



Qualitätsbewertung von Information-Retrieval-Systemen

Sind Synergieeffekte durch die Koordination verschiedener Bewertungsmethoden möglich?

Osman Bayraktar

IDS Prof. Scheer GmbH; Altenkesseler Str. 17; 66115 Saarbrücken
o.bayraktar@ids-scheer.de

Christa Womser-Hacker

Universität Konstanz; Informationswissenschaft; Postfach D 78; D-78457 Konstanz
cwh@inf-wiss.uni-konstanz.de

Inhalt

- 1 Einleitung
- 2 Zielrichtungen der Qualitätsbewertung
 - 2.1 Auswahl eines IRS
 - 2.2 Traditionelle IR-Evaluierung
- 3 Kombination: Ein Blick in die Zukunft
- 4 Fazit

Zusammenfassung

Ziel dieses Beitrags ist die Gegenüberstellung verschiedener Bewertungsmethoden und Zweckbestimmungen im Bereich von Information-Retrieval-Systemen. Es wird aufgezeigt, welche Ansätze vorliegen und hinterfragt, wie sie einander angepaßt werden können, um synergetische Effekte zu erzielen.

1 Einleitung

Zunehmender Wettbewerbsdruck und die Globalisierung von Märkten haben zur Folge, daß die Kundenorientierung im Produktionsmanagement, und hier vor allem im Qualitätsmanagement, stärker als je zuvor Berücksichtigung findet. Die Aufgaben unterliegen einem permanenten Wechsel; ihre Erfüllung hängt immer mehr von der Güte und Qualität der bereitgestellten Informationen ab. Der Einsatz von Informationstechnologien und das Management von Wissen haben sich daher für die Unternehmensführung zu einer strategischen Aufgabe entwickelt. Dabei liegt es am Informationsmanager, die strategische Rolle der Informationsfunktion und die inner- und außerbetrieblichen Bedingungen für die Umsetzung des Leistungspotentials der Informationsinfrastruktur zu analysieren und zu bestimmen. Nach einer Umfrage der Gartner Group liegen 90% alles Wissens in textueller Form vor. Deshalb werden Information-Retrieval-Systeme (IRS) zunehmend neben anderen Typen von Informationssystemen als integrale Bestandteile der Informationsinfrastruktur moderner Betriebswirtschaften angesehen, welche die nötigen Informationen (Informationsnachfrage) für die Bewältigung der Aufgaben wirtschaftlich auf der operativen, administrativen und strategischen Ebene bereitstellen. Dabei wird beim begrifflichen Verständnis von IRS über die traditionelle Sichtweise des Text- und Dokumentenretrieval hinausgegangen und auf Systeme erweitert, welche Information in verschiedenen Medien und Sprachen integrieren und auch Mischformen (wie Texte, Bilder, Fakten) zulassen.

Seit einigen Jahren existieren IRS mit komplexer Funktionalität (z.B. gewichtete Indexierung, Ranking, Relevance Feedback etc.) nicht nur als Prototypem im Bereich der Forschung. Einige Systeme haben einen Stand der Reife erreicht, so daß fertige Produktlösungen auch im kommerziellen Bereich zur Verfügung stehen. Durch diese Situation auf dem Informationsmarkt und durch die Neuorientierung vieler Organisationen und Unternehmen, der Informationsflut Herr zu werden, ergibt sich auch im Bereich der (vergleichenden) Qualitätsbewertung dieser Produkte eine zu überdenkende Situation im Hinblick auf die einzusetzenden Methoden.

2 Zielrichtungen der Qualitätsbewertung

Bei der Qualitätsbewertung von IRS oder IRS-Komponenten stehen sich unterschiedliche Zielrichtungen gegenüber, die wiederum in Kombination mit unterschiedlichen Methoden auftreten. Das Spektrum reicht hier von kontrollierten Experimenten, Untersuchungen, empirischen Langzeitstudien, Rapid-Prototyping-Verfahren und Simulationstests bis hin zu betriebswirtschaftlich orientierten Vorgehensweisen wie Benchmarking und der Analyse von Kosten-Nutzen-Faktoren oder (kritischer) Erfolgsfaktoren.

