



# Informationssuche in Multimedia-Dokumenten

---

Auf dem Weg  
zu einem Hochleistungs-Suchserver

Alois Knoll  
Ingo Glöckner  
Universität Bielefeld  
Technische Fakultät  
Arbeitsgruppe Technische Informatik

Hermann Helbig  
Fern-Universität Hagen

Wissen wird im Zeitalter des Internet und der multimedialen Information in immer größeren Mengen und in immer vielfältigeren Präsentationsformen verfügbar. Was aber gleichzeitig beständig schwieriger wird, ist die gezielte Suche nach bestimmten Informationen in der multimedialen Datenflut. Eine Forschungsgruppe mit Wissenschaftlern aus fünf nordrheinwestfälischen Hochschulen hat sich zum Ziel gesetzt, Methoden für die Konstruktion eines Hochleistungs-Suchservers zu entwickeln, der imstande ist, Benutzeranfragen so umzusetzen, daß eine gezielte Suche nach bestimmten Inhalten nicht nur in Textdokumenten, sondern etwa auch in Bildern, Filmen und Tondokumenten möglich ist. Darüber hinaus soll eine Bewertung und Verknüpfung der gefundenen Informationen möglich werden, und das Suchergebnis soll in einer für den menschlichen Wahrnehmungsapparat besonders geeigneten Form präsentiert werden. Erstmals wird hier eine durchgängige System-Architektur für die nutzerfreundliche Informationssuche entwickelt.

Wissen wird durch die immer komfortableren Möglichkeiten zur elektronischen Publikation zunehmend in Form von Volltext-Dokumenten verfügbar. Durch Fortschritte in der Multimedia-Technik bestehen die Dokumente heute immer seltener nur aus Text; vielmehr beinhalten sie eine Mischung ganz unterschiedlicher Präsentationsformen (Text, Formeln, Grafik, Bilder, Bildsequenzen, Ton) in den unterschiedlichsten Datenformaten. Über weltweit operierende Kommunikationsnetzwerke und -dienste werden diese in ihrer Zahl rasch wachsenden Dokumente zudem für jedermann immer leichter verfügbar. Nicht mehr der Zugang zu restriktiv verwalteten Daten, sondern der Suchprozeß in einer Überfülle angebotenen Materials und seine Aufbereitung für eine Nutzung, die auf den natürlichen Kommunikationsfähigkeiten des Menschen basiert, stellen deshalb das eigentliche Problem bei der Informationsbeschaffung dar.

Der potentielle Nutzer steht vor dem Problem, aus der riesigen Fülle des angebotenen Materials zielgerichtet das für ihn relevante auszuwählen. Dies wird für das Gros der Nutzer zukünftig nur möglich sein, wenn ihnen leicht bedienbare Werkzeuge an die Hand gegeben werden, die den Inhalt einer großen Zahl von Dokumenten automatisch soweit analysieren können, wie es für das Auffinden der vom Nutzer nachgefragten Information notwendig

ist. Daß sich die interessante Information innerhalb eines Dokumentes in jeder der erwähnten Präsentationsformen befinden kann (für die jeweils wieder sehr unterschiedliche Datenformate existieren), gibt dem Suchproblem zusätzliche Schärfe. Um die Datenflut für den Informationsnutzer beherrschbar zu machen, sind also völlig neuartige Werkzeuge erforderlich

- zur inhaltsbezogenen Formulierung von Anfragen auch durch nicht geschulte Benutzer,
- zum schnellen Aufspüren von relevanten Teilinformationen sowie interessanten Strukturen und Beziehungen in immens ausgedehnten, möglicherweise völlig unstrukturierten und in keiner Weise indexierten Datensammlungen, wie aber auch schon in kleinen Sammlungen heterogener Informationen,
- zur Vermittlung dieser aufgespürten Informationen an alle in Frage kommenden Benutzer (in der Form eines „brokers“),
- zur Bewertung, Deutung und Verknüpfung der aus verschiedenen Quellen stammenden Informationen,
- zur Aufbereitung der gefundenen Informationen in einer für Menschen leicht erfaßbaren Form (z.B. Visualisierung),

- zur Gewährleistung der Sicherheit der angebotenen Dienste, insbesondere im Hinblick auf Wahrung der Vertraulichkeit und der Integrität der Information, bei bezahlten Diensten auch im Hinblick auf die nachweisbare Nutzung von Informationen.

Systeme, die über die reine Stichwortsuche hinaus neben der Analyse des Inhalts von Text auch weitere Repräsentationsformen (bis auf weiteres sind dies hauptsächlich Bilder und Bildfolgen) in einen inhaltlichen Bezug zu den Informationswünschen des Nutzers bringen können, sind – wenn überhaupt – erst in Ansätzen erkennbar. Die gängigen Suchsysteme (etwa *OPAC* für Bibliotheksanfragen oder speziell die bekannten Internet-Suchmaschinen) können ganz offensichtlich nicht zu diesen Ansätzen gezählt werden. Im Gegenteil: Zumindest ungeübten Benutzern wird der Weg zu gezielten Antworten durch kryptische Anfragesprachen von vornherein versperrt.

