

## HYPertext - INFORMATION RETRIEVAL - MULTIMEDIA

### HIM'95, 5.-7. APRIL 1995 IN KONSTANZ

*Gerhard Knorz*

#### 1 HIM - ein rundum gelungenes Experiment

Anforderungen an moderne Informationssysteme lassen aus System- und Nutzer sieht eine isolierte Betrachtung von Problemen häufig nicht mehr zu 1. Deshalb unternahm es die GI-Fachgruppen „Hypertext“, „Information Retrieval“ und „Multimediale elektronische Dokumente“, zusammen mit der *Österreichischen Computer Gesellschaft*, der *Schweizer Informatiker Gesellschaft* und dem *Hochschulverband Informationswissenschaft* eine gemeinsame Tagung durchzuführen. Die Universität Konstanz, für ihr wissenschaftliches Engagement in den Bereichen Hypertext und Information Retrieval bestens ausgewiesen, hat das Risiko möglicher Komplikationen nicht gescheut und hat die Tagung in jeder Hinsicht erfolgreich vorbereitet und durchgeführt. Dies ist vor allem ein Verdienst von *Rainer Kuhlen* als Programmkomitee-Vorsitzender, von *Marc Rittberger*, bei dem alle Fäden der Organisation zusammenliefen sowie von der erfreulichen Resonanz auf den Call for Papers, die eine Selektion von 50% sehr hochwertiger Beiträge erlaubte.

#### Ambiente

Die Tagung fand mit ca. 200 Teilnehmern in einem lichtdurchfluteten separaten Gebäudeteil auf mehreren Etagen statt, was den dort präsentierten Postern und Rechnerdemonstrationen viel Resonanz bescherte. Die grüne Umgebung der Universität, der Bodensee in Blickweite und das

<sup>1</sup> Zitat aus dem gedruckten Tagungsprogramm.

z.T. sonnige Wetter trug auf ihre Weise zum Tagungserfolg bei, der durch keinerlei technische oder organisatorische Pannen getrübt wurde.

#### 2 Eingeladene Vorträge

Jede der veranstaltenden GI-Fachgruppen hatte ein *Invited Paper* beigesteuert. Die fachliche Eröffnung der Tagung hatte *Roy Rada* als Hauptredner für die Fachgruppe *Hypermedia* übernommen, mit einem Vortrag über „*publishing World Wide Web Material on CD-ROM: Coordination and CD-ROM*“. Ich will freimütig bekennen, daß ich im Hinblick auf persönliche Konstitution und Entfernung zur Projektionsfläche sehr ungünstige Voraussetzungen hatte, dem Vortrag zu folgen. So mag es an mir liegen, daß ich die recht kurze Plauderei von *Rada* (Universität Liverpool) über die Projekte seiner Studierenden als jene Art eingeladener Vorträge empfand, mit denen man nicht so recht glücklich ist - aus meiner Sicht der einzige Schwachpunkt des Tagungsprogramms.

#### 2.1 Evaluierung von Information Retrieval Systemen

Am Donnerstag rettete *Donna Harmann* nach meiner Einschätzung das Ansehen eingeladener Prominenter mit einem Vortrag über „*The TREC Conferences*“. Sie erläuterte den Anspruch der TREC-Konferenzen, IR-Systeme an Fragestellungen und Daten realer Größenordnungen unter kontrollierten Wettbewerbsbedingungen zu evaluieren und kommentierte den Weg,

der mittlerweile mit 3 Konferenzen zurückgelegt wurde. Die Proceedings von TREC 3 werden im übrigen in den nächsten Wochen im Internet präsent sein:

<http://potomac.ncsl.nist.gov/TREC>

Aufbau und Struktur der Aufgaben sowie das Vorgehen bei der Evaluierung wurden besprochen. Als klares Ergebnis stellte Harmann heraus, daß sich vollautomatisch ermittelte Queries unzweifelhaft manuell entwickelten Suchfragen überlegen gezeigt haben, sofern Feedback verfügbar ist. Dies ist immer dann der Fall, wenn immer wiederkehrende Fragen bearbeitet werden sollen (Routing, Filtering, Profildienste). Sie machte an einem Beispiel auch einsichtig, warum dies so ist: Die statistische Optimierung an Feedback-Daten findet Suchterms, an die ein Mensch kaum jemals denken würde, wohingegen die inhaltlich orientierten Suchstrategien einfach nicht so effektiv sind wie erhofft.

