# Forschungsbericht 2006

Institut für Wissens- und Sprachverarbeitung



Otto-von-Guericke-Universität Magdeburg
Fakultät für Informatik

# Institut für Wissens- und Sprachverarbeitung

Universitätsplatz 2, 39106 Magdeburg
Tel. +49 (0)391 67 18343, Fax +49 (0)391 67 12018
office@iws.cs.uni-magdeburg.de

# 1. Leitung

Prof. Dr. rer. nat. habil. Jürgen Dassow (geschäftsführender Leiter seit: 15. November 2005) Prof. Dr. rer. nat. habil. Dietmar Rösner (geschäftsführender Leiter bis: 14. November 2005)

# 2. Hochschullehrer

Prof. Dr. rer. nat. habil. Jürgen Dassow Prof. Dr. rer. nat. habil. Rudolf Kruse Prof. Dr. rer. nat. habil. Dietmar Rösner

Prof. Dr. Andreas Nürnberger

# 3. Forschungsprofil

- 1. Wissensbasierte Systeme und Dokumentverarbeitung
  - Teilprojekt Integration von Werkzeugen zur Wissensakquisition der DFG-Forschergruppe Informationsfusion
  - Promotionsvorhaben Informationsextraktion aus wissenschaftlichen Abstracts
  - XML-Technologie für Lehr-/Lernsysteme
  - · Aufbereitung von Dokumenten für das Semantic Web
- 2. Fuzzy-Systeme und Neuronale Netze
  - Neuro-Fuzzy-Systeme
  - Information Mining
  - Fuzzy-Daten-Analyse
- 3. Formale Sprachen und Automaten
  - · Grammatiken mit gesteuerter Ableitung
  - Kompetenz-basierte Kooperation in Grammatiksystemen
  - Syntaktische Komplexität von Lindenmayer-Systemen
  - Grammatikalische Bilderzeugung
  - Spezifikation von XML-Dokumenten
- 4. Wissensentdeckung und Maschinelles Lernen
  - Aktives Lernen
  - Boosting-Ansätze für ILP-Verfahren
  - ILP Lernverfahren für Spiele
  - Transformierende ILP-Verfahren für große Datenbestände

# 4. Forschungsprojekte

Projektleiter: Dr. Klaus Benecke

**Projektbearbeiter:** Klaus Benecke, Martin Schnabel **Förderer:** Haushalt; 01.01.2005 - 01.01.2007

### Die Endnutzprogrammiersprache ottoVonG

Für ottoVonG wurden allgemeine Operationen entworfen, die es gestatten Datenbankanfragen und Anfragen an Dokumente zu stellen sowie Berechnungen durchzuführen und Grafiken zu generieren. Die Sprache ist für breite Anwendungsgebiete geeignet insbesondere für die Wirtschaft und die Schule. Der bisherige Prototyp gestattet es Anfragen an XML-Dokumente zu stellen. Er kann unter

http://picasso.cs.uni-magdeburg.de/otto/web/index.html

getestet werden. ottoVonG-Programme zeichnen sich durch eine einfache (sequentielle) Programmlogik und durch leistungfähige universelle Operationen und durch kurze Programme aus.

Projektleiter: Prof. Dr. Rudolf Kruse

Projektbearbeiter: Christian Borgelt, Christian Döring, Georg Ruß

**Förderer:** Industrie; 01.10.2005 - 30.11.2006

## **Advanced Intelligent Data Analysis**

Die Intelligent Systems Research Unit der British Telecom Research

Laboratories (BT Labs) befasst sich mit der Entwicklung intelligenter Lösungen im Bereich der Telekommunikation. In dem Projekt werden gemeinsam verschiedene Datenanalyseprobleme bearbeitet. Unter anderem geht es um die Vorhersage der Dauer von Wartungsarbeiten, um die Wege der Techniker optimal planen zu können. Hierfür wurden statistische Verfahren und verschiedene Soft-Computing-Verfahren eingesetzt, u.a. neuronale Netze und Entscheidungs bzw. Regressionsbäume. Daneben wurde eine Software entwickelt, die interaktive, auf die vorliegenden Fragestellungen spezialisierte Datenanalysen ermöglicht.

