

# Tagungsberichte

## **Hypertext – Information Retrieval – Multimedia '97**

Theorien, Modelle und Implementierungen integrierter elektronischer  
Informationssysteme

**Universität Dortmund, Lehrstühle Informatik I und VI**  
**29. September – 2. Oktober 1997**

*Die Fachtagung wurde gemeinsam vom Hochschulverband Informationswissenschaft, den GI-Fachgruppen Hypertext, Information Retrieval und Multimediale elektronische Dokumente, der Österreichischen Computer Gesellschaft sowie der Schweizer Informatiker Gesellschaft veranstaltet.*

*Gerhard Knorz*

### **1 Drei Fachgruppen – eine Tagung**

Kaum eine andere Idee hat wie das World Wide Web neben den Perspektiven auch die Gewohnheiten so vieler Menschen weltweit in so kurzer Zeit grundlegend verändert. In die Phase der dynamischen Entwicklung fällt auch die Zeitspanne zwischen der ersten gemeinsamen Fachtagung der 3 Fachgruppen der Gesellschaft für Informatik, *Hypertext, Information Retrieval* und *Multimediale elektronische Dokumente* unter dem Acronym HIM '95 in Konstanz und nun der zweiten Tagung dieser Reihe in Dortmund. Wie wenig es Sinn macht, elektronische Informationssysteme nur aus *einem* Blickwinkel heraus zu konzipieren, wie stark also eine interdisziplinäre Sicht die richtige Perspektive auf einen Gegenstandsbereich ist, der sich noch nie zuvor so weit an die Bürger einer sich tatsächlich herausbildenden Informationsgesellschaft herangetreten ist, ist 2 Jahre nach einem ersten Experiment zu einem selbstverständlichen Credo aller Beteiligten geworden. Die HIM '97 war angetreten, sowohl für eine theorie- und methodenorientierte Diskussion wie auch für anwendungsorientierte Fragen eine anspruchsvolle Plattform zu bieten.

Die Tagung wurde von den Lehrstühlen 1 und 6 der Universität Dortmund organisiert. Vom Hauptbahnhof Dortmund liegen Universität – und damit auch der Tagungsort – im Prinzip einfach mit der S-Bahn zu erreichen. In den langen Zeiten des Wartens am Bahnhof hatte ich genügend Muße, die besondere Eignung des Tagungsortes für sein Thema wahrzunehmen. Das Interface der Bahn

zu ihren Kunden ist eine Fundgrube für jeden, der den Gründen mißlungener Interaktion und Nutzung nachgehen will: Von einem Fahrplan mit Fußnoten bis zur eigenen Logik der Fahrkartenautomaten und all den anderen Hindernissen auf dem Weg zu einem Platz im Abteil. Und seitdem ich die Unwägbarkeiten der Abfahrtszeiten wiederholt erfahren durfte, habe ich meine eigene Theorie darüber, warum Dortmund in der IR-Szene als das deutsche Mekka probabilistischer Ansätze gilt.

Die Tagung selbst bestand aus einem breitgefächertem Angebot wissenschaftlicher Beiträge auf gutem Niveau (WWW-Anwendungen, Interfaces, Digital Libraries, automatische Methoden der Hypertexterstellung u. a.) und einer begleitenden Ausstellung, z. T. korrespondierend zu einzelnen Vorträgen. *Eine* – die einzige – Enttäuschung darf man nicht beiseite lassen: Daß zu diesem unbestritten aktuellen Thema mit Mühe und Not vielleicht 100 Teilnehmer der Einladung von 3 kooperierenden Fachgruppen nach Dortmund zu locken waren, bedarf einer Erklärung, die ich selbst auch nicht anbieten kann. Statt Spekulationen hier der Hinweis für Daheimgebliebene: Der Tagungsband, herausgegeben von Norbert Fuhr, Gisbert Dittrich und Klaus Tochtermann ist als Band 30 der Schriften zur Informationswissenschaft beim Universitätsverlag Konstanz (UVK) erschienen.