## 2.2 Multimedia Trends

Das fachliche Programm der Tagung wurde mit dem Beitrag von *Simon Gibbs* (GMD) abgeschlossen: „*Directions in Multimedia Programming.II*“ Er diskutierte zunächst das Problem, zwischen Anwendung und Hardware eine für Multimedia-Entwicklungen geeignete Abstraktion zu finden und zu definieren, so wie dies etwa mit GKS (Graphisches Kernsystem) und anderen Modellen/Standards im Bereich des CAD bereits eingeführt sei. Dann widmete er sich seinem Schwerpunkt der Video-Konferenzen, deren Vor- und Nachteile er auf verschiedenen technischen Realisierungsniveaus verglich: von der gegenwärtig so teuren, vorbereitungs- und betreuungsintensiven Nutzung von Spezialräumen über die Versammlung einzelner Köpfe in Fenstern von Workstations mit integrierter Kamera bis zur Virtual Reality-Lösung, deren Technik mit der des virtuellen Filmstudio nahezu zusammenfällt: Ein kleiner Raum wird durch eine Projektionsfläche geteilt. Sensoren nehmen Erscheinungsbild, Stimme und aktuellen Standort des Diskussionsteilnehmers auf und sorgen über aufwendige Graphikrechner dafür, daß die anderen Diskussionsteilnehmer in der aktuell richtigen

Perspektive auf der Projektionsfläche dargestellt werden - so wie dies zeitgleich auch bei diesen geschieht. Das alles wurde anschaulich demonstriert anhand eines Videos. In diesem vom WDR moderierten und ausgestrahlten Film unter Verwendung von und über diese GMD-Technik, bei dem erst ein Graphikrechner den in einer Bluebox aufgenommenen Moderator mit einer computergenerierten Kulisse verrechnet und bei allen Kamerabewegungen perspektivenrichtig darstellt, konnte die bereits erreichbare "Illusionsqualität" unter Beweis gestellt werden.

## 2.3 Hypermedia in a Gambling Casino Setting

Neben den eingeladenen Vorträgen fiel ein weiterer bereits durch Thematik und Plazierung aus dem Rahmen: Zwischen Abendspaziergang auf der Insel Mainau mit anschließenden Cocktail im Palmenhaus und dem Abendessen im Comturey-Keller hielt *Hermann Maurer* (Graz) einen engagierten Vortrag über „*Hypermedia in a Gambling Casino Setting*“<sup>11</sup>. Mit der Idee zu diesem Arrangement hatte sich das Programmkomitee über eine Schwierigkeit hinweggerettet: Was tun, wenn ein renommierter Wissenschaftler ein seriös bis futuristisch anmutendes Paper über eine nicht im traditionellen Wissenschaftsbereich verankerte Anwendungsidee einreicht - und man gar nicht sicher ist, ob auch alles so ganz ernst gemeint ist? Nun, Maurer lies keinen Zweifel, daß es ihm *sehr* ernst ist und daß für ihn Hypermedia eben grundsätzlich mehr und innovativer ist und sein muß als *Multimedia mit Hyper*. Klar, daß auch das WWW einige bange Minuten der begründeten Polemik zu überstehen hatte und letztlich gegen Maurers *Hyper-G* ohne Chancen blieb. Ja, und dann entwickelte Maurer seine Vision, die jedoch wohl ganz auf dem Boden erstaunlicher Tatsachen aufbaut: Es geht um ein Spielkasino, dessen Hotelräume samt und sonders mit hochwertigen Graphic-Workstations ausgestattet werden sollen, und für die innovative und creative Anwendungen zu schaffen sind. Wollten Sie nicht schon immer mal Videoconferencing in der Gestalt von

Spider (Wo) man mit ihrer - letztlich unbekanntem - Barbekanntschaft von gestern abend machen und anschließend nach geeigneter Verwandlung - inspiriert durch einen Teleshopping-Ausflug - als virtuelle MitspielerIn beim realen Roulette teilnehmen und gewinnen?

### 3 Vorträge in den einzelnen Sektionen

Die Vorträge profitierten davon, daß bis auf eine Ausnahme am Freitag vormittag das Programm ohne parallele Sektionen auskam. Der Tagungsband, herausgegeben von Rainer Kuhlen und Marc Rittberger, erschien als Band 20 der H12-Reihe Schriften zur Informationswissenschaft beim Universitätsverlag Konstanz lag rechtzeitig zu Tagungsbeginn vor. Von allen TeilnehmerInnen wurde sowohl bei "öffentlichen" Meinungsäußerungen wie auch im informellen Gespräch die Qualität der Beiträge hervorgehoben. Aus den ca. 20 angenommenen Beiträgen will ich - nach ganz persönlichen Gesichtspunkten - nur einige wenige herausgreifen.