#### ■ Das Ziel: ein Hochleistungs-Suchserver

Die Gruppe II des vom Ministerium für Wissenschaft und Forschung des Landes Nordrhein-Westfalen geförderten Forschungsverbunds „Die virtuelle Wissensfabrik“ hat es sich zur Aufgabe gemacht, Methoden zu entwickeln, die die Konstruktion eines Hochleistungs-Suchservers (*High Performance Query Server*) ermöglichen, welcher die oben genannten Anforderungen erfüllt. Neben der in den einzelnen Teilprojekten geleisteten Forschungsarbeit bezieht dieses Projekt seinen Innovationsgehalt aus der erstmaligen Konzeption einer durchgängigen Architektur für einen solchen Server zur nutzerfreundlichen Informationsextraktion. Die Gruppe II umfaßt Arbeitsgruppen aus fünf Hochschulen des Landes: der Rheinisch-Westfälischen Technischen Hochschule Aachen, der Universität Dortmund, der Fernuniversität Hagen, der Universität-Gesamthochschule Paderborn und der Universität Bielefeld.

Die Architektur beinhaltet den gesamten Weg von einer Anfrageschnittstelle in natürlicher Sprache bis hin zur Analyse mit Hilfe von Methoden der Datenfusion sowie ihrer Implementierung für eine Beispiel-domäne. (Diese kann prinzipiell gegen jede andere

Anzeige

Anwendungsdomäne ausgetauscht werden, sofern ein Experte behilflich ist, das domänenabhängige Wissen zu formalisieren.)

Dieses System ist ein Schritt in Richtung eines einfach benutzbaren Assistenten zur Informationsgewinnung auch für ein großes Publikum, welches nach anfänglicher Euphorie für die Technik des Internet bzw. des *WorldWideWeb* zunehmend den zielgerichteten Zugang und die geeignete Aufbereitung des Rohstoffs Information nachfragen wird.

#### ■ Umfeld, Zielsetzungen und Systemarchitektur

Ein Prototyp des Hochleistungs-Suchservers wird zunächst Dokumente bearbeiten können, welche neben Text auch Farbbilder und Farbbildfolgen enthalten können. Es gibt bereits eine kleine Anzahl von Systemen, die eine inhaltsbasierte Abfrage von Bilddokumenten bieten. Bei einigen dieser Systeme kann der Nutzer abfragen, ob es Regionen innerhalb des Bildes gibt, die bestimmte Farben, Texturen oder Formen aufweisen.

Bei anderen Systemen gibt der Nutzer bei seiner Anfrage ein Bild vor, und das System sucht daraufhin nach Bildern in seinem Bestand, die eine hohe Ähnlichkeit mit dem vorgegebenen Bild aufweisen. Der Nutzer kann dabei den Suchprozeß steuern, indem er den einzelnen Ähnlichkeitsparametern bestimmte Gewichte zuweist. Es existieren Systeme, die die Bildsuche mit Hilfe einer formalen Abfragesprache ermöglichen, sowie andere, die vom Nutzer eine Strichskizze verlangen.

Alle diese Systeme können ein Bild nicht in der Weise analysieren, wie es notwendig ist, um eine Anfrage nach dem Bedeutungsinhalt eines Bildes beantworten zu können. Mit anderen Worten:

Regionen, die aus Punkten einer bestimmten Farbe bestehen, können gefunden werden, die Anfrage nach dem Konzept „Blume“ ist nicht möglich.

Es gibt ein einziges – inzwischen auch kommerziell als *MediaMiner* erhältliches – System zur inhaltsbasierten Bildsuche, welches die Eingabe von Stichworten erlaubt. Es transformiert Bilder in einen beschreibenden Text und führt die Stichwortsuche auf diesem Text aus. Es arbeitet auf CAD-Bildern und auf Landschaftsbildern mit eindeutigen Merkmalen (Seen, Gebirgen, Wäldern etc.). Auch für andere sehr begrenzte Inhaltsdomänen gibt es einfache Bildsuchsysteme.

Die Architektur unseres Hochleistungs-Suchservers soll demgegenüber ohne prinzipielle Einschränkung auf eine bestimmte Domäne natürlichsprachlich formulierte Anfragen an eine Datenbank aus Objekten aller Repräsentationsformen verarbeiten. Dazu ist es notwendig, den Informationsgehalt von multimediale Dokumenten zu analysieren und mit der Anfrage auf der Ebene der Bedeutung zu vergleichen. Es sollen ferner verschiedene Dokumentklassen miteinander fusioniert werden, d.h. Bezüge unter ihnen sollen ausgewertet werden. Das System soll außerdem Anfragen beantworten können, die nicht mit Hilfe eines Index zu beantworten sind, welcher bei der Einstellung des Dokuments in das System erzeugt wird. Es muß dazu also über die Fähigkeit zur on-line-Suche verfügen.