Die Zweckbestimmungen dieser Ansätze reichen von Anschaffungsentscheidungen, Funktionalitätsprüfungen, Zertifizierungsbestrebungen z.B. im Rahmen von Vermittlungsdiensten bis hin zu Initiativen genereller Natur, bei welchen das wissenschaftliche Interesse an der Funktionsweise des Retrievalprozesses dominiert. All diese Vorgehensweisen sind derzeit nur schwierig übertragbar und für andere

Zwecke zu verwenden, da sie äußerst unterschiedlichen Bewertungsphilosophien folgen.

Exemplarisch werden hier zwei relativ weit auseinander liegende Zielrichtungen (die Anschaffungsentscheidung eines IRS und die TREC-Studie) gegenübergestellt. Anschließend soll geprüft werden, an welchen Stellen Veränderungen möglich sind, die synergetische Effekte für eine Mehrfachnutzung und –interpretation erwarten lassen.

2.1 Auswahl eines IRS

Bei der Neuanschaffung von IRS stellt sich die Frage nach der Vorgehensweise, mit der prognostiziert werden kann, wie sich ein System im gegebenen Kontext verhalten wird. Idealerweise besteht die Möglichkeit eines Retrievaltests, welcher die auf dem Markt vorhandenen Systeme mit den tatsächlichen Gegebenheiten des intendierten Einsatzes in Beziehung setzt. Dabei wird auch die vorliegende Infrastruktur sowohl in bezug auf Soft- und Hardwareausstattung (mit Blick auf die Kompatibilität) als auch die intendierte Nutzungssituation berücksichtigt. Diese Strategie wurde mit der PADOK-Studie zur Ermittlung eines Indexierungsverfahrens für das Deutsche Patentinformationssystem verfolgt (cf. KRAUSE/WOMSER-HACKER 1990). Aufgrund der Aufwendigkeit und Kostspieligkeit derartiger Tests kommen solche Verfahren selten zum Einsatz, obwohl sie optimale Ergebnisse erwarten lassen.

