom Industrie-Workshop des Innovationslabors in Leipzig (April 2009)

Projektbeschreibungen

Themen

- Semantische Integration von Webdaten
- Dynamische Fusion verteilter Webdaten
- Objekt-Matching
- Affiliation Matching und semantische Annotierung von Webdaten
- Cloud-basiertes Objekt-Matching
- Schema- und Ontologie-Matching
- Evolution von Ontologien und Mappings
- MediGRID: Grid Computing für die biomedizinische Forschung - Modul Ontologie-Werkzeuge

Semantische Integration von Webdaten

E. Rahm, S. Maßmann, A. Thor



Anwendungen und Systeme, die einen integrierten Zugang zu Daten und Diensten verschiedenartiger Websites ermöglichen, werden zunehmend wichtiger. So sind bei Internetnutzern inzwischen etwa Portallösungen gefragt, welche einen Preisvergleich über unterschiedliche Produktarten verschiedenster Unternehmen (z.B. Elektronikartikel, Flugreisen) sowie daran angeschlossene Bestell- bzw. Buchungsmöglichkeiten anbieten. Die Lösung dieser Aufgabe erfordert eine semantische Integration der Webdaten. Dies erfolgt in derzeitigen Systemlösungen meist durch aufwändige Spezialprogrammierungen, die oft nur eine begrenzte Qualität erreichen. Zudem erfordert dabei jede Änderung an den Schemas oder Schnittstellen einzelner Anbieter eine aufwändige

Neuprogrammierung. Zur Beseitigung dieser Schwachstellen werden mächtige Werkzeuge zur weitgehenden Automatisierung von Aufgaben der semantischen Datenintegration benötigt, insbesondere zum Schema/Ontologie-Matching sowie zum Objekt-Matching. Dieses Projekt im Rahmen des BMBF ForMaT-Förderprogrammes untersuchte, wie zur Verbesserung der Situation in früherer Forschung entwickelte Prototypen für den Markteinsatz weiter entwickelt werden können. Diese Weiterentwicklungen sollen im Rahmen eines eigenen Innovationslabors, dem WDI-Lab, ab 2010 voran getrieben werden.

Dynamische Fusion verteilter Webdaten

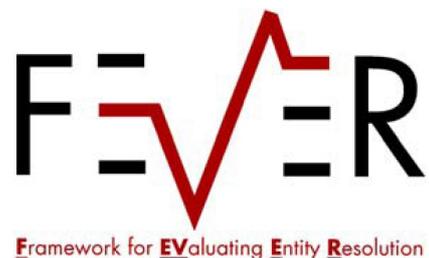
E. Rahm, S. Endrullis, H. Köpcke, A. Thor

Interaktive Webanwendungen, z.B. sogenannte Mashups, erfordern oft eine schnelle Fusion heterogener Daten zur Laufzeit. Eine solche dynamische Datenfusion ist jedoch angesichts der Heterogenität von Datenquellen sowie Qualitätsproblemen von Webdaten eine große Herausforderung. Wichtige Teilprobleme dabei sind die Bestimmung von Anfragestrategien bei der Nutzung von Webdatenquellen (z.B. Suchmaschinen, Datenbanken) sowie der semantisch korrekte Abgleich zwischen heterogenen Objektbeschreibungen (Objekt-Matching). Im Rahmen dieses DFG-Projektes wird ein flexibles, service-orientiertes Framework zur Unterstützung der dynamischen Fusion verteilter Webdaten entwickelt und evaluiert, das sowohl Query-Dienste für den Zugriff auf Datenquellen als auch Objekt-Matching-Dienste für den Abgleich von Objektbeschreibungen bereitstellt. Zusätzlich sollen Self-Tuning-Verfahren zur möglichst automatisierten Auswahl und Konfiguration von Query- und Match-Diensten untersucht werden. Innerhalb des Berichtszeitraums wurde dazu ein erstes Framework zur Definition und Evaluation von Anfragestrategien entwickelt und auf bibliografische Suchmaschinen angewandt. Des Weiteren konnten erste Self-Tuning-Verfahren zur Generierung effektiver und effizienter Anfragen entwickelt werden.