**Projektleiter:** Prof. Dr. Rudolf Kruse

**Projektbearbeiter:** Silvana Runow, Christian Kolbe, Matthias Steinbrecher

**Kooperationen:** DaimlerChrysler Research and Technology, Ulm

**Förderer:** Industrie; 01.05.2005 - 31.12.2006

# **Data Mining in Business and Banking**

Um die Produktqualität von Fahrzeugen der DaimlerChrysler-Gruppe über den gesamten Einsatzzeitraum nachzuvollziehen, finden Data Mining-Methoden Anwendung, die in einer komponentenbasierten Applikation zusammenarbeiten. Im Rahmen des Projektes wird diese Applikation um weitere Verfahren wie z. B. das Lernen graphischer Modelle ergänzt. Einen weiteren Schwerpunkt bildet die Entwicklung intuitiver Visualisierungsmethoden für die jeweiligen Datenanalyseergebnisse.

Projektleiter: Prof. Dr. Rudolf Kruse

**Förderer:** Industrie; 01.06.2005 - 01.06.2006

## Datananalyse von Finanz- und speziell Kreditwürdigkeitsdaten

Die Erstellung qualitativ hochwertiger Vorhersagen zur Kreditwürdigkeit zählt zu den bedeutenden Problemen im Bereich des Finanzwesens. Das Ziel dieses Projektes besteht in der Entwicklung von Werkzeugen, die eine umfassenden Analyse und Auswertung der

verfügbaren Datenquellen sowie die für Vorhersagen notwendige Modellgenerierung ermöglichen.

Im Rahmen dieses Projektes wird von der Arbeitsgruppe ein Datenanalysetool bereitgestellt und durch die Entwicklung zusätzlicher Module entsprechend der speziellen Anforderungen im gegebenen Problemfeld erweitert.

**Projektleiter:** Prof. Dr. Rudolf Kruse **Projektbearbeiter:** Xiaomeng Wang

**Förderer:** Land (Sachsen-Anhalt); 01.07.2003 - 30.06.2006

## Datenanalyse in Kommunikationssystemen

In der Telekommunikation, speziell in Mobilfunknetzen, entsteht eine große Menge von Daten, die Alarme und Fehler beschreiben. Die effiziente Auswertung und Anzeige dieser Alarme ist eine wichtige Aufgabe, da die Sendestationen in Mobilfunknetzen natürlich an vielen Orten verteilt sind und es daher mir einigem Aufwand verbunden ist, einen Techniker zur überprüfung und Wartung vor Ort zu bringen. Insbesondere müssen mehrere Alarme zusammengefasst werden, da jede Einheit auch dann einen Alarm auslöst, wenn eine ihrer Teileinheiten einen Alarm auslöst. Die Schwierigkeit besteht dabei darin, dass durch unterschiedliche Signallaufzeiten und ausfallende übertragungsstrecken die Alarme in unterschiedlicher Reihenfolge und unvollständig ankommen. In diesem Projekt wird versucht über Clusteringverfahren und die Erzeugung unscharfer Regeln, Vorschriften für die automatische Zusammenfassung von Alarmen zu finden, die z.B. ... mehr

**Projektleiter:** Prof. Dr. Rudolf Kruse

Projektbearbeiter: Christian Döring, Frank Rügheimer, Matthias Steinbrecher

**Förderer:** Haushalt; 01.01.2006 - 31.12.2007

## **Datenanalyse Plattform "InfoMiner"**

Gegenstand des Projektes ist die Fortführung der Ideen aus dem DFG Forschungsvorhaben (KR 521/4-1) und deren Implementierung in Form der flexiblen Data-Mining Plattform "InfoMiner". Die aus Vorarbeiten hervorgegangene Software ist komponentenbasiert und erlaubt in ihrer weiterentwickelten Form die Erstellung von Konfigurationen des Miners mit unterschiedlichem Funktionsumfang je nach Anwendungsfeld. Miner ist ein vertikales System: Die Menge von verfügbaren Datenzugriffs-, Vorverarbeitungs-, und Datenanalyseverfahren kann dem jeweiligen Sachbereich angepasst werden. Aktuelle Forschungsergebnisse auf dem Gebiet der Datenanalyseverfahren sowie den Sachbereichen angepasste Softwarelösungen können als neue Komponenten mit geringem Aufwand in die bestehende Architektur integriert und im Rahmen einer zugeschnittenen Konfiguration des Miners ausgeliefert werden.

