

Ein adaptives Information-Retrieval-Modell für Digitale Bibliotheken

Thomas Mandl – Christa Womser-Hacker

Informationswissenschaft – Universität Hildesheim
Marienburger Platz 22 – 31141 Hildesheim
mandl@rz.uni-hildesheim.de
womser@cl.uni-hildesheim.de

Zusammenfassung

In diesem Beitrag wird ein adaptives Information-Retrieval-Modell im Kontext Digitaler Bibliotheken vorgestellt, das während des Informationsprozesses lernt, Beziehungen zwischen Repräsentationsformen und Informations-Objekten herzustellen. Die aus großen Evaluierungsstudien abgeleiteten Erkenntnisse, dass Relevance Feedback und Polyrepräsentation echte Mehrwerte darstellen und zu besserer Retrievalqualität führen, dienen dabei als Grundlage. Um eine stärkere Adaption zu leisten, wird das Modell um Strategien erweitert, die Benutzer-, Dokument- und Anfrageeigenschaften einbeziehen. Im Rahmen Digitaler Bibliotheken spielen insbesondere Transformationen unterschiedlich erschlossener Objekte eine wichtige Rolle zur Kompensation semantischer Heterogenität. Auch in diesem Rahmen kann das adaptive Information-Retrieval-Modell eingesetzt werden.

Abstract

This paper presents an adaptive information retrieval system for digital libraries which learns to model the relationship between representation mechanisms and information objects during an information process. The evaluation results which form the foundation of the model show that relevance feedback and

and poly-representation provide value added services and lead to better retrieval quality. Features of users, documents, and queries are included to enhance the model and to guarantee a higher level of adaptation. In the context of digital libraries, transformations are an important strategy in order to resolve semantic heterogeneity. The adaptive information retrieval model can be applied to these transformations.

1 Herausforderungen für Digitale Bibliotheken

Das Information Retrieval (IR) befindet sich derzeit in einer Umbruchsituation, da aufgrund innovativer Anwendungsbereiche (Digitale Bibliotheken, Electronic Commerce, Suche im WWW etc.) modifizierte Anforderungen entstanden sind. Nicht nur die Art der informationellen Objekte ist heterogen, sondern auch die Domänen, die Benutzer, die Infrastrukturen etc. Neue Technologien wie Multimedia und das Internet beeinflussen den Informationsprozess in starkem Maße. Aufgrund dieser Komplexität und Heterogenität der Kontexte ist die Anwendung einer generellen Lösung nicht mehr möglich. In der adäquaten Selektion und Adaption von IR-Verfahren und ihrer Kombinationen sowie in den hinzukommenden Mehrwerten liegt der Schlüssel zum Erfolg.

Der Bereich der Digitalen Bibliotheken (cf. Fuhr 1999) bezieht neben einer Vielfalt unterschiedlicher Objekttypen (traditionelle Publikationen, multimediale Objekte) und neuen Möglichkeiten des Elektronischen Publizierens eine sehr heterogene Klientel ein, deren Know-how sehr unterschiedlich geprägt sein kann. Oft handelt es sich nicht um spezialisierte, mit informationeller Kompetenz ausgestattete Benutzer. Die Systeme müssen mit entsprechender Intelligenz ausgestattet werden, um diese Situation bewältigen zu können. Die Erkenntnisse aus langjähriger Erfahrung im Bereich der Evaluierung von IR-Systemen haben gezeigt, dass Benutzerverhalten und Kontextfaktoren den Informationsprozess weitaus stärker beeinflussen als dies so manche technische Parameter tun (cf. Tague-Sutcliffe & Blustein 1995). Dennoch werden diese Einflüsse nur sehr selten berücksichtigt, was in erster Linie daran liegt, dass sich derartige Kriterien nur schwierig fassen und in operationalisiertes Systemverhalten umsetzen lassen. Damit steht man vor der Problematik der Generalisierbarkeit von Aussagen und des Treffens der richtigen Auswahl.

Ein weiterer wichtiger Punkt kommt hinzu. Die technischen Möglichkeiten der Vernetzung lassen Digitale Bibliotheken auf die Informationsprodukte

verschiedener Anbieter zugreifen, da die Erweiterung der Datengrundlage die Suche für den Benutzer erleichtert. Die in einer virtuellen Bibliothek enthaltenen Objekte sind heterogen und auch ihre inhaltliche Erschließung und Repräsentation ist in der Regel unterschiedlich (cf. Krause 1998). Zum einen liegen eventuell automatisch und manuell indexierte Dokumente gemeinsam in einer Digitalen Bibliothek. Auch mit verschiedenen Thesauri manuell indexierte Dokumente führen zu Problemen bei der semantischen Konsistenz des Vokabulars. So ist z. B. das Vorkommen eines speziellen Fachbegriffs in einem Korpus von Zeitungstexten anders zu bewerten als in einer Sammlung von Dokumenten zu dem Fachgebiet. Da aus Kostengründen diese semantischen Probleme größtenteils nicht durch intellektuelle Nachbearbeitung gelöst werden, sind automatische Verfahren für die Transformation zwischen verschiedenen Repräsentationen erforderlich.