Objekt-Matching

E. Rahm, H. Köpcke, A. Thor

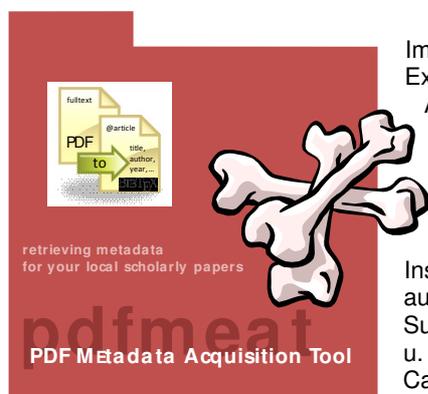
Objekt-Matching oder Objektkonsolidierung ist ein entscheidender Aspekt der Datenintegration. Es bezeichnet das Problem, Instanzen in verschiedenen Datenquellen zu finden, die dasselbe Objekt der realen Welt beschreiben. Im Berichtszeitraum wurde mit FEVER (Framework for EValuating Entity Resolution) ein flexibles Objekt-Matching-Evaluierungsframework geschaffen, das auf unseren bisherigen Frameworks MOMA (Mapping-based



Object Matching) und STEM (Self-tuning Entity Matching) aufbaut und ihre Funktionalität erweitert. FEVER erlaubt die Konstruktion sogenannter Match-Workflows, welche die Ausführung und Kombination verschiedener Matcher erlauben. Zur Konstruktion von Match-Workflows stehen verschiedene Matcher zur Verfügung, die sowohl Ähnlichkeiten zwischen Attributen als auch Kontextinformationen, d.h. Beziehungen zwischen Objekten, für das Matching nutzen. Verschiedene Operatoren erlauben die Kombination von Ergebnissen mehrerer Matcher. Um die Anzahl der notwendigen Vergleiche zwischen den zu matchenden Objekten zu reduzieren, stehen verschiedene sogenannte Blocking-Methoden zur Auswahl. Ein methodischer Arbeitsschwerpunkt lag in der Untersuchung von Möglichkeiten des Self-Tunings, um den erforderlichen Aufwand bei der Bestimmung geeigneter Match-Workflows zu reduzieren. FEVER stellt verschiedene trainingsbasierte maschinelle Lernverfahren zur Verfügung, um Matcher im Rahmen eines Match-Workflows 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. Ein weiterer methodischer Arbeitsschwerpunkt bestand darin, komparative Evaluierungen hinsichtlich Effektivität und Effizienz von verschiedenen Objekt-Matching Ansätzen mit unterschiedlicher Parameteranzahl durchzuführen. Dazu wurden Heuristiken entwickelt, die für einen Match-Ansatz bei gegebenem maximalem Aufwand eine optimale Parameterkonfiguration ermitteln.

Affiliation Matching und semantische Annotierung von Webdaten

E. Rahm, D. Aumüller



Im Web stehen semistrukturierte Informationen zur Verfügung, die mittels Extraktion zur weiteren Analyse genutzt werden können. Für bibliometrische Analysen, z.B. zu Publikations- und Zitierungshäufigkeiten, kommt neben der Bewertung einzelner Autoren auch die Herkunft der Papiere auf Ebene der Forschungsinstitutionen (Autor-Affiliations) eine wachsende Bedeutung zu. Die automatische Erfassung der Affiliations aus Volltextdokumenten sowie deren Weiterverwendung erfordert ein Matching heterogener Affiliation-Bezeichnungen, z. B. werden Institutsnamen auf recht unterschiedliche Arten abgekürzt. Hierzu wurden automatische Verfahren entwickelt, welche auch das Wissen von Suchmaschinen ausnutzen. Zur 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. Ein Service ermöglicht z. B. PDF-Volltexte mit bibliographischen Metadaten aus Webdatenquellen anzureichern. Dieser Abgleich führt zu einem Daten-Metadaten-Mapping, das neben der automatischen Erstellung einer Bibliographie (z. B. Bibtex) aus PDFs auch eine angereicherte Indexierung innerhalb Desktopsuchmaschinen ermöglicht.