## 2 Vorträge

Im folgenden werde ich schlaglichtartig auf einige der Vorträge eingehen, wobei die Auswahl von meinem eigenen Interesse, z. T. aber auch von Randbedingungen bestimmt ist, die unabhängig von der Tagung selbst sind (z. B. von der Abfahrtswahrscheinlichkeit der S-Bahn).

### ***2.1 Multimedia: Technologien und Strategien im Rundfunkbereich***

Der erste Hauptvortrag von L. Danilenko (Technische Direktion des WDR, Köln) geht auf eine fesselnde und vergnügliche Art der Frage nach, welche Zielgruppen der öffentliche Rundfunk mit welchen Strategien bedienen sollte. Die Mediennutzer beschreibt Danilenko als Mischung aus 3 Grundtypen mit völlig unterschiedlichen Anforderungen an und Potential für Medienanbieter:

- Der *homo sapiens* benötigt Informationen als Basis seiner Urteilsfähigkeit. Stichwörter sind Bildung, Wissen, Mitgestaltung. Spartenprogramme können eine der Antworten auf Anforderungen dieses Typs sein.

- Für den *homo economicus* steht der Nutzen im Vordergrund. Für Spielereien hat er keinen Sinn. („Erst das Fressen, dann die Moral“, [Brecht]). Eine thematische Antwort der Anbieter geht in Richtung Börse, Umwelt, Gesellschaft oder Testergebnisse. Auf der technischen Seite ist Videotext ein Angebot. Wenn es Vorteile bringt, toleriert der *homo economicus* auch eine komplexe Bedienung.
- Der *homo ludens* baut auf dem ältesten Verhaltenserbe des Menschen auf. Schließlich ist das Spielen der Tiere das überlebenswichtige Äquivalent für die Schule der Menschen. Wenn auch die Kontrolle der Medien beim *homo economicus* angesiedelt ist, erfunden wurden Sie vom *homo ludens*. „Niemals Langeweile“ lautet die Anforderung dieses Konsumententyps. Der *homo ludens* akzeptiert auch eine teure Ausrüstung, er ist ein Umsatz-Intensivierer.

Vom gesetzlichen Auftrag an den Öffentlichen Rundfunk her („landesweit und gleichwertig“) ergibt sich die Konsequenz eines Vollprogramms, das keinen der Bereiche Information, Bildung und Unterhaltung ausschließt. Die von Danilenko vertretene Strategie des öffentlichen Rundfunks muß die eines integrierten Ansatzes sein: Vernetzen statt Versparung! Herstellen von Kontext, Bezügen und Zusammenhängen. Unter dem Diktat knapper Kassen geht es zunächst um sparsame Mehrwert-Angebote wie MUX (Hintergrundberichte, neue Sendereihenfolgen) und Festival (wiederholte Nutzung von Eigenproduktionen). Ebenfalls vor der Tür: Filter für die persönliche Programmauswahl.

Im folgenden vergleicht Danilenko Fernseh- und Internet-Nutzung quantitativ und geht auf den Wechsel im Medienverständnis ein, den der Übergang von der Pull- zur Pushtechnologie mit sich bringt, so wie er gegenwärtig von Microsoft und Netscape vorangetrieben wird. Eine lebhaft diskutierte Diskussion im Anschluß belegt ein intensives Interesse der Tagungsteilnehmer.

## ***2.2 Navigation und Suche – zwei komplementäre Ansätze für multimediales Information Retrieval***

Der große Name bei den eingeladenen Vorträgen ist Yves Chiaramella mit seinem Beitrag „Browsing and Querying: Two Complementary Approaches for Multimedia Information Retrieval“. Wie kaum ein anderer steht Chiaramella für Arbeiten, die man allen verschiedenen Schwerpunkten der HIM zuordnen kann. Sein Vortrag ist nicht dazu da, revolutionäre Erkenntnisse einem überraschten Auditorium zu präsentieren. Das Verdienst des Beitrags liegt vielmehr in der didaktischen

Aufbereitung der Botschaft, daß die gegenwärtige Kombination von Suche und Navigation bisher über eine Addition nicht hinausgeht und daß sehr wohl ein Bedarf und auch ein Potential für integrierte Ansätze besteht. Der Weg dazu führt weg von der Sicht, die Dokumente als atomare Einheiten betrachtet, hin zu einem flexiblen Dokumentenbegriff, der strukturelle Differenzierungen innerhalb von (Web-)Seiten und über Seitengrenzen hinaus vorsieht.