Ziel des Projekts ist die Realisierung der modularen Architektur nach Bild 1, die es erstmalig erlauben wird, die folgenden Leistungen anzubieten:

- Extraktion semantischer Information aus Anfragen und Dokumenten,
  - Vergleich des Bedeutungsgehalts der Anfrage und der Dokumente im Dokumentenbestand,
  - Auswertung der Bezüge zwischen Text und Bildinformation innerhalb eines Dokuments oder über Dokumentgrenzen hinweg,
  - effiziente on-line-Suche auch in komplexen Dokumentenbeständen,
  - Abliefern von Suchergebnissen in der gleichen Form, wie sie als Dokumente in der Datenbank gespeichert sind oder Erzeugung neuer Dokumente zur spezifischen Beantwortung der Anfrage.
- Auf nationaler und auch internationaler Ebene sind uns direkt konkurrierende Projekte nicht bekannt. In der in der ersten Projektphase festgelegten Modulstruktur in Bild 1 des Hochleistungs-Suchservers werden diese Eigenschaften durch fünf Hauptmodule realisiert:

#### ■ Natürlichsprachlicher Zugang

Dieser Zugang zum System ist im Gegensatz zu den meisten traditionellen Nutzerschnittstellen auf ein tiefes Sprachverstehen ausgerichtet, das eine morphosyntaktische Analyse und eine semantische Interpretation umfaßt. Das Analysesystem hat die Aufgabe, die beiden teilweise im Konflikt miteinander stehenden Ziele der Robustheit und der semantischen Tiefe bei der Analyse natürlicher Sprache zu kombinieren. Der Konflikt besteht darin, daß Robust-

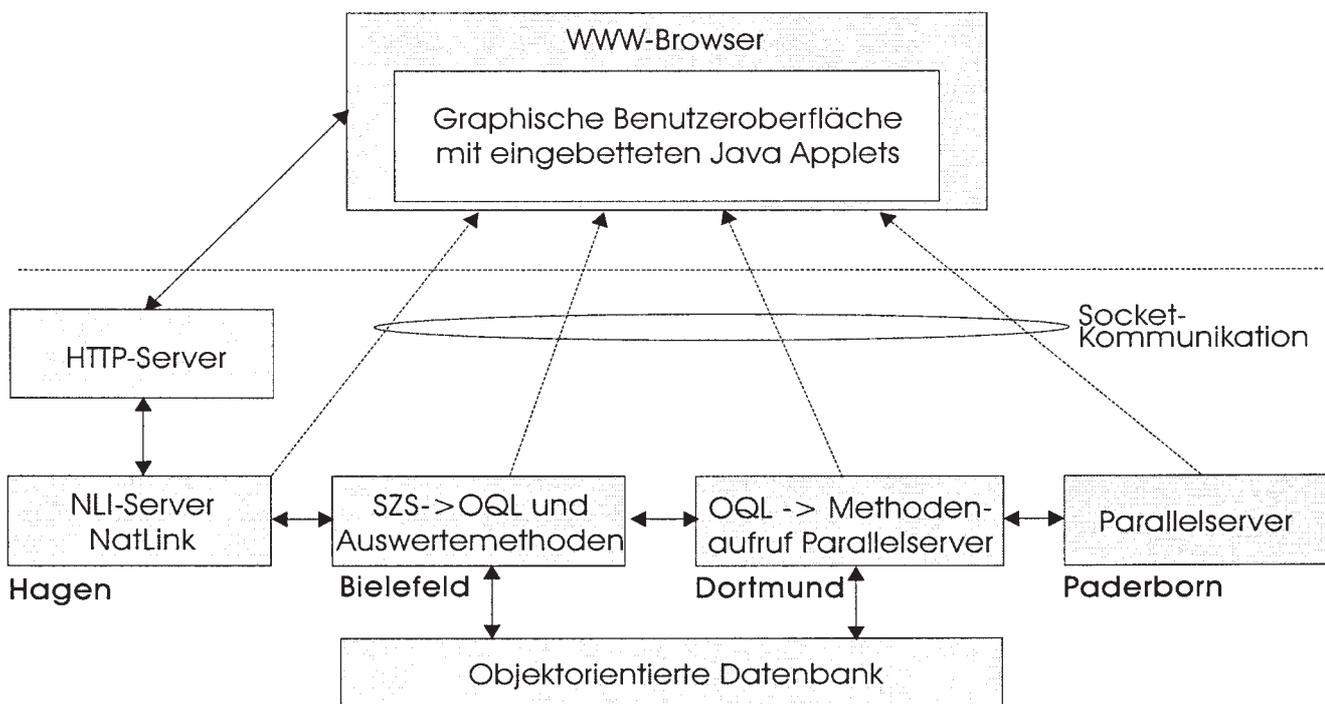


Bild 1: Die Architektur des Hochleistungs-Suchservers.

heit oft zu unvollständiger Information führt (zum Beispiel bei der Behandlung unbekannter Wörter, Komposita, Ad-Hoc-Formulierungen), während semantische Tiefe gerade vollständige semantische Information anstrebt. Um dieses Problem zu lösen, müssen verschiedene neuartige Methoden entwickelt werden. Dies sind etwa mehrschichtige erweiterte semantische Netze, welche die Semantik von Äußerungen mit unterschiedlicher Granularität und Spezifität darstellen können, zum Beispiel durch hierarchisierte semantische Relationen und Funktionen. Resultat des Analyseschritts, bei dem ein Fachlexikon und Thesauri sowie Hintergrundwissen über die Domäne verwendet werden, ist die Repräsentation der natürlichsprachlichen Anfragen in einer semantischen Zwischensprache.