Die Anforderungen Digitaler Bibliotheken bilden den Ausgangspunkt. In Kapitel 2 wird das MIMOR-Modell beschrieben und formalisiert. Das dritte Kapitel ist den Erweiterungen gewidmet, Kapitel 4 und 5 führen Transformationen ein, zeigen deren Rolle in Digitalen Bibliotheken und stellen ein Integrationskonzept in das MIMOR-Modell vor. Im Fazit werden die Ergebnisse zusammengeführt.

2 Das MIMOR-Modell

Das im vorliegenden Beitrag dargestellte MIMOR-Modell (Mehrfach-indexierung zur dynamischen Methoden-Objekt-Relationierung; cf. Womser-Hacker 1997) verfolgt das Ziel, unterschiedliche Objekteigenschaften mit speziellen Benutzerkontexten zu kombinieren.

2.1 Ausgangslage

Innerhalb von Evaluierungsstudien ist selten Raum, um auf der Basis realer Kollektionen des Kontexts das adäquateste Verfahren zu ermitteln. Meist besteht also nicht die Möglichkeit, die erzielten Ergebnisse mit isolierten Funktionen (z. B. mit der Wirksamkeit linguistischer Analysen) innerhalb eines Systems zu verknüpfen. Die Maßzahlen haben Gültigkeit für den gesamten Systemansatz, wobei Wechselwirkungen zwischen den Funktionen schwer zu

kontrollieren sind. Dies gilt auch für die TREC¹-Studie, die derzeit größte und aktuellste Evaluierungsinitiative. Hier werden in mehreren Runden Entwicklungsprototypen aber auch kommerzielle IR-Systeme auf ihre Effektivität hin untersucht. Die teilnehmenden Institutionen durchlaufen mit ihren Systemen zeitlich kontrolliert einen Test mit vorgegebenen Aufgaben (sog. Topics) und Testdokumenten. Am NIST erfolgt die Bewertung, deren Ergebnisse zusammen mit den Systempräsentationen auf einer jährlichen Konferenz vorgestellt und anschließend publiziert werden (cf. Harman 1995, 1996; Voorhees & Harman 1998).

Auch nach mehrjähriger Laufzeit zeigt sich, dass statistische Maßzahlen wie Recall und Precision aufgrund ihrer vieldimensionalen Abhängigkeiten ohne weiterführende qualitative Analysen oder Experimentwiederholungen nur pauschale Erkenntnisse wiedergeben. So zeigte sich z. B., dass die Unterschiedlichkeit der statistischen Retrievalmodelle nur geringen Einfluss auf die Effektivität der Ergebnisse hat. Die Konsequenz daraus ist die Überprüfung der Güte der zusätzlichen Komponenten und Verfahrensweisen, die beim Informationsprozess eine Rolle spielen.

Das MIMOR-Modell (cf. Womser-Hacker 1997) baut v. a. auf einigen Ergebnissen der Serie der TREC-Konferenzen auf (cf. Voorhees & Harman 1998). Zwei der wichtigsten Ergebnisse von TREC haben MIMOR besonders beeinflusst:

Relevance Feedback durch den Benutzer erwies sich als eine sehr erfolgreiche Methode, um die Qualität von Retrieval-Ergebnissen zu verbessern. Die inhärente Interaktivität des Retrieval-Prozesses führt zur Erarbeitung von Wissen über den Benutzer, das das System z. B. für die Optimierung der Anfrage einsetzt (cf. Belkin 1993).

Viele Systeme erzielen Ergebnisse ähnlicher Qualität, d. h. vergleichbare Precision-Werte. Jedoch sind die Ergebnisse sehr verschieden, d. h. die Systeme führen zu unterschiedlichen Ergebnisdokumenten. Die Schnittmenge aus den Treffern verschiedener Systeme ist meist relativ klein. Einige Forscher setzen daher Fusionsverfahren ein. Sie kombinieren dabei die Ergebnisse verschiedener Ansätze mit dem Ziel, in dem Gesamtergebnis mehr Treffer und damit

¹ TREC (Text Retrieval Conference) ist beim National Institute of Standards and Technology (NIST) angesiedelt und hat nunmehr eine fast zehnjährige Laufzeit (cf. Harman & Voorhees 1998).

eine höhere Qualität zu erreichen (cf. Fox & Shaw 1994). Die Schwierigkeit bei Fusionsverfahren besteht in der Kombination der Ergebnisse.