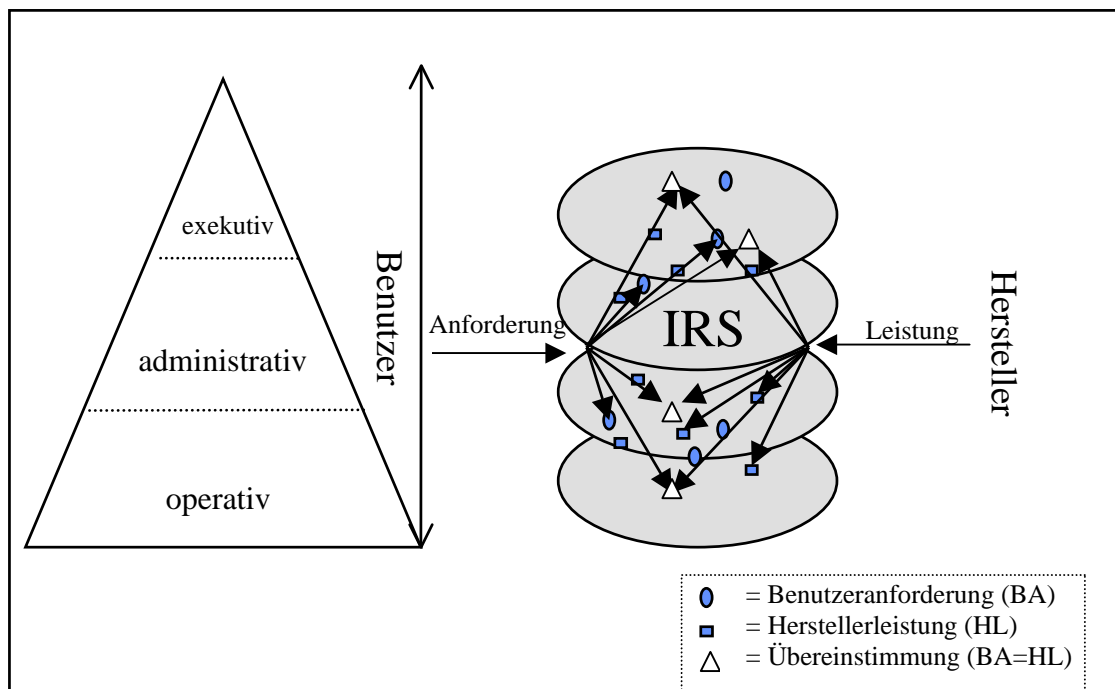


Abb. 1: Benutzersicht/-anforderung vs. Herstellersicht/-leistung

Abb.1 stellt die Situation dar, deren Ziel es ist, ein IRS auszuwählen, bei welchem sich möglichst viele Anforderungen von Seiten der Benutzer mit Herstellerleistungen decken. Dabei können die „Benutzer“ den verschiedenen Ebenen einer Orga-

nisation angehören oder was den häufigeren Fall darstellt, die entscheidende Gruppe nicht identisch mit der das IRS hauptsächlich nutzenden Gruppe sein. Letzteres erfordert einen ebenenübergreifenden Kommunikationsprozeß. Als methodisches Instrumentarium bietet sich eine Erfolgsfaktorenanalyse an um zu klären, welche IRS-Eigenschaften von den betroffenen Mitgliedern der unterschiedlichen organisatorischen Ebenen als relevant eingestuft werden. Für diese Eigenschaften werden Qualitätsaussagen benötigt, mit dem Ziel, sie miteinander zu kombinieren, um eine die Prioritäten berücksichtigende Gesamtbewertung zu erhalten.

2.2 Traditionelle IR-Evaluierung

Dem gegenüber stehen traditionelle Evaluierungskonzepte des IR, die meist nur in sehr geringem Maße an der Berücksichtigung von Benutzerinteressen orientiert sind. Im Vordergrund steht die Effektivität der Retrievalmaschinen, d.h. des Retrievalkerns, der für den Match zwischen Informationsobjekten und Anfragen verantwortlich ist (z.B. Vector Space Model vs. Probabilistisches Retrievalmodell) und das Ziel verfolgt, relevante Objekte zu selektieren und gleichzeitig irrelevante zurückzuhalten.

Im Information Retrieval wird die Entwicklung der Funktionalität seit jeher begleitet von Evaluierungsaktivitäten (cf. z.B. die Arbeiten zum SMART-System in SALTON 1971). Zu diesem Zwecke hat sich ein komplexer Methodenkanon herausgebildet, der in erster Linie auf statistischen Verfahren basiert, die auf den Information-Retrieval-Bereich übertragen wurden (Basismaße *recall* und *precision*, *single-number-measures* wie das E-Maß nach VAN RIJSBERGEN etc.). Neben der Forschung im Bereich der mathematischen Meßtheorie (cf. BOLLMANN 1977 und 1981, BOLLMANN/CHERNIAVSKY 1981) wurden Benutzeraspekte nur sehr rudimentär in die Maßzahlen integriert. Ein Beispiel ist die Formalisierung von sog. Benutzerstandpunkten z.B. im *normalisierten recall*, welche die Position von Ergebnis-Objekten in Ranking-Listen miteinbeziehen. Als Benutzerstandpunkt gilt z.B. eine spezifische Benutzerverhaltensweise, die Auswertung der Ranking-Liste nach drei aufeinander folgenden irrelevanten Dokumenten abzuberechnen. Für die Berechnung der Maße werden nur die Objekte bis zum festgelegten Cut-off einbezogen. Daß eine derartige Formalisierung aufgrund ihrer Pauschalität das reale Benutzerverhalten nur sehr bedingt abbilden kann, versteht sich von selbst.

Als derzeit größte Evaluierungsinitiative und als aktueller Endpunkt der langen Evaluierungstradition im IR sind die Text-Retrieval-Konferenzen (TREC) anzusehen (cf. HARMAN 1995 und 1998). Sie sollen dargestellt und in diesem Rahmen diskutiert werden (auch in bezug auf ihre Tauglichkeit für Anschaffungsentscheidungen).