Cloud-basiertes Objekt-Matching

E. Rahm, L. Kolb

Cloud Computing bezeichnet u.a. die Bereitstellung von IT-Infrastruktur durch externe Dienstleister. In den letzten Jahren hat dieses Konzept die IT-Welt massiv verändert. Durch die Möglichkeit, bei Bedarf in sehr kurzer eine große Menge von Rechenleistung, Speicherplatz und Bandbreite ohne Vorabinvestitionen nutzen zu können, entsteht die Illusion unendlicher on-demand verfügbarer Ressourcen. Gleichzeitig steigt in Unternehmen die Menge der zu verwaltenden und zu analysierenden Daten stetig an. Dies wird am Beispiel von Social Networks, wie z.B. Facebook deutlich, zu deren Datenbasis nach eigenen Angaben jeden Tag ca. vier TB neu hinzukommen. Die Verwaltung solcher Datenmengen bedingt eine verteilte Speicherung und Auswertung in Clustern mit Tausenden von einzelnen Knoten. Als Forschungsschwerpunkt wurde begonnen, die Nutzung des MapReduce-Konzepts zur automatischen Parallelisierung großer Objekt-Matching- Aufgaben zu untersuchen. Ausgangspunkt hierfür sind Match-Workflows, die mit dem an der Abteilung entwickelten FEVER-Framework entwickelt wurden. Durch die Erweiterung von FEVER um eine MapReduce-Schnittstelle, konnte bereits die sehr zeitintensive Anwendung von maschinellen Lernverfahren über mehrere Rechner parallelisiert werden.

Schema- und Ontologie-Matching

E. Rahm, A. Groß, M. Hartung, T. Kirsten, S. Maßmann, A. Thor, E. Peukert

Der früher erstellte Prototyp COMA++ zum Schema- und Ontologie-Matching wurde um ein Instanz-basiertes Matching erweitert, wozu Attributwerte sowie Constraints verwendet werden. Die Motivation für diese Erweiterung ist, dass die unterstützten Metadaten-Ansätze teilweise Schwierigkeiten im Umgang mit semantischer Heterogenität, wie z.B. mehrdeutigen Konzeptnamen und unterschiedlichen Konzeptgranularitäten, haben. Erste Evaluationen erfolgten mit Hilfe des OAEI-Benchmarks sowie von Produktkatalogen im eBusiness. Weitere Studien evaluierten nun bis zu welchem Grad diese Technik zum automatischen Berechnen von Directory-

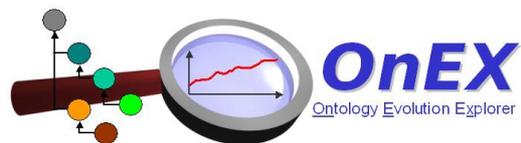
Mappings verwendet werden kann. Dabei wurden spezielle Instanz-basierte Matchtechniken vorgestellt, welche die URL, den Namen und die Beschreibung von kategorisierten Webseite verwenden. Die Instanz-basierten Ansätze wurden unter der Anwendung verschiedener Ähnlichkeitsmaße und in Kombination mit Metadaten-basierten Ansätzen anhand von realen Daten evaluiert. Zur Optimierung von Match-Workflows wurde ein erster regelbasierter Ansatz entwickelt und untersucht, mit dem die sequentielle Ausführungsreihenfolge von Matchern und Selektionsschritten angepaßt werden kann.

Im Bereich der Lebenswissenschaften wurde insbesondere die Stabilität von erstellten Ontologiemappings innerhalb mehrerer Versionen untersucht. Dazu wurden erste Metriken zur Selektion stabiler Match-Korrespondenzen konzipiert und an realen Daten evaluiert. Mit Hilfe dieser und weiterer Stabilitätsmetriken soll die Robustheit verschiedener Matchverfahren bezüglich der Evolution getestet, verglichen und zur Optimierung der Verfahren genutzt werden. Die Metriken wurden in den Prototyp GOMMA (Generic Ontology Matching and Mapping Analysis) integriert, der insbesondere eine Versionsverwaltung von Ontologien und Mappings sicherstellt. Unter Nutzung verschiedener Match-Workflows wurden Mappings zwischen heterogenen Anatomie-Ontologien erstellt. Die Mappings sollen künftig mit Hilfe des OAEI-Benchmarks evaluiert werden. Darüber hinaus dient das GOMMA-System zur Suche nach Duplikaten von Personendaten, die pseudonymisiert in unterschiedlichen deutschlandweiten, multizentrischen Studien aufgenommen wurden. Die Aufdeckung und Bereinigung von Duplikaten dient als wichtiger Schritt bei der integrierten Datenversorgung und ist Voraussetzung für eine hochqualitative Analyse. Hierzu wurden mehrere Algorithmen zur Ähnlichkeitssuche entworfen und realisiert, die in zukünftigen Arbeiten evaluiert werden sollen.