MIMOR greift auf beide Erkenntnisse zurück und integriert sie in einem adaptiven Modell. Anhand von Relevance Feedback lernt das Modell, die Einzelergebnisse zu fusionieren. Dabei liegt folgendes Basiskonzept zugrunde. Während einer anfänglichen Einsatzphase wird das System einem adaptiven Selbsttest unterzogen. Die vorhandenen Deskribierungswerkzeuge werden parallel angewandt (Polyrepräsentation), wobei der Benutzer dies nach außen hin nicht wahrnimmt. Das einzige, was dem Benutzer abverlangt wird, ist eine Stellungnahme in Form von Relevance Feedback über die Relevanz der Ergebnisse. Dieses Feedback ermöglicht eine zweifache Nutzung: einerseits in seiner traditionellen Form zur Optimierung der Anfrage im Sinne eines interaktiven Retrieval, andererseits um eine kontextabhängige Qualitätseinschätzung der Deskribierungsverfahren zu erreichen. Ein systeminternes Bewertungsschema sammelt, verwaltet und analysiert die Urteile. Eine nachgeordnete Klassifikation versucht, eine Korrelation zwischen den Verfahren und bestimmten Eigenschaften der Objekte herzustellen. Ziel ist, dass sich während des Einsatzes des Information-Retrieval-Systems die geeignetsten Verfahren durchsetzen. Ein adaptierbares Lernkonzept kann zu einem ausgewählten Zeitpunkt eine Neuorganisation der Systemkomponenten vornehmen.

2.2 Modellbildung

MIMOR kann als eine zusätzliche Schicht in einem Information-Retrieval-System betrachtet werden, welche die Kombination verschiedener Ergebnisse verwaltet. Zu Beginn erhalten alle Verfahren das gleiche Gewicht. Damit beeinflussen alle Verfahren das Ergebnis gleich stark. Im Lauf des Lernprozesses adaptiert MIMOR die Gewichte so, dass Verfahren, die das Ergebnis positiv beeinflussen, gestärkt werden. Die Entscheidung, welche Dokumente im Ergebnis besonders relevant sind, trifft der Benutzer durch das Relevance Feedback. Die Verfahren, die den positiv bewerteten Dokumenten hohe System-Relevanz zuweisen, erhalten höheres Gewicht.

MIMOR ist nicht auf Text-Objekte beschränkt, sondern eignet sich auch für Multimedia-Retrieval oder strukturierte Objekte.

2.3 Formalisierung

Jedes Information-Retrieval-System berechnet für alle Dokumente eine System-Relevanz oder Retrieval Status Value (RSV) in bezug auf eine Anfrage. MIMOR kombiniert nun die Ergebnisse verschiedener Verfahren zu einem Gesamtergebnis. Tabelle 1 skizziert die Ausgangslage für MIMOR und andere Fusionsverfahren.

Tabelle 1: Ausgangssituation für MIMOR

	Dok. 1	Dok. 2	Dok. 3	Dok. 4	...
System A	RSV_{A1}	RSV_{A2}	RSV_{A3}	RSV_{A4}	...
System B	RSV_{B1}	RSV_{B2}	RSV_{B3}	RSV_{B4}	...
System C	RSV_{C1}	RSV_{C2}	RSV_{C3}	RSV_{C4}	...
...

⇓

MIMOR	RSV_1	RSV_2	RSV_3	RSV_4	...
-------	---------	---------	---------	---------	-----

Jedes System trägt mit einem Gewicht ω_s zu der RSV jedes Dokuments im MIMOR-Ergebnis bei. Folgende Formel liefert dieses Ergebnis für ein Dokument j bei einer gegebenen Anfrage:

$$(1) \quad RSV_j = \frac{\sum_{i=1}^N (w_i RSV_{ij})}{N}$$

RSV_{ij} Retrieval Status Value des Verfahrens i für Dokument j

Zu Beginn sind alle Gewichte gleich. Im Einsatz adaptiert MIMOR diese Gewichte abhängig von den Relevanz-Bewertungen der Benutzer. In der Regel kann der Benutzer Dokumente als relevant oder nicht relevant markieren. Im formalen Modell ordnet er jedem Dokument einen Relevance-Feedback-Wert R zu. Für Dokumente, die ohne explizites Urteil bleiben, nimmt R den Wert

Null an. Die Gewichte der einzelnen Verfahren ändern sich nach folgender Formel:

$$(2) \quad \mathbf{w}_i = \mathbf{w}_i + (e \text{ RSV}_{ij} R_j)$$

e *Lernrate*

R_j *Relevanz-Bewertung für Dokument j*

Das Gewicht eines Retrieval-Systems wächst, wenn es einem positiv bewerteten Dokument eine hohe RSV zugewiesen hat. Nach der Anwendung von Formel 2 auf alle Gewichte werden diese normalisiert, so dass ihre Summe stets konstant bleibt.

Das MIMOR-Modell besteht aus dem Vektor von Gewichten, den ein Administrator des Retrievalsystems einsehen und modifizieren kann. Dieser Administrator kann die Gewichte der einzelnen Verfahren überwachen und z.B. sehr schlechte Verfahren aus Performanzgründen aus dem Retrievalsystem nehmen.

3 Erweiterungen des MIMOR-Modells

Zahlreiche Untersuchungen im Information Retrieval haben gezeigt, dass die Qualität von Information-Retrieval-Verfahren vom Korpus abhängig ist. Die eingesetzten Algorithmen eignen sich dann mehr oder weniger gut für die Eigenschaften der Dokumente im Korpus. Ebenso beeinflussen die Eigenschaften der Anfragen die Wahl eines Information-Retrieval-Systems. Spezifische Ähnlichkeitsfunktionen entstanden z. B. für kurze Anfragen (cf. Kwok & Chan 1998, Wilkinson et al. 1996).

