

# Forschungsbericht 2004/2005

## 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
Wiss. Mitarbeiter	Böhme, Timo
Wiss. Mitarbeiter (DFG / IZBI)	Dr. Do, Hong Hai (bis 8/05)
Stipendiat (Graduiertenkolleg)	Golovin, Nick
Wiss. Mitarbeiter (DFG)	Dr. Greiner, Ulrike (bis 9/04)
Wiss. Mitarbeiter (BMBF)	Hartung, Michael (seit 11/05)
Sekretärin	Hesse, Andrea
Programmierer	Jusek, Stefan
Wiss. Mitarbeiter (DFG / IZBI)	Kirsten, Toralf
Wiss. Mitarbeiter (Krebshilfe / IZBI)	Lange, Jörg (seit 9/04)
Wiss. Mitarbeiter (BMBF)	Massmann, Sabine (seit 4/05)
Wiss. Mitarbeiter	Dr. Sosna, Dieter
Stipendiat (Graduiertenkolleg)	Thor, Andreas



Abt. Datenbanken im Feb. 2005. V.l.n.r.: T. Böhme, H. Do, S. Jusek, J. Lange, S. Massmann, T. Kirsten, A. Hesse, A. Thor, D. Aumüller, D. Sosna, N. Golovin, E. Rahm

## Highlights

Im Berichtszeitraum 2004/2005 sind folgende Ereignisse hervorzuheben:

- Die in der Abteilung entstandene und im Juli 2004 verteidigte Dissertation von Sergey Melnik wurde im März 2005 durch die Gesellschaft für Informatik (Fachbereich Datenbanken und Informationssysteme) mit dem BTW-Dissertationspreis für eine der beiden besten Informatik-Dissertationen im deutschsprachigen Raum der Jahre 2003/2004 im Gebiet Datenbanken ausgezeichnet.
- 2005 wurden zwei weitere Dissertationen durch Ulrike Greiner und Hai-Hong Do fertig gestellt.
- Im Berichtszeitraum konnten drei neue Drittmittelprojekte begonnen werden. Die Deutsche Krebshilfe unterstützt in einem Verbundvorhaben die Erforschung molekularer Mechanismen maligner Lymphome; die Datenhaltung dazu erfolgt durch uns am Interdisziplinären Zentrum für Bioinformatik (IZBI). Das BMBF fördert ein Innoregio-Projekt zu E-Commerce im Musikinstrumentenbau sowie ein D-Grid-Projekt zu Ontologiewerkzeugen.
- Im März 2004 wurde mit einem erfolgreichen Workshop in Leipzig die neue internationale Workshop-Serie „Data Integration in the Life Sciences (DILS)“ ins Leben gerufen. 70 Teilnehmer aus 11 Ländern kamen zu dem 2-tägigen Programm.
- Das Oberseminar der Abteilung fand im Juli 2004 bereits zum vierten Mal an der Uni-Außenstelle in Zingst/Ostsee statt.
- Unsere Publikation zum Recommendation-System AWESOME wurde als eine der 4 besten Forschungsarbeiten der Top-Datenbankkonferenz VLDB 2004 in Toronto aus über 500 Einreichungen ausgewählt. Die Langfassung des Beitrages ist im VLDB Journal, Nov. 2005, erschienen.
- David Aumüller erhielt im Mai 2005 auf der 2nd European Semantic Web Conference in Heraklion eine Auszeichnung für die beste Demo (Best Demo Award)
- Der Abteilungsleiter verbrachte im Sommersemester 2005 ein Sabbatical in der Datenbankgruppe von Microsoft Research in Redmond, WA.
- Mit unserem Datenintegrationssystem iFuice wurde eine umfassende Zitationsanalyse von Datenbankpublikationen unter Nutzung von Google Scholar vorgenommen (Sigmod Record, Dec. 2005). Dabei wurde der Impact factor auch für Tagungen ermittelt. Zudem wurde ein Nationenranking bestimmt, bei dem Deutschland nach den USA auf Platz 2 kam.