Evolution von Ontologien und Mappings

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

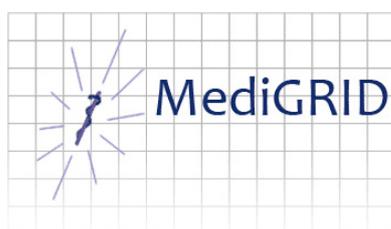
Ontologien werden in verschiedenen Wissenschaftsdisziplinen sowie in kommerziellen Anwendungen zur einheitlichen und semantischen Annotation von Objekten verwendet. Die oft rasante Entwicklung in den Anwendungsbereichen spiegelt sich auch in der Zunahme von Ontologien und ihrer beständigen Anpassung an neue Erkenntnisse und Erfordernisse wider. Das DFG-Projekt "Evolution von



Ontologien und Mappings" strebt die Konzeption und Realisierung eines generischen Frameworks für Evolutionsuntersuchungen von Ontologien und Mappings an. In einer eingehenden Analyse wurde die quantitative Evolution von Ontologien, Annotationen und Ontologie-Mappings vergleichend untersucht. Um derartige Evaluierungen für große, häufig geänderte Ontologien effizient und systematisch durchzuführen, wurde die GOMMA-Infrastruktur entwickelt. Diese Infrastruktur wird von der Webapplikation OnEX (www.izbi.de/onex) zur Analyse der Evolution von biomedizinischen Ontologien verwendet. Darüber hinaus fanden erste Untersuchungen zur Stabilität von Annotations- und Ontologiemappings statt. Zukünftig werden weiterhin Forschungsarbeiten zur Erkennung interessanter Ontologieregionen und zur Bestimmung des vollständigen und semantisch reichhaltigen Unterschieds zweier Ontologieversionen verfolgt. Außerdem werden die Auswirkungen von Evolution auf abhängige Algorithmen (z.B. Match-Algorithmen) und Applikationen detailliert untersucht.

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

E. Rahm, M. Hartung



Das im Februar 2009 erfolgreich abgeschlossene 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, verfolgte die Bereitstellung von Werkzeugen und Grid-Diensten für die Nutzbarmachung bereits existierender Ontologien für MediGRID-Anwendungen an. Während der Projektlaufzeit wurde die Realisierung einer Grid-Middleware für die Integration unterschiedlicher biomedizinischer Ontologien in das MediGRID verfolgt. Laufende Grid-Applikationen in MediGRID nutzen realisierte Ontologieservices 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. Das System ist unter <http://aprilia.izbi.uni-leipzig.de/dgo> erreichbar.

Veröffentlichungen / Graduierungsarbeiten

Zeitschriften

- **Aumüller, D.; Thor, A.:** Mashup-Werkzeuge zur Ad-hoc-Datenintegration im Web. *Datenbank-Spektrum*, 8. Jahrgang, Heft 26, Sep. 2008
- Bornmann, L.; Marx, W.; Schier, H.; **Rahm, E.; Thor, A.;** Daniel, H.-D.: Convergent validity of bibliometric Google Scholar data in the field of chemistry. *Journal of Informetrics*, Volume 3, Issue 1, Pages 27-35, January 2009
- **Hartung, M.; Kirsten, T.; Groß, A.; Rahm, E.:** OnEX: Exploring changes in life science ontologies. *BMC Bioinformatics*, 10:250, 2009
- **Hartung, M.;** Loebe, F.; Herre, H.; **Rahm, E.:** Management of Evolving Semantic Grid Metadata Within a Collaborative Platform. *Information Sciences 2009 (In Press)*
- **Köpcke, H.; Rahm, E.:** Frameworks for entity matching: A comparison. *Data & Knowledge Engineering 2009 (In Press)*
- Krefting, D.; Bart, J.; Beronov, K.; Dzhimova, O.; Falkner, J.; **Hartung, M.;** Hoheisel, A.; Knoch, T.A.; Lingner, T.; Mohammed, Y.; Peter, K.; **Rahm, E.;** Sax, U.; Sommerfeld, D.; Steinke, T.; Tolxdorff, T.; Vossberg, M.; Viezens, F.; Weisbecker, A.: MediGRID: Towards a user friendly secured grid infrastructure. *Future Generation Computer Systems*, Volume 25, Issue 3, Pages 326-336, March 2009
- Marz, M.; **Kirsten, T.;** Stadler, P.F.: Evolution of Spliceosomal snRNA Genes in Metazoan Animals. *Journal of Molecular Evolution*, Springer New York, 2008
- **Massmann, S.:** Ontologie-Matching von Produktkatalogen. *Datenbank-Spektrum*, 8. Jahrgang, Heft 24, Feb. 2008