**Projektleiter:** Prof. Dr. Rudolf Kruse **Projektbearbeiter:** Christian Borgelt

**Kooperationen:** Michael Berthold, Altana Lehrstuhl für angewandte Informatik, Universität

Konstanz

Förderer: Haushalt; 01.10.2003 - 31.05.2006 Finden relevanter Fragmente in Moleküldatenbanken

Viele Datenanalyseaufgaben in der Bioinformatik bestehen in der Analyse großer Sammlungen von Molekülen mit dem Ziel, Regelmäßigkeiten in den Molekülen einer bestimmten Klasse zu finden. Zum Beispiel möchte man in der Medikamentenentdeckung neue Kandidaten finden,

indem man hunderttausende von Moleküle einem Screening unterzieht, um experimentell ihre Aktivität in Bezug auf eine bestimmte Krankheit festzustellen. In diesem Projekt wird versucht, Fragmente (Teilstrukuren) von Molekülen zu finden, die in den Aktiven häufig und in den Inaktiven selten vorkommen. Als Ansatz wird eine Übertragung des Eclat-Algorithmus zum Lernen von Assoziationsregeln auf attributierte ungerichtete Graphen entwickelt.

**Projektleiter:** Prof. Dr. Rudolf Kruse **Projektbearbeiter:** Frank Rügheimer

**Förderer:** Haushalt; 01.06.2004 - 01.10.2007

## Repräsentation und Analyse impräziser Daten mittels Graphischer Modelle

Die Analyse von in Studien, Marktbeobachtungen oder Geschäftsprozessen anfallenden Daten spielt heute eine wichtige Rolle bei der Entscheidungsfindung und Planung. Data-Mining Methoden beinhalten ein breites Spektrum wirkungsvoller Verfahren um interessante und nutzbringende Zusammenhänge in Daten aufzudecken. Viele dieser Verfahren setzten jedoch eine hohe Präzision in den genutzen Datenbestände vorraus, die in Anwendungen häufig aus technischen Gründen nicht, oder nur mit hohem Aufwand erzielt werden kann.

Das Projekt widmet sich der effizienten Darstellung und Verwertung teilweise impräziser Daten durch Graphische Modelle. Insbesondere In Ergänzung zu bestehenden Verfahren sollen hierdurch neue Informationsquellen erschlossen und einer Nutzung beispielsweise im Bereich der Planung zugeführt werden.

**Projektleiter:** Jun. Prof. Dr. Andreas Nürnberger

Projektbearbeiter: Andreas Nürnberger, Korinna Bade, Ernesto William De Luca

**Förderer:** DFG; 01.05.2003 - 29.04.2008

# Benutzeradaptive Schnittstellen zur Suche und Navigation in Dokumentendatenbanken

Das Ziel des Forschungsprojektes ist die Entwicklung interaktiver, adaptiver

Benutzerschnittstellen für die Suche und Navigation in unstrukturierten Datenbanken (Information Retrieval Support Systems; IRS-Systeme). Hierbei sollen einige Hauptprobleme existierender Modelle behoben werden. Dies betrifft vor allem die unzureichende Einbindung von Benutzermodellen als auch die mangelhafte Integration der einzelnen Systemkomponenten in ein Gesamtsystem.

Im Unterschied zur Dokumentensuche ist das Ziel eines IRS-Systems nicht nur die Suche nach Dokumenten, die zu einer Anfrage passen, sondern eine umfassende Unterstützung eines Nutzers bei der Suche nach Informationen in einer Dokumentensammlung. Diese ist in der Regel nicht nach einem Schritt abgeschlossen, sondern besteht aus einem interaktiven Prozess. ... mehr

**Projektleiter:** Jun. Prof. Dr. Andreas Nürnberger **Förderer:** Industrie; 01.10.2005 - 01.05.2007 **Building Personalised off-line Self-Service Repositories** 

Ziel des Projektes ist die Entwicklung von Verfahren, die einen personalisierten Zugriff auf Informationssysteme ermöglichen, d.h. die vom System zur Verfügung gestellten Dokumente sollen abhängig vom Interessen-. und Kontextprofil eines Nutzers ausgewählt und strukturiert werden.