Bilder vom DILS2004 Workshop in Leipzig

## Projektbeschreibungen

### Themen:

- Peer-to-Peer-artige Datenintegration
- Schema- und Ontologie-Matching
- Data Warehousing zur Unterstützung biomedizinischer Analysen
- Integration molekular-biologischer Annotationsdaten
- MediGRID: Grid Computing für die biomedizinische Forschung - Modul Ontologie-Werkzeuge
- Adaptive Online-Empfehlungen
- Semantisches kollaboratives Content Management
- XML-Datenverwaltung
- Qualitätssicherung in Web-Service-basierten kooperativen Geschäftsprozessen
- E-Commerce im Musikinstrumentenbau
- Web-basiertes Lernen

### Peer-to-Peer-artige Datenintegration

E. Rahm, A. Thor, D. Aumüller, H. Do, N. Golovin, T. Kirsten

Datenintegrationssysteme setzen üblicherweise auf ein globales Schema, um eine einheitliche Sicht auf heterogene Datenquellen bereitzustellen. Im Bereich von Data Warehouses hat sich diese Herangehensweise als erfolgreich herausgestellt. Dabei ist jedoch der Aufwand zur Integration neuer Datenquellen sehr hoch und eine explorative Ad-hoc-Datenintegration wird nicht unterstützt. iFuice verkörpert einen neuen generischen Ansatz zur dynamischen Datenintegration, der dezentrale, Peer-to-Peer-artige Mappings zwischen verschiedenen Datenquellen nutzt. Die Mappings sind ausführbar und führen zu Korrespondenzen zwischen Instanzen der Datenquellen. Die Semantik der Quellen und der dazugehörigen Mappings spiegeln sich in einem Domänenmodell wider, das abstrakter als ein globales Schema modelliert ist, und somit einfacher zu handhaben ist. Mappings zwischen verschiedenen Datenquellen können auf Mengen von Objekten ausgeführt und aneinandergehängt werden. Dies ermöglicht eine effektive Bündelung von Informationen verschiedener Quellen. Eine Reihe von Skriptoperatoren steht zur Verfügung, um die Ausführung der verschiedenen Mappings und die Datenfusion zu steuern, sowie die Ergebnisse der Mappings weiterzuverarbeiten. Mappings werden z.B. durch spezielle Abfragen oder Web Services bereitgestellt. Die Ausführung dieser Mappings kann entweder skriptgesteuert oder explorativ, d.h. Schritt für Schritt, erfolgen.

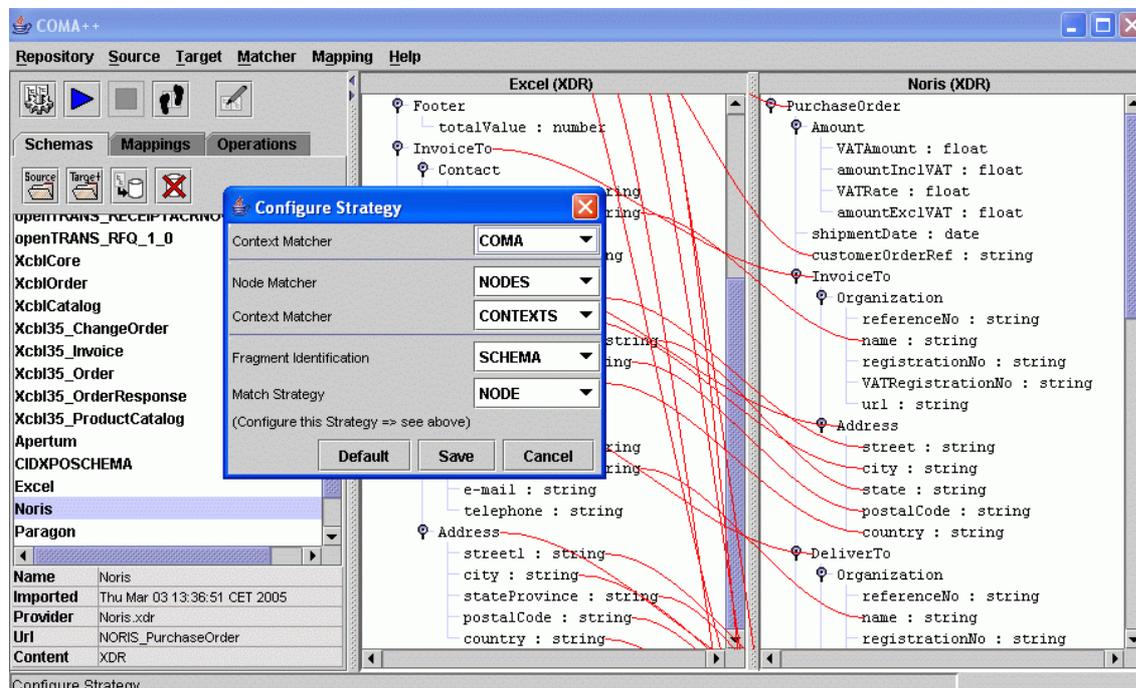
Basierend auf der generischen iFuice-Architektur wurde begonnen, mehrere domänenspezifische Integrationssysteme zu entwickeln. In *BiblioFuice* geht es um die Verknüpfung unterschiedlicher bibliographischer Datenquellen und Volltextarchive wie DBLP, ACM Digital Library und Google Scholar. Eine erste Anwendung ist die umfassende Zitationsanalyse der wichtigsten Tagungen und Zeitschriften im Datenbankbereich. *BioFuice* ermöglicht die Kopplung unterschiedlicher Datenquellen für Bioinformatikauswertungen. Neben öffentlichen Datenquellen wie Ensembl, NetAffx, oder HUGO können auch lokale Quellen eingebunden werden, z.B. private Genlisten oder Ergebnisdaten vorangegangener Analysen. BioFuice wird am Bioinformatikzentrum IZBI entwickelt und bereits in verschiedenen Forschungsprojekten genutzt. In *EC-Fuice* geht es um die semantische Integration von E-Commerce-Daten, um eine nahtlose Navigation zwischen Produktdaten und Produktontologien unterschiedlicher Anbieter zu erreichen. Der Ansatz ermöglicht auch verteilte Produktempfehlungen (recommendations). Die Realisierung nutzt das in der Abteilung Datenbanken entwickelte Ontologie-Matching-Tool COMA++.