TREC hat sich in den USA seit Anfang der neunziger Jahre als Initiative etabliert, die von staatlicher Seite (ARPA Advanced Projects Agency) am NIST (National Institute of Standards and Technology) gefördert wird. Auf der Basis großer Dokumentkollektionen (Dokumente, Topics und Relevanzurteile) wurden beispielsweise in der sechsten TREC-Runde 58 Retrievalsysteme verglichen (cf. VOORHEES/HARMAN 1998). Die teilnehmenden Systeme waren Forschungssysteme

aus Universitäten und wissenschaftlichen Industrieabteilungen, aber auch kommerzielle Systeme nahmen teil. Von allen Systemen wurden umfangreiche Angaben über die notwendigen Ressourcen (man and machine power) gefordert. Benutzer spielen bisher bei TREC noch keine Rolle, da die Erarbeitung der Anfragen aus den Topics von den jeweiligen Entwicklern geleistet wird. Diese versuchen, auf der Basis ihrer optimalen Kenntnis des jeweiligen Systems dessen Wirkung so positiv wie möglich zu steuern. Benutzer, die ihr System oft als Black Box wahrnehmen, reagieren folglich auf völlig andere Art und Weise. Da im Rahmen von TREC auch die Evaluierungsmethode auf die Probe gestellt wird, erfolgt über die Phasen hinweg eine gezielte Variation der Testgrundlagen (z.B. Verwendung stark elaborierter und wenig elaborierter Topics). Das aus TREC zur Verfügung stehende Material ist äußerst umfangreich, jedoch noch sehr wenig analysiert. Es beinhaltet neben den Testgrundlagen, den Fragebögen zu den einzelnen Systemen einen Leistungsvergleich der Systeme auf der Basis statistischer recall-precision-Werte. Die wichtigsten Maßzahlen bilden die Ergebnisse (inkl. ihrer graphischen Darstellung) der sog. frozen ranks method. Hier werden an vorher festgelegten Rängen Einschnitte vorgenommen und punktuelle precision-Werte berechnet, die anschließend einer Mittelung unterzogen werden. Als sehr aufschlußreiche, aber zeitaufwendige Informationsquelle sind in den TREC-Proceedings Beiträge der einzelnen Teilnehmergruppen enthalten, welche die jeweiligen Vorgehensweisen ihrer Gruppe im Detail beschreiben.

TREC ist auf Ranking-Systeme spezialisiert, d.h. Boolesches Retrieval nimmt am Test nicht teil. Insofern steht hier die Fähigkeit der Systeme im Vordergrund, möglichst viel relevantes Material auf guten Rangplätzen zu positionieren. Hier spielen neben der Gewichtungsfunktion in erster Linie Relevance-Feedback-Ansätze und Anfrageoptimierungstechniken eine entscheidende Rolle. In der Praxis liegt meist die notwendige gewichtete Indexierung nicht vor.

TREC ergänzt die klassische Evaluierungsform durch die Einbeziehung von Effizienzgesichtspunkten wie Systemressourcen, manueller und maschineller Aufwand, Input/Output-Operationen etc., jedoch sind diese Faktoren, die in einem Fragebogen ermittelt werden, bisher noch in keinster Weise Gegenstand weiterführender Analysen geworden.

Bei TREC wird die Hauptevaluierungslinie durch Nebenstränge (sog. special TREC tracks) ergänzt, die besondere Aspekte berücksichtigen (z.B. Multi- und Crosslingualität, Filtering, Verteilte Datenbanken, Spoken Language, Interaktivität etc.).

Die Interpretation der TREC-Ergebnisse kann unter verschiedenen Perspektiven geschehen. Zum einen werden die Entwicklungsgruppen die Veränderungen der Ergebnisse mit den Veränderungen im Systemdesign über die Phasen hinweg in Beziehung setzen. Wesentlich schwieriger ist es, bestimmte Werte mit isolierten Funktionen zu verbinden, d.h. z.B. eine Verbesserung in der precision mit einer aufwendigeren linguistischen Analyse zu begründen. Diese Herangehensweise ist möglicherweise äußerst versierten Kennern der TREC-Zusammenhänge möglich.