### **2.3 Ein Streifzug durch die Tagung**

Einen direkten thematischen Bezug zu Chiamellas Beitrag hatte der Vortrag von Marc Rittberger aufzuweisen: „*Kontextsensitive Visualisierung von Suchergebnissen*“ (Bekavac/Rittberger). Hintergrund ist das Konstanzer Projekt „*Virtuelle Informationsräume*“ für die Suche in elektronischen Marktplätzen. Es wird ein Verfahren vorgestellt, das Suchtreffer nicht als Liste isolierter Informationseinheiten (Dokumente) präsentiert, sondern das die Treffer in ihrem strukturellen Kontext darstellt.

Daß Retrieval und speziell die Frage nach der effektiven Unterstützung explorativer Suchen eine Problemstellung ist, auf die nicht nur IR-Spezialisten kommen, beweist der Beitrag „*Visualisierung zur Unterstützung der Suche in komplexen Datenbeständen*“ von Elzer/Krohn (Institut für Prozeß- und Produktionsleittechnik der TU Clausthal). Anlaß war die unerwartete Schwierigkeit, die Wiederverwendbarkeit von Softwareentwürfen durch Retrievalprozesse in entsprechenden Datenbanken effektiv zu unterstützen. Ein experimentelles Demonstrationssystem, das zur Suche und Ergebnisanzeige einen 3D-Raum verwendet, wurde auf der Basis von Smart (Salton), einem kommerziellen Statistikpaket (SAS) und einer kommerziellen Visualisierungssoftware entwickelt.

Einen ganz anderen Beitrag zur Gestaltung von Retrieval-Interfaces liefern Christian Wolf (Universität Leipzig) und Christa Womser-Hacker: „*Graphisches Faktenretrieval mit vager Anfrageinterpretation*“. Werkstoffinformation kann natürlichsprachig und durch Manipulation von Kurven in Liniengraphiken recherchiert werden. In beiden Fällen bietet Fuzzy Logic die Grundlage der adäquaten Behandlung vager Konzepte.

Mounia Lalmas stellt in sehens- und hörenswerter Weise „*A model for structured document retrieval: empirical investigations*“ (Lalmas/Ruthven) vor – ein Beitrag von der Universität Glasgow. Es geht darum, zu entscheiden, ob Dempster-Shafer's Evidenztheorie prinzipiell in der Lage ist, die Aggregation von Relevanzentscheidungen von isolierten Dokumentkomponenten zu modellieren.

In Ermangelung einer wirklich geeigneten Testkollektion basiert die „empirische Untersuchung“ auf einer sehr formal definierten Versuchsanordnung, in der das Ergebnis mehr von dieser Versuchsanordnung als von der untersuchten Evidenztheorie abhängt. Nun ja, bis auf diesen kleinen Schönheitsfehler hatte der Beitrag alles, was man von einem guten Beitrag erwartet.

Deutlich weiterführender erscheint mir der Dortmunder Beitrag *“Probabilistic Logical Information Retrieval for Content, Hypertext, and Database Querying”* (Rölleke/Blömer), in dem der programmatische Beitrag von Fuhr auf der HIM'95 über ein probabilistisches Datalog („Deduktive Datenbanken“) auf der Basis der Implementierung HySpirit weiterentwickelt und evaluiert wird

*“DVS – A System for Recording, Archiving and Retrieval of Digital Video in Security Environments”* (Herzner/Kummer/Thuswald) ist in dem Sinne spannend, als der Beitrag die praktisch zu lösenden Probleme von wirklich großen existierenden Anwendungen (Größenordnung von bis zu 1 000 installierten Analog-Video-kameras) schildert und eine technische Antwort darauf vorstellt.