3.1 Benutzer-Eigenschaften

Viele Information-Retrieval-Tests haben gezeigt, dass die Relevanz-Bewertungen unterschiedlicher Benutzer oft nicht übereinstimmen (cf. Voorhees 1998). Um eine optimale Ausnutzung zu gewährleisten, müsste demnach jeder Benutzer sein eigenes MIMOR-Modell mit seinen Relevanz-Bewertungen trainieren. Dazu sind aber viele Bewertungen erforderlich und nicht jeder Benutzer wird diese einbringen. Benutzer geben aufgrund des erforderlichen Zeitaufwands oft kein Relevance Feedback, obwohl dies eine der

effektivsten Methoden zur Verbesserung von Information-Retrieval-Ergebnissen ist (cf. Over 1998). Auch Benutzer, die dazu bereit sind, verfügen in der Trainingsphase noch über ein suboptimales Modell, das bei erst wenigen Beispielen im Lernprozess eventuell zu extrem in eine Richtung tendiert. Das andere Extrem besteht in einem MIMOR-Modell, das alle Benutzer integriert und somit zwar für eine hohe Anzahl von Trainingsbeispielen sorgt, das aber auch keine Möglichkeit zur Individualisierung einräumt.

Eine Lösungsmöglichkeit besteht in der gleichzeitigen Pflege eines privaten und eines öffentlichen Modells, die beide zum Gesamtergebnis beitragen (cf. Womser-Hacker & Mandl 1999). Beide Modelle bestehen aus einem Vektor von Gewichten, die den Einfluss der einzelnen Verfahren in dem jeweiligen Modell beschreiben. Der Einfluss jedes Modells wird ebenfalls durch ein Gewicht bestimmt. Das private Modell trägt mit dem Gewicht p zum Gesamtergebnis bei und das öffentliche Modell mit dem Gewicht $(1 - p)$. Im Zeitraum der Benutzung des Systems kann sich p verändern. Zu Beginn der Arbeit mit dem System steht das private Modell auf der Basis weniger Relevanz-Bewertungen und ist damit noch unzuverlässig. Daher ist es sinnvoll, p und damit den Einfluss des privaten Modells niedrig zu halten und sich mehr auf das bereits ausgewogene öffentliche Modell zu verlassen. Je mehr Bewertungen ein Benutzer im Lauf der Arbeit mit dem System vornimmt, desto größer kann der Einfluss des privaten Modells werden. Das Gewicht p könnte auch mit steigender Sättigung des privaten Modells wachsen. Ändert sich das private Modell nur noch wenig, so kann es als weitgehend vollständig und stabil gelten. Das Gewicht könnte für fortgeschrittene Benutzer auch manipulierbar sein.

Das Gesamtergebnis berücksichtigt nun zwei MIMOR-Modelle und damit zwei Vektoren von Gewichten:

- Das private Modell: $(\omega_{\text{private, A}}; \omega_{\text{private, B}}; \omega_{\text{private, C}}; \dots; \omega_{\text{private, N}})$
- Das öffentliche Modell: $(\omega_{\text{public, A}}; \omega_{\text{public, B}}; \omega_{\text{public, C}}; \dots; \omega_{\text{public, N}})$

Das Gesamtergebnis integriert die RSV des privaten und des öffentlichen Modells:

$$(3) \quad RSV_j = \frac{\sum_{i=1}^N ((p w_{\text{private},i} + (1-p) w_{\text{public},i}) RSV_{ij})}{N}$$

3.2 Dokument-Eigenschaften

Das MIMOR-Modell wird durch die Berücksichtigung von Dokument-Eigenschaften weiter verbessert. Als Kriterien kommen hier z. B. Länge, Dokument-Typ (Zeitungstext, Fachtext, Online-Diskussionsbeitrag, ...), Anzahl von Termen oder Schwierigkeit in Frage. In einer Digitalen Bibliothek müssen diese Eigenschaften automatisch erkannt werden. Für die Schwierigkeit eines Textes kommen als formal erkennbare Größen z. B. syntaktische Komplexität oder Anzahl von Wörtern in Frage (cf. z. B. Ballod 2000). Diese Eigenschaften können im System als Cluster von Dokumenten, die diese Eigenschaft besitzen, modelliert werden². Die Güte der Cluster bildet die Grundlage für den Erfolg dieses Ansatzes. Die Dokumente innerhalb eines Clusters sollten möglichst homogen und ähnlich sein, während Dokumente aus verschiedenen Clustern unterschiedlich sein sollten.

Tabelle 2: Dokument-Cluster mit verschiedenen Gewichten, $x \hat{I} \{private, public\}$

	Cluster 1			Cluster 2		
	Doc. 1.1	Dok. 1.2	...	Dok. 2.1	Dok. 2.2	...
System A	$w_{x, 1, A}$			$w_{x, 2, A}$		
System B	$w_{x, 1, B}$			$w_{x, 2, B}$		
System C	$w_{x, 1, C}$			$w_{x, 2, C}$		
...		