#### 3.1 Indexieren

Das Interesse am Thema der Dokumentenindexierung wurde durch zwei erfreuliche Beiträge von der ETH Zürich auch für solche belebt, denen dieser Bereich etwas ferner liegt. Bei beiden Papers geht es darum, Inhalte zu indexieren, die nicht in Text verpackt wurden und zwar, ohne daß man Voraussetzungen über den behandelten Gegenstandsbereich machen muß. Wenn also O. Lorenz und G. Monagan über *Automatisches Indexieren von Liniengraphiken* vortragen, dann meinen sie nicht, daß man vorab ein Inventar semantischer Primitive zu definieren hat, um etwa aus Symbolen für Wände und Tische Zeichnungen von Inneneinrichtungen zu interpretieren. Stattdessen bringen sie elementare Techniken aus Bilderkennung und Information Retrieval zusammen, um Strichzeichnungen, wie man sie etwa im Duden Bildwörterbuch findet, anwendungsneutral durch Indexierungsmerkmale zu beschreiben und auffinden zu können.

<sup>2</sup> Hochschulverband für Informationswissenschaft.

Die Query besteht dabei aus einem Bild bzw. Teilbild und die Retrievalantwort aus einem Ranking solcher Bilder, die sich auf der Ebene von Linienverläufen, Parallelen, Winkeln und auch Beschriftungen als ähnlich qualifizieren.

Die Besonderheit des Beitrags von M.

Wechsler: *Eine neue Indexierungsmethode für Information Retrieval auf Audiodokumenten* liegt in vergleichbarer Weise darin, daß die Verarbeitung ohne inhaltliche Vorarbeit greift: Es wird also kein allgemeines oder anwendungsspezifisches Wörterbuch vorausgesetzt, um Audiotexte zu indexieren. Vielmehr legt das auf Phonemerkennung nach Standardtechnik (Hidden Markov Modelle, Bi-gramme) basierte System selbst a posteriori fest, welche Phonemketten sich im Hinblick auf Länge, Abdeckung, Häufigkeit und Erkennungssicherheit eignen, um als Indexierungsmerkmale zu fungieren. Das System konnte in den Pausen auf der Basis eines Korpus mit gesprochenen Radionachrichten erprobt werden.

Beide Ansätze wurden intensiv diskutiert, und vermutlich haben die Vortragenden einige weitere anregende Ideen mit nach Hause genommen.

#### 3.2 Modellierung von Hypermedia Retrieval mittels Datalog

Aus Dortmund kam der Vorschlag, die Sprache *Datalog* zur Modellierung komplexer Retrievalstrategien für strukturierte Hypermediadokumente einzusetzen: „*Modelling Hypermedia Retrieval in Datalog*“ (N. Fuhr). Unter Datalog kann man sich ein um Berechnungsprobleme bereinigtes Prolog zur Realisierung deduktiver Datenbanken vorstellen. Ziel war es, die grundsätzliche Eignung von Datalog, was die Ausdrucksmächtigkeit betrifft, nachzuweisen. Weiterführende Arbeiten werden sich mit der Implementierung von Datalog beschäftigen, um damit im Gigabyte-Bereich Information Retrieval zu erproben - und natürlich wird es ein probabilistisches Datalog werden müssen, wenn es in Dortmund Bestand haben will.

### 3.3 Informationssuche in Faktenbanken

Zwei recht verschiedene Beiträge beschäftigten sich mit dem Thema, die Informationssuche in Faktenbanken zu unterstützen. STN Karlsruhe stellte Überlegungen vor, die in ihren Faktenbasen implizit vorhandenen Links für eine komfortable Hypermedia-Oberfläche für Datenbanken wie DETHERM zu nutzen: *Concept of a Hypertext Interface for the Information Retrieval in Complex Factual Databases* (L. Helmes, M. Razum, A. Barth). Es ist zu hoffen, daß im STN-internen Abstimmungsprozeß diese Überlegungen nicht auf dem Weg zur Realisierung steckenbleiben.

Der Beitrag „Softcomputing“- Verfahren zur Behandlung von Ähnlichkeiten und Vagheit in objektorientierten Informationssystemen (Th. Mandl und Ch. Womser-Hacker) motivierte zunächst das Problem vager Anfragen und stellte die Ähnlichkeitssuche als Möglichkeit vor, das Problem von Datenlücken zu überwinden. Der Regensburger Prototyp für ein multimodales Werkstoff-Informationssystem integriert Lösungen auf der Basis von Fuzzy Sets und von neuronalen Netzen. Er stellt eine menügesteuerte Möglichkeit zur Verfügung, linguistische Variablen mit Werten zu belegen (z.B. Temperatur mit *hoch*, *niedrig*, etc.), wobei eine Kontextabhängigkeit der Zugehörigkeitsfunktion (das Problem der *großen Maus*) noch nicht verwirklicht ist.

Die Ähnlichkeit zwischen Werkstoffen (hinsichtlich gemeinsamen Anwendungsmöglichkeiten) wurde nach dem Scheitern eines regelbasierten Ansatzes mittels eines neuronalen Netzes trainiert und bestimmt. Nach zwei Anläufen war schließlich bei ca. 20 bis 30 Eingangsparametern und *einer hidden layer* ein befriedigendes Resultat erreicht.