#### ■ Retrieval-Modul

Aufgabe dieses Moduls ist es, aus der in der semantischen Zwischensprache vorliegenden Nutzeranfrage mit Hilfe von Transformations- und Interpretationswissen eine Anfrage auf der Ebene der formalen Retrieval-Repräsentation zu erzeugen und deren Abarbeitung zu steuern. Dies geschieht durch eine Erweiterung des Objektdatenbankschemas um generische Klassen, die fuzzy-logische Verknüpfungen, unscharfe Methoden zur Dokumentauswertung sowie informationskombinierende Operatoren zur Verfügung stellen.

Die Verwendung von Fuzzy-Methoden dient zum einen dazu, die inhärente Unschärfe der natürlichsprachlichen Anfragen zu erfassen, die meist nur ungefähr ausdrücken, was der Nutzer letztendlich sucht. Zum anderen ist durch die Fuzzy-Methoden eine graduelle Entsprechung zwischen den abstrakten Konzepten der natürlichen Sprache und den low-level-Attributen der multimedialen Datenbank bzw. der verfügbaren elementaren Auswertungsmethoden modellierbar. Sie ermöglichen zudem die Einbeziehung assoziativer Suche (über Fuzzy Thesauri) und approximativer Suche (etwa zur robusten Verarbeitung von Schreibfehlern).

#### ■ Query-Mediator-Modul

Als lokale Komponente eines föderierten Systems benötigt der Hochleistungs-Suchserver ein anpaßbares und kommunikationsfähiges Multimedia-Schema zur (Selbst-)Beschreibung der vorliegenden Multimedia-Strukturen, der zu bewahrenden semantischen Bedingungen und der zu beachtenden Sicherheitsbedingungen. Der Query-Mediator realisiert ein solches Schema als Instanz eines Metakonzeptes für die Komponenten eines Multimedia-Systems. Er transformiert die Anfragen des Retrieval-Moduls in parallele (Teil-) Abfragen bzw. Methodenaufrufe an die objektorientierte Datenbank, d.h. er erzeugt Pläne für die Durchführung von Anfragen auf lokalen und/oder externen Systemen. Er ist ferner zuständig für die Durchsetzung von Sicherheitsbedingungen, die in einem föderierten System eine wesentliche Rolle spielen.

#### ■ Parallelservers

Die Aufgabe des parallelen Mediaservers besteht darin, sowohl die Massendatenobjekte zu verwalten und auszuliefern, als auch die in anderen Teilprojekten entwickelten Verfahren zur schnellen Suche auf den Massendatenobjekten zu integrieren. Er stellt somit die wesentliche Ausführungsplattform des Hochleistungs-Suchservers dar. Der Mediaserver basiert auf einem skalierbaren Prozessornetzwerk. Durch die Parallelisierung der Auslieferung von Daten sowie die Parallelisierung der Suchverfahren kann ein wesentlich höherer Durchsatz von Anfragen erzielt werden. Dabei ergeben sich eine Reihe von Fragestellungen, die durch den Zugriff auf gemeinsame Ressourcen (Festplatten, Kommunikationsverbindungen) innerhalb des Parallelrechnersystems bedingt sind. Diese Problemstellungen sind durch geeignete Scheduling- und Lastbalancierungsverfahren zu lösen und stellen aufgrund ihrer Komplexität neue Herausforderungen dar, die in der parallelen Datenverarbeitung bislang noch nicht behandelt wurden.

## ■ Suchprozessoren

Die on-line-Suche in großen und komplexen Datenbeständen läßt sich nur verwirklichen, wenn die häufig wiederkehrenden Teilberechnungen bei der Textsuche bzw. bei der Bildanalyse in dedizierter Hardware ausgeführt werden. Es sind zu diesem Zweck zwei unterschiedliche Prozessortypen vorgesehen: ein spezieller Bildprozessor und ein spezieller Textprozessor. Beide können über Hochleistungsverbindungen direkte Speichertransfers in die Prozessoren des Parallelservers ausführen. Prozesortechnik und optimierte Speichertechnik beschleunigen die Textsuche gegenüber einer Lösung mit Standard-Prozessoren etwa um den Faktor 1000. Etwa den gleichen Geschwindigkeitsvorteil gegenüber Software-Lösungen ermöglicht der Bildprozessor. Er realisiert die wichtigsten Algorithmen für die Farbbildverarbeitung (unter anderem zweidimensionale Faltung, Medianfilter und Histogrammbildung). Steht entsprechende Übertragungsbandbreite zur Verfügung (etwa über zwei PCI-Bus-Verbindungen), dann kann ein Prozessor etwa 25 Bilder pro Sekunde mit der Bildqualität von hochauflösendem Fernsehen filtern.

## ■ Aufbereitung und Darstellung der Ergebnisse

Dieses Modul hat die Aufgabe, die Benutzereingaben zu akzeptieren, vor allem aber die Ergebnisse in kognitiv adäquater Weise darzustellen. Dies bedeutet in Abhängigkeit von der Anfrage entweder, daß ein Dokument – abgesehen von ansprechender graphischer Aufbereitung – in genau der Form präsentiert wird, in der es sich in der Datenbank befindet. Alternativ wird aus dem Suchergebnis (das aus mehreren Teildokumenten berechnet wurde) ein Ergebnisdokument erzeugt, das spezifisch die Nutzeranfrage beantwortet, aber eine ganz andere Repräsentationsform als die gefundenen Dokumente hat. Mit anderen Worten: Es wird neue, auf die Anfrage zugeschnittene Information erzeugt, also in gewisser Form neues Wissen generiert.