Jedes Cluster verfügt über ein eigenes, veränderbares Gewicht für jedes Information-Retrieval-Verfahren. Damit kann ein Verfahren in einem Cluster ein hohes Gewicht und in einem anderen ein niedriges Gewicht haben. Dieses Verfahren eignet sich dann gut für Dokumente mit der Eigenschaft des ersten Clusters und weniger gut für Dokumente mit der Eigenschaft des zweiten

² Momentan sind die Cluster überschneidungsfrei modelliert. Da in der Praxis sicher häufig Mehrfachzuordnungen auftreten, ist eine Erweiterung auf sich überlappende Fuzzy-Cluster geplant.

Clusters. Tabelle 2 skizziert die Verteilung unterschiedlicher Gewichte auf die Dokument-Eigenschaften.

Je nach Zugehörigkeit des in Frage stehenden Dokuments wird für die Berechnung der RSV ein anderes Gewicht verwendet. Gehört das Dokument j zum Cluster C , dann ergibt sich in Abwandlung von Formel 1:

$$(4) \quad RSV_{j \in C} = \frac{\sum_{i=1}^N (w_{x,c,j} RSV_{ij})}{N}$$

R_{cj} Relevanz-Bewertung für Dokument j in Cluster C

RSV_{ij} Retrieval Status Value von Verfahren i für Dokument j

Die Lernformel für die Veränderung der Gewichte ist fast identisch mit der Lernformel (2) für MIMOR. Die Adaption wirkt nun aber nur für die Gewichte der Verfahren für das Cluster, in dem das bewertete Dokument liegt.

$$(5) \quad w_{x,c,i} = w_{x,c,i} + (e R_{cj} RSV_{ij})$$

e Lernrate

R_{cj} Relevanz-Bewertung für Dokument j in Cluster c

RSV_{ij} Retrieval Status Value von Verfahren i für Dokument j

Für diese Erweiterung des Modells benötigt das Lernverfahren mehr Trainingsbeispiele, da die Anzahl der zu trainierenden Parameter mit der Anzahl der Cluster wächst. Die Eignung bestimmter Cluster bzw. Eigenschaften kann nur empirisch überprüft werden. Es ist auch unwahrscheinlich, dass sich allgemeingültige Regeln finden, vielmehr dürften gute Kandidaten für Dokument-Eigenschaften stark von der jeweiligen Kollektion abhängen.

3.2 Anfrage-Eigenschaften

Da auch die Eigenschaften der Anfragen bei der Bewertung der Qualität von Information-Retrieval-Systemen eine Rolle spielen (cf. Voorhees & Harman 1998), sollte ein Fusions-Verfahren darauf Rücksicht nehmen. MIMOR berücksichtigt daher analog zu den Dokument-Eigenschaften auch die Eigen-

schaften von Anfragen, die wiederum durch die Zusammenfassung in Cluster modelliert werden. Dabei sind weniger Cluster zu erwarten als bei den Dokumenten, da bei den Anfragen eine geringere Varietät bezüglich ihrer formal bestimmbaren Eigenschaften vorliegt. In der Regel gibt es wesentlich weniger Anfragen als Dokumente, die Anfragen sind meist erheblich kürzer und verschiedene Texttypen wie bei den Dokumenten sind bei den Anfragen kaum denkbar.

Die Berechnungsformel für das Endergebnis und die Lernformel für die Adaption der Gewichte der Cluster ergeben sich analog zu den Formeln für die Dokument-Eigenschaften.

4 Transformationen in Digitalen Bibliotheken

Die semantische Heterogenität in umfassenden Digitalen Bibliotheken ist eine Folge der Integration verschiedener Objekte (Volltext, Abstracts, Fachliteratur, multimediale Dokumente, ...) und deren heterogener Erschließung (automatische Indexierung, Thesauri, Klassifikationen, Verschlagwortung durch Autoren oder Dokumentare, etc.).

Ein Beispiel für ein derartiges System ist die geplante Virtuelle Fachbibliothek Sozialwissenschaften, in der Bibliotheken, Verlage, die Friedrich-Ebert-Stiftung und das Informationszentrum Sozialwissenschaften eine gemeinsame Plattform für die Integration ihrer bisher getrennten Datenbestände schaffen und so die Versorgung der Fachwissenschaftler mit Literatur verbessern (cf. Kluck et al. 2000).

Um mit der semantischen Heterogenität der unterschiedlich erschlossenen Dokumente umzugehen, sind entsprechende Komponenten erforderlich. Da die meisten Retrieval-Verfahren auf einer homogenen Repräsentation aufsetzen, müssen automatische Verfahren gefunden werden, die Dokumente zwischen verschiedenen Repräsentationen transformieren, um so eine homogene Grundlage für das Retrieval zu schaffen. Kuhlen etwa spricht von einer „postkoordinierenden Ordnung durch transformierende Anpassung“ (Kuhlen 1999:138).