Eine Anwendung, deren Nutzen sich mir erst beim zweiten Nachdenken überzeugend erschlossen hat, kommt von der Universität Alcal: *“A User Interface for the Design of Human Figures Multimedia Animations”*. Es geht darum, wie man virtuelle Menschenpuppen effizient zum Tanzen bringt, und in Ermangelung einer breit akzeptierten Notation für Choreographie hat man mit einem entsprechenden System gleich auch ein nachgefragtes Kompositionswerkzeug für Choreographen.

Der Beitrag *„Transformationelle Multimedia-Softwareentwicklung“* (Boles/Wüterich) wird mit viel Engagement vorgetragen. Der Vortragende kritisiert die Dominanz der Informatiker bei der Entwicklung von Multimedia-Produkten und überzeugt dann aber vermutlich nur wenige, daß genau diese Dominanz sich nicht in dem vorgestellten Prinzip der transformationellen Softwareentwicklung wiederfindet.

Für mich einer der besten Beiträge der Tagung kommt von Christoph Baumgarten: *„Probabilistische Modellierung der effizienten Informationssuche in verteilten multimedialen Dokumentbeständen durch Einschränkung des Suchraumes“*. Ich nutze hier den Titel gleich als ein Abstract und beschränke mich auf den Hinweis, daß das behandelte Thema im Kontext von WWW und Digital Libraries keinesfalls nur von theoretischer Bedeutung ist. Ein verwandtes Problem behandelt Norbert Gövert (Universität Dortmund): *„Evaluierung eines entscheidungstheoretischen Modells zur Datenbankselektion“*. Negatives Ergebnis dieser (kleineren) Untersuchung ist die Tatsache, daß die Selektion geeigneter Datenbanken

für ein Suchproblem die Schätzung der Anzahl relevanter Dokumente der Datenbank erfordert, und daß diese Schätzung mit einfachen Mitteln sich als nicht möglich herausstellt.

Ein sehr interessanter Ansatz aus Frankreich verwendet Ansätze von Information Extraction (Eigennamen, Ereignisse, Rollen), um aus Agenturmeldungen automatisch einen zusammenhängenden Hypertext zu konstruieren: *“High Precision Hypertext Navigation based on NLP Automatic Extractions”* (Vichot/Tomeh/Dillet/Wolinski/Guennou/Aydjian). Ebenfalls mit dem Thema „Hypertext und Agenturmeldungen“ setzt sich der Beitrag *“Automatic Construction of News Hypertext”* (Dalamagas/Dunlop) von der Universität Glasgow auseinander. Clustertechniken werden eingesetzt, um Links zwischen Dokumenten zu erzeugen. Durch Eliminieren bestimmter Links entstehen unter Berücksichtigung der zeitlichen Reihenfolge der Nachrichten zusammenhängende Stories.

Einen Einblick in die Alltagswelt eines Suchmaschinenbetreibers liefert der Beitrag *„Realisierung und Optimierung der Informationsbeschaffung von Internet-Suchmaschinen am Beispiel von www.crawler.de“*. Mit Hilfe umfangreicher statistischer Erhebungen wird belegt, daß tatsächlich ein Zusammenhang besteht zwischen der Hierarchietiefe eines Dokumentes und der Wahrscheinlichkeit, daß es geändert wird. Fazit für die Suchmaschine: Je kürzer der Pfad eines Dokumentes, desto häufiger muß es auf Änderung überprüft werden.

### **3 Zusammenfassung**

Eine Tagung zu veranstalten, ist für viele beteiligte Personen ein Unternehmen mit vielen zusätzlichen Sorgen und Arbeitsaufträgen. Was das wissenschaftliche Programm der HIM '97 anbelangt, so hat sich dieser Aufwand ganz sicher gelohnt. Auch die Organisation der Tagung lief völlig reibungslos. Ein Kompliment an die Dortmunder Veranstalter! Was an Wünschen bleibt? Man hätte der Tagung noch deutlich mehr Teilnehmer gewünscht. Und für 1999 wünsche ich mir eine Tagung, deren Scope dem der HIM entspricht – mindestens, denn die Designer stehen (zu recht) immer höher im Kurs, aber auf Tagungen wie dieser immer noch außen vor. Aber bis dahin ist ja noch Zeit. Zum Beispiel dafür, den Tagungsband nochmals etwas genauer zu studieren!