## Schema- und Ontologie-Matching

E. Rahm, H. Do, D. Aumüller, S. Maßmann

Der früher erstellte Schema-Matching-Prototyp COMA (Composite Match System) wurde zu einer einheitlichen Plattform für sowohl Schema- als auch Ontologie-Matching weiter entwickelt. Basierend auf einer flexiblen Kombination von Match-Algorithmen bietet das neue System COMA++ eine umfassende Infrastruktur zur Lösung auch großer Match-Probleme an. Eine graphische Nutzeroberfläche erlaubt dem Nutzer vielfältige Interaktionsmöglichkeiten und unterstützt ihn bei der Konfiguration der Match-Algorithmen sowie der Match-Strategie. Unter Nutzung eines generischen Datenmodells ist COMA++ in der Lage, Schemata sowie Ontologien verschiedener Formate wie z.B. der W3C XML Schema und OWL-Standards zu importieren. Weiterhin können schon vorhandene Match-Ergebnisse miteinbezogen und wiederverwendet werden (Re-Use). Eine sog. fragment-basierte Match-Strategie erhöht die Performanz des kompletten Prozesses, in dem exklusiv ähnliche Fragmente zwischen großen Schemata identifiziert und abgeglichen werden. COMA++ und die darin unterstützten Strategien wurden im Rahmen einer umfassenden Evaluation quantitativ bewertet.

Die Abbildung zeigt die Oberfläche von Coma++.



## Data Warehousing zur Unterstützung biomedizinischer Analysen

E. Rahm, T. Kirsten, J. Lange

Zur Unterstützung datenintensiver molekular-biologischer und biomedizinischer Analysen wurden am Interdisziplinären Zentrum für Bioinformatik (IZBI) mehrere Data-Warehouse-basierte Analyseplattformen entwickelt. Zur Genexpressionsanalyse auf Basis der Microarray-Technologie wurde die GeWare-Plattform weiterentwickelt, welche verschiedenste Methoden der Datenvorverarbeitung (Normalisierung, Hintergrundbereinigung etc.) sowie der Analyse, wie der differentiellen Genexpression, unterstützt. Die Plattform ist vollständig web-basiert und wird derzeit von unterschiedlichen Gruppen der Universität Leipzig genutzt. Eine Weiterentwicklung des Warehouse dient als Datenplattform für die bundesweiten Verbundprojekte zur Untersuchung von molekularen Mechanismen in „malignen Lymphomen“ und „Gliomen“ der Deutschen Krebshilfe. Die Erweiterungen betreffen zusätzliche experimentelle Daten (z.B. Matrix-CGH Daten), die Kopplung mit anonymisierten Patientendaten sowie zusätzliche Analyseverfahren. Das Projekt wird am IZBI in enger Zusammenarbeit mit dem IMISE der Universität Leipzig (Prof. Löffler) sowie den Studienpartnern an anderen Standorten durchgeführt.

