

**Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft
Band 14**

Automatische Indexierung unter
Einbeziehung semantischer Relationen:
Ergebnisse des Retrievaltests zum
MILOS II-Projekt

Elisabeth Sachse
Martina Liebig
Winfried Gödert

Dezember 1998

Sachse, Elisabeth; Liebig, Martina; Gödert, Winfried:

Automatische Indexierung unter Einbeziehung semantischer Relationen:

Ergebnisse

des Retrievaltests zum MILOS II-Projekt /

von Elisabeth Sachse, Martina Liebig, Winfried Gödert. -

Köln : Fachhochschule Köln, Fachbereich Bibliotheks- und

Informationswesen, 1998. -

(Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; 14)

ISSN (Print) 1434-1107

ISSN (elektronische Version) 1434-1115

Die Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft berichten über aktuelle Forschungsergebnisse des Fachbereichs Bibliotheks- und Informationswesen der Fachhochschule Köln. Veröffentlicht werden sowohl Arbeiten der Dozent/inn/en, als auch herausragende Arbeiten der Studierenden. Die Kontrolle der wissenschaftlichen Qualität der Veröffentlichungen liegt bei der Schriftleitung. Jeder Band erscheint parallel in Printversion und in elektronischer Version (über unsere Homepage: <http://www.fbi.fh-koeln.de/papers/index/titel/htm>).

Fachhochschule Köln Fachbereich Bibliotheks- und Informationswesen Claudiusstr.1 D-50678

Köln

Tel.: 0221/8275-3376 Fax: 0221/3318583

Schriftleitung: Sabine Schäfer, Prof. Dr. Wolfgang G. Stock

© by FH Köln 1998

| | | |
|----------|---|------|
| 1 | Vorbemerkung | 5 |
| 2 | Retrievaltests: Geschichtlicher Abriss und untersuchte Fragestellungen | 7 |
| 2.1 | ASTIA-Uniterm und Cranfield-Uniterm-Tests (1953) | 7 |
| 2.2 | Cranfield I (1957) | 8 |
| 2.3 | Cranfield II (1966) | 9 |
| 2.4 | Medlars (1966) | 11 |
| 2.5 | Inspec-Test (1969) | 11 |
| 2.6 | Aberystwyth Index-Languages Test (1972) | 12 |
| 2.7 | ONTAP ERIC (1975) | 133 |
| 2.8 | TREC Conferences (1992-) | 133 |
| 2.9 | GIRT (1996-) | 155 |
| 3 | Der MILOS-II-Retrievaltest | 168 |
| 3.1 | Datenbasis | 168 |
| 3.2 | Suchfragen | 20 |
| 3.3 | Suchprozess | 233 |
| 3.4 | Retrievalsoftware | 244 |
| 3.5 | Messzahlen | 277 |
| 3.6 | Relevanzbewertung | 299 |
| 4 | Ergebnisse des MILOS-II-Retrievaltests | 31 |
| 4.1 | Ermittelte Werte | 31 |
| 4.2 | Suche über 'unbehandelte' Titelstichwörter | 333 |
| 4.3 | Suche über IDX-Titelstichwörter | 333 |
| 4.4 | Suche über RSWK-Schlagwörter | 344 |
| 4.5 | Vergleich der Retrievalergebnisse von IDX-Stichwörtern und RSWK-Schlagwörtern | 355 |
| 4.6 | Suche über RSWK-Ketten | 355 |
| 4.7 | Bestätigung der Ergebnisse durch Untersuchung des DDB-Altdatenpools | 366 |
| 5 | Zusammenfassung der Ergebnisse und Empfehlungen | 377 |
| 6 | Anhang | 399 |
| 6.1 | Tabelle 1: Suchfragen mit Ergebnissen | 399 |
| 6.2. | Tabelle 2: Relationencodes | 444 |
| 6.3 | Praktische Hinweise zur Durchführung von Retrievaltests | 455 |
| 6.4 | Verzeichnis der zitierten Literatur | 499 |
| 6.5 | Auswahlbibliographie zu durchgeführten Retrievaltests | 5151 |

Zusammenfassung

Im Rahmen von MILOS II wurde das erste MILOS-Projekt zur automatischen Indexierung von Titeldaten um eine semantische Komponente erweitert, indem Thesaurusrelationen der Schlagwortnormdatei eingebunden wurden. Der abschließend zur Evaluierung des Projekts durchgeführte Retrievaltest und seine Ergebnisse stehen im Mittelpunkt dieses Texts. Zusätzlich wird ein Überblick über bereits durchgeführte Retrievaltests (vorwiegend des anglo-amerikanischen Raums) gegeben und es wird erläutert, welche grundlegenden Fragestellungen bei der praktischen Durchführung eines Retrievaltests zu beachten sind.

1 Vorbemerkung

Zwischen 1993 und 1996 führte die Universitäts- und Landesbibliothek Düsseldorf gemeinsam mit der Fachrichtung Informationswissenschaft der Universität des Saarlandes (Prof. Dr. Harald H. Zimmermann) die zwei von der Deutschen Forschungsgemeinschaft geförderten Erschließungsprojekte MILOS I und MILOS II durch¹.

Ziele von MILOS I waren die Weiterentwicklung von IDX, einem Verfahren zur automatischen Indexierung, das von Harald H. Zimmermann entwickelt wurde, und die Anpassung des Systems an die spezielle Arbeitsumgebung einer wissenschaftlichen Universalbibliothek. Während der Projektlaufzeit wurden in großem Umfang Titeldaten der ULB Düsseldorf automatisch indexiert, aus dem praktischen Einsatz heraus zahlreiche Systemverbesserungen programmiert und auf der Grundlage der indexierten Daten neue Wörterbücher aufgebaut bzw. bereits bestehende stark erweitert.

Im Rahmen des Projekts wurde abschließend ein Retrievaltest durchgeführt mit dem Ziel, die Auswirkung der automatischen Indexierung auf das OPAC-Retrieval zu untersuchen². Auf der Basis einer Teilmenge aus dem Datenbestand der ULB Düsseldorf (40.000 Titel) wurden 50 Suchfragen in drei verschiedenen Registern (Titelstichwortregister, Titelstichwortregister ergänzt um die Ergebnisse der automatischen Indexierung, Titelstichwortregister ergänzt um die verstichworteten Begriffe aus lokalen Schlagwörtern) gesucht. Zusätzlich wurden sämtliche in der Datenbank enthaltenen relevanten Titel ermittelt, so dass sowohl Recall als auch Precision bestimmt werden konnten (vgl. auch 3.5).

¹ Lepsky, K.: RSWK - und was noch?: Stellungnahme zum Bericht 'Sacherschließung in Online-Katalogen' der Expertengruppe Online-Kataloge. In: Bibliotheksdienst. 29 (1995), Nr. 3, S. 500-519. Lepsky, K.: Maschinelles Indexieren zur Verbesserung der sachlichen Suche im OPAC: DFG-Projekt an der Universitäts- und Landesbibliothek Düsseldorf. In: Bibliotheksdienst. 28 (1994), Nr. 8, S. 1234-1242. Lepsky, K., J. Siepman u. A. Zimmermann: Automatische Indexierung für Online-Kataloge: Ergebnisse eines Retrievaltests. In: Zeitschrift für Bibliothekswesen und Bibliographie. 43 (1996), Nr. 1, S. 47-56. Niggemann, E. u. K. Lepsky (Hrsg.): Zukunft der Sacherschließung im OPAC: Vorträge des 2. Düsseldorfer OPAC-Kolloquiums am 21. Juni 1995. Düsseldorf: Universitäts- und Landesbibliothek 1996. (Schriften der Universitäts- und Landesbibliothek Düsseldorf; Bd. 25). Lepsky, K.: Automatische Indexierung und bibliothekarische Inhaltserschließung: Ergebnisse des DFG-Projekts MILOS I. In: Zukunft der Sacherschließung im OPAC: Vorträge des 2. Düsseldorfer OPAC-Kolloquiums am 21. Juni 1995. Hrsg.: E. Niggemann u. K. Lepsky. Düsseldorf 1996. (Schriften der Universitäts- und Landesbibliothek Düsseldorf; Bd. 25). Softex GmbH Saarbrücken (Hrsg.): Universitäts- und Landesbibliothek Düsseldorf (Hrsg.): MILOS: Automatische Indexierung für Bibliotheken: Handbuch. Stand: Juni 1996. Düsseldorf: Universitäts- und Landesbibliothek 1996. Gödert, W., M. Liebig: Maschinelle Indexierung auf dem Prüfstand: Ergebnisse eines Retrievaltests zum MILOS-II-Projekt. In: Bibliotheksdienst. 31 (1997), Nr. 1, S. 59-68. Sacherschließung. In: Jahresbericht 1996. Hrsg.: Die Deutsche Bibliothek. Frankfurt: Die Deutsche Bibliothek 1997. Lepsky, K.: Auf dem Weg zur automatischen Inhaltserschließung?: Das DFG-Projekt MILOS und seine Ergebnisse. In: Mitteilungen der Gesellschaft für Bibliothekswesen und Dokumentation des Landbaus. (1997), Nr. 53, S. 46-52. Lepsky, K.: Sacherschließung ohne RSWK?: Neue Praxis an der Universitäts- und Landesbibliothek Düsseldorf. In: ProLibris. 3 (1998), Nr. 2, S. 112-114.

² Lepsky, K., J. Siepman u. A. Zimmermann: Automatische Indexierung für Online-Kataloge: Ergebnisse eines Retrievaltests. In: Zeitschrift für Bibliothekswesen und Bibliographie. 43 (1996), Nr. 1, S. 47-56.

Der Retrievaltest machte deutlich, dass die alleinige Suche über unbehandelte Titelstichwörter für die thematische Suche im OPAC ungeeignet ist. Gleichzeitig zeigte der Test, dass sich durch das Einspielen von zusätzlichen Deskriptoren, die auf der Basis von Titelstichwörtern und verstichworteten Schlagwörtern durch automatische Verfahren gewonnenen wurden, das Suchergebnis signifikant verbessern lässt.

Aufbauend auf den Ergebnissen von MILOS I wurde innerhalb von MILOS II an einer Ausweitung der Funktionalität von IDX gearbeitet. Konventionelle Methoden der inhaltlichen Erschließung - verbale Sacherschließung nach den Regeln für den Schlagwortkatalog (RSWK), klassifikatorische Erschließung - wurden mit den Möglichkeiten der automatischen Indexierung zusammengeführt. In Zusammenarbeit mit der Deutschen Bibliothek wurde IDX durch die Einbindung von Thesaurusrelationen der Schlagwortnormdatei (SWD) in das Wörterbuchkonzept um semantische Funktionalitäten erweitert.

Wie schon bei MILOS I wurde auch für MILOS II ein abschließender Retrievaltest durchgeführt, um die Wirksamkeit des eingesetzten Verfahrens zu evaluieren³. Der vorliegende Bericht ist daher zunächst Ergebnisbericht zu diesem Retrievaltest MILOS II. Darüber hinaus haben aber die Arbeiten im Rahmen der Vorbereitung und Durchführung des Tests nur allzu deutlich gemacht, dass das allgemeine Thema Retrievaltest komplex und die konkrete praktische Realisierung mit einer Reihe von Problemen und offenen Fragen konfrontiert ist. Deutsche Bibliotheken haben sich bislang nicht um dieses Gebiet gekümmert, obwohl die Effektivität ihrer Informationsinstrumente täglich auf dem Prüfstand steht. Wir haben daher versucht, über das eigentliche Thema des Berichts hinausgehend, den Gesamtkomplex Retrievaltest etwas eingehender darzustellen, nicht zuletzt auch verbunden mit der Hoffnung, so Hilfestellung und Anstoß für weitere Tests zu liefern. Diesem Zweck dienen insbesondere die im Anhang zusammengestellten Hinweise zur Durchführung von Retrievaltests sowie die Literaturzusammenstellung.

Düsseldorf und Köln, November 1998

Prof. Winfried Gödert
Fachhochschule Köln
Fachbereich Bibliotheks-
und Informationswesen

Martina Liebig
Medical Relations GmbH
Langenfeld

Elisabeth Sachse
Universitäts- und Landes-
bibliothek Düsseldorf

³ Gödert, W., M. Liebig: Maschinelle Indexierung auf dem Prüfstand: Ergebnisse eines Retrievaltests zum MILOS-II-Projekt. In: Bibliotheksdienst. 31 (1997), Nr. 1, S. 59-68.

2 Retrievaltests: Geschichtlicher Abriss und untersuchte Fragestellungen

In den folgenden Abschnitten wird ein Überblick über die grundlegenden Konzeptionen bereits durchgeführter Retrievaltests gegeben. Es handelt sich dabei vor allem um Tests aus dem angloamerikanischen Raum; ein nennenswerter Test im deutschen Sprachraum wurde erst in allerneuester Zeit mit dem GIRT-Projekt verwirklicht. Schwerpunkt der untersuchten Datenbestände sind weniger OPACs als vielmehr Datenbanken.

2.1 ASTIA-Uniterm und Cranfield-Uniterm-Tests (1953)

Die ersten Retrievaltests fanden 1953 in den USA (ASTIA⁴) und in Großbritannien (Cranfield⁵) statt. Durch sie sollte das zu der Zeit kontrovers diskutierte Uniterm-System der Firma Documentation Incorporated evaluiert werden. Dieses System, in dem Dokumente durch einzelne, dem Titel oder Abstract entnommene Begriffe repräsentiert wurden, sollte verglichen werden mit konventionellen Ansätzen der Indexierung und des Retrievals.

Zwei Gruppen, die eine bestehend aus Mitarbeitern des ASTIA (Armed Services Technical Information Agency), die andere aus Mitarbeitern von Documentation Incorporated, indexierten unabhängig voneinander den circa 15.000 Dokumente umfassenden Datenbestand von ASTIA und setzten dann 93 Suchfragen darüber ab. Die ASTIA-Mitarbeiter verwendeten die auch in der Praxis eingesetzte alphabetische *Subject Heading List* für die Indexierung, während Documentation Incorporated sog. *Uniterms* einsetzte. Kriterium für die Effektivität des Systems war die Relevanz der gefundenen Dokumente, womit Relevanz zum ersten Mal als Kriterium für die Beurteilung eines Retrieval-Systems genutzt wurde.

Gerade über die Relevanzbeurteilung kam es jedoch zu Differenzen zwischen den beiden Testgruppen, da sie sich nicht darüber einigen konnten, welche Dokumente für welche Frage relevant waren, und deshalb zu konträren Aussagen hinsichtlich der Testergebnisse kamen. Die Tester konnten sich weder auf ein einheitliches Relevanzkriterium noch auf einen gemeinsamen Entscheidungsfindungsprozess einigen. Beide Gruppen, die natürlich nicht unvoreingenommen waren, beanspruchten für sich, die besseren Testergebnisse erzielt zu haben.

Im selben Jahr wurde in Großbritannien von Cleverdon, Mills und Keen der Cranfield-Uniterm-Test am Cranfield-College durchgeführt, ebenfalls mit dem Ziel, ein traditionelles Indexierungssystem, eine modifizierte Version der Universal Decimal Classification (UDC), mit dem Uniterm-System zu vergleichen. Im Gegensatz zum ASTIA-Test war der Testaufbau des Cranfield-Tests sehr viel stärker kontrolliert. Er beruhte auf 200 Dokumenten zum Thema Aeronautik, denen eine bestimmte Anzahl an sog. Source Documents entnommen wurden, aus denen 40 Fragen entnommen wurden. Die 40 Fragen wurden daraufhin über die 200 Dokumente abgefragt. Als Kriterium für die Effektivität galt die Fähigkeit des Systems, die 40 Source Documents zu finden. Die Ergebnisse

⁴ Vgl.: Gull, C. D.: Seven years of work on the organization of materials in the special library. In: American Documentation. 7 (1956), S. 320-329.

