



Implementierung und Evaluierung des MIMOR-Modells: Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval

Ralph Kölle, Thomas Mandl, Christa Womser-Hacker

Informationswissenschaft - Universität Hildesheim
Marienburger Platz 22 - 31141 Hildesheim
{koelle, mandl, womser}@uni-hildesheim.de

1 Das MIMOR Modell

Das MIMOR-Modell (Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval) ist ein Ansatz zur Erhöhung der Adaptivität auf einer Meta-Ebene. MIMOR arbeitet mit Mehrfachindexierung, also der Repräsentation der Objekte (Texte, Bilder usw.) aus mehreren Perspektiven. Große Evaluierungs-Studien haben u.a. gezeigt, dass die besten IR-Verfahren sich in der Qualität nur unwesentlich unterscheiden. Ihre Treffermengen weisen jedoch nur wenig Überschneidung auf. Das bedeutet, dass gute Verfahren etwa gleich viele relevante Dokumente finden, aber eben verschiedene (cf. WOMSER-HACKER 1997). Sogenannte Mehrfachindexierungs- und Fusionsansätzen versuchen dies auszunutzen und die relevanten Dokumente mehrerer Ansätze zu vereinigen. Das Problem besteht darin, die relevanten Dokumente in der Gesamtergebnismenge zu finden und stärker zu gewichten. MIMOR lernt, die Fusion durch Relevanz-Feedback und Benutzerpartizipation zu steuern. Das System kombiniert die Ergebnisse der einzelnen Systeme im Lauf des Lernprozesses zunehmend besser. Jedes System trägt mit einem bestimmten Gewicht zum Gesamtergebnis bei. Dieses Gewicht erhöht sich, wenn ein einzelnes System häufig positiv bewertete Dokumente auf die gesamte Ergebnisliste bringt (cf. MANDL & WOMSER-HACKER 2000, WOMSER-HACKER 1997).

2 Evaluierung mit CLEF Daten

Das MIMOR-Modell wurde in JAVA implementiert und konnte somit evaluiert werden. Für die Evaluierung von Information Retrieval Systemen haben sich die großen Evaluierungsinitiativen etabliert. In Europa ist dies vor allem die CLEF Kampagne (Cross Language Evaluation Forum, cf. KLUCK ET AL.



2002). Innerhalb der CLEF Daten sind insbesondere die GIRT Daten (German Indexing and Retrieval Testdatabase) interessant. Dabei handelt es sich um ca. 80.000 sozialwissenschaftliche Fachtexte, die auch intellektuell erschlossen sind.

MIMOR wurde mit den GIRT Daten evaluiert (cf. HACKL ET AL. 2002). Für einen Fusionsansatz müssen zunächst mehrere einzelne Retrievalsysteme zur Verfügung stehen, deren Ergebnisse verschmolzen werden. Die Wahl fiel auf das *irf* Paket (Information Retrieval Framework) des NIST (National Institute of Standards and Technology, USA), da dieses ebenfalls in JAVA realisiert ist und als Quellcode vorliegt. Mit *irf* konnten zwei Systeme simuliert werden, da es zahlreiche Parametrisierungsmöglichkeiten bietet. Diese wurden einmal mit gleichen Gewichten und einmal mit optimierten Gewichten linear fusioniert. Die Optimierung basiert auf einer Analyse des gleichen Systems mit den Daten und den menschlichen Relevanzurteilen des Jahres 2001. Nach Bekanntgabe der Ergebnisse durch die CLEF Organisatoren lässt sich der Einfluss der Fusion und der Optimierung abschätzen. Im nächsten Jahr sind die Einbindung mehrerer Systeme und die Teilnahme am mehrsprachigen Retrieval geplant.

Die MIMOR Implementierung eignet sich aufgrund der Flexibilität des objekt-orientierten Ansatzes auch für die weitere Untersuchung der Validität der Testergebnisse von CLEF.

3 Literaturverzeichnis

- HACKL, René; KÖLLE, Ralph; MANDL, Thomas; WOMSER-HACKER, Christa (2002): Domain Specific Retrieval Experiments at the University of Hildesheim with the MIMOR System. Erscheint in: Cross Language Evaluation Forum. CLEF-Workshop Notes. 19.-20.9.2002. Rom.
- KLUCK, Michael; MANDL, Thomas; WOMSER-HACKER, Christa (2002): Cross-Language Evaluation Forum (CLEF): Europäische Initiative zur Bewertung sprachübergreifender Retrievalverfahren. In: nfd Information - Wissenschaft u. Praxis vol. 53 (2). S. 82-89.
- MANDL, Thomas; WOMSER-HACKER, Christa (2000): Ein adaptives Information Retrieval Modell für Digitale Bibliotheken. In: Knorz, Gerhard; Kuhlen, Rainer (Hrsg.): Informationskompetenz – Basiskompetenz in der Informationsgesellschaft. Proc 7. Intl Symposium für Informationswissenschaft (ISI 2000) 8.-10. Nov., Darmstadt. S. 1-16.
- WOMSER-HACKER, Christa (1997): Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitationsschrift. Universität Regensburg, Informationswissenschaft. 1997.