Ein Detailproblem der Faktensuche behandelten U. Pfeifer, Th. Poersch, N. Fuhr: *Searching Proper Names in Databases*, in dem sie mehrere Verfahren daraufhin verglichen, mit den Problemen von Schreibvarianten von Namen fertigzuwerden. Ergebnis: Alles ist besser, als nur auf Identität zu prüfen.

### 3.4 Multimedia und Hypertext

Von den in dieser Besprechung aus Zeitgründen zu kurz gekommenen Beiträgen über Multimedia und Hypertext haben mir insbesondere die Beiträge *"Towards a Family of Formal Models for Hypermedia"* (K. Tochtermann, G. Dittrich), *„Navigation via Similarity in Hypermedia and Information Retrieval“* (D. Tudhope, C. Taylor, P. Beynon-Davie) sowie *„Ein graphischer Browser für das Konstanzer Hypertext-System“* (S. Arens, R. Hammwöhner) gut gefallen.

## 4 Digitale Bibliotheken

Eine sehr lebhaft entwickelte sich auf dem von Hans-Peter Frei moderierten Panel über *Digitale Bibliotheken*. Die Panel-Teilnehmer zeigten zunächst, daß sie durchaus Unterschiedliches im Sinn haben, wenn sie über digitale Bibliotheken sprechen: Herrmann Maurer sprach über Probleme des Internet, über die Integration von Internet und LAN, von Internet und CDROM und über die Integration von Anwendungen und Produkten (wie z.B. von elektronischen Wörterbüchern). Im Gedächtnis ist mir der zitierte De Kemp mit seiner Antwort auf die Frage geblieben: *Wie kann man mit elektronischen Zeitschriften Geld verdienen?* - „Drucken Sie sie!“ (online 94).

Norbert Fuhr sprach über Probleme der (mangelnden) Ergonomie von Bildschirmen, der Archivierung und der Kosten. Er zitierte seinerseits Czermak mit dem Prinzip der *„flächendeckenden Entgeltlichkeit“* und forderte einen weiterhin freien Zugang zu Information, wobei eine Steigerung der Kosten gegenwärtig sicher nicht in die Landschaft paßt.

Jean-Frederic Jauslin (Direktor der Schweizer Landesbibliothek) lieferte für die spätere Diskussion im Plenum sicher die meisten Stichwörter. Er baute folgende Kette verkürzt wiedergegebener Thesen auf: Bibliotheken hatten bisher eine Schlüsselrolle, gegenwärtig sind die Dinge in Bewegung, den Bibliotheken ist vielfach nicht bewußt, was auf dem Spiel steht, die Gefahr ist groß, daß Bibliotheken künf-

tig nur noch wenige Wissenschaftler bedienen bzw. ganz verschwinden. *Jauslin* stellt fest, daß das Mehrsprachenproblem zwischen Produzenten - Benutzern - Archiv auch in der EG noch nicht so recht adressiert wird und daß es insgesamt nicht um die Überwindung sondern um die Ergänzung des Buches durch elektronische Formen geht.

*Rainer Kuhlen* zitiert aus einer Studie, die das Problembewußtsein und die Aufgeschlossenheit der Bibliotheksprofession doch zu belegen scheint und gibt auf die Forderung, daß das Internet Mechanismen des Qualitätsmanagements braucht, zu bedenken, daß das *sky writing* auch als Basisdemokratisierung der Information empfunden werden könne.

In der weiteren Diskussion fallen die Stichwörter der Informations- Umweltverschmutzung und -Ökologie, wird gefragt, ob Bibliotheken noch Mauern brauchen, werden ernsthafte Überlegungen aus den U.S.A. zitiert, die die Notwendigkeit für eine Bibliothek für eine neue Universität grundsätzlich hinterfragen, werden Aufgaben- und Prioritätsverlagerungen in Bibliotheken dargestellt und das papierlose Büro mit Hohn und Spott überschüttet.

## 5 Fazit

HIM'95 war eine Tagung auf gutem bis sehr gutem Niveau, das Modell einer integrierenden Tagung taugt zur Nachahmung. Daß die Synthese zwischen *Hypertext* und *Information Retrieval* besser gelungen war als die Einbeziehung der Fachgruppe *Multimedia*, war erkennbar, doch - so darf man hoffen - ein überwindbares Problem. Die Rahmenbedingungen der Tagung, die Vorbereitung und die Organisation waren hervorragend. Daß eine konkrete Vorplanung für eine HIM'97 noch nicht existiert, liegt eigentlich nur am globalen Wissenschaftsdorf und den Schwierigkeiten, die die ACM mit ihrer Unsicherheit über die nächste europäische ECHT in die deutschsprachige Tagungsszene hineintransportiert. Aber es wird schon werden, zumindest bis - zur zweiten HIM!