## Integration molekular-biologischer Annotationsdaten

E. Rahm, H. Do

Informationen über molekular-biologische Objekte, wie z.B. Gene und Proteine, werden kontinuierlich als so genannte Annotationen erzeugt, gesammelt und in zahlreichen öffentlichen Webdatenbanken zur Verfügung gestellt. Aufgrund derer Heterogenität stellt die Integration dieser Daten ein großes Problem in der Bioinformatik dar. Wir haben das GenMapper-System entwickelt, welches Annotationsdaten physikalisch in einer zentralen relationalen Datenbank integriert und diese auf einer flexiblen Art und Weise für interaktive Abfragen sowie automatische Analysen bereitstellt. GenMapper benutzt ein generisches Datenmodell, um Annotationen aus unterschiedlichen Quellen einheitlich zu repräsentieren. Beziehungen zwischen molekular-biologischen Objekten, welche in der Regel durch mühsame Kurationsarbeit der Forscher entstanden sind, werden beim Datenimport explizit extrahiert und bei der Integration verwendet. Häufige Abfragen und Analysen werden durch entsprechende Annotationssichten unterstützt, die flexibel aus der generischen Repräsentation mittels mächtiger Operatoren generiert werden können. GenMapper befindet sich bereits im praktischen Einsatz zur funktionellen Analyse großer Genmengen genutzt. Daneben wurde ein hybrider Integrationsansatz entwickelt, um Annotationsdaten für datenintensive Expressionsanalysen verwendbar zu machen. Dabei werden in einem Warehouse vorgehaltene Expressionsdaten über einen Mediator mit unterschiedlichen Web-Datenquellen verknüpft. Für die einheitliche Anbindung der Datenquellen wird das verbreitete Tool SRS (Sequence Retrieval System) genutzt. Die Systementwicklungen sind über die Website <http://www.izbi.de> zugänglich.

---

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

E. Rahm, M. Hartung, T. Kirsten

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, Ressourcenusion 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. Die Projektdurchführung erfolgt in Kooperation mit den IZBI (Interdisziplinäres Zentrum für Bioinformatik) und IMISE (Prof. Löffler).

---

## Adaptive Online-Empfehlungen

E. Rahm, A. Thor, N. Golovin

Web-Recommendation-Systeme sind ein weit verbreitetes Mittel zur Erhöhung der Nutzbarkeit von Web Sites. Innerhalb des Projekts „Adaptive Online-Empfehlungen“ haben wir eine Architektur entwickelt, welche die Adaption der Web-Empfehlungen bzgl. den Interessen der Benutzer erlaubt. Unser Ansatz basiert auf einem Web Usage Data Warehouse, in welchem Navigationsverhalten, Nutzerfeedback zu den gegebenen Empfehlungen und sonstige verfügbare Informationsquellen in einer multidimensioneller Struktur gespeichert werden. Die Data-Warehouse-Technologie ermöglicht uns große Datenmengen effektiv zu verwalten und unterschiedliche Empfehlungsalgorithmen zu unterstützen. Wir haben mehrere Machine-Learning-basierte Strategien für die Adaption des Empfehlungssystems untersucht, welche die Adaption sowohl auf der Ebene von Empfehlungsalgorithmen als auch auf der Ebene von einzelnen Empfehlungen unterstützen. Die prototypische Implementierung wurde im Rahmen kommer-

zieller als auch nichtkommerzieller Web Sites umgesetzt, um unsere Architektur und die angewendeten Adaptionstechniken unter unterschiedlichen Real-Life-Bedingungen zu evaluieren.