Proceedings

- **Aumüller, D.:** Towards web supported identification of top affiliations from scholarly papers. *Proc. of 13. GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web (BTW) , GI Lecture Notes in Informatics P-144, 2009*
- **Aumüller, D.:** Retrieving metadata for your local scholarly papers. *Proc. of 13. GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web (BTW), GI Lecture Notes in Informatics P-144, 2009*
- **Aumüller, D.; Rahm, E.:** Web-based Affiliation Matching. *Proc. Intl. Conference on Information Quality (ICIQ), 2009*
- **Endrullis, S.; Thor, A.; Rahm, E.:** Evaluation of Query Generators for Entity Search Engines. *Proc. of Intl. Workshop on Using Search Engine Technology for Information Management (USETIM), 2009*
- **Groß, A.; Hartung, M.; Kirsten, T.; Rahm, E.:** Estimating the Quality of Ontology-based Annotations by Considering Evolutionary Changes. *Proc. of 6th Intl. Workshop on Data Integration in the Life Sciences (DILS), Springer LNCS (Lecture Notes in Computer Science) 5647, 2009*
- **Hartung, M.:** Management von Ontologien in den Lebenswissenschaften. *Tagungsband zum 20. GI-Workshop über Grundlagen von Datenbanken (GvD), 2008*
- **Hartung, M.; Kirsten, T.; Rahm, E.:** Analyzing the Evolution of Life Science Ontologies and Mappings. *Proc. 5th Intl. Workshop on Data Integration in the Life Sciences (DILS), Springer LNCS (Lecture Notes in Computer Science) 5109, 2008*
- **Hartung, M.;** Loebe, F.; Herre, H.; **Rahm, E.:** A Platform for Collaborative Management of Semantic Grid Metadata. *Proc. of 2nd Intl. Symposium on Intelligent Distributed Computing (IDC), Springer SCI (Studies in Computational Intelligence) 162, 2008*
- **Kirsten, T.; Hartung, M.; Groß, A.; Rahm, E.:** Efficient Management of Biomedical Ontology Versions. *Proc. 4th Intl. Workshop on Ontology Content (Part of the OTM Conferences & Workshops), Springer LNCS (Lecture Notes in Computer Science) 5872, 2009*
- **Köpcke, H.; Rahm, E.:** Training Selection for Tuning Object Matching. *Proc. of the 6th Intl. Workshop on Quality in Databases and Management of Uncertain Data (QDB/MUD), 2008*
- **Köpcke, H.; Thor, A.; Rahm, E.:** Comparative evaluation of entity resolution approaches with FEVER. *Proc. 35th Intl. Conference on Very Large Databases (VLDB), Demo, 2009*

- **Massmann, S.; Rahm, E.:** Evaluating Instance-based Matching of Web Directories. *Proc. 11th Intl. Workshop on the Web and Databases (WebDB), 2008*
- **Peukert, E.;** Berthold, H.; **Rahm, E.:** Rewrite Techniques for Performance Optimization of Schema Matching Processes. *To appear in Proc. of 13th Intl. Conference on Extending Database Technology (EDBT)*
- **Rahm, E.:** Comparing the Scientific Impact of Conference and Journal Publications in Computer Science. *Proc. Academic Publishing in Europe (APE), 2008*
- **Rahm, E.:** Logging and Recovery. Definitional Entry for Encyclopedia of Database Systems, *Springer 2009*
- **Rahm, E.:** Instance-based matching of Large Ontologies. *Tagungsband zum 21. GI-Workshop über Grundlagen von Datenbanken (GvD), 2009 (invited contribution)*
- **Thor, A.; Hartung, M.; Groß, A.; Kirsten, T.; Rahm, E.:** An Evolution-based Approach for Assessing Ontology Mappings - A Case Study in the Life Sciences. *Proc. of 13. GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web (BTW), GI Lecture Notes in Informatics P-144, 2009*

