



2008

JAHRESBERICHT

Auszug



Fraunhofer Institut
Intelligente Analyse- und
Informationssysteme



Das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS erforscht und entwickelt innovative Systeme, um Daten zu analysieren und Informationen zu erschließen. Wir realisieren einerseits Anwendungslösungen aus den Bereichen Data Mining, Business Intelligence oder High Resolution Management zur Optimierung von Produkten, Dienstleistungen und Prozessen. Zum anderen entwickeln wir Systeme, die große Datenmengen (Data Warehouses) durch innovative mediale Präsentation erschließen helfen (Wissensextraktion, interaktive Exploration, Wissensmanagement, Medienanalyse und Visualisierung).

Mit seinen Mitarbeitern bündelt das Fraunhofer IAIS Kompetenzen und wissenschaftliche Qualitäten aller ingenieurwissenschaftlichen Disziplinen, insbesondere der Informatik sowie der Mathematik, Naturwissenschaften, Betriebswirtschaft, Geo- und Sozialwissenschaften mit profundem Branchenwissen.

Die Geschäftsfelder des Instituts werden auf die Bedarfe der Kunden, Auftraggeber und Entwicklungspartner ausgerichtet. Sie werden sowohl branchenorientiert (zum Beispiel auf Handel, Medien, Energiewirtschaft etc.) als auch funktionsorientiert (zum Beispiel auf Marktforschung, Controlling, Prozessoptimierung, Automotive etc.) definiert. Da die Kundenbedarfe einer mannigfaltigen Marktdynamik unterworfen sind, werden die Geschäftsfelder kontinuierlich beobachtet und je nach Bedarf für die Zukunft neu ausgerichtet. Derzeit bestimmen folgende Geschäftsfelder die Forschungs- und Entwicklungsarbeit des IAIS:

- Marketing, Marktforschung & Mediaanalyse
- Unternehmensplanung & Controlling
- Digital Media Asset Management
- Process Intelligence
- Preventive Security
- High-Tech-Erlebnisräume

Multimediale Daten (Text-, Ton-, Audio-, Bild- und Videodaten) gewinnen im Zeitalter vereinfachter Produktions- und Distributionsmöglichkeiten über Breitbandnetze in Unternehmen und Organisationen zunehmend an Bedeutung.

Hier einige Beispiele:

- User Generated Content – Nutzer werden zu Produzenten und stellen ihre vielfältigen Inhalte anderen Nutzern zur Verfügung. In das Videoportal Youtube werden täglich tausende von Videos hochgeladen und abgerufen
- Mediatheken – Im Zuge der vermehrten Nutzung des Internets bieten Verlage und Fernsehsender ihre audiovisuellen Inhalte über markenbildende Internetportale (zum Beispiel die Mediathek des ZDF) an
- Medienarchive – Die Digitalisierung hat in den verschiedensten Bereichen der Medienproduktion (Fernsehsender, Verlage, Internetanbieter) bereits Einzug gehalten. Um die Nutzung von analogen Daten zu ermöglichen, gilt es, die Archive in die digitalen Produktions- und Distributionsumgebungen einzubinden.

Das Geschäftsfeld Digital Media Asset Management des Fraunhofer IAIS richtet sich an Unternehmen und Einrichtungen, die ihre Mediendaten einer effizienten Nutzung zuführen wollen und vor den bereits genannten Herausforderungen stehen. Vor allem das Thema der Suchmaschinentechnologien und -anwendungen für multimediale Daten steht im Zentrum des Geschäftsfelds Digital Media Asset Management. Die immer größer und komplexer werdenden Medienbestände erfordern neue Technologien und Ansätze, um eine automatische Erschließung und Indizierung von Text-, Ton-, Audio-, Bild- und Videodokumenten zu ermöglichen. Dazu werden Verfahren aus der Mustererkennung für die Analyse der multimedialen Dokumente entwickelt und an die jeweiligen Anforderungen des Kunden angepasst. Das Leistungsspektrum des Geschäftsfelds reicht von der Lizenzierung innovativer Verfahren zur Medienererschließung über die Entwicklung kundenspezifischer Softwarelösungen bis hin zu Dienstleistungen wie die Digitalisierung und Erschließung umfangreicher Medienbestände.