---

## **Semantisches kollaboratives Content Management**

D. Aumüller

Benutzer legen Informationen aus den verschiedensten Datenquellen auf ihrem Desktop ab, nicht nur um auf diese zu einem späteren Zeitpunkt wieder zuzugreifen, sondern auch, um sie mit anderen Nutzern zu teilen. Dazu sind persönliche Anmerkungen zu einzelnen Informationen oft wünschenswert, um in der Fülle der Dokumente nicht die Übersicht zu verlieren. Solche Annotationen können Kommentare sowie Kategorisierung und Querverweise zwischen den Dokumenten beinhalten. Um den Wissensaustausch zu fördern, kann das Annotieren gemeinsam erfolgen und Inhalte bestimmten Nutzergruppen zur Verfügung gestellt werden. Des Weiteren sind zu vorhandenen weitere relevanten Informationen auf einfache Weise zur Verfügung zu stellen. Diese Ziele verfolgt das Framework zum semantischen kollaborativem Content Management. Aufbauend auf einem konventionellen Web-CMS wurde mit der Entwicklung eines Prototyps begonnen, der automatisiert Inhalte aus externen Datenquellen zu importieren ermöglicht und die integrierten Inhalte anhand von Ontologien kategorisiert. Dabei ist neben dem möglichst hohen Grad an Automatismus auch ein Hauptaugenmerk auf die Einfachheit der Bedienung zu legen, da die Integration heterogener Daten oft nicht vollautomatisch zu bewerkstelligen ist. Außerdem sollen eine manuelle Annotierung durch verschiedene Benutzer sowie die Integration bzw. Referenzierung lokaler Dokumente unterstützt werden. Die anhand eines im Vorfeld entwickelten semantischen Wiki-Systems gesammelten Erfahrungen bzgl. Eingabe strukturierter Annotationen, kontextabhängiger Navigation und semantischer Anfragemöglichkeit fließen in den neuen Prototypen ein.

---

## **XML-Datenverwaltung**

E. Rahm, T. Böhme

Die massive Verbreitung von XML verlangt zunehmend eine effiziente Verwaltung großer Mengen von XML-Daten. In einem ersten Schritt haben wir die Leistungsfähigkeit der verschiedenen Architekturen zur Verwaltung von XML Daten mittels des von uns entwickelten skalierbaren Multi-User Benchmarks XMach-1 (XML Data Management benchmark) untersucht. Auf der Grundlage dieser Ergebnisse erfolgt in einem zweiten Schritt die Entwicklung optimierter Strukturen für die effiziente XML-Datenverwaltung mit Schwerpunkt auf objektrelationalen Datenbanksystemen. Es sollen hierbei speziell die Potentiale dieser Datenbanksysteme für eine performante, skalierbare und generische Verwaltung dokumentorientierter XML-Daten untersucht werden. Als erstes Ergebnis entstand das hierarchische Nummerierungsschema DLN (Dynamic Level Numbering). Dieses unterstützt eine optimierte Anfrageverarbeitung, eine schnelle Rekonstruktion von XML-Dokumenten sowie Einfügeoperationen ohne Änderung bestehender Daten. Als nächstes wurden die Möglichkeiten zur automatischen Anfrageoptimierung mittels materialisierter Hilfsstrukturen untersucht. Hierfür wurden Parametrisierte XPath-Views sowie ein spezielles Kostenmodell auf Basis der Kostenmodelle der Datenbanksysteme entwickelt.

---

## **Qualitätssicherung in Web-Service-basierten kooperativen Geschäftsprozessen**

E. Rahm, U. Greiner

Interoperabilitätstechniken wie XML und Web-Services werden zunehmend zur Realisierung kooperativer Workflows und Geschäftsprozesse genutzt. Eine wesentliche Herausforderung dabei ist die Sicherstellung einer hohen Ausführungsqualität trotz der Autonomie und Heterogenität der beteiligten Kooperationspartner und Prozessschritte. Das entwickelte WebFlow-

System ermöglicht die Spezifikation und Überwachung unterschiedlicher Qualitätskriterien (u.a. zeitliche und inhaltliche Restriktionen) bei der Ausführung von Web Services. Verletzungen der Bedingungen werden über einen regelbasierten Ansatz automatisch erkannt, um eine frühzeitige Reaktion zur Qualitätssicherung zu ermöglichen. Die aktuelle Implementierung basiert auf IBM Alphaworks und BPEL4WS (Business Process Execution Language for Web Services).