Projektleiter: Jun. Prof. Dr. Andreas Nürnberger

Projektbearbeiter: Andreas Nürnberger

**Förderer:** EU - FPR; 01.02.2005 - 31.01.2008 **Nature-inspired Smart Information Systems (NiSIS)** 

NiSIS ist ein Coordination Action (CA) Projekt der EU. Das Projekt hat zum Ziel Forschung an intelligenten Verfahren im Bereich von Informationssystemen zu koordinieren. Ziel ist insbesondere die Betrachtung von Modellen, die auf in der Natur vorkommenden Systemen beruhen oder durch natürliche Systeme motiviert werden. Hierbei werden sowohl natürliche Prozesse auf biologischer Ebene (micro) als auch auf Verhaltensebene (macro) einbezogen. Das Projekt hat 31 Partnerinstitutionen aus 13 Ländern.

**Projektleiter:** Prof. Dr. Dietmar Rösner

Projektbearbeiter: Dipl.-W-Inf. Mario Amelung; Michael Piotrowski, M.A.; Wolfram Fenske

Förderer: Haushalt; 01.01.2005 - 31.12.2007 EduComponents -- E-Assessment in der Informatiklehre

Übungsaufgaben sowie Tests und Prüfungen spielen in der Lehre eine wichtige Rolle. Dabei geht es heute nicht mehr nur um die Benotung am Ende einer Lehrveranstaltung, sondern ebenso um die kontinuierliche Verfolgung des Lernprozesses. Lernende werden motiviert, indem ihr Lernfortschritt sichtbar wird und sie Schwachstellen identifizieren können. Lehrende erhalten einen Überblick, ob die Lernziele tatsächlich erreicht werden, wie erfolgreich sie die Lerninhalte präsentieren und wo evtl. eine intensivere Beschäftigung mit dem Lehrstoff nötig ist.

In diesem Zusammenhang werden von der Arbeitsgruppe bereits seit dem Wintersemester 2003/04 Softwarekomponenten entwickelt und erprobt, die sowohl die Erstellung und Verwaltung von Übungsaufgaben und Tests als auch die Einreichung und Bewertung von studentischen Lösungen elektronisch unterstützen. ... mehr

**Projektleiter:** Prof. Dr. Dietmar Rösner **Projektbearbeiter:** Dipl.-Inf. Manuela Kunze

**Kooperationen:** Prof. Dr. E. Hinrichs, Universität Tübingen Förderer: Haushalt; 01.10.2001 - 30.06.2006

# Linguistische Analysen für die semantische Auszeichnung natürlichsprachlicher Dokumente

Auch im Informationszeitalter stellt die natürliche Sprache das zentrale Mittel zur zwischenmenschlicher Kommunikation dar, ob nun in gesprochener oder geschriebener Form.

Die Anzahl von Dokumenten, sei es als Emails, Besprechungsnotizen, Firmendokumente etc. wächst von Jahr zu Jahr. Die Suche nach Informationen in diesen Dokumentenbeständen ist manuell kaum noch zu bewältigen. Systeme, des Informationen Retrieval und der Informationsextraktion werden eingesetzt, um den Nutzer bei der Suche zu unterstützen.

Die Grundlage für diese Systeme bildet die Auswertung des Inhalts eines

Dokumentes. Dazu werden die Dokumente z.B. nach phrasalen Mustern durchsucht oder eine tiefergehende Analyse der vorkommenden Wörter und linguistischen Strukturen vorgenommen.

Die vorliegende Arbeit beschäftigt sich mit Verfahren, die die semantische Anreicherung von Dokumenten unterstützen.