3 Kombination: Ein Blick in die Zukunft

Die Qualität von Systemen hängt in starkem Maße von den jeweiligen Bedingungen ihres zukünftigen Einsatzes ab. Insofern ist der Kontext und die Rahmenbe-

dingungen von großer Bedeutung für die Auswahl eines geeigneten Systems. Dies ist mit einer vielschichtigen Problematik verbunden.

Zunächst sind die Faktoren zu ermitteln, die für die Bewertung innerhalb einer Organisation oder eines Unternehmens entscheidend sind. Dies erfordert einen iterativen Prozeß, um die Anforderungen exakt zu erarbeiten bzw. herauszudestillieren und miteinander in Beziehung zu setzen. Wichtig ist, daß durch die Beteiligung von Personen mit verschiedenen Aufgaben und den sie einbeziehenden Kommunikationsprozeß sichergestellt wird, daß ein homogenes Verständnis vorliegt, das die Faktoren nicht nur in abstrakter Form anführt, sondern ihre operative Bedeutung miteinbezieht (z.B. Was bedeutet „Benutzerfreundlichkeit“, „Transparenz“, „Selbsterklärungsfähigkeit“ im IR-Kontext?¹). Hinzu kommt die Notwendigkeit, Wissen über die Potentiale aktueller IRS an dieser Stelle einzubringen. Als Methode wurde auf die Erfolgsfaktorenanalyse (EFA) (cf. HEINRICH 1996) zurückgegriffen, mit der es möglich ist, Qualitätsmerkmale von IRS aus der jeweiligen Benutzersicht und vor dem Hintergrund der jeweiligen Arbeitsanforderung zu identifizieren, um somit eine Grundlage für die Systembewertung bereitzustellen.

Die Darstellung der EFA wird durch die Ergebnisse einer Studie untermauert, die im Rahmen des Kurses „*Evaluation interaktiver, multilingualer und multimedialer Informationssysteme*“ an der Universität Konstanz vorgenommen wurde. Die Identifikation der Erfolgsfaktoren erfolgte im Rahmen eines dreiwöchigen Workshop. Teilnehmer² der Arbeitsgruppe waren sieben fortgeschrittene Studenten des Aufbaustudiums der Konstanzer Informationswissenschaft und ein Moderator. Die Teilnehmer verfügten über unterschiedlichen Background; ihre Erstausbildung erstreckte sich von Sozialwissenschaften über Naturwissenschaften bis zu den Geisteswissenschaften, wobei die Mehrzahl bereits über mehrere Jahre Berufspraxis verfügte. Einschränkend ist zu sagen, daß es sich nicht um eine „echte“ geschlossene IRS-Benutzergruppe handelte und daher die Studie als erster Versuch hinsichtlich der Bestimmung der Erfolgsfaktoren aus Benutzersicht zu verstehen ist. Sie stellt eine Orientierungshilfe für eine breit angelegte Studie dar.

Der erste Schritt bestand in einer Brainstorming-Sitzung, in der Qualitätsmerkmale gesammelt wurden. Die Teilnehmer wurden dabei durch die Präsentation verschiedener Beispiele sensibilisiert. Nachdem ein gewisses Vertrauen in die Methode des Brainstorming vorhanden war, hielt die Gruppe insgesamt 66 in Frage kommende Erfolgsfaktoren fest. Anschließend wurden die aufgelisteten Eigenschaften diskutiert und überarbeitet, wobei folgende Kontrollfragen im Mittelpunkt der Diskussion standen:

- *Welche der Eigenschaften sind für die Benutzerzufriedenheit entscheidend?*
- *Welche Eigenschaften fehlen und müssen ergänzt werden?*

Deutlich wurde hier die Notwendigkeit ausführlicher Diskussion, um eindeutige Bedeutungen und Definitionen zu gewährleisten und terminologische Vereinheitlichung zu erzielen. Die Arbeitsgruppe versah jeden Erfolgsfaktor mit einer genauen Beschreibung. Weiteres Ziel des Workshop war, eine Systematisierung und Kate-

¹ Auch die ISO-Normen (z.B. 9126, 9241 Teil 10) bleiben hier sehr vage und gehen nicht auf spezifische Ausprägungen von Software-Systemen ein.