Der große Umfang der Daten in Digitalen Bibliotheken spricht gegen eine ausschließlich intellektuelle Lösung. Basis von automatischen Verfahren für Transformationen ist meist ein Korpus, dessen Objekte in zwei Eigenschafts-

räumen repräsentiert sind. In der Regel bestehen diese Doppelkorpora aus Texten, die nach zwei Indexierungsverfahren erschlossen wurden.

Bisher leisten meist statistische Verfahren auf der Basis einer Kookkurrenz-Analyse die Transformation (cf. z. B. Ferber 1997, Lam & Yang 1998). Mandl 1998 schlägt auch neuronale Netze vor, wobei das häufig eingesetzte Backpropagation-Netzwerk als besonders erfolgversprechend gilt. Der Backpropagation-Algorithmus ist ein Lernverfahren, das anhand von Trainingsbeispielen eine Abbildung von einem n-dimensionalen in einen m-dimensionalen Raum lernt. Eine versteckte Schicht von Neuronen zwischen Input- und Output-Schicht erhöht die Mächtigkeit und erlaubt auch das Erlernen komplexer Funktionen (cf. Scherer 1997).

Ein Vergleich der Qualität von Transformationsverfahren in Mandl 2000 führte zu interessanten Ergebnissen. Dabei wurden ein statistisches Verfahren, das auf der Basis von Kookkurrenzen Assoziationswerte zwischen den Term-Räumen berechnet, und ein neuronales Backpropagation-Netzwerk verglichen, das auf mit Latent Semantic Indexing (cf. Berry et al. 1995) reduzierte Repräsentationen zugreift.

Testgrundlage bilden Daten des Informationszentrum Sozialwissenschaften (IZ) und der Universitäts- und Stadtbibliothek Köln (USB). Das IZ indexiert sozialwissenschaftliche Dokumente mit einem Thesaurus von Schlagwörtern und zusätzlich mit einer Klassifikation wissenschaftlicher Fachgebiete. Die USB verschlagwortet teilweise die gleichen Dokumente nach ihrer Bibliotheksklassifikation. In einem Experiment wurde eine Transformation vom IZ-Thesaurus auf die IZ-Klassifikation mit 12.000 Dokumenten trainiert und getestet und in einem zweiten Experiment wurde eine Abbildung von der USB-Klassifikation zu einem Teil des IZ-Thesaurus mit 15.000 Dokumenten trainiert und getestet. Im ersten Fall war die Qualität der Transformation beim statistischen Verfahren und beim Backpropagation-Netzwerk sehr ähnlich und im zweiten Experiment erwies sich das Backpropagation-Netzwerk als besser (cf. Mandl 2000).

Interessanterweise ist in beiden Fällen die Schnittmenge der Ergebnisse sehr klein. Im ersten Experiment führen beide Verfahren zu vergleichbarer Qualität, die Treffer sind aber sehr unterschiedlich. Damit ergibt sich in Mandl 2000 für Transformationen ein ähnliches Bild wie bei TREC für Retrieval. Als Konsequenz daraus sollte auch bei Transformationen mit Fusionsansätzen experimentiert werden, um die Resultate mehrerer Verfahren zu kombinieren.

5 Anwendung von MIMOR auf Transformationen

Das MIMOR-Modell lässt sich für Transformationen adaptieren. Liegen Relevance-Feedback-Informationen zu den Zuordnungen vor, dann lernt MIMOR, den Einfluss einzelner Transformations-Verfahren auf das Gesamtergebnis zu optimieren.

Transformationen bilden Dokumente aus einem Term-Raum in einen anderen Term-Raum ab. Dazu weisen sie einem Dokument für jeden Term im Ziel-Term-Raum ein Gewicht zu. Eine Transformations-Funktion ordnet also jedem Paar aus Dokument und Zielklasse ein Gewicht zu. Bei der Kombination mehrerer Verfahren muss der Fusionsalgorithmus aus den Gewichten aller Verfahren ein Gesamtgewicht bilden. Wie Tabelle 3 zeigt, ist damit die Ausgangssituation ähnlich zu der von MIMOR.

Tabelle 3: Ausgangssituation für MIMOR für die Transformation eines Dokuments

	Term 1	Term 2	Term 3	Term 4	...
Verfahren A	G_{A1}	G_{A2}	G_{A3}	G_{A4}	...
Verfahren B	G_{B1}	G_{B2}	G_{B3}	G_{B4}	...
Verfahren C	G_{C1}	G_{C2}	G_{C3}	G_{C4}	...
...