⁵ Vgl.: Thorne, R. G.: The efficiency of subject catalogues and the cost of information searches. In: Journal of Documentation. 11 (1955), S. 130-148.

waren sehr positiv für das Uniterm-System, das 85% der Source Documents fand gegenüber nur 50%, die durch die UDC-Klassifikation gefunden wurden. Das Testverfahren ist in zwei Aspekten kritisch zu betrachten. Erstens wurde die Effektivität basierend auf nur einem Meßwert ermittelt; Werte hinsichtlich zusätzlich gefundener Dokumente, ob relevant oder nicht-relevant, wurden nicht einbezogen. Zweitens wird bei dem hier eingesetzten Verfahren, das die Suchfragen in Anlehnung an in der Datensammlung enthaltene Dokumente formuliert, ein begriff-basiertes System grundsätzlich bessere Werte als ein konzept-basiertes System erzielen⁶.

2.2 Cranfield I (1957)

Finanziert von der National Science Foundation (USA) wurde 1957 von der Association of Special Libraries and Information Bureaux (Aslib) der Cranfield I-Test⁷ in Großbritannien durchgeführt. Ziel war es, vier verschiedene Indexierungssysteme zu testen. Dabei handelt es sich um die UDC, einen alphabetischen Subject-Index, eine Facettenklassifikation sowie das Uniterm-System. Der Testaufbau entsprach weitgehend dem vorangegangenen Cranfield-Uniterm-Test, wobei aber eine sehr viel größere Datenmenge eingesetzt wurde. 18.000 Dokumente zum Thema Luftfahrt-Ingenieurwesen wurden mit den vier Systemen indexiert. Darüber wurden 1.200 Fragen gestellt, die wie auch bei dem vorherigen Test nach der 'Source Document'-Methode ermittelt wurden. Nur diejenigen Fragen, die die Source-Dokumente fanden, galten als erfolgreich. Bei den nicht erfolgreichen Fragen wurde anschließend untersucht, ob die Gründe für das Nicht-Finden in der Suchfrage, dem Indexierungsverfahren, der Suche oder dem System lagen.

Insgesamt fiel das Ergebnis insbesondere für die Stichwortsuche positiv aus, wobei aber alle vier Systeme ähnlich gut abschnitten. Das Uniterm-Verfahren fand 82% der Source-Dokumente, an zweiter Stelle folgte die Erschließung mittels Schlagwörtern der alphabetischen Schlagwortliste (81,5%), an dritter Stelle die Klassifizierung mittels UDC (75,6%) und an letzter Stelle die Erschließung mit der Facettenklassifikation (73,8%). Das Ergebnis belegte wiederum die Überlegenheit der maschinell erzeugten Stichwörter gegenüber den traditionellen intellektuellen Verfahren.

Kritik war hinsichtlich derselben Punkte wie beim Cranfield-Uniterm-Test zu erwarten. Um dem vorzubeugen, wurden zwei zusätzliche Tests durchgeführt. Für den ersten Test wurden 100 der Suchfragen an Informationseinrichtungen und Bibliotheken, die in ähnlichen Fachgebieten tätig sind, geschickt mit der Bitte, Literaturlisten zu den Fragen zu erstellen. Die Titel der zurückgesandten 88 Listen wurden an der ursprünglichen Testdatenmenge abgeglichen. Die so ermittelten 359 Titel wurden anschließend auf ihre Relevanz hinsichtlich der Source-Dokumente und der Suchfragen untersucht. Nach der Einordnung in die drei Relevanzstufen 'genauso relevant wie das Source-Dokument', 'etwas relevant' und 'nicht relevant' wurden 41 Fragen ermittelt, die 120 entweder 'genauso

⁶ Ellis, D.: Progress and problems in information retrieval. London: Library Association Publishing 1996, S. 1-3.

⁷ Vgl.: Cleverdon, C. W.: Report on the first stage of an investigation into the comparative efficiency of indexing systems. Cranfield: College of Aeronautics 1960. Cleverdon, C. W.: Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems. Cranfield: College of Aeronautics 1962. Lancaster, F. W., J. Mills: Testing indexes and index language devices: the ASLIB Cranfield project In: American Documentation. 15 (1964), S. 4-13.

relevante' oder 'etwas relevante' Dokumente fanden. Diesen Fragen wurde nochmals in der Gesamtdokumentmenge gesucht, mit folgenden Ergebnissen: Uniterm 75%, alphabetische Schlagwort-Liste 75%, UDC 74% und Facettenklassifikation 60%.

Bei einem zweiten Zusatztest wurden 759 Dokumente als Ergebnismenge zu 79 zufällig ausgewählten Suchfragen wie im ersten Zusatztest auf ihre Relevanz überprüft. Auf diese Weise sollte herausgefunden werden, inwieweit die verschiedenen Systeme fähig sind, nicht-relevante Dokumente zurückzuhalten. Hierbei erzielte die alphabetische Schlagwort-Liste das beste Ergebnis (12,5%), gefolgt vom Uniterm-System (12%), der Facettenklassifikation (7,5%) und der UDC (7%).

Anhand der Testergebnisse schlußfolgerte Cyril Cleverdon, der Projektleiter, dass eine inverse Beziehung besteht zwischen der Fähigkeit, relevante Dokumente zu finden und der Fähigkeit, nicht relevante Dokumente zurückzuhalten. Starke Kritik wurde wiederum am Verfahren der Suchfragenformulierung geübt, das angeblich dem Stichwortverfahren entgegenkam. Bemängelt wurde ferner, dass in einer Praxis-Situation keine Source-Dokumente vorhanden seien, und dass die Beziehung zwischen Source-Dokument und Suchfrage zu eng sei. Swanson schlug daher vor, die Source-Dokumente für den Test auszuschließen⁸.

2.3 Cranfield II (1966)

1966 wurde für Cranfield II davon ausgegangen, dass Indexierungssprachen aus der Kombination mehrerer Indexierungsarten bestehen. Es wurden daher keine bereits existierenden Indexierungssysteme, sondern verschiedene Indexierungsarten getestet. Ziel des Tests war zum einen, die Auswirkung der verschiedenen Variationen von Indexierungsarten auf die Systemperformance zu testen und zum anderen, herauszufinden, welchen Einfluss diejenigen Indexierungsarten, die die Precision erhöhen und diejenigen, die den Recall erhöhen sollen, auf die Performance haben. Hierzu wurden erstmals die Werte Recall und Precision eingesetzt.

Für diesen Test wurden 33 verschiedene Indexierungssprachen geschaffen, die aus variierenden Terminologien und Strukturen bestanden. Sie unterschieden sich hinsichtlich des Ausmaßes, in dem sie Einzelwörter oder Komposita, Hierarchien und kontrolliertes Vokabular verwendeten. Die Datenmenge bestand aus 1.400 Dokumenten zum Thema Luftfahrt, an die 211 Suchfragen gestellt wurden. Als Suchfragen wurden die Fragen gewählt, die die Verfasser der in der Datenbank enthaltenen Titel als Ausgangspunkt für das Verfassen ihrer Texte verwendet hatten. Um den Prozess der Bewertung zu beschleunigen, wurde die Dokumentmenge in mehrere kleine Segmente aufgeteilt.

Im Gegensatz zu Cranfield I wurde die Effektivität nur anhand der Relevanz gemessen. Die Effektivität der verschiedenen Testsysteme wurde danach beurteilt, inwieweit vorher als relevant bewertete Dokumente gefunden wurden. Die Relevanz der Dokumente wurde wiederum anhand eines zweistufigen Verfahrens ermittelt. Zunächst recherchierten Studierende des Luftfahrtwesens die gesamte Dokumentensammlung. Die dabei als relevant eingestuftten Dokumente sowie die darin enthaltenen Literaturhinweise wurden dann dem Verfasser vorgelegt, der ein endgültiges Relevanzurteil vornahm. Zusätzlich wurden Literaturhinweise herausgesucht, in denen gemeinsame Zitate gefunden wurden (sog.

⁸ Ellis: Progress and problems in information retrieval, S. 3-6.

bibliographic coupling), und ebenfalls den Verfassern zur Beurteilung gegeben. Einer der Juroren beschreibt das Problem der Relevanzbeurteilung so:

„... the assessment of relevance categories seemed particularly difficult to interpret in relation to most of these additional documents. I believe that I have 'scored' the documents roughly in proportion to the degree of irritation I should feel if a librarian produced them in response to my original query. Whether this is a proper basis for measurement of relevance may be arguable.“⁹

Cranfield II erbrachte zusammenfassend folgende Ergebnisse:

- Einzelwort-Indexierungssprachen erzielten die beste Performance;
- werden zusätzlich zur Einzelwort-Indexierung Begriffsgruppen oder -klassen über Synonyme oder Wortformen hinaus gebildet, so verringerte sich die Performance;
- zufrieden stellende Ergebnisse wurden erzielt mit natürlicher Sprache kombiniert mit einfacher Koordinierung;
- sämtliche Ergebnisse bestärkten die Hypothese, dass eine inverse Relation zwischen Recall und Precision besteht.

Insbesondere die Feststellung, dass natürlichsprachige Suchen mit einfacher Wortkombination sehr gute Retrievalergebnisse erzielen, überraschte die Fachleute. Aus der Bestätigung einer inversen Relation zwischen Recall und Precision ließ sich ableiten, dass weitere Versuche, beides innerhalb eines Systems zu vereinigen, erfolglos sein mussten.

Keen kommt in seiner Analyse zu weiteren Erkenntnissen:

- increase in depth or exhaustivity of indexing increases recall and decreases precision;
- term specificity governs precision and recall;
- vocabulary cross references can help increase recall and decrease precision;
- levels of exhaustivity, specificity and linkage at the input stage can be balanced or their effect altered by what happens at the search stage;
- input levels which give the best desired trade-off between recall and precision require optima to be found, although determining the optimum level is not easy¹⁰.

Kritik richtete sich erneut vor allem auf das Verfahren der Relevanzbeurteilung. Sowohl Studierende als auch die abschließenden Relevanzbeurteiler hatten Probleme dabei. Beispielsweise wurden durch die Studierenden nicht alle Dokumente als relevant gewertet, die im Nachhinein durch das Bibliographic Coupling ermittelt wurden. Somit konnte eventuell ein nicht unerheblicher Anteil an relevanten Dokumenten unberücksichtigt geblieben sein. Grundsätzlich muss daher gefragt werden, inwieweit die Ergebnisse dieses Labortests auf die Praxis übertragen werden können¹¹.

⁹ Harter, S. P.: Variations in relevance assessments and the measurement of retrieval effectiveness. In: Journal of the American Society for Information Science. 47 (1996), Nr. 1, S. 40.

¹⁰ Keen, E. M.: Aspects of computer-based indexing languages. In: Computers in libraries international '91. Proceedings of the 5th annual conference on computers in libraries, held in London in February 1991. London 1991, S. 150.

¹¹ Ellis: Progress and problems in information retrieval, S. 6-11.

2.4 Medlars (1966)

Der Medlars-Retrievaltest¹² der National Library of Medicine in Washington ist ein Beispiel für den Test einer großen Datensammlung in Praxis-Umgebung. Für diesen Test wurden 302 'echte' Nutzeranfragen über die MEDLARS-Datenbank abgefragt, die zu diesem Zeitpunkte aus mehr als 700.000 Dokumenten bestand. Da es bei dieser Datenmenge unmöglich war, die Relevanz jedes der Dokumente festzustellen, wurde sie anhand der ausgegebenen Dokumente und ausgewählten Dokumenten aus anderen Quellen geschätzt.

Der Recall der Suchergebnisse lag zwischen 60% und 82%, die Precision zwischen 14% und 35%¹³. Im Durchschnitt wurde eine relativ hohe Precision von 50% bei relativ niedrigem Recall von 58% erzielt. Um die Gründe für diese Verteilung herauszufinden, wurden 4.000 'Misserfolge' 'analysiert, also nicht-relevante Dokumente, die gefunden wurden sowie relevante Dokumente, die nicht gefunden wurden. Dabei wurde deutlich, dass die größte Fehlerquelle in der Umsetzung der Nutzerinformationsbedürfnisse in eine adäquate Suchstrategie lag¹⁴.

2.5 Inspec-Test (1969)

Im 1969 durchgeführten Inspec-Test¹⁵ (Comparative evaluation of index languages) wurden Suchen mit fünf verschiedenen Indexierungsverfahren untersucht:

- Titel,
- Titel und Abstracts,
- Begriffe aus dem gedruckten Schlagwortindex der Science Abstracts,
- natürlichsprachige, intellektuelle Indexierung und,
- kontrolliertes Vokabular, das einem im Entwurf befindlichen Thesaurus entnommen wurde¹⁶.

Die Testmenge bestand aus 542 Dokumenten aus den Sachgebieten Physik und Elektrotechnik, die entweder nur Titel oder Titel und Abstracts umfassten, über die 97 Suchfragen abgefragt wurden. Bis zu drei unterschiedliche Suchformulierungen wurden pro Frage und System gewählt.

Der Test ergab insgesamt nur leichte Unterschiede zwischen den einzelnen Suchverfahren, wobei die Suche mit kontrolliertem Vokabular das beste Ergebnis erzielte¹⁷. Gleichzeitig bestätigte er die Ergebnisse des Cranfield II-Tests hinsichtlich einer möglichst erschöpfenden Indexierung sowie des überraschend guten Abschneidens der natürlichsprachigen Indexierung¹⁸.

12 Vgl.: Lancaster, F. W.: Evaluation of the MEDLARS demand search service. Washington: U.S. Government Print Office 1968.

13 Ellis: Progress and problems in information retrieval, S. 10-11.

14 Cleverdon, C. W.: Progress in documentation: evaluation tests of information retrieval systems. In: Journal of Documentation. 26 (1970), Nr. 1, S. 60.

15 Vgl.: Aitchison, T. M.: Comparative evaluation of index languages. Part I, Design. Part II, Results. London: Inspec 1969, 1970.

16 Rowley, J. E.: The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research. In: Journal of Information Science. 20 (1994), Nr. 2, S. 111.

17 Rowley, J. E., C. M. Turner: The dissemination of information. Worcester: André Deutsch Ltd. 1978, S. 193.

18 Cleverdon: Progress in documentation, S. 60.

2.6 Aberystwyth Index-Languages Test (1972)

Ziel des Aberystwyth Index-Languages Test¹⁹ war es, unter Laborbedingungen die Effektivität und Effizienz verschiedener zur Sacherschließung eingesetzter Indexierungssprachen zu vergleichen. Das für den Test genutzte Fachgebiet war Bibliotheks- und Informationswissenschaft, bei den Indexierungssprachen handelt es sich um drei postkoordinierende (Compressed term index language, uncontrolled index language und hierarchically structured index language) und zwei präkoordinierende Indexierungssprachen (hierarchically structured index language, relational indexing index language). Diese fünf Sprachen wurden im Hinblick auf acht verschiedene Kriterien getestet: post-co-ordinate index languages of measured specificity and linkage, indexing specificity, indexing exhaustivity, method of co-ordination, post-co-ordinate index languages including variations in indexing specificity and exhaustivity, precision device of partitioning, precision device of relational operators, provision of context in the search file.