² Hier sind Frauen und Männer gemeint.

gorisierung der Erfolgsfaktoren herbeizuführen. Das Ergebnis dieser Kategorisierung wurde mehrfach überarbeitet.

Das Ergebnis der Studie machte neben den Anforderungen an die Funktionalität die Bedeutung von Benutzerfaktoren, Service und Qualitätsmerkmalen deutlich, wobei in besonderem Maße die Benutzungsschnittstelle und ausführliche Hilfe- und Unterstützungsfunktionen betont wurden. Abbildung 2 zeigt exemplarisch die Ergebnisse auf. Im Detail sind diese im Abschlußbericht (cf. BAYRAKTAR/WOMSER-HACKER 1998) enthalten.

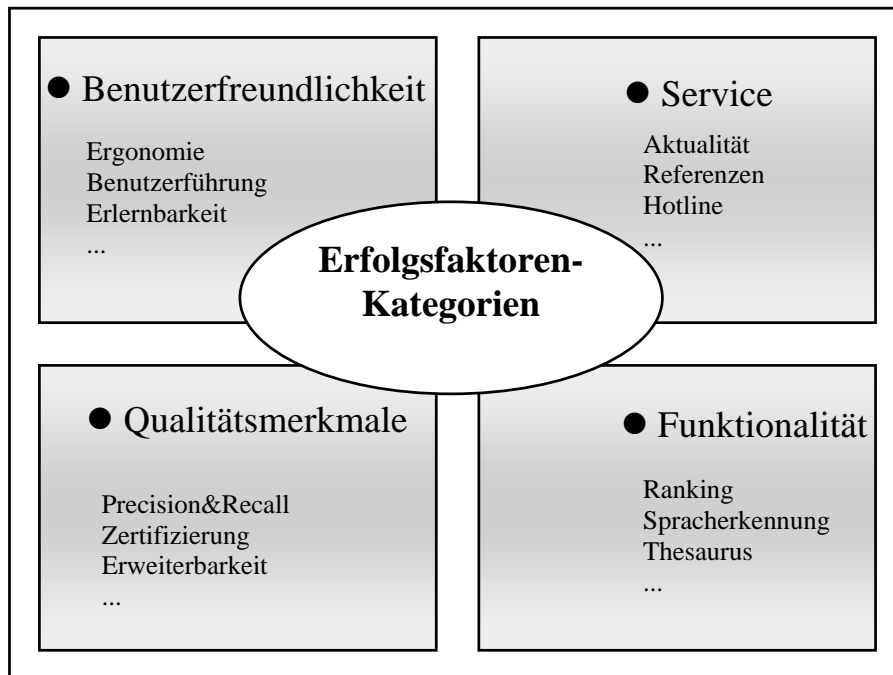


Abb. 2: Kategorisierung der Erfolgsfaktoren

Der nächste Schritt bestand in der Klassifizierung der Qualitätsmerkmale nach ihrer Priorität. Auch hier sollte die Sicht des Benutzers die Grundlage der Bewertung darstellen (Leitfragen: *Was will der Benutzer? Wofür ist er bereit zu warten oder zu bezahlen?*). Die Bewertung der Erfolgsfaktoren erfolgte mit Hilfe von Fragebögen, die einen Prioritätsteil umfaßten.

Die folgende Abbildung zeigt einen Ausschnitt aus den Fragebogen exemplarisch auf:

Bewerten Sie die Erfolgsfaktoren nach ihrer Bedeutung! Welche Bedeutung (Wichtigkeit oder Priorität) haben die Erfolgsfaktoren für das IRS?

Erfolgsfaktor (X)	Irrelevant		evtl. nützlich		wichtig	entscheidend	
	1	2	3	4	5	6	7
Ergonomie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Spracherkennung	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Transparenz der Systemfunktionalität	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Antwortzeit	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 3: Bewertung der Erfolgsfaktoren nach ihrer Bedeutung

Die Ergebnisse wurden anhand eines Portfolios visualisiert, aus dem der Handlungsbedarf deutlich wurde.

