

Information Retrieval

Christa Wormser-Hacker
Universität Regensburg

Tagungsbericht

Workshop "Information Retrieval" Gesellschaft für Informatik (GI) - GMD-IPSI 24. und 25. Juni 1991.

Am 24. und 25. Juni 1991 fand bei der GMD in Darmstadt ein Workshop mit dem generellen Thema "Information Retrieval" statt. Das große Interesse an diesem Workshop spiegelte sich in den nahezu 100 Teilnehmern wider, die sich zum größten Teil aus Wissenschaftlern und Vertretern der Industrie zusammensetzten. Die Teilnehmer kamen aus Deutschland, Österreich und der Schweiz. Die beiden eingeladenen Vortragenden (Prof. Keith van Rijsbergen und Prof. Bruce Croft) erweiterten dieses Spektrum auf Großbritannien und die USA.

Veranstalter des Workshops waren die Gesellschaft für Informatik und die Gesellschaft für Mathematik und Datenverarbeitung/Institut für Integrierte Publikations- und Informationssysteme GMD-IPSI, die auch ihre Infrastruktur zur Verfügung stellte. Der Workshop sollte den Start bilden für die neue Fachgruppe "Information Retrieval" innerhalb der GI.

Die zweitägige Veranstaltung gliederte sich in sieben Sitzungen mit unterschiedlichen inhaltlichen Schwerpunkten, in welchen insgesamt 15 Vorträge gehalten wurden. Zum Teil ergänzten entsprechende Systemvorführungen die Referate.

Höhepunkte des Workshops waren sicherlich die bei den eingeladenen Vorträge. Keith van Rijsbergen (Professor an der University of Glasgow) sprach über "Information Retrieval and its Logic: Inference versus Relevance". Ausgehend von der Tatsache, Information Retrieval mittlerweile nicht nur mit Dokumenten beschäftigt, die in Form von Texten vorliegen, sondern auch Bilder, Videos, Objekte aus der Büroumgebung etc.

(Hypermedia!) gespeichert und wiederaufgefunden werden sollen, müssen auch die Grundlagen des Information Retrievals, d.h. die Modelle, auf ihre Anwendbarkeit auf die multimedialen Anforderungen überprüft werden. Van Rijsbergen sieht bei der Verschmelzung eines Hypermedia-Systems mit dem probabilistischen IR-Modell das Problem darin, daß die Unsicherheit der Information durch das Hinzufügen von Information aus der Umgebung oder einem Thesaurus noch verstärkt wird. Van Rijsbergen zeigte, indem er auf das "logical uncertainty principle" zurückgriff, daß Information Retrieval durchaus als eine Form von Inferenz angesehen werden kann.

Bruce Croft (Professor an der University of Massachusetts, Amherst), der zweite eingeladene Redner, hielt einen Vortrag mit dem Titel "Retrieval from Large Databases". Zunächst stellte er heraus, daß sich Information Retrieval in den USA nach wie vor großen Interesses erfreut, vor allem aufgrund der neuen Größenordnungen im Bereich der Dokumentanzahl, auf seiten der Anfragen und last not least auch im Hinblick auf das individuelle Dokument ("What is large?"). Ausgehend von den amerikanischen IR-Forschungsaktivitäten der letzten Jahrzehnte zeigte er, inwieweit die erzielten Ergebnisse vereinbar seien mit den Anforderungen von Datenbanken, die den neuen Größenordnungen entsprechen. Er warf Fragen auf nach der Übertragbarkeit von Experimenten in bezug auf die Beurteilung von Effektivität und Effizienz bestimmter Algorithmen oder Funktionen, die auf der Grundlage kleiner Kollektionen bzgl. der einbezogenen Dokumente und Anfragen erzielt worden waren. Bruce Croft ging auf drei Projekte ein, an welchen er beteiligt ist (TIPSTER, WESTLAW und CONSTRUE), die beitragen sollen, dieses Defizit zu beheben. Durchaus

optimistisch für die Forschung seine abschließende Formel: "Large Databases Present Large Challenges" !

Die 13 regulären Vorträge des Workshops kreisten um folgende Schwerpunkte: Entwurf von IR-Systemen und -Komponenten, Repräsentation und Deskribierung im IR-Bereich, Evaluierung und Messung, Realisierung von IR-Systemen aus verschiedenen Anwendungsbereichen und Hypertext und IR.