---

## **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 von Händlern ausgewertet. Daraus resultieren Empfehlungen für den Inhalt und zu unterstützende Features für Erfolgsversprechende 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“ ([www.musiconvalley.de](http://www.musiconvalley.de)).

---

## **Web-basiertes Lernen**

E. Rahm, T. Böhme, D. Sosna

Das web-basierte SQL-Übungssystem wurde zu einer eigenständigen E-Learning-Plattform LOTS (Leipziger Online Training System, <http://lots.uni-leipzig.de>) entwickelt, welche die Erstellung, Bearbeitung und automatische Auswertung von Übungsaufgaben vom Multiple-Choice-Typ unterstützt. Das System wird sehr erfolgreich im realen Studienbetrieb eingesetzt und soll weiter ausgebaut werden. Im Berichtszeitraum wurde ferner die Realisierung eines internet-gestützten elektronischen Wörterbuchs Neugriechisch - Deutsch in Zusammenarbeit mit dem Institut für Klassische Philologie begonnen.

## Veröffentlichungen / Graduiierungsarbeiten

### Bücher

- **Rahm, E.** (Hrsg.): Data Integration in the Life Sciences. Proc. 1<sup>st</sup> International Workshop on Data Integration in the Life Sciences, Lecture Notes in Bioinformatics (LNBI) 2994, Leipzig, March 2004



### Zeitschriften

- **Greiner, U., Müller, R., Rahm, E., Ramsch, J., Heller, B., Löffler, M.:** AdaptFlow: Protocol-based Medical Treatment Using Adaptive Workflows. *Methods of Information in Medicine* 2005; 44: 80-88
- **Khaitovich, P., B. Mützel, X. She, M. Lachmann, I. Hellmann, J. Dietzsch, S. Steigele, H.-H. Do, G. Weiss, W. Enard, F. Heissig, T. Arendt, K. Nieselt-Struwe, E.E. Eichler S. Pääbo:** Regional patterns of gene expression in human and chimpanzee brains. *Genome Research* 14(8), 2004
- **Müller, R., Greiner, U., Rahm, E.:** AGENTWORK: A Workflow-System Supporting Rule-Based Workflow Adaptation. *Data & Knowledge Engineering* 51(2), Elsevier, November 2004.
- **Rahm, E., Do H.H., Massmann, S.:** Matching Large XML Schemas. *Sigmod Record* 33(4), December 2004.
- **Rahm, E., Thor, A.:** Citation Analysis of Database Publications. *ACM SIGMOD Record* 34(4), 2005
- **Thor, A., Golovin, N, Rahm, E.:** Adaptive Website Recommendations with AWESOME. *VLDB Journal* 14(4), 2005

### Proceedings

- **Aumüller, D.:** SHAWN: Structure Helps A Wiki Navigate, BTW 2005 Workshop WebDB Meets IR, Karlsruhe, Germany, March 2005
- **Aumüller, D., Do, H.H., Massmann, S., Rahm, E.:** Schema and Ontology Matching with COMA++. Proc. SIGMOD 2005 (Software Demonstration), Baltimore, June 2005
- **Aumüller, D.:** Semantic Authoring and Retrieval within a Wiki, 2nd European Semantic Web Conf. 2005 (ESWC2005), Heraklion, Greece, May 2005 - Best Demo Award
- **Aumüller, D., Auer, S.:** Towards a Semantic Wiki Experience - Desktop Integration and Interactivity in WikSAR, 1st Workshop on The Semantic Desktop: Next Generation Personal Information Management and Collaboration Infrastructure, Galway, Ireland, Nov. 2005
- **Böhme, T., Rahm, E.:** Supporting Efficient Streaming and Insertion of XML Data in RDBMS. Proc. 3rd Int. Workshop Data Integration over the Web (DIWeb), 2004
- **Do, H.H., Rahm, E.:** Flexible Integration of Molecular-biological Annotation Data: The GenMapper Approach. Proc. EDBT 2004, Heraklion, Springer LNCS, March 2004
- **Golovin, N., Rahm, E.:** Reinforcement Learning Architecture for Web Recommendations. Proceedings of the ITCC2004. IEEE, 2004.
- **Golovin, N., Rahm, E.:** Automatic Optimization of Web Recommendations Using Feedback and Ontology Graphs. Proc. 5th Int. Conf. on Web Engineering (ICWE), Springer LNCS, 2005: 375-386
- **Greiner, U., Rahm, E.:** Quality-Oriented Handling of Exceptions in Web-Service-Based Cooperative Processes, Proc. of EAI-Workshop 2004 - Enterprise Application Integration, Oldenburg. GITO-Verlag, Berlin, Feb. 2004.
- **Greiner, U., Ramsch, J., Heller, B., Löffler, M., Müller, R., Rahm, E.:** Adaptive Guideline-based Treatment Workflows with AdaptFlow. Proc. Symposium on Computerized Guidelines and Protocols (CGP 2004), Prague, IOS Press, April 2004.
- **Kirsten, T., Körner, C., Do, H., Rahm, E.:** Hybrid Integration of Molecular-biological Annotation Data. Proc. 2nd International Workshop on Data Integration in the Life Sciences (DILS), San Diego, Springer LNCS, 2005