Die nachfolgend beschriebenen Projekte zeigen die Lösungskompetenzen des Geschäftsfelds. In dem vom Bundesministerium für Wirtschaft und Technologie (BMWi) geförderten Theseus-Projekt Contentus werden Komponenten und Systeme zur automatischen Analyse, Erschließung von gedruckten Dokumenten und Sprachdaten erforscht, die für den Einsatz in der Deutschen Nationalbibliothek und in den Mediatheken der Rundfunksender zum Einsatz kommen. In dem EU Projekt Live Staging of Media Events werden Technologien und Konzepte für die Verarbeitung von live produzierten Fernsehinhalten entwickelt, die beim ORF während der Produktion der Olympischen Spiele 2008 verwendet wurden. Das Audiomining stellt eine lizenzierbare SW-Lösung zur automatischen Erschließung und Suche von umfangreichen Audio- und Sprachdaten in Mediatheken, Medienarchiven oder Aufzeichnungen in Callcentern dar. Innerhalb des Geschäftsfelds werden derzeit mehrere Projekte für Zeitungsverlage durchgeführt, um die bisher analog vorliegenden Zeitungsarchive an die digitalen Redaktionsprozesse anzubinden und dem Leser die Möglichkeit einzuräumen, in der Geschichte zu »surfen«.

CONTENTUS – TECHNOLOGIEN FÜR DIGITALE BIBLIOTHEKEN DER NÄCHSTEN GENERATION

In Bibliotheken, Archiven und Datenbanken liegen die Wissensschätze unserer Gesellschaft. Digitalisierung, Metadaten und das Internet haben deren Nutzung vereinfacht. Die inhaltliche Erschließung der Dokumente und damit die Unterstützung bei der inhaltlichen Recherche stellt für die Informationstechnik jedoch nach wie vor eine gewaltige Herausforderung dar. Im Rahmen des Projekts CONTENTUS arbeiten wir am Fraunhofer IAIS an Lösungen für eben diese Probleme.

Im Rahmen des Projekts wurden Verfahren für die automatische Aufbereitung und Analyse von archivierten multimedialen Inhalten entwickelt. Hierzu gehören gedruckte Dokumente, gesprochene Audiodokumente, Musikdokumente sowie Videosequenzen und Fotos. Ein weiterer Schwerpunkt lag in der semantischen Annotation und damit der eigentlichen inhaltlichen Erfassung der Bestände. Insgesamt ist ein modularer und effizienter Workflow von der Digitalisierung der Inhalte bis zu ihrer semantischen Vernetzung erstellt worden. Der Anwender wird bei der thematischen Recherche in den erfassten Beständen durch die Nutzung von Wissensmodellen unter Einbeziehung bestehender Systematiken weit über das derzeit in Bibliotheks- und Archivsystemen verfügbare Maß hinaus unterstützt.

Ergebnisse der aktuellen Projektphase zur semantischen Suche in gedruckten Dokumenten wurden auf der Frankfurter Buchmesse im Oktober 2008 demonstriert. Hier haben unsere

Wissenschaftler den automatischen Workflow zur Analyse von digitalisierten Büchern realisiert und auf Jahreschroniken der Jahre 2005 und 2006 angewendet. Dazu haben wir bestehende Module zur Dokumentenanalyse genutzt sowie neuartige Verfahren (zum Beispiel zur Artikel-separierung, Entitätenerkennung und Disambiguierung) entwickelt. Um eine Verknüpfung der Chroniken mit weiteren Hintergrundinformationen zu ermöglichen, wurden die Ergebnisse der Dokumentenanalyse mit elektronischen Artikeln der Neuen Zürcher Zeitung und Pressemitteilungen der DPA des Jahres 2006 verknüpft und mittels statistischer Verfahren die relevanten Entitäten wie Personen- und Ortsnamen detektiert. Dies ist wichtig, um eine automatische Verknüpfung der Entitäten mit Hintergrundwissen zu ermöglichen.