Das Testverfahren war gut durchdacht, um Manipulationen vorzubeugen. Zudem wurde versucht, die gesamte Vorgehensweise möglichst praxisnah zu gestalten, sie also nicht für diesen Test in unrealistischer Weise zu perfektionieren. Die verschiedenen Durchführungsschritte des Tests (Indexierung, Frageformulierung, Durchführung der Suche, Relevanzbeurteilung) wurden jeweils von unterschiedlichen Personen vorgenommen. Die Datensammlung bestand aus 800 Dokumenten (Zeitschriftenartikel und Reports), zu denen 63 Suchfragen gestellt wurden. Die Relevanz, unterteilt in die drei Stufen 'hoch relevant', 'teilweise relevant' und 'nicht relevant', beurteilten Mitarbeiter des College of Librarianship Wales. Als Werte zur Relevanzmessung wurden Recall und 'nicht-relevante gefundene Dokumente' gewählt. Die Precision wurde bewusst nicht ermittelt, weil sie zum einen eine bestimmte Auffassung eines Nutzers von Relevanz oder Nicht-Relevanz voraussetzt und zum anderen die absoluten Werte von relevanten und nicht-relevanten Dokumenten außer Acht lässt und nur Aussagen über das Verhältnis von relevanten und nicht-relevanten Dokumenten zulässt²⁰.

Insgesamt waren keine großen Unterschiede in der Effizienz und Effektivität der verschiedenen Indexierungssprachen festzustellen. Es zeigte sich jedoch, dass das kontrollierte Vokabular in verschiedener Hinsicht nicht ausreichend war; eine möglichst spezifische Indexierung war hierdurch nur bedingt zu gewährleisten. Das unkontrollierte Vokabular schnitt dagegen vergleichsweise gut ab, insbesondere wenn der hierfür nötige geringe intellektuelle Aufwand berücksichtigt wird. Eine Verbesserung der Indexierungsbreite (exhaustivity) ergab eine leichte Verbesserung des Recalls ohne einen Verlust an Precision. Im Vergleich zum postkoordinierten Verfahren erzielte das präkoordinierte Verfahren schlechtere Ergebnisse bei geringerem Recall²¹.

19 Vgl.: Keen, E. M., J. A. Digger: Report of an information science index language test. Part 1: Text. Part 2: Tables. Aberystwyth: Department of Information Retrieval Studies, College of Librarianship 1972; Keen, E. M.: The Aberystwyth index languages test. In: Journal of Documentation. 29 (1973), Nr. 1, S. 1-35.

20 Keen: The Aberystwyth index languages test, S. 9.

21 Rowley: The dissemination of information, S. 194.

2.7 ONTAP ERIC (1975)

Der ONTAP ERIC-Test²² von 1975 basiert auf dem Bestand der ERIC-Datenbank aus dem selben Jahr. Die vom Host DIALOG angebotene Datenbank dient als Trainingsdatenbank für Online-Researcher. Sie enthält 35.394 Dokumente und 29 Trainingsfragen, die in die Schwierigkeitsgrade 'einfach', 'mittelschwer' und 'schwer' eingeteilt sind. Gleichzeitig werden die für die Trainingsfragen relevanten Dokumente zur Verfügung gestellt. Auf diese Weise können Recall und Precision für individuelle Suchfragen und -prozeduren bestimmt werden.

ONTAP ERIC wurde unter anderem aufgrund nicht berücksichtigter relevanter Dokumente kritisiert. So wurde die Trainingsdatenbank beispielsweise bei zwei Doktorarbeiten eingesetzt, um das Verhalten von Online-Nutzern zu untersuchen. Hierbei kam es zu erheblichen Unterschieden in der Relevanzbeurteilung sowohl zwischen den beiden Doktoranden als auch zwischen den beiden und dem früheren Testergebnis.

2.8 TREC Conferences (1992-)

Seit 1992 wird das bisher größte Experiment zum Vergleich der Effektivität verschiedener Retrievalmethoden, TREC²³ (Text Retrieval Conference), durchgeführt. Es wird gefördert vom National Institute of Standards and Technology (NIST) sowie vom Information Technology Office of the Defense Advanced Research Projects Agency (DARPA) und findet jährlich in Form eines internationalen Wettbewerbs statt, in dem kontrollierte Testbedingungen (eine große Datensammlung, einheitliche Beurteilungsmethoden) zur Verfügung gestellt werden, anhand derer die teilnehmenden Institutionen (kommerzielle Unternehmen, Universitäten etc.) ihre unterschiedlichen, oftmals experimentellen Information Retrieval-Verfahren vergleichen können. Die eingesetzten Verfahren reichen von automatischen Thesauri, Gewichtungungsverfahren, natürlichsprachigen Techniken, Relevance Feedback bis hin zu Pattern Matching-Verfahren. Die Resultate der Teilnehmer werden im Rahmen einer internationalen Konferenz präsentiert.

An dem ersten TREC-Wettbewerb (TREC-1) 1992 nahmen 25 Gruppen teil, die ihre Verfahren an einer fünf Megabyte großen Textmenge testen konnten. An der letzten Konferenz TREC-6, die 1997 stattfand, nahmen schon 51 Gruppen aus 12 Ländern teil und testeten mit einer zwei Gigabyte großen Textmenge. Zwei Hauptaufgaben des Information Retrieval stehen zur Wahl: einerseits die sog. *ad hoc*-Aufgabe, bei der neue Fragen an eine statische Datenmenge gestellt werden, und andererseits die sog. *routing*-Aufgabe, in der dieselben Fragen an eine Datenmenge gestellt werden, die ständig aktualisiert wird. Bei beiden Varianten haben die Teilnehmer die Möglichkeit, entweder ihre Suchfragen oder die Datenmenge entsprechend aufzubereiten. Die 1000 am höchsten gewerteten Dokumente für jedes Thema werden dann vom NIST hinsichtlich Recall und Precision evaluiert und zu einem Vergleich der verschiedenen Systeme herangezogen.

Die Dokumente für den Test stammen aus elf verschiedenen Datenquellen, v.a. Artikeln und Abstracts von Zeitschriften und Nachrichtenagenturen. Die 350

22 Harter, S. P.: Variations in relevance assessments and the measurement of retrieval effectiveness. In: Journal of the American Society for Information Science. 47 (1996), Nr. 1, S. 42.

23 Weitere Informationen finden sich unter der TREC-Homepage: <http://trec.nist.gov>.

Suchthemen sollen 'wirklichen' Nutzer-Informationsbedürfnissen entsprechen und sind von unterschiedlicher Länge. Die Relevanz wird anhand des Pooling-Verfahrens ermittelt. Hierfür werden die 100 Dokumente, die von jedem der teilnehmenden Verfahren bereitgestellt wurden, zu einer Ergebnismenge vereinigt, auf der die späteren Relevanzbeurteilungen beruhen. Die vereinigten Trefferlisten werden intellektuell auf ihre Relevanz überprüft, wobei jedes Thema von einer einzigen Person bearbeitet wird, um eine konsistente Beurteilung zu gewährleisten. Bei TREC-6 wurden pro Thema durchschnittlich 1445 Dokumente beurteilt, von denen 6% als relevant eingeschätzt wurden.

Neben den zwei Hauptaufgaben gibt es acht Zusatzaufgaben (TREC Tracks):

| | |
|----------------|--|
| Chinese | <i>ad hoc</i> -Aufgabe mit Themen und Dokumenten auf Chinesisch |
| Cross-Language | <i>ad hoc</i> -Aufgabe mit Dokumenten und Fragen auf Deutsch, Englisch und Französisch, in der alle zu einem Thema gehörigen Dokumente gefunden werden sollen ungeachtet der Sprache |
| Filtering | <i>routing</i> -ähnliche Aufgabe, bei der Systeme eine binäre, und keine auf Ranking beruhende Relevanzentscheidung treffen sollen |

| | |
|---------------------------|--|
| High Precision User Track | <i>ad hoc</i> -Aufgabe bei der innerhalb von fünf Minuten Dokumenten zu einem Thema gefunden werden sollen unter Einsatz beliebiger Mittel |
| Interactive | mit dieser Aufgabe soll die Interaktion zwischen Nutzer und Text Retrieval-System untersucht werden |
| NLP | <i>ad hoc</i> -Aufgabe, die die Einsatzmöglichkeiten von Natural Language Processing-Verfahren untersucht |
| Spoken Document Retrieval | Retrieval in Berichten, die aus 50 Stunden gesprochenen Nachrichtentexten bestehen |
| Very Large Corpus (VLC) | <i>ad hoc</i> -Aufgabe, um die Fähigkeit von Retrieval-Systemen, mit sehr großen Datenmengen (ca. 20 GB) umzugehen, zu testen. |

Ein wichtiger Aspekt von TREC besteht in der leicht zugänglichen Datensammlung. Forschungsgruppen können hierdurch ihre experimentellen Systeme in einer nahezu mit der Praxis gleichzusetzenden Testumgebung testen und können diese gleichzeitig mit traditionelleren Verfahren vergleichen. Kommerzielle Unternehmen können die Daten zu internen Performance-Tests ihrer Suchmaschinen einsetzen.

Die TREC-Ergebnisse zeigen eine ständige Entwicklung hin zu komplexeren Retrievaltechniken mit immer höherer Effektivität. Außerdem fördert TREC den Austausch von neuen Techniken sowie die Forschung auf neuen Gebieten des Information Retrieval²⁴.

Auch bei TREC steht die Frage der Relevanzbeurteilung in der Kritik. Harter z.B. bezweifelt, dass sämtliche relevanten Dokumente auffindig gemacht werden konnten und dass die Relevanzbestimmung auch unter anderen Umständen genauso ausgefallen wäre:

„However, from what we know about the great many factors affecting relevance judgements [...], we can be sure that under many other conceivable conditions the set of relevant documents in the TREC experimental collection would be very different from those that served as the foundation for the actual TREC experiments. What this means in terms of the validity of the conclusions reached by TREC experimenters or of retrieval evaluation more generally is not so clear.“²⁵

2.9 GIRT (1996-)

Eine Entsprechung der TREC-Idee für den deutschsprachigen Raum ist das GIRT-Projekt (German Indexing and Retrieval Testdatabase) des InformationsZentrum Sozialwissenschaften (IZ)²⁶.

24 Harman, D.: The Text Retrieval Conferences (TREC): providing a test-bed for information retrieval systems. In: Bulletin of the American Society for Information Science. (1998) April/May, S. 11-13.

25 Ebd., S. 42.

26 Vgl.: Frisch, E., M. Kluck: Pretest zum Projekt German Indexing and Retrieval Testdatabase (GIRT) unter Anwendung der Retrievalsysteme Messenger und freeWAIS-sf. IZ-Arbeitsbericht. (1997), Nr. 10. Knorz, G.: Testverfahren für intelligente Indexierungs- und Retrievalsysteme anhand deutschsprachiger sozialwissenschaftlicher Fachinformation (GIRT): Bericht über einen Workshop am 12. September 1997 im IZ Sozialwissenschaften, Bonn. In: Nachrichten für Dokumentation. 49 (1998), Nr. 2, S. 111-116.

In Kooperation mit dem Hochschulverband Informationswissenschaft (HI) und der Gesellschaft für Informatik (GI) wird Forschern und Softwareherstellern die Möglichkeit gegeben, anhand einer einerseits begrenzten, andererseits nicht zu kleinen Menge von Dokumenten die Leistungsfähigkeit der unterschiedlichen Komponenten moderner Indexierungs- und Retrievalverfahren zu testen. [...] Insbesondere sollen die spezifischen Probleme der deutschen Sprache bzw. die Unterschiede zu Verfahren, die für das Englische entwickelt wurden, herausgearbeitet werden²⁷.

Die GIRT-Testdatenbank besteht aus insgesamt 11.000 Dokumenten (Titel-daten zuzüglich deutschsprachiger Inhaltsangabe oder Kurzreferat) der Datenbanken SOLIS (Sozialwissenschaftliches Literaturinformationssystem) und FORIS (Forschungsinformationssystem Sozialwissenschaften), die die Themenbereiche 'Frauenforschung', 'Industrie- und Betriebssoziologie' und 'Migration und ethnische Minderheiten' umfassen. Zusätzlich zu den Dokumenten mit Titel und Abstract stehen auch die Volltexte eines Teils der Testdatenmenge zur Verfügung, um Vergleiche zwischen der Indexierung von und Recherche in unterschiedlicher Dokumentkategorien zu ermöglichen.

Die Suchfragen werden zum größten Teil vom IZ Sozialwissenschaften formuliert, weitere Fragestellungen können aber auch von den Testern hinzugefügt werden. Sie können interaktiv oder vollständig automatisch an das entsprechende System gerichtet werden.

Die relevanten Dokumente werden intellektuell ermittelt, mit einer Relevanzbeurteilung versehen und ihrer Wichtigkeit entsprechend sortiert. Zur Testauswertung werden die Ergebnismengen der teilnehmenden Verfahren mit dieser Referenzmenge verglichen.

Für den Vergleich werden auf der einen Seite die Verfahren der Teilnehmer und auf der anderen Seite verschiedene Datenbanken des IZ bzw. der Hosts, bei denen Datenbanken des IZ aufliegen, herangezogen. So können verschiedene Retrievalsysteme miteinander verglichen werden²⁸.

3 Der MILOS-II-Retrievaltest

Dieses Kapitel stellt grundsätzliche Prinzipien von Retrievaltests dar und setzt sie in Zusammenhang mit den konkreten Fragestellungen, die sich bei der Planung des MILOS-II-Retrievaltests ergaben.

3.1 Datenbasis

Für die Beschreibung der eingesetzten Datenbank sind im wesentlichen die Aspekte Größe der Datenbank (Dokumentanzahl oder Größe in Bytes), enthaltenes Themenspektrum, Nutzungsart der Datenbank sowie Form und Art der Einträge von Interesse.

Die Datengrundlage für den Retrievaltest war eine Datenbank mit ca. 190 000 Titeln der Erscheinungsjahre 1991-1995 aus dem Datenbestand *Der Deutschen Bibliothek* (DDB). Die Titel der Testdatenbank verteilen sich auf alle Sachgruppen der Deutschen Nationalbibliographie mit Ausnahme der Belletristik, der Kinder- und Jugendliteratur sowie der Kalender. Die Datenbank *Der Deutschen*

27 Kluck, M.: Eine deutschsprachige Testdatenbank für moderne Erschließungs- und Retrievalsysteme (German Indexing and Retrieval Testdatabase - GIRT). In: Die digitale Revolution: Deutscher Dokumentartag 1996. Frankfurt a. M. 1996. S. 427-431.

28 Ebd., S. 428-429.

Bibliothek dient als bibliographisches Nachweisinstrument. Bei den enthaltenen Datensätzen handelt es sich um formale Titelaufnahmen von Monographien, Zeitschriften, Reihen und Serien, die zusätzlich mittels Schlagwörtern und Schlagwortketten, die gemäß der Regeln für die Schlagwortvergabe in Katalogen (RSWK) angesetzt wurden, sachlich erschlossen sind.

Im Rahmen von MILOS II wurden die Titeldaten zusätzlich erschlossen durch Deskriptoren, die durch automatische Indexierung mit dem System MILOS/IDX erzeugt wurden. MILOS/IDX ist ein wörterbuchorientiertes Verfahren, d.h. alle Prozesse zur Bearbeitung des Textmaterials laufen über einen Abgleich mit verschiedenen elektronischen Wörterbüchern²⁹. Diese Verfahrensweise basiert auf dem Konzept, sämtliche Relationen zwischen Wortformen und ihren zugehörigen Grundformen sowie auch semantische Beziehungen zwischen verschiedenen Begriffen (z.B. Synonyme, Teilwort/Kompositum, Oberbegriff/Unterbegriff) zu formalisieren. Hierfür gibt es die sogenannten Identifikationswörterbücher, in denen alle Grundformen samt ihrer möglichen Endungen enthalten sind, und die sogenannten Relationenwörterbücher, die neben semantischen Beziehungen zwischen den Wörterbucheinträgen Informationen zu Mehrwortgruppen, Dekomposition und Derivation festhalten. Durch die Bereitstellung fremdsprachiger Übersetzungswörterbücher können Texte auch in unterschiedlichen Sprachen indexiert werden (z.Zt. Englisch ⇒ Deutsch und Französisch ⇒ Deutsch).