Die Festlegung auf eine bestimmte Domäne geschieht ausschließlich über das Computerlexikon und die Sammlung der domänenspezifischen Transformations- und Auswertungsregeln. Die physikalische Kommunikation der Module erfolgt über IP-Verbindungen bzw. direkte Bus-Interfaces, die logische Anbindung über formal spezifizierte Schnittstellen.

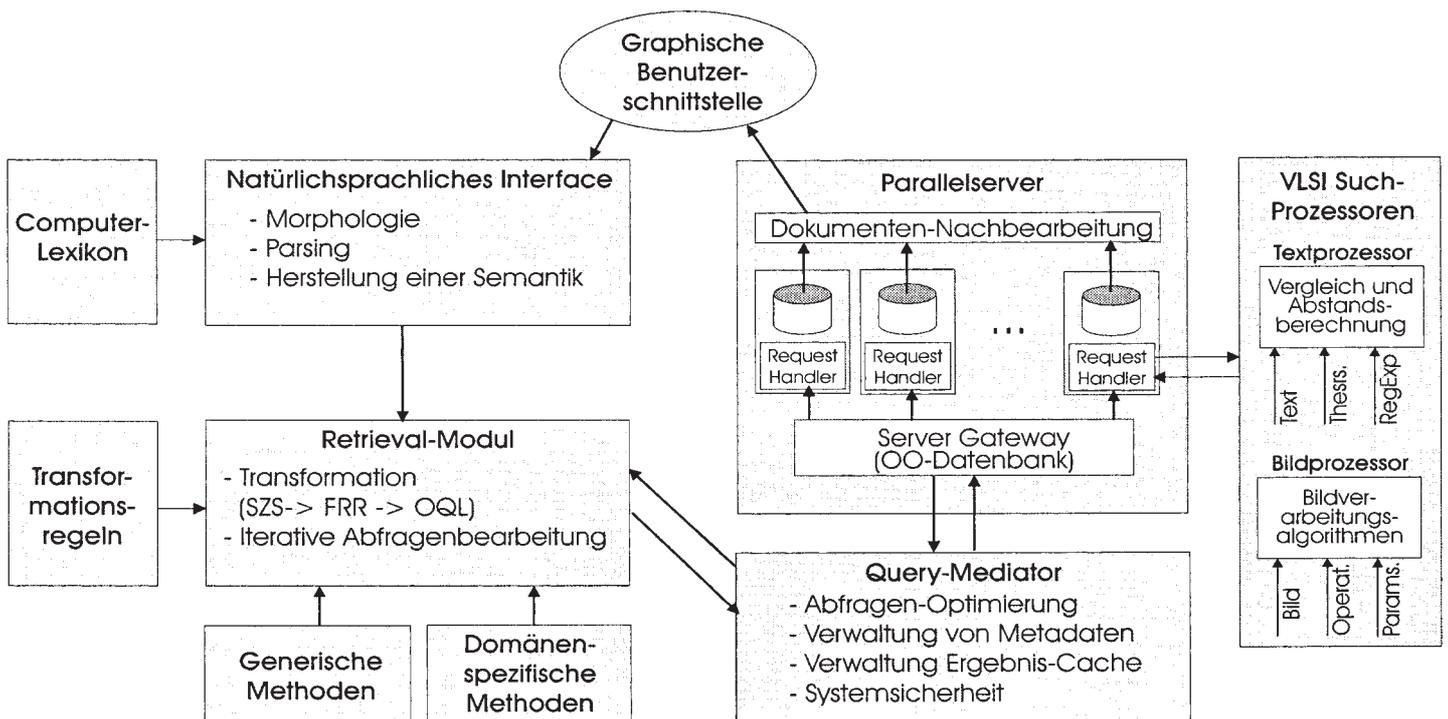


Bild 2: Die Struktur des Prototypen für den Hochleistungs-Suchserver

## Anzeige

### ■ Wissensdomäne

Nach umfangreicher Untersuchung verschiedener Möglichkeiten zur Festlegung einer Domäne für die Abfragen aus dem Bereich der Wirtschaft, der Medizin, der Geowissenschaften, der Werkstoffwissenschaften und der Paläontologie wurde als Hauptdomäne die Meteorologie festgelegt (die ursprüngliche Domäne aus der Paläontologie wird zu Kontrollzwecken weiterverfolgt). Der Hochleistungs-Suchserver wird dementsprechend darauf ausgerichtet, multimediale Information über Wetterlage und -entwicklung zu halten und darüber Anfragen zuzulassen. Neben der Tatsache, daß sich jedermann mit dem Wetter identifizieren und dementsprechend sinnvolle Anfragen formulieren kann, ohne Spezialist sein zu müssen, sprechen weitere Gründe für diese Wahl:

- Die Daten sind heterogen und reichhaltig. Berücksichtigt werden könnten prinzipiell Dokumente wie Satellitenbilder, Wetterkarten, grafische Wetteranalysen, Wetterberichte und tabellarische Wetterdaten, Bildfolgen (Fernseh-Wetterberichte, Satellitenfilme) und Audio-Daten (Radio-Wetterbericht).
- Es gibt Korrespondenzen zwischen Text (Wetterlage, Vorhersage, Temperaturangaben) und Bild

(Karten, Zahlenkolonnen), die nicht redundant sind (im Bild befindet sich mehr spezifische Information als im Text und umgekehrt).