Auch generelle Probleme derartiger Analysemethoden konnten in diesem Zusammenhang aufgedeckt werden. Ein besonderes Problem besteht in der Komplexität aktueller IRS, deren Systemfunktionalität sich nur auf einer sehr detaillierten Ebene systematisch darstellen, kommunizieren und vergleichbar machen läßt.

Es stellt sich nun die Frage, inwieweit großangelegte Studien wie TREC einen Beitrag leisten können um zu prüfen, in welchem Maße die erarbeiteten Erfolgsfaktoren erfüllt sind.

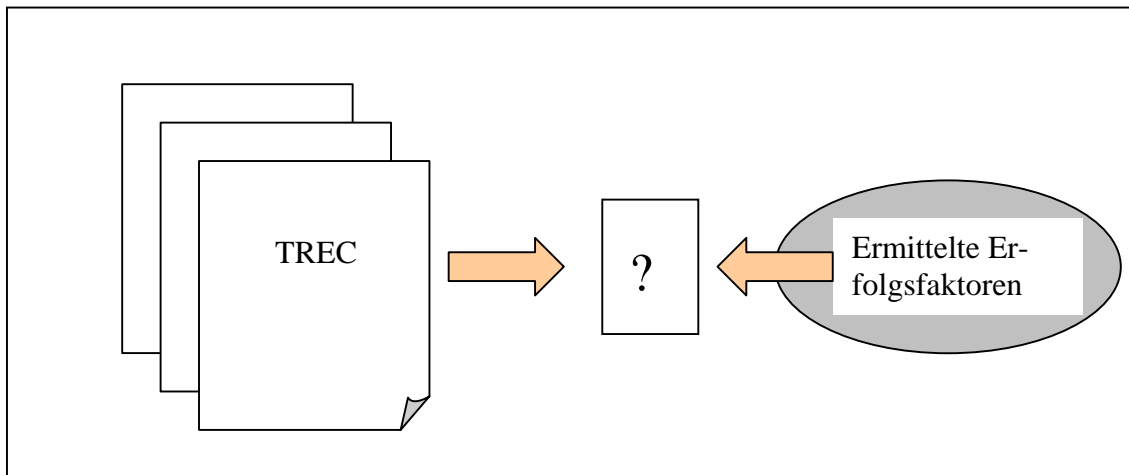


Abb. 4: Kombination von Methoden

Trotz der Fülle des Materials, das durch die sechs bisherigen TREC-Phasen geliefert wurde, fehlen weiterführende Analysen, welche die Ergebnisse auf eine Metaebene stellen, die zweckorientierten Interpretationen zugänglich ist. Noch erfordert dies eine intensive Kenntnis der TREC-Zusammenhänge, um das Material optimal verwerten zu können.

Ein wichtiger Punkt besteht z.B. in der Aufteilung der bei TREC getesteten Systeme, in solche die erworben werden können und somit für einen realen Einsatz überhaupt in Frage kommen können, und reinen Forschungssystemen, die über keine Infrastruktur zur Vermarktung und Pflege verfügen. TREC stellt sich die Aufgabe, auch in diesem Sinne Transferarbeit zwischen Industrie und Forschung zu leisten.

Hinzu kommt die Notwendigkeit, Benutzerfaktoren in die TREC-Studie einzubringen. Da sich der interaktive TREC-Teil im wesentlichen mit den Auswirkungen von Relevance-Feedback-Komponenten befaßt, sind Fragestellungen wie Visualisierung komplexer IR-Funktionalität, Einbringung von softwareergonomischen Aspekten etc. derzeit nicht inhaltlich vertreten. Auch die Erarbeitung der Anfragen im Test wird derzeit – wie oben erwähnt - durch die Entwickler vorgenommen, die über bestes Wissen im Hinblick auf die Funktionalität verfügen. Dies hat mit realen Situationen im späteren Einsatz nur wenig gemein.

Ein weiteres Problem ist in der Lückenhaftigkeit der TREC-Studie zu sehen, d.h. daß keine Vollständigkeit bzgl. der potentiellen IR-Funktionalität bzw. möglicher kombinatorischer Ansätze vorhanden ist. Die Teilnehmer entscheiden in jeder Phase darüber, mit welcher Systemvariante sie teilnehmen. Aus diesem Grund ist es schwierig, generelle Fragestellungen wie z.B. „Lohnt sich die Kombination von NLP-Techniken mit statistischen Verfahren“? zu beantworten.