Technische Berichte und Poster

- Bermudez-Santana, C.; Attolini, C.S.-O.; **Kirsten, T.;** Sonja Prohaska, S.; Steigele, S.; Engelhardt, J.; Stadler, P.F.: tRNA Clusters. *Poster at German Conference on Bioinformatics (GCB), 2009*
- **Groß, A.; Hartung, M.; Kirsten, T.; Rahm, E.:** Evolution-based analysis of functional protein annotation. *7th Leipzig Research Festival for Life Sciences, 2008*
- **Hartung, M.; Kirsten, T.; Groß, A.; Rahm, E.:** Exploring changes in life science ontologies with OnEX. *Poster at 6th Intl. Workshop on Data Integration in the Life Sciences (DILS), 2009*

Dissertationen, Graduiierungsarbeiten

- **Thor, A.:** Automatische Mapping-Verarbeitung auf Webdaten. Dissertation, Univ. Leipzig, 2008
- **Märtens, H.:** Beiträge zur dynamischen Lastbalancierung in parallelen Datenbanksystemen. Dissertation, Univ. Leipzig, 2008
- **Golovin, N.:** Web Recommendations for E-Commerce Websites (eingereicht)



Nach der erfolgreichen Verteidigung: A. Thor (Mai 2008) und H. Märtens (Jun. 2008) mit Doktorhut

- Azagra, X.: Comparison of HTML-aware tools for web data extraction. Diplomarbeit, Univ. Leipzig, 2008
- Ao, Q.: Entwicklung einer Object-Matching-Komponente für ein Data-Warehouse zur Durchführung von Zitierungsanalysen. Diplomarbeit, Univ. Leipzig, 2008
- Chu, T.K.: Dokumentenserver 2.0: Aktuelle Navigationsverfahren, Suchverfahren und Methoden zum Metadaten austausch. Bachelorarbeit, Univ. Leipzig, 2008
- Do, V.H.: Extraktion und Integration von Produktdaten aus Onlineshops. Bachelorarbeit, Univ. Leipzig, 2009
- Engsterhold, R.: Automatische Integration von Publikationen am Institut für Informatik in den Dokumentenserver. Bachelorarbeit, Univ. Leipzig, 2009
- Fiedler, M.: Ein hybrider Ansatz zur flexiblen Speicherung von XML-Daten in RDBMS am Beispiel von PubMed, Diplomarbeit, Univ. Leipzig, 2009

- Groß, S.: Konzeption und Realisierung eines webbasierten Informationssystems zur Erfassung, Darstellung und Verwaltung von Kooperationen zwischen Kompetenzträgern in Sachsen und Mittel- und Osteuropa. Diplomarbeit, Univ. Leipzig, 2008
- He, S.: Eine graphische Benutzerschnittstelle für ausgewählte Anwendungsfälle der Ontologie- und Mappingverwaltung sowie -generierung mit GOMMA. Diplomarbeit, Univ. Leipzig, 2009
- Kolb, L.: Entwurf und Umsetzung eines Frameworks zum fairen Vergleich von Object-Matching-Algorithmen. Diplomarbeit, Univ. Leipzig, 2009
- Kotlarz, D.: Entwurf und Implementierung einer serviceorientierten Architektur für die Bereitstellung von Object Matching Web Services. Bachelorarbeit, Univ. Leipzig, 2009
- Landsmann, I.: Semantische Feature-Extraktion zur Kategorisierung von Publikationen. Bachelorarbeit, Univ. Leipzig, 2009
- Liegl, M.: Datenschutz in computerforensischen Veranstaltungen. Diplomarbeit, Univ. Leipzig, 2009
- Müller, T.; Horn, S.: Evaluation von Object Matching Frameworks. Diplomarbeit, Univ. Leipzig, 2009
- Schilling, S.: Analyse zum Einsatz eines Master Data Management Systems. Diplomarbeit, Univ. Leipzig, 2009
- Schnerwitzki, T.: Entwicklung eines Data Warehouses zur Durchführung von Zitierungsanalysen. Diplomarbeit, Univ. Leipzig, 2008
- Tappert, E.: Aufbau einer Forschungsinstitut-Datenbank - Daten- und Informationsextraktion zu wissenschaftlichen Publikationen. Masterarbeit, Univ. Leipzig, 2008
- Wartner, C.: Evaluierung von Querystrategien zur Informationsfusion von bibliografischen Daten. Masterarbeit, Univ. Leipzig, 2009