- **Körner, C., Kirsten, T., Do, H., Rahm, E.:** Hybride Integration von molekularbiologischen Annotationsdaten. Proc. 11th Conf. of Database Systems for Business, Technology and Web (BTW), Karlsruhe, 2005
- Melnik, S., Bernstein P.A., Halevy, A., **Rahm, E.:** Supporting Executable Mappings in Model Management. Proc. SIGMOD 2005, Baltimore, June 2005
- **Rahm, E., Thor, A., Aumüller, D.; Do, H.-H., Golovin, N., Kirsten, T.:** iFuice - Information Fusion utilizing Instance Correspondences and Peer Mappings. 8<sup>th</sup> International Workshop on the Web and Databases (WebDB), 2005
- **Thor, A., Rahm, E.:** AWESOME - A Data Warehouse-based System for Adaptive Website Recommendations. Proc. 30th Intl. Conference on Very Large Databases (VLDB), Toronto, Aug. 2004

#### Technische Berichte und Poster

- **Böhme, T.; Rahm, E.; Sosna, D.:** Konzeption und Einsatz eines Online-Übungssystems an der Universität Leipzig. Techn. Report, Univ. Leipzig, 2005
- **Kirsten, T., Do, H.H., Rahm, E.:** A Data Warehouse for Multidimensional Gene Expression Analysis. Leipzig Bioinformatics Working Paper 04-01, 2004
- **Kirsten, T.; Rahm, E.:** BioFuice: A decentralized approach to integrate molecular biological annotation data. 3rd Research Festival, Univ. Leipzig, 2005
- **Lange, J; Kirsten, T.; Rahm, E.; Berger, H.; Löffler, M.:** An integrated platform for analyzing clinical and molecular-biological data in collaborative research projects. 3rd Research Festival, Univ. Leipzig, 2005

#### Dissertationen, Graduierungsarbeiten

- **Do, H.:** Schema Matching and Mapping-Based Data Integration. Dissertation, Univ. Leipzig, 2005
- **Greiner, U.:** Quality-oriented Execution and Optimization of Cooperative Processes: Model and Algorithms. Dissertation, Univ. Leipzig, 2005
- **Melnik, S.:** Generic Model Management: Concepts and Algorithms. Dissertation, Univ. Leipzig, 2004
- Berr, T.: Definition and Monitoring of Quality Constraints for Web-Service-Based Cooperative Workflows. Diplomarbeit 2004
- Eckardt, J.: Konzeption und Implementierung eines offenen, ontologiebasierten Übungssystems. Diplomarbeit, 2005
- Fritzsche, V.: Konzeption und Implementierung einer Logging-Komponente zur Bestimmung von Aktionen und Recommendations für die dynamische Ausnahmebehandlung in Web-Service-basierten, kooperativen Prozessen. Diplomarbeit, 2005
- Heinemann, C.: Entwurf und Implementierung eines Recherchesystems zur Web-Service-basierten Integration digitaler Bibliotheken. Diplomarbeit 2004
- Kappel, T.: Evaluation von Content Management Systemen und Einrichtung eines CMS für den Internetauftritt der Abteilung Datenbanken. Bachelorarbeit, 2005
- Körner, C.: Integration von Genannotations aus Web-Datenquellen. Diplomarbeit 2004
- Lange, J.: Flexible Verwaltung und Spezifikation von Experimentannotations für Genexpressionsanalysen. Diplomarbeit 2004
- Maßmann, S.: Analyse der Wiederverwendung zum Schema-Matching. Diplomarbeit, 2005
- Stoll, S.: Terminologieorientiertes Website-Matching mit COMA++. Diplomarbeit 2005