- Die Bilddokumente sind verhältnismäßig leicht analysierbar, damit entfällt die Notwendigkeit, von vornherein einen Experten hinzuziehen zu müssen.
- Bei Anfragen sind alle Fragenkategorien denkbar.

In Bielefeld wird seit August 1996 eine umfangreiche Datenbasis aus Wetterdokumenten (Satellitenbilder, Temperatur-, Druck-, Feuchtigkeitswerte) unterhalten und ständig weiter gepflegt.

Es stehen ferner nach Verhandlungen mit dem Deutschen Wetterdienst historische Daten von 18 Wetterstationen aus dem Zeitraum 1992 bis 1996 zur Verfügung, die zumindest die Überprüfung der prinzipiellen Korrektheit der Analysemethoden und eine Schätzung von Methodenparametern zulassen.

Um die auf dieser Domäne möglichen Fragen zu illustrieren, seien im folgenden einige Fragen aus einem Katalog verschiedenster Fragenklassen (die wiederum auf unterschiedliche Weise parametrisiert sein können) genannt:

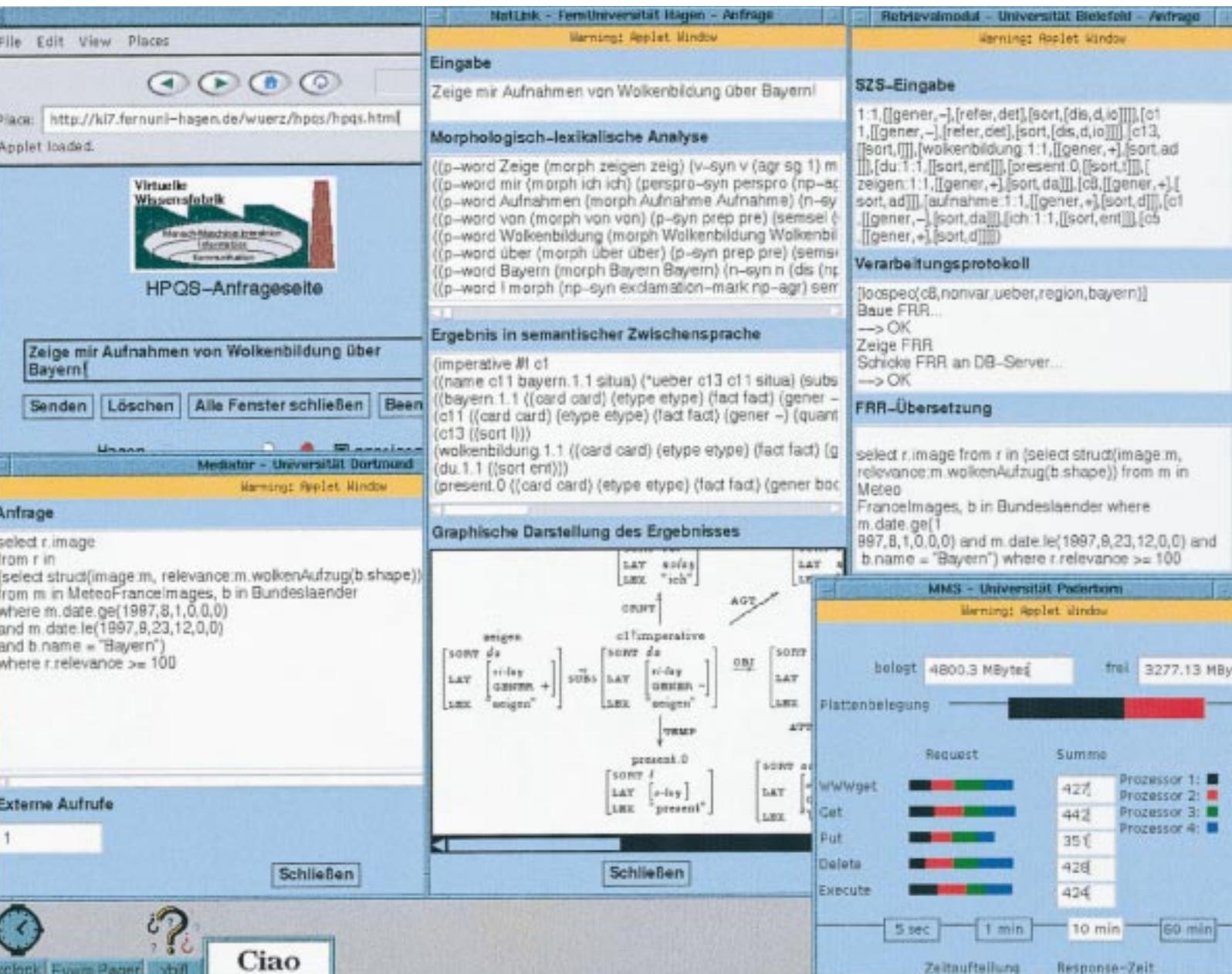


Bild 3: Die Kontrollausgaben auf dem Bildschirm zu einer Anfrage nach der Woleknbildung über Bayern.