Für die Indexierung des Datenpools der DNB wurde folgender Funktionsumfang eingesetzt:

- Die *Lemmatisierung* führt grammatikalisch unterschiedliche Wortformen im Text (Pluralformen, flektierte Formen) auf ihre Grundform zurück und stellt diese für ein späteres Retrieval bereit.
- *Stoppwörter* werden markiert bzw. eliminiert anhand von erweiterbaren Wortlisten.
- *Dekomposition* und *Derivation* zerlegen Komposita in ihre sinnvollen Bestandteile, die als Einzelbegriffe für das Retrieval bereitstehen.
- Durch die *Wortrelationierung* werden semantische Beziehungen zwischen Wörterbucheinträgen hergestellt.³⁰
- Die *Mehrwort-Erkennung* ermöglicht das Retrieval feststehender Wendungen und die *Wortbindestrichergänzung* löst das für die Stichwortsuche auftretende Problem der Tilgung von Wortbestandteilen.

Für den Einsatz in einem OPAC bietet MILOS dadurch folgende Funktionserweiterungen:

- Retrieval mit Grundformen und sinntragenden Teilwörtern (das Trunkierungsproblem entfällt weitgehend)
- Retrieval mit sinntragenden Mehrwortbegriffen (damit entfällt das Adjacency-Problem)
- Retrieval unter Einbeziehung von Übersetzungsäquivalenten.

Die Testdaten standen im MAB-Format zur Verfügung, so dass die für die Indexierung zu nutzenden Kategorien anhand ihrer Kategoriennummern selektiert

²⁹ Vgl. Fußnote 1: Lepsky, K.: Maschinelles Indexieren zur Verbesserung der sachlichen Suche im OPAC: DFG-Projekt an der Universitäts- und Landesbibliothek Düsseldorf. In: Bibliotheksdienst. 28 (1994), Nr. 8, S. 1234-1242.

³⁰ Eine Liste aller semantischer Relationen und ihrer Relationencodes findet sich im Anhang in der Tabelle 'Relationencodes', S. 42.

werden konnten. Es handelt sich dabei um die Kategorien Hauptsachtitel, Zusatz zum Hauptsachtitel sowie die Schlagwortkategorie.

Insgesamt besteht der Indexierungsablauf aus den Phasen Rechtschreibkontrolle und Wörterbuchpflege und der eigentlichen automatischen Indexierung. Diesen Phasen voran geht die Sprachselektion der Titel, durch die deutsche Titel von englischen, französischen und sonstigen Titeln getrennt werden. Dies geschieht durch das Schnittstellenprogramm MILSEL.

Während der Rechtschreibkontrolle können nicht nur Rechtschreibfehler im Textmaterial korrigiert werden, sondern auch in den Wörterbüchern fehlende (v.a. neue, unbekannte) Begriffe festgestellt und anschließend im Rahmen der Wörterbuchpflege nachgetragen werden. Sie erfolgt mit dem Programm WBTOOL, mit dem das Wortmaterial hinsichtlich seiner grammatikalischen Merkmale kodiert wird. Hier wird entschieden, ob die zuvor markierten Begriffe dem Vokabular hinzugefügt werden sollen, oder nicht, da es sich nicht um wichtige Begriffe handelt. Diese Begriffe bleiben zwar für die Indexierung unberücksichtigt, stehen jedoch weiterhin als Stichwörter für die Suche zur Verfügung. Die anschließende automatische Indexierung erfolgt in zwei Stufen mit dem Programm MILIDX, durch das die Grundformermittlung, die Kompositazerlegung, die Aufbereitung von Teilworttilgungen, die Bereitstellung von Wortrelationierungen und das Erkennen von Mehrwortbegriffen durchgeführt werden. Die hieraus entstehenden Ergebnisdateien, die jeweils die Identnummer der Datensätze und die Ergebnisse des ersten Indexierungslaufs enthalten, werden auf Fehler wie sinnlose Zerlegungen (z.B. 'Haltbarkeit' in 'Halt' und 'Barkeit') oder nichterkannte Komposita untersucht. Die nötigen Eintragungen im Relationenwörterbuch werden manuell eingefügt, anschließend kann der zweite Indexierungslauf folgen, in dem möglichst alle Wörter erkannt werden sollten. Die Ergebnisse des zweiten Indexierungslaufs, die neu gebildeten Wörter (Indexate) aus vier verschiedenen Ergebnisdateien (jeweils eine für die deutschen, die englischen und die französischen Titel sowie eine für die RSWK-Schlagwörter), werden in das *allegro*-Format umgesetzt und in der *allegro*-Datenbank zu den entsprechenden Datensätzen in zwei neue Kategorien (IDX (Titel) und IDX(RSWK)) dazugespielt. Die Verknüpfung der Indexate mit den einzelnen Titelaufnahmen geschieht über die Identnummern. Im folgenden Diagramm ist der Ablauf der Indexierung dargestellt:

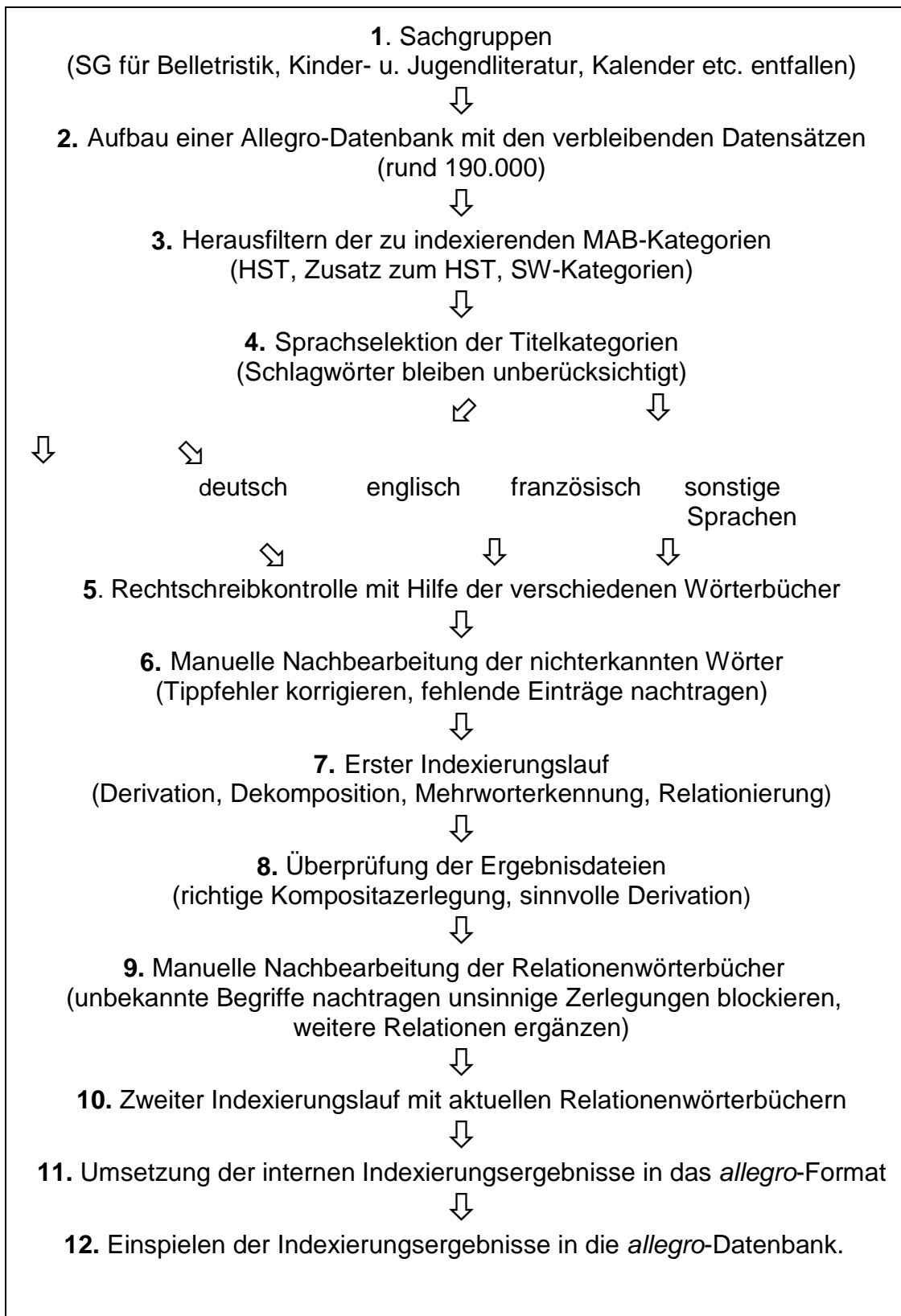


Abb. 1: Ablauf der Indexierung in Einzelschritten

Die folgende Abbildung zeigt einen Titelsatz mit den Indexierungsergebnissen:

| | |
|--|---|
| Dem Schlaganfall vorbeugen : Durchblutungsstörungen des Gehirns ; Risikofaktoren, Warnsignale, Untersuchungs- und Behandlungsmöglichkeiten / Günter Krämer. [Textzechn.: Friedrich Hartmann]. - Stuttgart: TRIAS Thieme Hippokrates Enke, 1993. - 322 S. | |
| Literaturverz. S. 296 - 299 | |
| ISBN 3-89373-227-6 | |
| IDX (Titel) | : Anfall <3>; Apoplektischer Insult <24>; Apoplexia cerebri <24>; Apoplexie <24>; Behandlung <3>; Behandlungsmöglichkeit; Cerebrum <24>; Durchblutung <3>; Durchblutungsstörung; Encephalon <24>; Faktor <3>; Gehirn; Gehirnschlag <24>; Hirn <24>; Hirn <4>; Hirnschlag <24>; Möglichkeit<3>; Risiko <3>; Risikofaktor; Schlag <3> Risikofaktor; Schlag <3>; Schlaganfall; Signal <3>; Störung <3>; Suchung; Vorbeugung <4>; Warnsignal; ZVI <24>; Zerebrovaskulärer Insult <24>; des; |
| IDX (RSWK) | : Anfall <3>; Apoplektischer Insult <24>; Apo- plexia cerebri <24>; Apoplexie <24>; Berater <25>; Gehirnschlag <24>; Hirnschlag <24>; Rat <3>; Ratgeber; Schlag <3>; Schlaganfall; ZVI <24>; Zerebrovaskulärer Insult <24> |
| Swk. (RSWK) | : Schlaganfall / Ratgeber |

Abb. 2: Titelsatz mit Indexierungsergebnissen zur Frage 'Behandlung von Schlaganfällen'³¹

3.2 Suchfragen

Unter einer Suchfrage ist das verbalisierte Informationsbedürfnis eines Nutzers zu verstehen. Für den Einsatz bei einem Retrievaltest ist die thematische Auswahl der Fragen wichtig und ob die Fragen der Praxis entnommen oder eigens für den Test formuliert wurden. Die Umsetzung einer Suchfrage in die Struktur einer bestimmten Retrievalsprache wird als Such-Statement oder Suchformulierung bezeichnet; eine Abfolge von Such-Statements oder Suchformulierungen, die sich alle auf dieselbe Suchfrage beziehen, ist ein Suchprozess.³² Die Fragen für den MILOS-Retrievaltest sollten einerseits den typischen Suchgewohnheiten eines Bibliotheksbenutzers entsprechen und andererseits auch die positiven wie negativen Effekte der automatischen Indexierung aufdecken. Für den Test wurde eine Liste von 100 Suchfragen zugrundegelegt. Es handelt sich dabei um die 50 Fragen, die bereits im Retrievaltest zum MILOS I-Projekt verwendet worden waren und in Anlehnung an die Ergebnisse einer in Düsseldorf durchgeführten Benutzerstudie³³ ausgewählt wurden, sowie um 50 weitere Fragen, die entsprechend den in der Literatur statistisch beobachteten Be-

³¹ Eine Liste aller semantischer Relationen und ihrer Relationencodes findet sich im Anhang in der Tabelle 'Relationencodes', S. 42.

³² Tague-Sutcliffe, J.: The pragmatics of information retrieval experimentation, revisited. In: Information Processing and Management. 28 (1992), Nr. 4, S. 471.

³³ Dreis, G.: Benutzerverhalten an einem Online-Publikumskatalog für wissenschaftliche Bibliotheken: Ergebnisse und Erfahrungen aus dem OPAC-Projekt der Universitätsbibliothek Düsseldorf. Frankfurt am Main: Klostermann 1994.

nutzergewohnheiten zur Formulierung von Suchfragen zusammengestellt wurden. Sie lassen sich folgendermaßen gruppieren³⁴:

1. Fragen nach einfachen Sachverhalten mit einem Suchbegriff (einschließlich Komposita);
2. Fragen nach Sachverhalten mit zwei Suchbegriffen;
3. Fragen, in denen Beziehungen zwischen mehreren Begriffen vorhanden sind;
4. Fragen, die Adjektiv-Substantiv-Verbindungen für einen festen Begriff enthalten;
5. Fragen, die Eigennamen mit einer Mehrwort-Verbindung oder Zählung enthalten.

Durch die Verwendung der 50 Fragen für den MILOS I-Retrievaltest ist ein Vergleich der in beiden Tests erzielten Ergebnissen möglich.

Inhaltlich wurden Fragen bevorzugt, die dem bibliothekarischen Alltag, d.h. der OPAC-Praxis entstammen (könnten). Da bislang kein allgemeinverbindlicher Kanon derartiger Fragen vorliegt, ist eine solche Zusammenstellung natürlich nicht repräsentativ.

Ein zusätzliches Kriterium bei der Auswahl der Testfragen lag darin, ob und wenn ja in welcher Höhe die jeweilige Suchfrage Treffer erwarten ließ. Es sollte differenziert werden können zwischen Nulltreffer-Mengen, die aufgrund des Nichtvorhandenseins von relevanten Dokumenten vorkommen und solchen, die durch nicht ausreichende Erschließung der Dokumente entstehen.

Im folgenden eine Auflistung der 100 Suchfragen³⁵:

3.2.1 Fragen nach einfachen Sachverhalten mit einem Suchbegriff (einschließlich Komposita)

- | | |
|----------------------------|-----------------------------|
| 1. Altersversorgung | 14. Konjunkturpolitik |
| 2. Antriebstechnik | 15. Kriegsberichterstattung |
| 3. Aromatherapie | 16. Kurden |
| 4. Betriebspsychologie | 17. Lichtbogenschweißen |
| 5. Bundestagswahl | 18. Magersucht |
| 6. Chaostheorie | 19. Marktanalyse |
| 7. Direktmarketing | 20. Nomaden |
| 8. Eherecht | 21. Qumranrollen |
| 9. Energiewirtschaft | 22. Sprachwandel |
| 10. Ergonomie | 23. Straßenkinder |
| 11. Festkörperphysik | 24. Theatergeschichte |
| 12. Gerontologie | 25. Tierexperimente |
| 13. Gesprächsführung | 26. Umweltgifte |
| 27. Heilfasten | 32. Verkehrsgeographie |
| 28. Homebanking | 33. Wirtschaftsethik |
| 29. Interkontinentalrakete | 34. Wissensorganisation |
| 30. Internet | 35. Zimmerpflanzen |
| 31. Karate | |

³⁴ Gödert, W.: Sacherschließung in EDV-Katalogen: Probleme der Sacherschließung unter Einsatz von PC-Systemen. In: Mitteilungen der Arbeitsgemeinschaft der Parlaments- und Behördenbibliotheken. (1992), Nr. 73, S. 5-37.

³⁵ Die Suchfragen und die dazugehörigen Suchformulierungen sowie die Ergebnisse der Suche können im Anhang der Tabelle 1: 'Suchfragen mit Ergebnissen', S. 37, entnommen werden.