Anhand von exemplarischen Implementierungen werden Verfahren vorgestellt, die für die Auszeichnung einfacher (token-basierter) Konzepte sowie auch komplexerer Sachverhalte genutzt werden können. ... mehr

**Projektleiter:** Prof. Dr. Dietmar Rösner

Förderer:

Projektbearbeiter: Milan Gnjatovic, M.Sc; Wolfram Fenske, Mirko Hannemann

Kooperationen: - Dr.-Ing. habil. Ulrich Schmucker, IFF, Intelligente Sensor-Systeme, - PD

Dr. Frank Ohl, Leibniz-Institut für Neurobiologie,

BioFuture-Nachwuchsgruppe Neuroprothesen, - Prof. Dr. med. Henning

Scheich, Direktor IfN, Auditorisches Lernen und Sprache, Prof. Dr.

Andreas Wendemuth, OVGU-FEIT, Kognitive Systeme und

Sprachverarbeitung, Prof. Jochen Braun, Ph.D., FNW, Prof.- Prof. Dr.-Ing.

Bernd Michaelis, OVGU-FEIT, Technische Informatik Land (Sachsen-Anhalt); 01.12.2005 - 31.12.2007

## NIMITEK: TP1 Wissensrepräsentation und die Rolle der Prosodie im Dialog

Das vorgeschlagene Teilprojekt soll systematisch die Rolle prosodischer Information im Zusammenhang mit Dialogen der Informationsnachfrage bzw. der Informationsdarbietung untersuchen. Auf dieser Basis sollen dann Algorithmen entwickelt, implementiert und in Demonstratorsysteme integriert werden sowohl für die Erkennung der mit prosodischer Information zusätzlich kommunizierten Information bei der Analyse wie für den Einsatz prosodischer Information bei von Systemen generierten Äußerungen (Antworten, Informationsdarbietung). In enger Zusammenarbeit mit den Arbeiten der Neurobiologen zum Thema `Prosodieverarbeitung" sollen u.a. die folgenden Fragen untersucht werden:

- Welche zusätzlichen Informationen werden über Prosodie vermittelt? (z.B. Merkmale des Sprechers, Intentionen, ...)
- Wie stehen sie im Zusammenhang mit den lexikalisch, syntaktisch und semantisch vermittelten Informationen?
- Inwieweit ist Prosodie nutzbar, um die Verstehbarkeit und die Akzeptanz generierter sprachlicher Äußerungen zu verbessern?

# 5. Eigene Kongresse und wissenschaftliche Tagungen

- Prof. Rösner, Prof. Tsujii German-Japanese Workshop on natural Language Processing Tokyo 31.05. - 01.06.2006
- G. Della Riccia, D. Dubois, Prof. R. Kruse, H-J. Lenz International Workshop on Preferences and SimiliaritiesUdine, Italien, 05.10. 07.10.2006
- E. Hüllermeier, Prof. R. Kruse, A. Nürnberger, J. Strackeljan Symposium on Fuzzy Systems

in Computer ScienceMagdeburg, 27.09. - 28.09.2006

- 1st Workshop on Learning the Semantics of Audio Signals (LSAS 2006), December 6, 2006,
- AthensGreece Symposium on Fuzzy Systems in Computer Science 2006 (FSCS 2006), September 27-28, 2006
- Magdeburg 17th European Conference on Machine Learning (ECML) and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), September 17-22, 2006
- Berlin4th International Workshop on Adaptive Multimedia Retrieval (AMR 2006), Juli 27-28, 2006.
- University of Geneva, Switzerland Symposium on Network Analysis in Natural Sciences and Engineering (part of AISB'06 - Adaptation in Artificial and Biological Systems, April 3-6, 2006, University of Bristol, Bristol, UK

# 6. Veröffentlichungen

Originalartikel in begutachteten internationalen Zeitschriften

Barbaiani, Madalina; Bibire, Cristina; Dassow, Jürgen; Delaney, Aidan; Fazekas, Szilárd; Ionescu, Mihai; Liu, Guangwu; Lodhi, Atif; Nagy, Benedek
The power of programmed grammars with graphs from various classes
In: Journal of applied mathematics & computing: JAMC; an international journal. - Seoul: Society, ISSN 1598-5865, Bd. 22 (2006), 1/2, S. 21-38

## Csuhaj-Varjú, Erzsébet; Dassow, Jürgen; Holzer, Markus

CD grammar systems with competence based entry conditions in their cooperation protocols In: International journal of computer mathematics. - London [u.a.]: Taylor and Francis, ISSN 0020-7160, Bd. 83 (2006), 2, S. 159-169 [Imp.fact.: 0.254]