⇓

MIMOR	G_1	G_2	G_3	G_4	...
-------	-------	-------	-------	-------	-----

Jedes Verfahren erhält zu Beginn des Prozesses den gleichen Einfluss auf das Gesamtergebnis, das sich aus folgender Formel ergibt:

$$(6) \quad G_j = \frac{\sum_{i=1}^N w_i G_{ij}}{N}$$

G_j	<i>Gewicht von Term j für das Dokument im Gesamtergebnis</i>
G_{ij}	<i>Gewicht von Term j für das Dokument von Verfahren i</i>
w_i	<i>Gewichtung von Verfahren i</i>

Im Laufe der Interaktion adaptiert MIMOR die Gewichte so, dass die Verfahren, die positiv auf das Gesamtergebnis wirken, ein höheres Gewicht erhalten. Um die Wirkung zu bestimmen, beurteilt der Benutzer Dokumente als relevant oder nicht relevant für die Zielklasse. Das Relevance Feedback bezieht sich jetzt also nicht mehr auf die Relevanz zu einer Anfrage, sondern auf die Adäquatheit einer Zuordnung zu einem Term. Die Gewichte der einzelnen Verfahren ändern sich ähnlich wie in der Lernformel (2):

$$(7) \quad w_i = w_i + (e G_{ij} R_j)$$

R_j *Relevanz-Bewertung für Dokument j bezüglich der Zielklasse*

Ein Problem besteht in der Gewinnung von Relevance-Feedback-Informationen für die Transformation zwischen Term-Räumen, da Benutzer sinnvollerweise nur das Endergebnis eines Suchprozesses und damit Dokumente bewerten können. Eine direkte Bewertung der Transformation kommt evtl. für Fachleute in Frage, jedoch ist die Abschätzung des Retrievalerfolgs aufgrund einer Liste von Termen sehr schwierig. Die Relevance-Feedback-Bewertung von Dokumenten muss also in eine Bewertung der Terme umgesetzt werden. Dabei werden die Transformations-Verfahren belohnt, die Terme hoch gewichtet haben, die wiederum zu positiv bewerteten Dokumenten führten. Dazu wird aus den Relevanz-Feedback-Informationen ein optimaler Term-Vektor gewonnen. Die Term-Vektoren der bewerteten Dokumente werden mit dem jeweiligen, numerisch umgesetzten Relevanz-Urteil multipliziert und anschließend summiert:

$$(8) \quad \vec{o}_j = \sum_{i=1}^B (R_i \vec{d}_i)$$

B Anzahl der bewerteten Dokumente

Die Lernformel berechnet den Erfolg eines Verfahrens als Mittel des Erfolgs bei allen Termen im Zielvokabular:

$$(9) \quad \mathbf{w}_I = \mathbf{w}_I + \mathbf{e} \sum_{j=1}^{Dim} G_{ij} (o_j - Erg_j)$$

<i>Dim</i>	Anzahl Vektoren im Zielvokabular
<i>o_j</i>	optimaler Wert für Term <i>j</i> laut Relevanz Feedback
<i>Erg_j</i>	berechneter Wert für Term <i>j</i>

Völlig analog lassen sich auch die Ansätze zur Erweiterung von MIMOR auf die Transformationen übertragen. Auch die Transformation kann in ein privates und öffentliches Modell gespalten werden, die jeweils mit einem eigenen Gewicht ins Ergebnis eingehen. Weiterhin können die Dokumente und Zielklassen je nach ihren Eigenschaften in Cluster unterteilt werden, von denen jedes ein eigenes Gewicht für alle Verfahren erhält. Damit erhält z. B. ein Verfahren, das sich für bestimmte Dokument-Eigenschaften bewährt, für diese Dokumente ein höheres Gewicht.

6 Fazit

Digitale Bibliotheken verändern die Anforderungen an Information-Retrieval-Systeme und erfordern zusätzliche Komponenten. Im vorliegenden Zusammenhang wurde besonderes Augenmerk auf die unterschiedlichen Repräsentationsformen informationeller Objekte gelegt. Deren Integration in einem Fusionsverfahren erlaubt es, dass die Stärken und Schwächen einzelner Repräsentationsformen sehr differenziert zur Geltung kommen. Eine Implementierung und anschließende Evaluierung von MIMOR in JAVA ist geplant. Dabei sollen die Komponenten so realisiert werden, dass sie soweit möglich auch für Transformationen einsetzbar sind.