3.2.2. Fragen nach einfachen Sachverhalten mit zwei Suchbegriffen

- | | |
|---|---------------------------------------|
| 36. Alarmanlagen für das Auto | 53. Hyperaktivität bei Kindern |
| 37. Anleitung zum Videofilm | 54. Koedukation in der Schule |
| 38. Außenpolitik der USA | 55. Kunst in der Antike |
| 39. Behandlung bei Schlaganfällen | 56. Ökologie der Gewässer |
| 40. Das Individuum in der Gesellschaft | 57. Pflanzen im Aquarium |
| 41. Datenbanken in der Dokumentation | 58. Philosophie der Gegenwart |
| 42. Denken und Lernen | 59. Politik und Massenmedien |
| 43. Einführung in die Volkswirtschaftslehre | 60. Reaktorunfall und Tschernobyl |
| 44. Einwanderer in Amerika | 61. Reiseführer für die Toskana |
| 45. Elektronik im Kraftfahrzeug | 62. Selbstbewusstsein stärken |
| 46. Erkrankungen der Lunge | 63. Sexualerziehung und Schule |
| 47. Frau und Beruf | 64. Software zur Textverarbeitung |
| 48. Geschichte der Plastik | 65. Strahlenschutz beim Röntgen |
| 49. Geschichte des Mittelmeerraumes | 66. Studium im Ausland |
| 50. Geschichte Israels | 67. Widerstand im Nationalsozialismus |
| 51. Gründung eines Vereins | 68. Wirtschaft und Wachstum |
| 52. Hemmung von Enzymen | 69. Wirtschaftswunder in Deutschland |
| <i>'oder'-Verknüpfung</i> | |
| 70. Umweltökonomie / -ökonomik | 72. Kindesmissbrauch /-misshandlungen |
| 71. Abfallentsorgung / -beseitigung | 73. Bedeutung von Ying und Yang |

3.2.3 Fragen, in denen Beziehungen zwischen mehreren Begriffen vorhanden sind; 'Und'-Verknüpfung

- | | |
|---------------------------------------|--|
| 74. Folgen einer Scheidung für Kinder | 77. Spurenelemente in der Ernährung des Menschen |
| 75. Krankheit infolge von Stress | |
| 76. Medizin im Dritten Reich | 78. Vergesslichkeit und Übungen zum Gedächtnistraining |

3.2.4 Fragen, die Adjektiv-Substantiv-Verbindungen für einen festen Begriff enthalten

- | | |
|-----------------------------------|----------------------------------|
| 79. Alles über die Zeugen Jehovas | 87. Lateinamerikanische Tänze |
| 80. Alternative Energiequellen | 88. Pluralistische Gesellschaft |
| 81. Homoöpathische Mittel | 89. Politischer Skandal |
| 82. Interkulturelle Erziehung | 90. Psychologische Diagnostik |
| 83. Islamischer Fundamentalismus | 91. Psychosomatische Krankheiten |
| 84. Italienische Kochbücher | 92. Shakespeares Dramen |
| 85. Kritische Theorie | 93. Wissenschaftliches Arbeiten |
| 86. Künstliche Intelligenz | |

3.2.5 Fragen, die Eigennamen mit einer Mehrwort-Verbindung oder Zählung enthalten

94. Weiße Rose

3.2.6 Eigennamen, die in verschiedenen Schreibweisen möglich sind

95. Baby-Massage
96. Excel 5.0
97. Risk-management

3.2.7 Komplexere Suchfragen, die aus drei Suchbegriffen bestehen und miteinander verknüpft werden (Kombination mit 'und' und 'oder')

98. Herstellung von Zeitungen und Zeitschriften
99. Angst, Angststörungen und Angsterkrankungen
100. Entstehung und Entwicklung der Schrift.

Für die Suchfragen wurden jeweils verschiedene Suchformulierungen gewählt, um herauszufinden, mit welchen Formulierungen die besten Suchergebnisse erzielt werden können und so die Problematik der verschiedenen Interpretationsmöglichkeiten von Suchfragen zu verdeutlichen. Gleichzeitig sollten die Suchformulierungen für die verschiedenen Typen von Fragestellungen möglichst ähnlich sein, damit die Vergleichbarkeit der einzelnen Fragen gewährleistet ist. Bei der Frageformulierung wurde auf Trunkierung verzichtet.

3.3 Suchprozess

Dieselben natürlichsprachigen Fragestellungen können in dem Prozess der Umsetzung in eine Retrievalsprache von verschiedenen Personen unterschiedlich interpretiert werden:

„Differiert das Themenverständnis der Versuchsperson von dem des Jurors, wirkt sich das nicht nur in der Beurteilung der gefundenen Dokumente aus, sondern es bestimmt durch die entsprechende Suchfrageformulierung auch, welche Dokumente überhaupt gefunden werden und somit für die Beurteilung vorliegen. Darüber hinaus führt dieses Beispiel vor Augen, wie groß der Einfluss der Versuchspersonen bei derartigen Tests ist und wie schwierig es ist, letztlich die Leistung des Retrievalsystems von den Fähigkeiten der recherchierenden Person zu trennen.“³⁶

Die Umsetzung der Fragestellung in eine bestimmte Retrievalsprache ist nicht nur inhaltlichen Interpretationsmöglichkeiten unterworfen, sondern hängt auch von der fachlichen Kompetenz des Nutzers auf dem Gebiet der Retrievalsprache ab. Erfahrene Informationsvermittler können doppelt so hohe Recall-Werte erzielen wie ungeschulte Benutzer³⁷.

Bessere Kenntnis einer Datenbank und der Retrievalsprache müssen aber nicht zwingend auch zu besseren Retrievalergebnissen führen. Ein unvoreingenommenes Herangehen und Formulieren der Suchfragen kann ebenso zu guten Retrievalergebnissen führen: „Gleichzeitig haben weniger rechercheerfahrene Versuchspersonen teilweise bei einzelnen Fragestellungen bessere Ergebnisse erzielt als erfahrenere Versuchspersonen. Die weniger erfahrenen Versuchspersonen haben in diesen Fällen die Frage unmittelbarer in Suchbegriffe umgesetzt und weniger ausgefeilte Strategien angewendet“³⁸.

Dieser Aspekt ist bei Überlegungen zum zukünftigen Design von Datenbanken und Benutzeroberflächen zu berücksichtigen. Es ist davon auszugehen, dass Recherchen in Zukunft immer häufiger von retrieval-unerfahrenen Nutzern selbst durchgeführt werden und weniger unter Inanspruchnahme von Daten-

³⁶ Frisch: Pretest zum Projekt German Indexing and Retrieval Testdatabase, S. 39.

³⁷ Glöckner-Rist, A.: Suchfragen im Information Retrieval: eine empirische Untersuchung zum Rechercheverhalten von Informationsvermittlern und Endbenutzern. Konstanz: Universitätsverlag 1993.

³⁸ Frisch: Pretest zum Projekt German Indexing and Retrieval Testdatabase, S. 42.

bankexperten. Deshalb ist es wichtig, diese Nutzerschnittstellen möglichst nutzerfreundlich und einfach handhabbar zu gestalten.

Während des Suchprozesses kann Einfluss auf die Retrievalergebnisse genommen werden. Modifikationen der Suchfragen können zum einen nach Ansicht der ersten Suchergebnisse und anschließendes Hinzufügen oder Ausschließen bestimmter Suchbegriffe geschehen oder andererseits durch ein Relevance-Feedback-Verfahren. Das Relevance Feedback stellt einen dynamischen Suchprozess dar, in dem Suchfragen unter Einbeziehen von Informationen aus vorangegangenen Suchen automatisch umformuliert werden. Der Einfluss von Fragemodifikationen auf die Retrievalergebnisse stellt - wie jedes 'Eingreifen' von Menschen in den Retrievalprozess - einen schwer zu messenden Faktor dar.

Im Rahmen dieses Tests wurde keine Modifikation der Suchfragen während des Suchprozesses vorgenommen, um klare und eindeutige Aussagen treffen zu können.

3.4 Retrievalsoftware

Für den Retrievaltest war es notwendig, die Daten der Testdatenbank so aufzubereiten, dass die aus der automatischen Indexierung gewonnenen Indexate zu den bereits vorhandenen Titelaufnahmen *Der Deutschen Bibliothek* hinzugespielt werden konnten. Da MILOS keine Datenbank- und Retrievalfunktion bietet, wurde hierfür die Bibliothekssoftware *allegro* ausgewählt, die differenzierte Suchen in einer Online-Umgebung ermöglicht. *Allegro* wurde gewählt, weil zum einen mit dieser Software die Anbindung der Indexierungsdaten an die vorhandenen OPAC-Titeldaten am einfachsten zu bewerkstelligen war und sich zum anderen die Register für die Suche individuell generieren lassen. Über diese Register werden die Suchfragen eingegeben und einzelne Suchbegriffe daraus können anhand der Boole'schen Operatoren *und*, *oder* und *nicht* miteinander verknüpft werden. Da die Suchanfrage durch Auswählen des gesuchten Begriffs in einem Register ausgeführt wird, kann direkt überprüft werden, ob überhaupt Eintragungen zu dem Suchbegriff vorhanden sind. Ist dies nicht der Fall, wird der dem gesuchten Begriff nächstliegende Begriff angezeigt. So kann festgestellt werden, ob es Begriffe in anderen Wortformen zu dem gesuchten Sachverhalt gibt, oder ob eine andere Schreibweise nötig ist.

Die Besonderheiten einer Retrievalumgebung beeinflusst die Durchführung von Suchen. Da Suchen in *allegro* über (meist wortinvertierte) Register abgewickelt werden, ist bei einer Suche einerseits das alphabetische Umfeld eines Suchbegriffs ständig sichtbar, andererseits sind z.B. Adjektiv-Substantiv-Verbindungen teilweise nur durch kombinierende Verknüpfung der Wörter suchbar. Es musste also die Suchworteingabe bzw. die Eingabe einer komplexen Mehrwortfrage seitens eines fiktiven Benutzers mit den Gegebenheiten alphabetisch geordneter Listen in Übereinstimmung gebracht werden. Die Alternative einer Eingabe über Suchmasken, ohne gleichzeitig das Register aufgeblättert zu bekommen - wie dies beispielsweise in vielen WWW-OPACs der Fall ist - stand nicht zur Verfügung. So waren verschiedene Abwägungen im Hinblick darauf zu treffen, dass bei Eingabe einer Frage in eine Suchmaske nicht die gleiche Transparenz des Vokabulars (z.B. Singular-Pluralformen oder Flektionsendungen) gegeben ist wie in einer alphabetisch geordneten Liste. Soweit wie möglich und sinnvoll wurden die Suchfragen in verschiedene Eingabevarianten transformiert, um die Treffermengen zu bilden.

Die Testfragen wurden in den folgenden fünf Registern bearbeitet:

- Register der Titel-Stichwörter,
- Register der Indexierungsergebnisse,
- Register der verstichworteten RSWK-Ketten,
- Register der RSWK-Ketten,
- Basic Index (Titel-Stichwörter, Indexierungsergebnisse und verstichwortete RSWK-Ketten).

Um im Basic Index die Begriffe aus den verschiedenen Feldern voneinander unterscheiden zu können, wurden sie mit bestimmten Kennungen versehen. Allen Begriffen, die aus zerschlagenen Schlagwortketten stammen, wurde ein Stern angefügt (*), während die IDX-Indexate mit einem Gitter (#) versehen wurden. Eintragungen ohne Kennung stammen aus der Gesamtmenge von Stichwort, IDX-Indexat oder RSWK-Schlagwort.

```

41  stress *           => konventionelles Schlagwort
    1  stress-belastung
    1  stress-belastung #
    3  stress-management
    3  stress-management #
    5  stress-situation
    5  stress-situation #
    5  stress-situationen
    1  stress-test     => Vereinigung aus Stichwort, IDX, Schlagwort
    1  stress-test #   => Ergebnis der automatischen Indexierung
Andere Stelle? Suchwort eintippen.      <Sh+F 9> : Ergebnis-Kurzliste
-+ = Titelanzeige    ? = HILFE                                     Ergebn.: 16

```

Abb. 3: Auszug aus dem Basic Index

Als Suchergebnis einer Suchanfrage wird eine Kurztitelliste der Ergebnismenge angezeigt. Von dort aus kann die Vollanzeige aufgerufen werden.

Ein Titeleintrag enthält neben den formalen Daten einer RAK-Titelaufnahme zusätzlich als inhaltserschließende Kategorien RSWK-Schlagwörter und -Ketten, sowie die Indexate der Titel, RSWK-Ketten und -Schlagwörter.

Anhand der Suche nach der Frage 'Tierexperimente' soll das Suchverfahren in *allegro* beispielhaft dargestellt werden.

Zunächst wurde der Suchbegriff 'Tierexperimente' im Stichwortregister eingegeben. In den Registern öffnet sich das Fenster für die Eingabe automatisch, sobald der erste Buchstabe des gesuchten Begriffs eingetippt wird. Über das Stichwortregister konnte die Suchfrage nicht beantwortet werden, da 'Tierexperimente', also die Pluralform zu 'Tierexperiment', in keinem Titel als Stichwort vorkommt und nur die Singularform des Suchbegriffs zu Treffern führen würde.

Der selbe Suchvorgang wurde auch im Schlagwort- und im RSWK-Ketten-Register durchgeführt; da 'Tierexperimente' keine Vorzugsbenennung ist, gab es dort keine Treffer. Literatur zu diesem Thema wäre nur über den Begriff 'Tierversuch' zu finden, was 22 relevante Titel ergeben hätte.

Anschließend wurde die Suche im Basic-Index wiederholt:

```

3 : Basic Index (StW/IDX-StW/RSWK-SW, IDX-StW (#), RSWK-SW (*))
38  tierexperiment
38==>tierexperiment #
12  tierexperimentell
12  tierexperimentell #
12  tierexperimentelle
1  tierexperimentellen
1  tierexperimenten
7   tierfabel
7   tierfabel #
3   tierfang
3   tierfang #
3   tierfang *
Andere Stelle? Suchwort eintippen. <Sh+F 8> = Titel-Kurzliste
-+ = Titelanzeige      ? = HILFE

```

Abb. 4: Anzeige im Basic Index zur Frage nach 'Tierexperimente'

Von der Liste der Registereinträge gelangt man durch 'Enter' zu der dazugehörigen Kurztitelliste:

| Titel | Verf. | Jahr |
|---|------------------------|--------------------|
| 17. Erste Untersuchungen über Biodistribution u/ | Steinmann, | 1993 |
| 18. Sterben für die Forschung | /Kracht, Ol | 1993 |
| 19. Leben ohne Seele? | /Schmidt, W | 1991 |
| 20. Kritik und Verteidigung des Tierversuchs | /Maehle, An | 1992 |
| 21. Einwachsverhalten zementfrei implantierter H/ | Mockwitz, | 1994 |
| 22. Möglichkeiten und Grenzen der Reduktion von | /Schöffl, H | 1992 |
| 23. Kollagenolyse und Ovulation | /Hirsch, Bu | 1994 |
| 24. Sepsis bei Splenektomie | /Izbicki, J | 1991 |
| 25. Gesundheit für Mensch und Tier | /Küsters, G | 1993 |
| 26. Asbest, Fibrose, Krebs | /Friemann, | 1992 |
| 27.=> Leiden der Versuchstiere! | /Hannak, El | 1990 |
| 28. Tierschutz durch Alternativen | /Hardegg, W | 1993 |
| 29. Tierversuche in der Forschung | | 1993 |
| 30. Über Leichen zum Examen? | /Rieg, Timo | 1993 |
| <Bild> <Bild> : blättern | <Esc> : Normalregister | |
| -+ = Titelanzeige | ? = HILFE | Ergebn.: 38 |

Abb. 5: Kurztitelliste zur Frage nach 'Tierexperimente'

Bereits anhand der Kurztitelliste ist zu erkennen, dass sich unter den aufgelisteten Titeln einige relevante Treffer finden.