#### Originalartikel in begutachteten nationalen Zeitschriften

#### Schmitt, Ingo; Nürnberger, Andreas

Inhaltsbasiertes multimedia Retrieval: Überblick und Herausforderungen In: Datenbank-Spektrum: Zeitschrift für Datenbanktechnologie; Organ der Fachgruppe Datenbanken der Gesellschaft für Informatik e.V.. - Heidelberg: dpunkt.verl., ISSN 1618-2162, Bd. 6 (2006), 18, S. 6-13

#### Originalartikel in begutachteten zeitschriftenartigen Reihen

## Amelung, Mario; Piotrowski, Michael; Rösner, Dietmar

EduComponents: experiences in e-assessment in computer science education In: Conference on Innovation and Technology in Computer Science Education: Proceedings of the 11th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education: June 26 - 28, 2006, Bologna, Italy. - New York, NY, (2006), S. 88-92 (SIGCSE bulletin; 38.2006,3)

Bade, Korinna; De Luca, Ernesto W.; Nürnberger, Andreas; Stober, Sebastian

CARSA: an architecture for the development of context adaptive retrieval systems In: Adaptive multimedia retrieval: user, context, and feedback: third international workshop, AMR 2005, Glasgow, UK, July 28-29, 2005; revised selected papers. - Berlin [u.a.]: Springer, (2006), S. 91-101

## Dassow, Jürgen; Vaszil, György

Ciliate bio-operations on finite string multisets

In: Developments in language theory: 10th international conference, DLT 2006, Santa Barbara, CA, USA, June 26 - 29, 2006; proceedings. - Berlin [u.a.]: Springer, (2006), S. 168-179 (Lecture notes in computer science; 4036)

## Kruse, Rudolf; Döring, Christian

Computational Intelligence: Stand und Perspektiven

In: Sport und Informatik IX: Bericht zum 6. Workshop Sportinformatik der DVS-Sektion Sportinformatik an der Otto-von-Guericke-Universität Magdeburg vom 22. - 24. Juni 2006.

- Aachen: Shaker, (2006), S. 29-36 (Berichte aus der Sportwissenschaft)

## Rösner, Dietmar; Amelung, Mario; Piotrowski, Michael

E-Learning-Komponenten zur Intensivierung der Übungen in der Informatik-Lehre: ein Erfahrungsbericht

In: Fachtagung Hochschuldidaktik der Informatik : HDI 2006, Hochschuldidaktik der Informatik: Organisation, Curricula, Erfahrungen; 2.GI-Fachtagung, 7. - 8. 12. 2006 in München, Germany.

- Bonn: Ges. für Informatik, (2006), S. 89-102 (GI-Edition: Proceedings; 100)

## Wissenschaftliche Monografien

#### Kunze, Manuela

Linguistische Analysen für die semantische Auszeichnung natürlichsprachlicher Dokumente In: München: Verl. Dr. Hut, 2006. - VII, 220 S. : graph. Darst.Zugl.: Magdeburg, Univ., Fak. für Informatik, Diss., 2006

#### Herausgeberschaften

# Spiliopoulou, Myra; Kruse, Rudolf; Borgelt, Christin; Nürnberger, Andreas; Gaul, Wolfgang

From data and information analysis to knowledge engineering: proceedings of the 29th annual conference of the Gesellschaft für Klassifikation e.V., University of Magdeburg, March 9-11, 2005

Berlin [u.a.]: Springer, 2006. - XIX, 761 S.: Ill.. - (Studies in classification, data analysis, and knowledge organization)

#### Buchbeiträge

### Dassow, Jürgen; Martin-Vide, Carlos; Mitrana, Victor

Some remarks on homogeneous generating networks of free evolutionary processors In: Formal models, languages and applications. - Singapore [u.a.]: World Scientific, (2006), S.

# 65-78 (Series in machine perception and artificial intelligence; 66)

# Andere Materialien

# **Gnjatovic, Milan**

DietmarRösner

In: Proceedings of the International Symposium on Linguistic Patterns in Spontaneous Speech, 17-18 Nov. 2006, Academia Sinica, Taipei, (2006), S. 55-66