■ Typische Fragen, die der Benutzer dem System stellen kann

- Wo war es gestern in Nordrhein-Westfalen am wärmsten?
- In welchem Bundesland war es gestern am kältesten?
- Wie hoch war im Juli die durchschnittliche Tagestemperatur in Aachen?
- Berechne die Zeitreihe der Tageshöchsttemperaturen in Aachen für April 1997!
- Wo regnete es heute vor einem Jahr?
- An wie vielen Tagen war es diesen Monat in Bielefeld sonnig?
- Zeige mir alle Wolken auf dem Infrarot-Satellitenbild von gestern 12 Uhr!
- Ist es in Bielefeld öfter bewölkt als in Aachen?
- Zeige mir Wetterkarten zum letzten DWD-Wetterbericht!

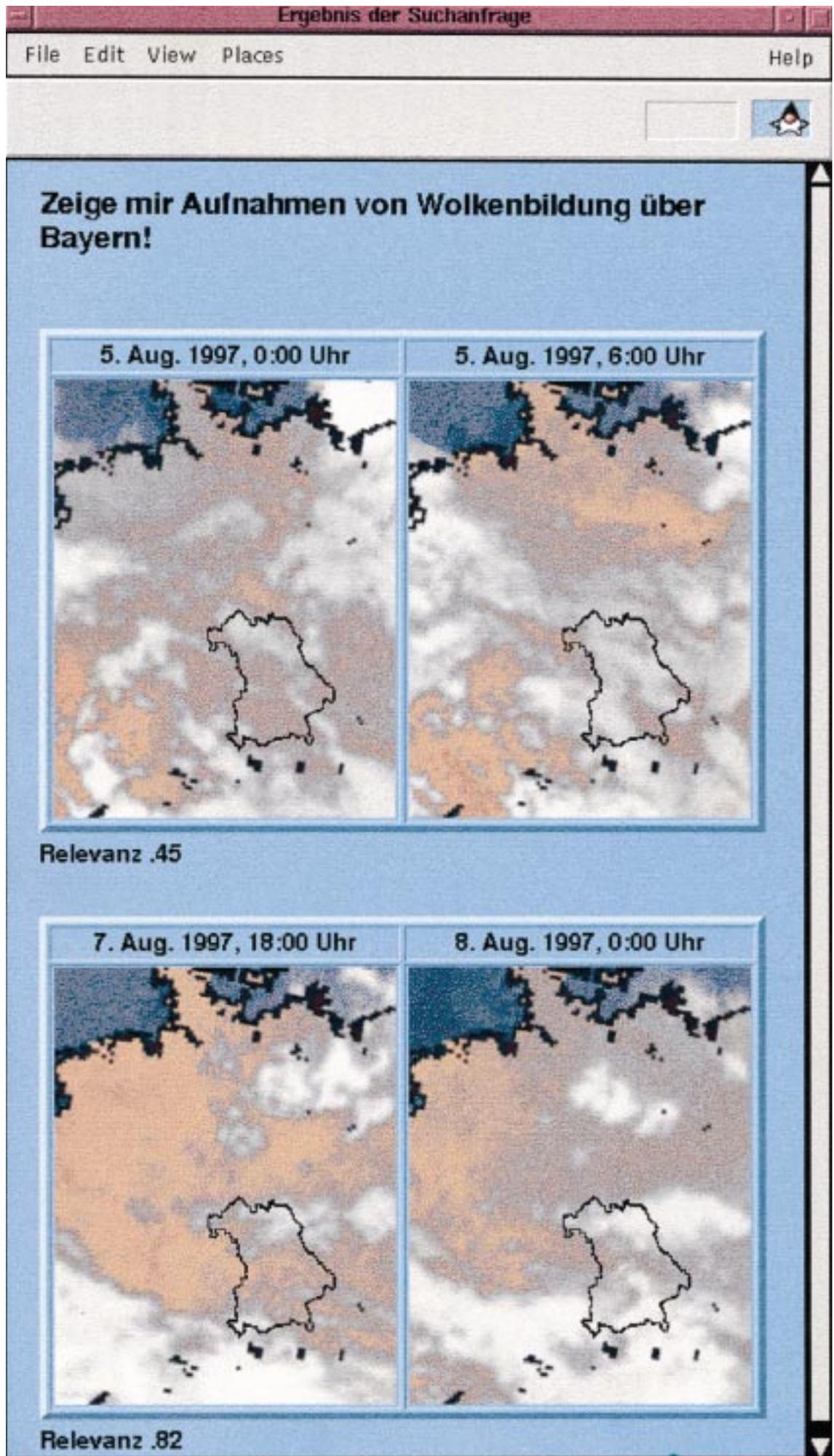


Bild 4: Das Ergebnis der Anfrage zur Wolkenbildung über Bayern, wie es anhand des Bielefelder Datensatzes ermittelt wurde. Die Umrisse von Bayern sind eingezeichnet. Graue und weiße Zonen bedeuten Wolken, braune Bereiche sind wolkenfrei.

## ■ Prototyp

Gegenwärtig befindet sich der erste Prototyp in der Phase der Integration. Er soll unter anderem die oben angegebenen Beispielfragen verarbeiten können. Diese Beispielfragen stellen an alle Module, speziell aber an die Sprachanalyse und insbesondere das Retrieval-Modul bereits hohe Anforderungen, weil sinnvolle Antworten nur gegeben werden können, wenn bereits ein großer Satz ausgetesteter Methoden zur Bildinterpretation und zur Abfragenverarbeitung zur Verfügung steht.