Von der Kurztitelliste ausgehend gelangt man mit 'Enter' zu der Vollanzeige des Titels:

| | |
|---|--|
| Die Sepsis bei Splenektomie : tierexperimentelle Befunde zum Milzerhalt und zur Immunaktivierung ; mit 15 Tabellen / Jakob R. Izbicki. - Berlin ; Heidelberg ; New York ; London ; Paris ; Tokyo ; Hong Kong ; Barcelona: Springer, 1991. - XI, 102 S. (Hefte zur Unfallheilkunde ; 210) | |
| Literaturverz. S. 87 - 97 | |
| ISBN 3-540-53180-7 | |
| IDX (Titel) | : Aktivierung <3>; Befund; Diagramm <15>; Erhalt <3>; Immunaktivierung; Milz <3>; Milzerhalt; Milzexstirpation <24>; Sepsis; Splenektomie; Tabelle; Tier <3>; Tierexperiment <4> ; experimentell <3>; immun <3>; tierexperimentell; zu; |
| IDX (RSWK) | : Modell <3>; OPSI; OPSI-Syndrom; Overwhelming postsplenectomy infection <24>; Postsplenektomie-Sepsis <24>; Sepsis <15>; Syndrom; Tier <3>; Tiermodell |
| Swk. (RSWK) | : OPSI-Syndrom / Tiermodell |
| -> = zurück zum Register ? = HILFE | |
| Ergebn.: 38 | |

Abb. 6: Titelsatz zur Frage nach 'Tierexperimente'

Interessant ist nun, wie die Differenz der 38 durch IDX gefundenen und nur 22 durch Schlagwort gefundenen Titel zu erklären ist. Bei näherer Untersuchung der 16 differierenden Titel, die über die Suchfrage 'Tierexperiment not Tierversuch' ermittelt wurden, zeigt sich, dass es sich um Titel handelt, in denen Tierversuche nur als Mittel zum Zweck der medizinischen Forschung dienen, und dass also die RSWK zwar den Gegenstand der Untersuchung, nicht aber die angewandte Methode erfassen.

3.5 Messzahlen

Messzahlen dienen dazu, die Effektivität eines Retrievalsystems in möglichst objektiv nachvollziehbaren Werten darzustellen und eine Vergleichsmöglichkeit zu bieten. Die am häufigsten genutzten Werte sind Recall und Precision³⁹. Recall definiert die Anzahl der relevanten nachgewiesenen Dokumente im Verhältnis zu allen in der Dokumentmenge enthaltenen relevanten Dokumente, Precision die Anzahl der nachgewiesenen relevanten Dokumente im Verhältnis zu sämtlichen in der Dokumentmenge enthaltenen Dokumente. Zusätzlich kann auch der Fallout hinzugezogen werden, der die Anzahl an nichtrelevanten nachgewiesenen Dokumente wiedergibt. Bei Unterscheidung einer Datenkollektion in folgende Teile

- a Anzahl der nachgewiesenen und relevanten Dokumente
- b Anzahl der nachgewiesenen und nicht relevanten Dokumente
- c Anzahl der nicht nachgewiesenen und relevanten Dokumente
- d Anzahl der nicht nachgewiesenen und nicht relevanten Dokumente

ergeben sich als Formeln

Recall:
$$r = \frac{a}{a + c}$$

39 Salton, G.: The state of retrieval system evaluation. In: Information Processing and Management. 28 (1992), Nr. 4, S. 441.

Precision: $p = \frac{a}{a+b}$

Fallout: $f = \frac{b}{b+d}$

Ein optimales Retrievalsystem weist maximalen Recall bei minimalem Fallout auf.

Schwierigkeiten bei der Ermittlung des Recalls bestehen darin, alle in einer Dokumentmenge enthaltenen relevanten Dokumente zu ermitteln sowie die bereits beschriebene Variabilität des Relevanzkriteriums.

Neben den beschriebenen rechnerisch zu ermittelnden und daher (eher) objektiven Maßzahlen kann die subjektive Relevanz beschrieben werden anhand der Kriterien

- Neuigkeitsrate: nachgewiesene Dokumente, die relevant und vorher nicht bekannt waren,
- Abdeckungsrate: nachgewiesene relevante Dokumente, die bereits bekannt waren und
- gesuchter Recall: Anteil der nachgewiesenen relevanten Dokumente unter allen gesuchten relevanten Dokumenten⁴⁰.

Verschiedene Lösungsansätze beschäftigen sich mit dem Problem der Recall-Bestimmung⁴¹:

1. Eine bestimmte Menge an relevanten Dokumenten wird im Voraus ermittelt.
2. Eine kleine Dokumentmenge wird für den Test genutzt, und die Relevanz aller Dokumente für die Suchfragen wird durch Benutzer oder Systembetreuer ermittelt.
3. Eine zufällige Auswahl aus den nicht nachgewiesenen Dokumenten wird im Hinblick auf ihre Relevanz untersucht. Da aber in großen Datenbanken die Menge an relevanten Dokumenten sehr gering ist, muss eine sehr große Auswahl getroffen werden, um auf diese Weise zu aussagekräftigen Ergebnissen zu kommen.
4. Bei vergleichenden Test kann der relative statt der absolute Recall gemessen werden. Dies kann allerdings nur angewandt werden, wenn auf der Basis einer Dokumentmenge und eines bestimmten Fragenkatalogs verschiedene Retrievaltechniken verglichen werden sollen.
5. In sehr großen Datenbanken kann eine sehr weit gefasste Suche ausgeführt werden und die so ermittelte Dokumentmenge als die Menge der in der Datenbank enthaltenen relevanten Dokumente zugrunde gelegt werden.
6. Bei einem vergleichenden Test mit denselben Voraussetzungen wie den unter Punkt 4 erläuterten wird durch das sog. Pooling-Verfahren die Gesamtanzahl aller durch verschiedene Retrievaltechniken ermittelten relevanten Dokumente als Grundlage für die Recall-Bestimmung genommen. Kritik richtet sich bei diesem Verfahren darauf, dass es die subjektiven Relevanzentscheidungen einzelner Nutzer außer Acht lässt:

⁴⁰ Salton: Information Retrieval, S. 188.

⁴¹ Tague-Sutcliffe: The pragmatics of information retrieval, S. 474.

„However, it is ill-suited to a user-oriented research agenda, because it obliterates the environmental differences among users, queries, and their associated relevance judgements.“⁴²

Kritisch zu sehen ist, dass für die Berechnung der Werte Recall und Precision andere wichtige Messwerte wie Größe der Datenkollektion und Anzahl der über eine Suchfrage nachgewiesenen Dokumente nicht berücksichtigt werden⁴³. Eine Alternative für die Relevanzbeurteilung stellt das Verfahren von Cooper dar, in dem die Relevanz anhand der *expected search length*, der erwarteten Suchlänge, ermittelt wird. Hierbei wird gemessen, wie viele Dokumente einer Ergebnismenge durchschnittlich durchsucht werden müssen, bis eine bestimmte Anzahl an relevanten Dokumenten gefunden wurde⁴⁴. Problematisch dabei ist allerdings, dass im Vorhinein eine gewünschte Anzahl an relevanten Dokumenten festgelegt werden muss. Diese Zahl ist aber nur schwer zu bestimmen und wiederum abhängig von bestimmten Nutzergruppen und -bedürfnissen. Bei dem Umfang der für diesen Retrievaltest eingesetzten Datenbank war es von vornherein ausgeschlossen, alle enthaltenen Dokumente zu kennen. Insofern war es auch nicht möglich, für alle Suchfragen alle in der Datenbank vorhandenen relevanten Dokumente zu bestimmen. Es wurde darauf verzichtet, durch Einsatz aller zur Verfügung stehenden Retrieval-Hilfsmittel eine Annäherung an diese Dokumentmenge zu finden, da nicht sichergestellt werden konnte, dass dies gleichmäßig für alle 100 Fragen möglich sein würde. Die Angabe einer rechnerischen Größe Recall war damit nicht möglich. Die vorgenommene quantitative Analyse beschränkt sich dementsprechend auf die Angabe der Werte der gefundenen Titel, die in Beziehung gesetzt werden zu den gefundenen relevanten Titeln.

3.6 Relevanzbewertung

Unterschiedliche Informationsbedürfnisse und Wissensstand der Nutzer führen zu unterschiedlichen Relevanzeinschätzungen, so dass zwischen objektiver und subjektiver Relevanz zu unterscheiden ist⁴⁵. Bei objektiver Relevanz wird ein Dokument auch dann als relevant angesehen, wenn es dem Nutzer bereits bekannt ist. Subjektiv relevant sind dahingegen nur solche Dokumente, die das aktuelle Informationsbedürfnis eines Nutzers befriedigen. Dies entspricht auch der Relevanzunterscheidung von Cleverdon in *stated relevance* und *user relevance*, je nachdem, ob ein Relevanzurteil aufgrund eines konkreten Informationsbedürfnis gefällt wird oder nicht. Letzteres kann von jeder Person, die ein gewisses Fachwissen hinsichtlich der Suchfrage besitzt, beantwortet werden, während die *user relevance* allein von dem jeweiligen Fragesteller eingeschätzt werden kann⁴⁶. Ein Dokument kann generell gesehen relevant, im Hinblick auf die konkrete Fragesituation aber nutzlos sein.

⁴² Harter: Variations in relevance assessments, S. 46.

⁴³ Salton: The state of retrieval system evaluation. S. 442.

⁴⁴ Cooper, W. S. : Expected search length: a single measure of retrieval effectiveness based on weak ordering action of retrieval systems. In: Journal of the American Society for Information Science. 19 (1968), S. 30-41.

⁴⁵ Salton, G.; M. J. McGill: Information Retrieval - Grundlegendes für Informationswissenschaftler. Hamburg: McGraw-Hill 1987. S. 174.

⁴⁶ Cleverdon, C. W., J. Mills u. E. M. Keen: Factors determining the performance of indexing systems. Vol. 1: Design. Cranfield: Aslib 1966.

Um den Einfluss der Konsistenz und Variabilität von Relevanzfeststellungen zu untersuchen, wurden eigene Tests durchgeführt. Es wurde festgestellt, dass auch äußere Faktoren eine Beurteilung beeinflussen:

„... These included the type of document, its subject matter, level of difficulty, style and condensation, the conditions of the relevance judgement in terms of the time available, order of presentation of material, definition of relevance employed, the content and specificity of the stated information requirements, and the type of relevance judge employed in terms of experience, background, attitude, knowledge, intelligence, error preference and user orientation.“⁴⁷

Zudem wurde durch Experimente mit Psychologiestudenten belegt, dass die von den Relevanzbeurteilern vermutete Gebrauchsabsicht für ein Dokument individuelle Relevanzbeurteilungen beeinflusst. Dies bedeutet, dass allein durch bestimmte Suchthemenformulierungen oder Instruktionen an die Testpersonen die Relevanzurteile stark variieren können.

Der Einfluss der Relevanz auf Maßzahlen für Retrievaltests wurde von Lesk und Salton untersucht⁴⁸. Die Schwierigkeit bei Relevanzurteilen in Retrievaltests besteht demnach darin, dass einerseits ein an der Praxis orientiertes, realistisches Informationsbedürfnis repräsentiert werden soll, und dieses andererseits nach objektiv nachvollziehbaren Kriterien beurteilt werden können muss, also auf intellektuellen Prozessen basierende Kennzahlen für die Errechnung der Systemeffektivität zugrunde gelegt werden⁴⁹.

Relevanz kann in verschiedenen graduellen Abstufungen oder binär in relevante und nicht relevante Dokumente unterschieden werden. Graduelle Abstufungen werden häufig in eine vierwertige Skala unterteilt, die die Werte 'sehr relevant - relevant - nicht zutreffend - völlig irrelevant' umfasst. Es können aber auch fünf- oder sechswertige Skalen genutzt werden. Zu beachten ist dann allerdings, dass es zu Ungenauigkeiten in der Auswertung kommen kann. Graduelle Abstufungen erlauben präzisere Aussagen über die Relevanz, sind aber umstritten hinsichtlich der Frage, ob differenzierte Relevanzurteile schwerer zu fällen und mit größerer Unsicherheit verbunden sind. Salton stellt fest, dass „... in der Praxis [...] mit graduellen Relevanzabstufungen viel mehr Schwierigkeiten verbunden [sind] als mit der einfachen Unterscheidung zwischen relevanten oder teilweise relevanten und irrelevanten Dokumenten. Sollen die Nutzer zwischen absolut relevanten, möglicherweise relevanten, wenig relevanten und nicht relevanten Dokumenten unterscheiden, nehmen Fehlerrate und Unsicherheit in der Beurteilung zu“⁵⁰. Dahingegen äußern Frisch und Kluck in ihrer Darstellung des Pretests zum Projekt GIRT, dass „mit einer mehrwertigen Skala [...] Entscheidungsunsicherheiten eher aufgefangen werden [können]“⁵¹.

Für den vorliegenden Retrievaltest wurde folgende Festlegung für die Relevanzbeurteilung getroffen: jeder Titel, der nach Ansicht aller Daten der biblio-

⁴⁷ Cuadra, C. A., R. V. Katter: Experimental studies of relevance judgements. Santa Monica, Calif. 1967. Cuadra, C. A., R. V. Katter: Opening the black box of relevance. In: Journal of Documentation. 23 (1967), S. 291-303.

⁴⁸ Lesk, M. E., G. Salton: Relevance assessments and retrieval system evaluation. In: Information Storage and Retrieval. 4 (1968), S. 343-359.

⁴⁹ Ellis: Progress and problems in information retrieval, S. 18.

⁵⁰ Salton: Information Retrieval, S. 185.

⁵¹ Frisch: Pretest zum Projekt German Indexing and Retrieval Testdatabase, S. 17.

graphischen Beschreibung einschließlich der zugeteilten Schlagwörter nicht als von vornherein irrelevant erschien, bei dem also ein Interesse vermutet werden konnte, sich das Originaldokument genauer anzusehen, wird als relevant gewertet. Zu unterscheiden war in relevante und nicht relevante Dokumente. Das sehr weit gefasste Relevanzkriterium wurde gewählt, da keine Fachwissenschaftler zur Verfügung standen, um Relevanzbeurteilungen zu treffen und dies auch aufgrund der Größe der Datenbank und der Anzahl der Suchfragen nicht durchführbar gewesen wäre.

4 Ergebnisse des MILOS-II-Retrievaltests

In diesem Kapitel werden detailliert die Ergebnisse der Suche über die einzelnen Register dargestellt, miteinander verglichen und teilweise anhand von Beispielen erläutert.

4.1 Ermittelte Werte

Die Grundlage für die Berechnung von Recall und Precision bilden die Treffermengen der einzelnen Suchfragen in den verschiedenen Registern. Die Gegenüberstellung der insgesamt gefundenen Treffer innerhalb der einzelnen Register zeigt, daß durch die automatische Indexierung insgesamt gesehen mehr relevante Titel gefunden wurden, wenn auch mehr Ballast erzeugt wurde. In den folgenden Abbildungen sind die Ergebnisse der Suchen in den einzelnen Registern abgebildet.