Im Beitrag von H.P. Frei und S. Meienberg ging es um ein Verfahren, das automatisch Boolesche Anfragen aus Anfragen mit gewichteten Suchtermen generiert und die Antwortmenge durch Rankingverfahren ordnet, um so das Retrieval effizienter zu gestalten. Die Vorträge von K. Tzeras und U. Pfeifer beschäftigten sich mit automatischer Indexierung. K. Tzeras referierte über die Abschätzung des Aufwands in bezug auf die Entwicklung eines Indexierungswörterbuchs. U. Pfeifer stellte neue Ansätze vor, die Experimenten zufolge erfolgversprechende Wege zu besseren automatischen Indexierungsfunktionen, die über die Deskriptorzuteilung entscheiden, darstellen. Beide Vorträge sind im Zusammenhang mit dem Darmstädter Indexierungsansatz zu sehen. Anforderungen an das **IR** aus der Praxis stellten H. Amstutz und B. Holländer/Thönssen durch ihren Vortrag "Elektronische Ablage und Archivierung auf der Basis eines Database Management Information Retrieval Systems" dar. K. Englmeier präsentierte mit dem wissensbasierten System WIDAB (Wirtschaftsinformationen und Datenbank) ein System zur Informationsorganisation zur effizienten Nutzung von Wirtschaftsinformation, die bei statistischen Ämtern und Wirtschaftsforschungsinstituten gehalten wird. Ein weiterer Vortrag aus der anwendungsorientierten Richtung von M. Münzenberger stellte ein Beispiel für eine IR-Anwendung im Rahmen des "Forensischen Informationssystems Handschriften" (entwickelt beim Bundeskriminalamt) vor. Durch Kombination von Mustererkennungs- und Clustering-Verfahren soll eine effektive Recherche ermöglicht werden. Der Beitrag von W. Augsburg, H.K. Rieder und J. Schwab kam aus dem Bereich des Faktenretrievals. Hier ging es um ein System, das Unternehmenszahlen zu bewertbaren Kennzahlen umformt und eine automatische Vorbewertung vornimmt. Mit dem Berliner ATLAS-Projekt befaßten sich die Vorträge von J. Willenborg und T. Muhr.

Im ersten Referat ging es um einen Ansatz zur Unterstützung bei der Thesauruserstellung und -pflege, im zweiten um ein

Interpretations- Unterstützungs-System, das die Arbeit mit Texten z.B. im Hinblick auf Vergleich, Kommentierung, Kategorisierung etc. unterstützt. H.P. Frei, S. Meienberg und P. Schäuble stellten eine neues Effektivitätsmaß vor, das die Nachteile der klassischen Maße recall und precision kompensieren soll. Drei Vorträgen waren der Koppelung von Hypertext und IR gewidmet. J. Herrmann und D. Meiser stellten die Realisierung eines Thesaurus durch Hypertextverknüpfungen vor. D. Merkl, A. Min Tjoa und S. Vieweg setzten ein Hypertext-System ein zur Navigation in juristischen Dokumenten. Um HyperQuery, ein Anfragesystem mit graphischer Benutzeroberfläche am Beispiel eines Reiseauskunftsystems ging es im letzten Beitrag von J. Herczeg, H. Hohl und M. Ressel.

Die Vorträge werden in einem Tagungsband publiziert, der im Herbst bei Springer erscheinen wird (Fuhr, N. (Hrsg.) (1991), Information Retrieval. Proceedings des Workshops "Information Retrieval", 24.-25.6.1991 in Darmstadt).

Hinzuweisen ist noch auf die Gründung der neuen Fachgruppe "Information Retrieval" innerhalb der GI, die sich an Forscher, Entwickler und Anwender von Informationssystemen richtet. Dieser Workshop ist als erste Veranstaltung dieser neuen Fachgruppe zu sehen, der in regelmäßigen Abständen weitere Tagungen folgen sollen. Themenschwerpunkte liegen u.a. in folgenden Bereichen: Schnittstellengestaltung, Information System Management, Retrievalmodelle, Evaluierung von Retrievalverfahren usw.

Kontaktadresse für Interessierte: Prof. Dr.

Norbert Fuhr Lehrstuhl Informatik VI Universität Dortmund W-4600 Dortmund 50