4 Fazit

Derzeit bestehen zwischen den dargestellten methodischen Ansätzen noch sehr wenig Anknüpfungspunkte, um unmittelbare Verbindungen aufzubauen. Jedoch ist die TREC-Methode kein starres Konzept, sondern flexible genug, um relevante, neue Fragestellungen einbringen zu können. Durch die Methode der Erfolgsfaktorenanalyse lassen sich für verschiedene Bereiche entscheidende Kriterien erarbeiten. Die dargestellte Studie an der Universität Konstanz zeigte, daß gerade Faktoren besondere Bedeutung zukommt, die derzeit bei TREC noch vernachlässigt werden. Längerfristiges Ziel muß es sein, daß sich Studien wie TREC an diesen Faktoren orientieren, um ihre Aussagen den tatsächlichen Bedürfnissen anzupassen. Ein erster Schritt wäre in der Einbeziehung realer Benutzer zu sehen, um auch in diesem Punkt Realitätsnähe zu beweisen. Auch in bezug auf das Spektrum der TREC-Nebenstränge würde sich eine Integration der Bewertung von Benutzungsschnittstellen, Visualisierungskomponenten und multimedialen Interaktions- und Dialogformen anbieten. Dies macht auch im Bereich der Bewertungsmethoden neue Wege erforderlich.

Literatur

- BAYRAKTAR, O., WOMSER-HACKER, CH. (1998), Erfolgsfaktorenanalyse von Information-Retrieval-Systemen. Arbeitsbericht. Lehrstuhl Informationsmanagement, Universität Konstanz.
- BOLLMANN, P. (1977), Untersuchung von Effektivitätsmaßen für Dokumenten-Retrieval-Systeme. Berlin.
- BOLLMANN, P. (1981), Metrische Verfahren des Vergleichs und der Auswahl von Datenbanken. In: DEUTSCHE GESELLSCHAFT FÜR DOKUMENTATION E.V. (ED.), Deutscher Dokumentartag 1980. München et al.
- BOLLMANN, P., CHERNIAVSKY V.S. (1981), Measurement-theoretical Investigation of the MZ-metrics. In: ODDY, R.N., ROBERTSON, S.E., VAN RIJSBERGEN, C.J., WILLIAMS, P.W. (EDS.), Information Retrieval Research, London et al.
- HARMAN, D. (1998), The Text Retrieval Conferences (TREC) and the Cross-Language Track. In: RUBIO, A., GALLARDO, N., CASTRO, R., TEJADA, A. (EDS.), Proceedings of the First International Conference on Language Resources & Evaluation, Granada, pp. 517-522.
- HARMAN, D. (1995), The TREC Conferences. In: KUHLEN, R., RITTBERGER, M. (EDS.), Hypertext – Information Retrieval – Multimedia. Synergieeffekte elektronischer Informationssysteme. Proceedings HIM 95, pp. 5-28.
- HEINRICH, J.-L. (1996), Informationsmanagement. Planung, Überwachung und Steuerung der Informationsinfrastruktur. München.
- HORVÁTH UND PARTNER (EDS.) (1997), Qualitätscontrolling. Ein Leitfaden zur betrieblichen Navigation auf dem Weg zum TQM. Stuttgart.
- KRAUSE, J., WOMSER-HACKER, CH. (HRSG.) (1990), Das Deutsche Patentinformationssystem. Entwicklungstendenzen, Retrievaltests und Bewertungen. Köln et al.
- SALTON, G. (1971), The SMART Retrieval System – Experiments in Automatic Document Processing. Englewood Cliffs, NJ.
- VAN RIJSBERGEN, C.J. (1979), Information Retrieval. London et al.
- VOORHEES, E., HARMAN, D. (1998), Overview of the sixth Text Retrieval Conference (TREC-6). Proceedings of the Sixth Text Retrieval Conference (TREC-6). In press.