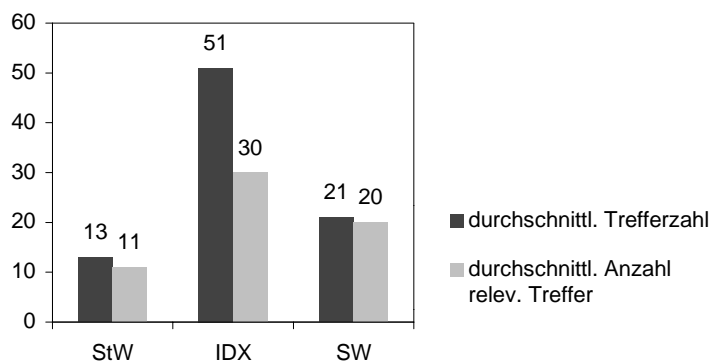


Abb. 7: Anzahl Gesamttreffer / Anzahl relevanter Dokumente

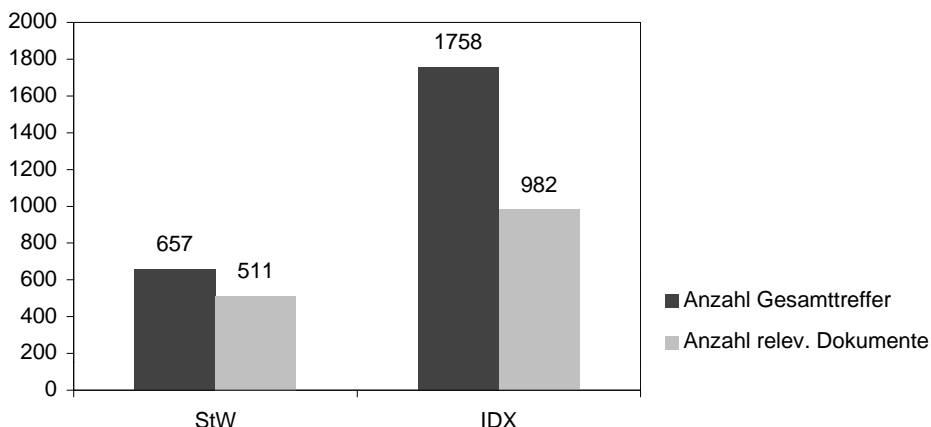


Abb. 8: durchschnittl. Trefferzahl / durchschnittl. Anzahl relevanter Treffer

Die Suche im IDX-Register erzielt fast dreimal so viele relevante Treffer wie die Suche im Titelstichwort-Register und um die Hälfte mehr relevante Treffer als die Suche im Schlagwort-Register. Da in die automatische Indexierung die sachliche Erschließungskomponente der SWD integriert wurde, wird nicht ganz deutlich, inwieweit die Suche im IDX-Register auch dann zu positiven Retrievalergebnissen führt, wenn dieser Aspekt nicht einbezogen werden kann. Die relativ hohe Anzahl von nicht relevanten, gefundenen Titeln durch die Suche im IDX-Register ist u.a. durch die Auswahl der Suchfragen zu begründen: bei den Fragen nach 'Interkontinentalrakete' und 'Medizin im Dritten Reich' war bereits im voraus abzusehen, daß sie negativ für IDX ausfallen würden, da aufgrund der einbezogenen Homonymenzusätze nahezu alle gefundenen Treffer irrelevant sind.

Eine wesentliche Verbesserung stellt die Vermeidung von Null-Treffermengen dar:

| Register | Anzahl der Null-Treffermengen |
|---------------------|-------------------------------|
| Stichwort-Register | 15 |
| IDX-Register | 3 |
| Schlagwort-Register | 30 |

Die Gründe für die nahezu vollständige Vermeidung von Null-Treffermengen im IDX-Register liegen darin, dass die Anzahl an Deskriptoren, die die einzelnen Titel beschreiben, durch die automatische Indexierung erhöht wurden und somit die Wahrscheinlichkeit, mit einer Suchfrage einen Titel zu finden, steigt. Dagegen kommt es bei der Suche im Schlagwort-Register zu sehr viel mehr Null-Treffermengen, da durch die intellektuelle Erschließung den Titeln zwar sehr präzise, terminologisch kontrollierte Deskriptoren zugewiesen werden, die Art der Zuordnung den Nutzern in der Regel aber nicht bekannt ist und deshalb nicht für das Retrieval eingesetzt werden kann.

Bei der Ergebnisauswertung spielte die Formulierung der einzelnen Suchfragen eine wichtige Rolle. Für die Berechnung von Recall und Precision wurde bei jeder Suchfrage jeweils die Frageformulierung zugrunde gelegt, die für das jeweilige Register das beste Ergebnis erzielte. Die Berechnung der Werte für

die einzelnen Register kann daher auf unterschiedlichen Frageformulierungen beruhen.

| Register | Precision |
|----------------------|-----------|
| Stichwort-Register | 0,82 |
| IDX-Register | 0,75 |
| RSWK-Ketten-Register | 0,95 |

Der Precisionwert für die Suche im IDX-Register ist zwar geringer als der in den beiden anderen Register. Dennoch ist die Precision von 0,75 bemerkenswert, da insgesamt gesehen erheblich mehr Treffer erzielt wurden als über die Suche in den anderen Registern.

4.2 Suche über 'unbehandelte' Titelstichwörter

Unter 'unbehandelten' Titelstichwörtern werden diejenigen Titelstichwörter verstanden, die in der Flektionsform, in der sie im Titel stehen, direkt in ein Register geschrieben und für das Retrieval zur Verfügung gestellt werden.

Zwei Aspekte sind bei der Suche mit Titelstichwörtern verglichen mit der Schlagwortsuche problematisch:

- Die Ergebnisse einer Suche mit Titelstichwörtern ist immer abhängig von der Qualität der Titelformulierungen. Allerdings ist davon auszugehen, dass dieser Aspekt nicht in allen Wissenschaftsgebieten und Literaturgattungen gleichermaßen ins Gewicht fällt. Auch die fachliche Spezialisierung einer Datenbank spielt eine Rolle. So sind in einem OPAC einer Universitätsbibliothek mit sehr heterogenen Beständen schlechtere Ergebnisse zu erzielen als in einer thematisch eng begrenzten Fachdatenbank.
- Titelstichwörter stehen nur in der Form, in der sie im Titel benutzt wurden, für die Suche zur Verfügung, so dass ein Suchbegriff theoretisch in sämtlichen Flektionsformen gesucht werden müsste. Eine mögliche Lösung ist der Einsatz von Trunkierung. Dies ist aber in der Praxis wenig hilfreich, da nur wenige Nutzer diese Funktion anwenden.

Der Test belegt, dass die alleinige Suche über den Titelstichwortindex aufgrund der oben genannten Gründe keine zufrieden stellenden Ergebnisse bringt. Die Ergebnistabelle im Anhang zeigt die Ergebnisse der Suche mit Titelstichwörtern: sie brachte bei jeder Suchfrage weniger Treffer.

4.3 Suche über IDX-Titelstichwörter

Bei den IDX-Titelstichwörtern handelt es sich um Titelstichwörter und Schlagwortketten, die mittels der automatischen Indexierung auf grammatikalische Grundform, Einzelbestandteile etc. zurückgeführt wurden. Aufgrund der Zusammenfassung der beiden Kategorien kann nicht exakt differenziert werden, welche Ergebnisse durch die Schlagwortvergabe und welche durch die automatische Indexierung erzielt wurden. So sind unter den Ergebnissen sowohl Titel zu finden, die aufgrund der Indexierungsergebnisse gefunden wurden als auch solche, die ohne Verschlagwortung auch durch die Indexierung nicht gefunden worden wären.

Auch hier wurde deutlich, wie sehr die Retrievalergebnisse von dem Titelmateriale abhängen. So kam es teilweise aufgrund bestimmter Titelformulierungen zu irreführenden Indexierungsergebnissen. Beispielsweise erbrachte die Frage nach 'Abfallentsorgung und -beseitigung' unter anderem diesen Titel:

```

Bauer, Ingrid:
Glück kann man lernen : Mental-Recycling aktiv / Ingrid & Kurt
Bauer. - Neuhausen: Urania-verl., [1992]. - 181 S.
(Urania-Lebenshilfe)
ISBN 3-908644-80-1
IDX (Titel)      : Abfallbeseitigung <15>; Abfallrecycling <24>;
Ab-
fallverwertung<24>; Abfallwirtschaft <15>;
Lernen
<4>; Mental-Recycling; Müllverwertung <24>;
Recy-
cling; Recycling <23>; Rückgewinnung <24>;
Wiederaufbereitung <24>; Wiedergewinnung
<24>;
aktiv <4>; aktivisch; mental; mental <23>;
IDX (RSWK)      : Berater <25>; Denken; Denkfähigkeit <24>;
Denk-
prozess <24>; Denkvermögen <24>; Denkvorgang
<24>; Führung <3>; Informationsverarbeitung
<15>;
Leben <3>; Lebensführung; Lebensstandard
<25>;
Lebenstechnik <24>; Positiv; Rat <3>; Ratgeber;
positiv
Swk. (RSWK)     : (5.5) Positives Denken / Lebensführung / Rat-
geber

```

Abb. 9: Titelanzeige zur Frage nach 'Abfallentsorgung und -beseitigung'

Durch die automatische Indexierung liegt die Trefferquote höher als bei der konventionellen Titelstichwortsuche. Dies ist zum einen darauf zurückzuführen, dass durch die Einbindung der SWD-Relationen 'Ober-/Unterbegriff' und 'Schlagwort/nicht zugelassenes Schlagwort' zusätzliche Begriffe für das Retrieval zur Verfügung standen. Zum anderen wurden auch solche Titel als Retrievalergebnis ausgegeben, die zwar nicht direkt auf die Fragestellung zutreffen, aber das weitere Umfeld der Frage mit abdeckten. Dies kann vorteilhaft sein, wenn die Suchformulierung der Nutzer nicht genau dem Informationsbedürfnis entspricht oder sie sich noch nicht ganz im Klaren über ihr Informationsbedürfnis sind. Es können Titel ausgegeben werden, die Teilaspekte miteinbeziehen, welche durch den eingegebenen Suchbegriff nicht gefunden worden wären. Ein entscheidender Vorteil bei der Suche über die IDX-Titelstichwörter ist darin zu sehen, dass die Anzahl der Null-Treffermengen deutlich reduziert wird. Null-Treffermengen zu vermeiden stellt eines der wichtigsten Probleme des Information Retrieval dar, da sie dem Nutzer keine Erkenntnis darüber geben, ob der Suchansatz falsch gewählt war oder ob keine relevanten Titel in der Datenbank vorhanden sind.

4.4 Suche über RSWK-Schlagwörter

Die Testergebnisse für die Suche mit RSWK-Schlagwörtern fielen eher negativ aus. Gründe hierfür sind vor allem in den engen Ansetzungsregeln für RSWK-Schlagwörter zu sehen, die die Nutzer nicht kennen und bei der Formulierung von Suchfragen nicht berücksichtigen können. Zudem entsprechen die von der SWD vorgeschriebenen Vorzugsbenennungen oftmals nicht dem Vokabular, das Nutzer für die Formulierung ihrer Fragen verwenden, oder sie geben den Begriff im Plural ein, obwohl er in der Singularform von der SWD benutzt wird. Sinnvoll wäre es daher, den Nutzern bei der Eingabe der Suchbegriffe die zu

dem Suchbegriff dazugehörigen Thesauruseinträge zur Präzisierung oder Erweiterung der Suchfrage anzubieten.

Positiv ist festzustellen, dass die Precision-Werte der mit den RSWK-Schlagwörtern erzielten Retrievalergebnisse teilweise höher als bei den IDX-Retrievalergebnissen waren, da die über eine Suche im Schlagwortregister gefundenen Titel überwiegend relevant waren. Dies konnte allerdings nur bei exakter Übereinstimmung (Matching) zwischen Suchbegriff und RSWK-Schlagwort erreicht werden.

Ein generelles Problem der RSWK-Schlagwörter liegt darin, dass nicht sämtliche Titel einer Datenbank erschlossen werden. Dadurch ist es nicht möglich, vollständige Suchergebnisse zu erzielen. Ist jedoch eine möglichst spezifische Erschließung einer Datenbank angestrebt, ist die intellektuelle Indexierung nach wie vor unumgänglich.

Problematisch für den Einsatz im Rahmen von automatischer Indexierung gestalten sich die in den RSWK-Ketten gebrauchten Forms Schlagwörter, da für jedes Forms Schlagwort die zugehörigen SWD-Relationen bei der automatischen Indexierung erzeugt werden. Beispielsweise ist 'Literaturverzeichnis' das Synonym zu 'Bibliographie' und wird bei jedem Vorkommen von 'Bibliographie' zusätzlich als Deskriptor ausgegeben.

4.5 Vergleich der Retrievalergebnisse von IDX-Stichwörtern und RSWK-Schlagwörtern

Bei vielen Suchfragen ist an den unterschiedlichen Ergebnismengen zu erkennen, dass durch die Suche mit IDX-Stichwörtern oftmals insgesamt gesehen mehr relevante Dokumente gefunden wurden als über die intellektuelle Erschließung durch RSWK-Schlagwörter. Dies resultiert daraus, dass bei der automatischen Indexierung das Suchvokabular durch die Einbeziehung der SWD-Relationen thesaurusartig angereichert wird. Bei den so erzielten Treffern handelt es sich jedoch häufig um Treffer, die aufgrund der Relation Schlagwort - Unterbegriff gefunden wurden, wodurch zwar der Recall erhöht, die Precision aber verringert wird. Ein Beispiel hierfür ist die Frage nach 'Frau und Beruf'. Über die Suche mit IDX-Stichwörtern wurden 234 Treffer erzielt, von denen 118 als relevant gewertet wurden. Dahingegen wurde über die Suche mit RSWK-Schlagwörtern nur elf Titel gefunden, von denen jeder relevant war.

Suchen nach sehr speziellen Sachverhalten sind über die IDX-Stichwörter schwerer durchzuführen als über die Suche mit RSWK-Schlagwörtern. Am Beispiel der Frage nach 'Energiewirtschaft' ist aber zu sehen, dass durch die Suche mit IDX-Stichwörtern durchaus auch Darstellungen von Einzelaspekten gefunden werden. Dies ist insbesondere deshalb hilfreich, weil Nutzer in der Annahme, dass ihr Thema so speziell ist, dass darüber nichts in den Beständen einer Bibliothek zu finden ist, häufig dazu tendieren, Suchfragen eher weit und unspezifisch zu formulieren.

4.6 Suche über RSWK-Ketten

Suchfragen über das Register der RSWK-Schlagwortketten zu lösen erwies sich als nicht einfach, da die Suche über dieses Register generell anders ablaufen muss als über die restlichen Register. Verknüpfungen innerhalb der einzelnen Schlagwortketten sind nur schwer möglich. Gezieltes Suchen nach bestimmten Ketten ist kaum durchführbar, da im Vorhinein nicht bekannt ist, welche Ketten gebildet wurden, um einen Dokumentinhalt zu beschreiben. Daher müssen Suchfragen durch Browsing im Register beantwortet werden.

Das RSWK-Kettenregister besteht aus terminologisch kontrolliertem Vokabular und im Gegensatz zu dem konventionellen Titelstichwort- und dem IDX-Stichwortregister wird der Inhalt eines Titels sehr präzise wiedergegeben. Gleichzeitig entspricht aber das dafür verwendete Vokabular nicht unbedingt dem Suchvokabular eines Nutzers. Da die Testfragen also nicht die genau den Registereinträgen entsprechenden Suchbegriffe enthielten, kam es häufig zu Null-Treffermengen. Sinnvoll wäre deshalb eine Verweisung vom Nichtdeskriptor auf den von der SWD vergebenen Hauptdeskriptor.

RSWK-Ketten können hilfreich sein für den Fall, dass ein Nutzer eine nur vage Vorstellung vom gesuchten Thema hat und durch Browsing innerhalb des Registers das Thema enger eingrenzen kann.