Literaturverzeichnis

- Ballod, Matthias (2000): Comprehensible Science – Computer-Supported Text Analysis of Publications on Gene Technology. In: Global Dialogue. Science and Technology – Thinking the Future at EXPO 2000 Hannover. 11. – 13. 7. 2000. <http://www.shaping-the-future.de/pdf_www/207_paper.pdf>.
- Belkin, N. (1993): Interaction with Texts: Information Retrieval as Information Seeking Behavior. In: Knorz, G.; Krause, J.; Womser-Hacker, C. (Eds.): Information Retrieval '93. Von der Modellierung zur Anwendung. Proc. d. 1. Tagung Information Retrieval. Konstanz. S. 55 – 66.
- Berry, M.; Dumais, S.; Letsche, T. (1995): Computational Methods for Intelligent Information Access. In: Proc. of ACM Supercomputing '95. San Diego, CA. S. 1 – 38.
- Croft, B.; Moffat, A.; Rijsbergen, K. van; Wilkinson, R.; Zobel, J. (1998) (Eds.): Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '98). Melbourne 24. – 28 .8. 1998.
- Ferber, R. (1997): Automated Indexing with Thesaurus Descriptors: A Cooccurrence Base Approach to Multilingual Retrieval. In: Peters, C.; Thanos, C. (Eds.): Research and Advanced Technology for Digital Libraries. 1st European Conf. ECDL '97. Pisa, 1. – 3. 9. 1997. Berlin et al.. S. 233 – 252.
- Fox, E.; Shaw, J. (1994): Combination of Multiple Searches. In: Harman, D. (Ed.): The Second Text Retrieval Conference (TREC-2). NIST Publ. 500-215. S. 243 – 252.
- Fuhr, N. (1999): Information Retrieval in Digitalen Bibliotheken. In: Schmidt, Ralph (Hrsg.): Proc. 21. Online Tagung der DGI. Aufbruch ins Wissensmanagement. Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis. Frankfurt, 18. – 20. 5. 1999. S. 93-102.
- Harman, D. (1995): The TREC conferences. In: Kuhlen, Rainer; Rittberger, Marc (Eds.) (1995): HIM '95. Hypertext, Information Retrieval, Multimedia. Synergieeffekte elektronischer Informationssysteme. Konstanz. 5. – 7. 4. 1995. Konstanz. S. 9 – 28.
- Harman, D. (Ed.) (1996): The Fourth Text Retrieval Conference (TREC-4). NIST Special Publication 500-236. National Institute of Standards and Technology. Gaithersburg, Maryland, 1. – 3. 11. 1995. <http://trec.nist.gov/pubs/trec4/t4_proceedings.html>.
- Kluck, M.; Krause, J.; Müller, M.; in Kooperation mit: Schmiede, R.; Wenzel, H.; Winkler, S.; Meier, W. (2000): Virtuelle Fachbibliothek Sozialwissenschaften: IZ-Arbeitsbericht 19, IZ Sozialwissenschaften, Bonn. <<http://www.bonn.iz-soz.de/publications/series/working-papers/index.htm#Virtuelle>>.
- Krause, Jürgen (1998): Innovative Current Research Information Systems in the Information Society. In: CRIS '98 Current Research Information Systems. Luxemburg, 12-14.3.1998. <<ftp://ftp.cordis.lu/pub/cybercafe/docs/krause.zip>>.

- Kuhlen, Rainer (1999): Die Konsequenzen von Informationsassistenten: Was bedeutet informationelle Autonomie oder wie kann Vertrauen in elektronische Dienste in offenen Informationsmärkten gesichert werden? Frankfurt a. M.
- Kwok, K. L.; Chan, M. (1998): Improving Two-Stage Ad-Hoc Retrieval for Short Queries. In: Croft et al. (1998). S. 250 – 256.
- Lam, Wai; Ho, Chao Yang (1998): Using a Generalized Instance Set for Automatic Text Categorization. In: Croft et al. (1998).
- Mandl, T. (1998): Vague Transformations in Information Retrieval. In: Zimmermann, H.; Schramm, V. (Eds.): Knowledge Management und Kommunikationssysteme: Workflow Management, Multimedia, Knowledge Transfer. Proc. 6. Int. Symposium für Informationswissenschaft. (ISI '98). 3. – 7. 11. 1998, Prag, Konstanz. S. 312 – 325.
- Mandl, T. (2000): Einsatz neuronaler Netze im Information Retrieval. Arbeitsbericht. Informationswissenschaft, Universität Hildesheim.
- Over, P. (1998): TREC-6 Interactive Track Report. In: Voorhees/Harman (1998). S. 73–81.
- Scherer, A. (1997): Neuronale Netze: Grundlagen und Anwendungen.
- Tague-Sutcliffe, J.; Blustein, J. (1995): A Statistical Analysis of the TREC-3 Data. In: Harman, D. (Hrsg.): The Third Text REtrieval Conference (TREC-3). NIST Special Publication 500-225. National Institute of Standards and Technology. Gaithersburg, Maryland, 2.-4.11.1994. S. 385.
- Voorhees, E. (1998): Variations in relevance judgments and the measurement of retrieval effectiveness. In: Croft et al. 1998. S. 315-323
- Voorhees, E.; Harman, D. (1998): The Sixth Text Retrieval Conference (TREC-6). NIST Special Publication 500-240. August 1998.
<http://trec.nist.gov/pubs/trec6/t6_proceedings.html>.
- Wilkinson, R.; Zobel, J.; Sacks-Davis, R. (1996): Similarity Measures for Short Queries. In: Harman (1996).
- Womser-Hacker, C. (1997): Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitationsschrift. Universität Regensburg.
- Womser-Hacker C.; Mandl, T. (1999): Adapting Meta Information Retrieval to User Preferences and Document Features. In: Bullinger, H.-J.; Ziegler, J. (Eds.): Human-Computer Interaction: Communication, Cooperation and Application Design. Proceedings of the HCI International '99 (8th International Conference on Human-Computer Interaction), Munich, Germany, 22 – 27. 8. 1999. Lawrence Erlbaum Associates: Mahwah, NJ; London. Vol. 2. S. 604 – 608.