Vorträge

- Aumüller, D.: Towards web supported identification of top affiliations from scholarly papers. 13. GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web (BTW), Münster, März, 2009
- Aumüller, D.: Web-based Affiliation Matching. Intl. Conference on Information Quality (ICIQ), Potsdam, November, 2009
- Endrullis, S.: Evaluation of Query Generators for Entity Search Engines. Intl. Workshop on Using Search Engine Technology for Information Management (USETIM), Lyon (France), August, 2009
- Groß, A.: Untersuchungen von Annotation-Mappings am Beispiel der GeneOntology. BBZ Kolloquium, Leipzig, Dezember, 2008
- Groß, A.: An Evolution-based Approach for Assessing Ontology Mappings – A Case Study in the Life Sciences. 13. GI-Fachtagung Datenbanksysteme für Business, Technologie und Web (BTW), Münster, März, 2009
- Groß, A.: Estimating the Quality of Ontology-Based Annotations by Considering Evolutionary Changes. 6th Intl. Workshop on Data Integration in Life Sciences Workshop (DILS), Manchester (UK), Juli, 2009
- Groß, A.: Quality of Functional Annotations in Life Science Data Sources. 1. Workshop des GI-Arbeitskreises Ontologien in Biomedizin und Lebenswissenschaften (OBML), Leipzig, November, 2009
- Hartung, M.: Management von Ontologien in den Lebenswissenschaften. 20. GI-Workshop über die Grundlagen von Datenbanken (GvD), Apolda, Mai, 2008
- Hartung, M.: Analyzing the Evolution of Life Science Ontologies and Mappings. 5th Intl. Workshop on Data Integration in the Life Sciences, Evry/Paris (France), Juni, 2008
- Hartung, M.: A platform for collaborative management of semantic grid metadata. 2nd Intl. Symposium on Intelligent Distributed Computing, Catania (Italy), September, 2008
- Hartung, M.: Ontologie-Evolution in den Lebenswissenschaften. BBZ Kolloquium, Leipzig, Dezember, 2008
- Hartung, M.: Evolution of Life Science Ontologies. 1. Workshop des GI-Arbeitskreises Ontologien in Biomedizin und Lebenswissenschaften (OBML), Leipzig, November, 2009
- Kirsten, T.: Integration and annotation concepts for the analysis of high throughput data. Workshop in Computational Approaches to Stem Cell Biology, Leipzig, Januar, 2009
- Kirsten, T.: Efficient Management of Biomedical Ontology Versions. 4th Intl. Workshop on Ontology Content (Part of the OTM Conferences & Workshops), Algarve (Portugal), November, 2009
- Kirsten, T.: Matching Large Life Science Ontologies. 1. Workshop des GI-Arbeitskreises Ontologien in Biomedizin und Lebenswissenschaften (OBML), Leipzig, November, 2009
- Köpcke, H.: Training Selection for Tuning Object Matching. 6th Intl. Workshop on Quality in Databases and Management of Uncertain Data (QDB/MUD), Auckland (New Zealand), August, 2008
- Massmann, S.: Evaluating Instance-based Matching of Web Directories. 11th Intl. Workshop on the Web and Databases (WebDB), Vancouver (Canada), Juni, 2008

- Rahm, E.: Comparing the Scientific Impact of Conference and Journal Publications in Computer Science. Invited Talk Academic Publishing in Europe (APE) Conf., Berlin, Januar, 2008
- Rahm, E.: Instance-based Matching of Large Ontologies. Invited Talk GvD Workshop, Rostock, Juni, 2009
- Rahm, E.: Object Matching for Improving Information Quality. Talk at Karlsruhe institute of Technology (KIT). November 2009
- Thor, A.: Automatische Mapping-Verarbeitung auf Webdaten. Information Systems Group Forschungsseminar Hasso Plattner Institut, Potsdam, Januar, 2008
- Thor, A.: Evaluierung von Data-Warehouse-Werkzeugen. Congress on e-Government, Leipzig, Mai, 2009

Mitgliedschaften in Gremien/Redaktionskollegien, Herausbergremien u.ä.

Rahm, E.:

- 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. SIGMOD 2009, VLDB 2008, VLDB 2010, ICDE 2008 [Vice-PC-Chair], BTW 2009, DILS 2008, DILS 2009, ICIQ 2009)
- Vorstandsmitglied IZBI (Interdisz. Zentrum für Bioinformatik, Leipzig)
- Vorsitzender Prüfungsausschuß Informatik
- Mitglied im Fakultätsrat

Aumüller, D.:

- Programmkomitee/Reviewer verschiedener Workshops/Journals (Scripting for the Semantic Web 2008/2009, SemWiki 2008/2009, Journal of Web Semantics)