Wie in Bild 2 gezeigt, werden die Einzelmodule des Prototypen an den Orten betrieben, an welchen sie entwickelt wurden; die Paderborner Massendatenbank wird regelmäßig von Bielefeld aus auf den neuesten Stand gebracht. Die Benutzung des Prototypen erfolgt über eine WWW-Browser-Schnittstelle. Diese in Hagen entwickelte Schnittstelle stellt jedem Modul über Java-Applets eine Socket-Kommunikationsschnittstelle zur Verfügung, mit deren Hilfe es seine spezifischen Meldungen und Ergebnisse präsentieren kann.

Bild 3 stellt den aus einer typischen Frage (hier nach der Wolkenbildung über Bayern) resultierenden Aufbau der Kontrollausgaben auf einem Bildschirm dar. Man erkennt das WWW-Browser-Fenster für die Eingabe der Anfrage und rechts daneben verschiedene Fenster, in welchem das Resultat der syntaktisch-semanticen Analyse der Anfrage sowohl in textueller als auch in graphischer Form dargestellt wird. Im Fenster ganz rechts wird der Fortschritt der Übersetzung der formalisierten Anfrage in die Formale-Retrieval-Repräsentation angezeigt. Diese Repräsentation ist geeignet, als Anfrage auf dem Parallelserver abgearbeitet zu werden; die dynamische Auslastung dieses Server zeigt das Bildschirm-Fenster rechts unten auf dem Schirm.

Bild 4 schließlich gibt das Ergebnis der Anfrage wieder, wie es anhand des in Bielefeld vorhandenen Datensatzes ermittelt wurde. Eingezeichnet sind die Umrisse von Bayern; braune Bereiche sind wolkenfrei, während graue und weiße Zonen Wolken darstellen. Es ist gut zu erkennen, daß das System tatsächlich automatisch aus dem gesamten Datensatz zwei jeweils im sechsstündigen Abstand aufgenommene Satellitenaufnahmen gefunden hat, bei denen ein besonders hohes Maß an Wolkenaufzug zu beobachten ist. Beim oberen Bildpaar ist der „Gradient“ geringer als beim unteren, entsprechend ist die Relevanz (Bewertung der Retrieval-Präzision durch das System selbst) unten höher.

## ■ Ausblick

Zur Zeit ist es noch viel zu früh, um an eine Kommerzialisierung des Gesamtsystems zu denken. Der dringende Bedarf nach intuitiv bedienbaren Werkzeugen zur gezielten Suche nach multimedialen Dokumenten wird jedoch mit der rasanten Ausweitung des *WorldWideWeb* noch weiter anwachsen. Einzelne im Rahmen dieses Projekts geleistete Arbeiten könnten bereits jetzt zu einer Verbesserung bestehender Produkte beitragen. Zum einen könnten mit der entwickelten Technologie im Bereich natürlichsprachlicher Schnittstellen bestehende Anfragesysteme um leistungsfähige intuitiv handhabbare Anfragemöglichkeiten erweitert werden. So ist etwa geplant, eines der Informationssysteme im Intranet der Universität Bielefeld mit dieser Technologie auszurüsten. Zum anderen wurde bei der Spezifikation der Spezialprozessoren für Bild- und Textsuche der universellen Einsetzbarkeit dieser Prozessoren besondere Beachtung geschenkt. Der Textsuchprozessor etwa ermöglicht nicht nur eine extrem schnelle approximative und assoziative Volltextsuche, sondern kann auch in traditionelleren, indexbasierten Systemen zur extrem schnellen approximativen Indexsuche verwendet werden. Der Matching-Algorithmus erlaubt eine Kaskadierung der Prozessoren zum Matching längerer Symbolketten und wäre daher etwa auch für Anwendungen im Bereich der DNA-Analyse geeignet.



**Prof. Dr. Alois Knoll** (links) studierte Elektrotechnik an der Universität Stuttgart und war anschließend Entwickler bei der Siemens AG. Danach war er Wissenschaftlicher Mitarbeiter am Institut für Technische Informatik der Technischen Universität Berlin. Er promovierte 1988 über Verfahren zur Objekterkennung mit Ultraschall. Nach seiner Habilitation 1993 wurde er auf eine Professur für Technische Informatik an die Universität Bielefeld berufen. **Ingo Glöckner, M.A.**, (rechts) studierte Computerlinguistik, Informatik und Mathematik in Osnabrück und arbeitet seit 1996 in der Arbeitsgruppe Technische Informatik der Universität Bielefeld. Arbeitsschwerpunkte sind: multimediale Informationssysteme und Wissensverarbeitung mit Fuzzy-Methoden. **Prof. Dr. Hermann Helbig** studierte Physik in Leipzig und promovierte 1976 an der TH Chemnitz. Er war Forschungsgruppenleiter im Kombinat Robotron und Honorarprofessor für Künstliche Intelligenz an der TU Dresden. 1986 habilitierte er sich; 1992 wurde er auf eine ordentliche Professur für Praktische Informatik an der Fernuniversität Hagen berufen (Forschungsschwerpunkte: Computerlinguistik und automatische Wissensverarbeitung).