Ergebnis der automatischen logischen Segmentierung einer Buchseite für den Demonstrator auf der Buchmesse 2008 in Frankfurt. Die Beiträge auf der Seite sind mit unterschiedlichen Farben markiert.



Grafisches Benutzer-Interface für den CONTENTUS Demonstrator zur Frankfurter Buchmesse 2008.



Über **KOHL** höhnte **STRAUSS**: Er wird nie Kanzler werden Die Zeit 18.07.03



Mit dieser Erkennung kann eine gezielte Personensuche realisiert werden. Zur Vermeidung von Doppeldeutigkeiten bei Wörtern wurden spezielle Verfahren entwickelt und verwendet. Methoden des Text Minings und des Semantic Webs verknüpfen relevante Informationen aus der Wikipedia und der DBpedia mit dem Datencorpus für den Buchmessedemonstrator. Dieser Demonstrator war ein integraler Bestandteil des CONTENTUS-Stands innerhalb des Messeauftritts der Deutschen Nationalbibliothek auf der Buchmesse. Die Projektpartner haben hier die Vorteile der semantisch suchbaren elektronischen Version dem physikalischen Buch direkt gegenüber gestellt.

Begriffe der Sprache sind häufig mehrdeutig. Für eine semantische Suche ist daher die korrekte Identifizierung der Bedeutung eines Wortes wichtig.

Projektpartner des Fraunhofer IAIS sind:

- empolis GmbH
- Deutsche Nationalbibliothek (DNB)
- Deutsche Thompson OHG (DTO)
- Institut für Rundfunktechnik GmbH (IRT)
- m2any GmbH
- moresophy GmbH
- Mufin GmbH

Das Projekt CONTENTUS wird vom Bundesministerium für Wirtschaft und Technologie (BMWi) im Rahmen des Forschungsprogramms THESEUS gefördert.



PROJEKTMITWIRKENDE

Doris Baum, Lars Bröcker, **Stefan Eickeler**, Iuliu Konya, Lukas Molzberger, Stefan Paal, Gerhard Paaß, Jochen Schwenninger, Christoph Seibert, Thomas Tikwinski, Michael May, Joachim Köhler

Materialauswahlmöglichkeiten in CONTENTUS.
© Deutsches Musikarchiv der Deutschen Nationalbibliothek in Berlin

REDAKTION

Hanne Bendel
Dipl.-Ing. Sabine Nebelung (Leitung)
Presse und Public Relations

GESTALTUNG

Daniela Plath
Annika Zeyen
Mediengestalterinnen Digital- und Printmedien Fraunhofer IAIS

FOTOGRAFIEN

Oliver Diez
Mediengestalter Bild und Ton Fraunhofer IAIS

BILDNACHWEIS

Bild Seite 47: Deutsches Musikarchiv der Deutschen Nationalbibliothek in Berlin
Bild Seite 79: Maurer, United Architects
Alle weiteren Fotos: Fraunhofer IAIS

DRUCK

Warlich Druck Meckenheim GmbH
www.warlich.de

ANSCHRIFT DER REDAKTION

Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
Schloss Birlinghoven
53754 Sankt Augustin
Telefon: 02241-14 2252
Fax: 02241-14 2381
info@iais.fraunhofer.de
www.iais.fraunhofer.de

Bestellung von Publikationen bitte per E-Mail oder per Fax.

Bei Abdruck ist die Einwilligung der Redaktion erforderlich.

© Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS,
Sankt Augustin, 2009