Nach der erfolgreichen Verteidigung: S. Melnik (Juli 2004) und U. Greiner (Dez. 2005) mit Doktorhut

## Vorträge

- Aumüller, D.: SHAWN: Structure Helps A Wiki Navigate, BTW 2005 Workshop WebDB Meets IR, Karlsruhe, März 2005
- Aumüller, D.: Semantic Authoring and Retrieval within a Wiki, Demo-Vorführung. 2nd European Semantic Web Conference 2005 (ESWC2005), Heraklion, Mai 2005
- Böhme, T.: Supporting Efficient Streaming and Insertion of XML Data in RDBMS. 3rd Int. Workshop Data Integration over the Web (DIWeb), Riga, Juni 2004
- Do, H.H.: Flexible Integration of Molecular-biological Annotation Data: The GenMapper Approach. EDBT 2004, Heraklion, März 2004
- Do, H.H.: Schema and Ontology Matching with COMA++. Demo-Vorführung, ACM SIGMOD 2005, Baltimore, Juni 2005
- Greiner, U.: Quality-Oriented Handling of Exceptions in Web-Service-Based Cooperative Processes, EAI-Workshop, Oldenburg, Feb. 2004
- Greiner, U.: Adaptive Guideline-based Treatment Workflows with AdaptFlow. Symposium on Computerized Guidelines and Protocols (CGP 2004), Prag, April 2004
- Golovin, N.: Automatic Optimization of Web Recommendations Using Feedback and Ontology Graphs. Int. Conf. on Web Engineering, Sydney, Aug. 2005
- Kirsten, T.: GeWare – Data Warehouse for Gene Expression Analysis. 1<sup>st</sup> Int. Workshop on Data Integration in the Life Sciences (DILS), Leipzig, März 2004
- Kirsten, T.: Hybrid Integration of Molecular-biological Annotation Data. 2nd Int. Workshop on Data Integration in the Life Sciences (DILS), San Diego, Juli 2005
- Rahm, E.: Adaptive Web Recommendations, Firmenpräsentation bei Quelle AG, Fürth (Feb. 2004) und Intershop, Jena (März 2004)
- Rahm, E.: Web Recommendations, Informatik-Kolloquium, HU Berlin, Okt. 2004
- Rahm, E.: Fusion verteilter Daten. Hauptvortrag „Tag der Informatik“, RWTH Aachen, Dez. 2004
- Rahm, E.: Adaptive Web Recommendations, Informatik-Kolloquium, TU München, Jan. 2005
- Rahm, E.: Schema Matching with Coma++. Database Seminar, Microsoft Research, Redmond, WA, März 2005
- Rahm, E.: Schema Evolution. Database Seminar, Microsoft Research, Redmond, WA, Juli 2005
- Rahm, E.: Fusion verteilter Daten. Workshop Institut für Versicherungswissenschaft, Leipzig, Sep. 2005
- Thor, A.: AWESOME - A Data Warehouse-based System for Adaptive Website Recommendations. 30th Intl. Conf. on Very Large Databases (VLDB), Toronto, Aug. 2004
- Thor, A.: iFuice - Information Fusion utilizing Instance Correspondences and Peer Mappings. 8<sup>th</sup> Int. Workshop on the Web and Databases (WebDB), Baltimore, Juni 2005

## Mitgliedschaften in Gremien / Redaktionskollegien, Herausgebergremien u. ä.

Rahm, E.:

- Vorstandsmitglied IZBI (Interdisz. Zentrum für Bioinformatik, Leipzig)
- Vorsitzender Prüfungsausschuß Informatik
- Leitungsgremium Fachbereich "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
- Tagungsleiter und Program Chair 1st Int. Workshop Data Integration in the Life Sciences, Leipzig, März 2004
- Programmkomitee verschiedener Konferenzen (u.a. VLDB2004, DILS2004, Data Warehousing 2004, Data Engineering 2005, BTW 2005, VLDB2005, DILS 2005, WebDB 2005, EDBT 2006)