Sollen RSWK-Ketten für das Retrieval stärker eingesetzt werden, ist es nötig, den Nutzern die Syntax der Ketten zu verdeutlichen. Aber auch dann bleibt es fraglich, ob RSWK-Ketten gezielt eingesetzt werden können, da thematische Suchen eher über eine Benutzeroberfläche mit verschiedenen Eingabefeldern durchgeführt werden als über das Browsing in Registern. Interessant wäre es zu untersuchen, inwieweit die über Listen zugänglichen Kettenregister von Nutzern für ihre Suche in anderen DDB-Datenbanken genutzt werden.

4.7 Bestätigung der Ergebnisse durch Untersuchung des DDB-Altdateipools

Im Anschluss an den Retrievaltest über die DNB-Datenbank von 1991 bis 1995 wurde zusätzlich der sogenannte Altdateipool der DDB von 1945-1965 mit 377.336 Datensätzen indexiert und getestet. Für die automatische Indexierung wurden auch hierbei bestimmte SWD-Relationen hinzugezogen.

Die Besonderheit dieses Datenbestandes liegt darin, dass er in keiner Weise intellektuell sachlich erschlossen wurde, das Retrieval bisher also nur auf Basis einer Titelstichwortsuche möglich ist. Daher sollte durch den Retrievaltest festgestellt werden, inwieweit die automatische Indexierung hier eine sinnvolle Ergänzungsfunktion haben kann. Von besonderem Interesse war das Ergebnis der Indexierung unter Integration der SWD.

Ausgangsbasis waren die 100 Fragen des zuvor durchgeführten Tests. Zwangsläufig mussten aus diesem Fragenkatalog aufgrund des zeitlich eingeschränkten Themenumfangs im Altdateipool einige Fragen herausgenommen werden. Beispielsweise konnten zum Thema 'Internet' oder 'Software zur Textverarbeitung' noch keine Titel enthalten sein. Der Fragenkatalog umfasste schließlich noch 70 Fragen. In der folgenden Abbildung sind die Ergebnisse der Suchen in den einzelnen Registern abgebildet.

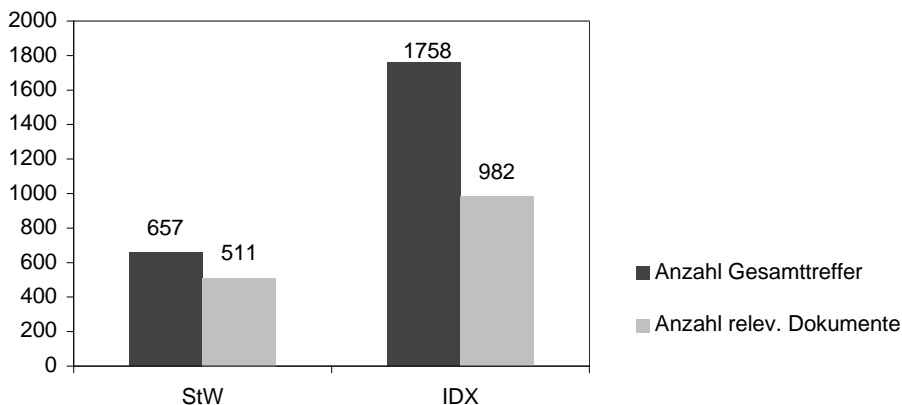


Abb. 10: Anzahl Gesamttreffer / Anzahl relevanter Dokumente

Durch die Suche im IDX-Register wurden insgesamt die 2,7-fache Anzahl an Titeln gefunden, unter denen fast doppelt so viele relevante Treffer waren wie unter den Treffern durch die Suche im Stichwort-Register (92% mehr). Die Precision-Werte (Mittelwert der einzelnen Precision-Werte über alle Suchanfragen) liegen, ähnlich wie im Haupttest zu MILOS II, im Stichwort-Register bei 0,79 und im IDX-Register bei 0,74. Es zeigt sich, dass ein gewisses Maß an Ballast bei der Suche im IDX-Register nicht zu vermeiden ist.

5 Zusammenfassung der Ergebnisse und Empfehlungen

Durch den Einsatz der automatischen Indexierung entstehen deutlich bessere Retrievalmöglichkeiten, da das für das Retrieval zur Verfügung stehende, die einzelnen Titel beschreibende Vokabular sowohl durch Funktionalitäten wie Grundformermittlung, Kompositazerlegung, Derivation etc. als auch durch die Einbeziehung der SWD-Relationen erweitert wird. Die Retrievalergebnisse ergeben durchweg einen erhöhten Recall bei nur unwesentlich gesunkener Precision. Insgesamt gesehen wird eine erhöhte Anzahl an relevanten Titeln gefunden.

Angemessen in der Praxis eingesetzt werden können die Vorteile der automatischen Indexierung aber erst, wenn auch die entsprechenden Voraussetzungen auf Seite der Retrievalsoftware geschaffen werden.

Ungünstige Auswirkungen auf die Retrievalergebnisse entstehen durch Homonyme und Polyseme. Da die automatische Indexierung wörterbuch- und nicht konzeptorientiert arbeitet, werden sämtliche Bedeutungsvarianten zu den gesuchten Begriffen als Ergebnisse mit ausgegeben. Vermieden werden kann dies durch entsprechende Funktionalitäten bereits bei der Sucheingabe, indem bei mehrdeutigen Begriffen unter den verschiedenen Bedeutungen die gewünschte ausgewählt werden muss.

Es sei allerdings im Zusammenhang solcher durch die automatische Indexierung erzeugter Fehldeskriptoren nochmal darauf hingewiesen, dass eine Beurteilung der automatischen Indexierung nicht aufgrund der einzelnen den Dokumenten zugeteilten Deskriptoren vorgenommen werden kann, sondern dass vielmehr das Gesamtretrievalergebnis zur Beurteilung herangezogen werden muss.

Der Retrievaltest hat gezeigt, dass mit dem MILOS-Verfahren ein einsatzberechtigtes und positiv zu bewertendes Indexierungssystem für den Einsatz in der

Praxis zur Verfügung steht. Auf Erschließungsseite sind demnach weitgehende Verbesserungsansätze entwickelt worden. Wichtig ist es nun, die durch den Einsatz von MILOS sich bietenden Möglichkeiten auch auf der Retrievalseite zu unterstützen. Hierzu lassen sich einige Prinzipien der Gestaltung von Oberflächen von OPACs nennen, die bei der Konzeption eine Rolle spielen sollten:

- Nutzerunterstützung hinsichtlich des Verständnisses der Suchoperationen und des Suchprozesses:
 - durch Bereitstellen von Kategorien, die den Nutzer bei der Frageformulierung unterstützen
 - durch übersichtliche Gestaltung der Oberfläche (wichtige Funktionen sollten besonders auffallen) hinsichtlich Anzahl, Größe und Form der 'Icons'
 - durch Vermeiden von zu viel Navigation auf dem Suchbildschirm, beispielsweise um Suchbegriffe anzeigen zu lassen oder hinzuzufügen
- Nutzer sollten darüber informiert werden, ob sich das Fortsetzen des Suchprozesses lohnt, indem z.B. alphabetische oder hierarchisch gegliederte Listen der Suchbegriffe angeboten werden oder indem die Anzahl an Dokumenten, die mit einem Suchbegriff zusammenhängen, gezeigt wird; die Trefferlisten sollten klar strukturiert sein, und die Suchanfrage sollte sichtbar bleiben, während die Trefferlisten durchgesehen werden
- Nutzer sollten bei der Präzisierung der Suchfrage oder des Informationsbedürfnisses unterstützt werden, durch klare Anweisungen, Hilfeseiten, Anzeige der Suchgeschichte oder als Antwort eine Liste von verwandten Suchthemen
- die Aktionen des Systems sollten explizit erläutert werden und für den Nutzer nachvollziehbar sein
- Icons sollten mit eindeutigen, verbalen Bezeichnungen versehen sein⁵².

Auf der Retrievalebene sind folgende zusätzliche Funktionalitäten für einen zukünftigen Einsatz sinnvoll:

- Ranking
- Aufforderung seitens des Systems, die Suche zu spezifizieren, sobald eine bestimmte Trefferanzahl überschritten ist; Vorgabe von konkreten Vorschlägen (z.B. Unter-/Oberbegriffe)
- system-interne automatische Trunkierung der Suchbegriffe
- Thesaurus an der Stelle des Suchbegriffs aufschlagen, weitere Begriffe auswählbar machen
- Homonyme / Polyseme anzeigen und auswählbar machen.

52 Denning, R., M. Shuttleworth u. P. Smith: Interface design concepts in the development of a web-based information retrieval system. In: Bulletin of the American Society for Information Science. (1998) April/May, S. 17-20.

6 Anhang

6.1 Tabelle 1: Suchfragen mit Ergebnissen⁵³

| Suchformulierung | StW-Register | | IDX-Register | | SW-Register | | Basic-Index | | Precision | | | |
|---|--------------|------------|--------------|------------|-------------|------------|-------------|------------|-----------|-----------|----------|-------------|
| | gef. Titel | rel. Titel | gef. Titel | rel. Titel | gef. Titel | rel. Titel | gef. Titel | rel. Titel | STW-Index | IDX-Index | SW-Index | Basic-Index |
| <i>Wachstum + Wirtschaft</i> | 1 | 0 | 52 | 23 | 0 | 0 | 52 | 23 | 0,00 | 0,44 | | 0,44 |
| <i>„Kritische Theorie“</i> | 0 | 0 | 0 | 0 | 41 | 41 | 0 | 0 | | | 1,00 | |
| <i>kritische + Theorie</i> | 26 | 17 | 0 | 0 | 41 | 41 | 72 | 30 | 0,65 | | | |
| <i>Kritisch + Theorie</i> | 0 | 0 | 102 | 48 | 0 | 0 | 104 | 48 | | 0,47 | | 0,46 |
| <i>Ökologie + Gewässer</i> | 0 | 0 | 30 | 20 | 1 | 1 | 30 | 20 | | 0,67 | 1,00 | 0,67 |
| <i>Hemmung + Enzymen</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| <i>Hemmung + Enzym</i> | 0 | 0 | 6 | 5 | 0 | 0 | 6 | 5 | | | | |
| <i>Hemmung + Enzyme</i> | 0 | 0 | 6 | 5 | 0 | 0 | 6 | 5 | | 0,83 | | 0,83 |
| <i>Gerontologie</i> | 9 | 8 | 37 | 28 | 8 | 7 | 37 | 28 | 0,89 | 0,76 | 0,88 | 0,76 |
| <i>Frau + Beruf</i> | 7 | 5 | 234 | 118 | 11 | 11 | 236 | 118 | 0,71 | 0,50 | 1,00 | 0,50 |
| <i>Wirtschaftswunder + Deutschland</i> | 0 | 0 | 4 | 4 | 0 | 0 | 4 | 4 | | 1,00 | | 1,00 |
| <i>Kunst + Antike</i> | 7 | 2 | 48 | 7 | 1 | 1 | 48 | 7 | 0,29 | 0,15 | 1,00 | 0,15 |
| <i>Angst/Angststörungen/Angsterkrankungen</i> | 291 | 207 | 0 | 0 | 0 | 0 | 459 | 0 | | | | |
| <i>Angst/Angststörung/Angsterkrankung</i> | 290 | 207 | 455 | 284 | 90 | 82 | 459 | 284 | 0,71 | 0,62 | 0,91 | 0,62 |
| <i>Datenbanken + Dokumentation</i> | 2 | 2 | 0 | 0 | 0 | 0 | 7 | 6 | 1,00 | | | 0,86 |
| <i>Datenbank + Dokumentation</i> | 2 | 1 | 13 | 9 | 1 | 0 | 13 | 9 | | 0,69 | 0,00 | |
| <i>Spurenelemente + Ernährung</i> | 2 | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 1,00 | | | |
| <i>Spurenelement + Ernährung</i> | 0 | 0 | 3 | 3 | 1 | 1 | 3 | 3 | | 1,00 | 1,00 | 1,00 |
| <i>Festkörperphysik</i> | 14 | 14 | 20 | 20 | 18 | 18 | 20 | 20 | 1,00 | 1,00 | 1,00 | 1,00 |
| <i>Tierexperimente</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| <i>Tierexperiment</i> | 1 | 0 | 38 | 20 | 0 | 0 | 38 | 20 | 0,00 | 0,53 | | 0,53 |
| <i>„Psychologische Diagnostik“</i> | 0 | 0 | 31 | 6 | 22 | 22 | 0 | 0 | | | 1,00 | |
| <i>Psychologische + Diagnostik</i> | 10 | 5 | 38 | 9 | 22 | 22 | 64 | 27 | 0,50 | | | |
| <i>psychologisch + Diagnostik</i> | 0 | 0 | 46 | 31 | 0 | 0 | 46 | 31 | | 0,67 | | 0,67 |
| <i>„Interkulturelle Erziehung“</i> | 0 | 0 | 0 | 0 | 53 | 52 | 0 | 0 | | | 0,98 | |
| <i>Interkulturelle + Erziehung</i> | 12 | 10 | 0 | 0 | 53 | 52 | 58 | 55 | 0,83 | | | 0,95 |
| <i>Interkulturell + Erziehung</i> | 1 | 1 | 59 | 55 | 0 | 0 | 59 | 55 | | 0,93 | | |
| <i>Widerstand + Nationalsozialismus</i> | 12 | 12 | 124 | 109 | 112 | 98 | 124 | 109 | 1,00 | 0,88 | 0,88 | 0,88 |
| <i>Erkrankungen + Lunge</i> | 3 | 3 | 0 | 0 | 0 | 0 | 4 | 3 | 1,00 | | | 0,75 |
| <i>Erkrankung + Lunge</i> | 0 | 0 | 14 | 13 | 0 | 0 | 15 | 11 | | 0,93 | | |
| <i>„Psychosomatische Krankheiten“</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| <i>Psychosomatische Krankheit</i> | 0 | 0 | 1 | 0 | 141 | 141 | 0 | 0 | | | 1,00 | |
| <i>Psychosomatische + Krankheiten</i> | 2 | 2 | 0 | 0 | 0 | 0 | 9 | 9 | | | | 1,00 |
| <i>Psychosomatische + Krankheit</i> | 3 | 3 | 1 | 1 | 143 | 141 | 151 | 142 | 1,00 | 1,00 | | |

⁵³ Sofern eine Suchfrage in verschiedenen Frageformulierungen gesucht wurde, sind alle Varianten aufgeführt. Eine Phrasensuche ist durch Anführungszeichen gekennzeichnet; eine 'und'-Verknüpfung wird durch das Pluszeichen dargestellt. Im Anschluß an die Tabelle (S. 42) sind in der Legende die Inhalte der einzelnen Spalten erläutert.

| Suchformulierung | StW-Register | | IDX-Register | | SW-Register | | Basic-Index | | Precision | | | |
|--|--------------|------------|--------------|------------|-------------|------------|-------------|------------|-----------|-----------|----------|-------------|
| | gef. Titel. | rel. Titel | gef. Titel | rel. Titel | gef. Titel | rel. Titel | gef. Titel | rel. Titel | STW-Index | IDX-Index | SW-Index | Basic-Index |
| <i>Konjunkturpolitik</i> | 7 | 7 | 23 | 15 | 11 | 11 | 23 | 15 | 1,00 | 0,65 | 1,00 | 0,65 |
| <i>Wissensorganisation</i> | 1 | 1 | 7 | 7 | 6 | 6 | 7 | 7 | 1,00 | 1,00 | 1,00 | 1,00 |
| <i>Umweltökonomie / Umweltökonomik</i> | 18 | 18 | 48 | 46 | 46 | 44 | 48 | 46 | 1,00 | 0,96 | 0,96 | 0,96 |
| <i>Abfallentsorgung / Abfallbeseitigung</i> | 30 | 19 | 225 | 64 | 37 | 37 | 227 | 65 | 0,63 | 0,28 | 1,00 | 0,29 |
| <i>Shakespeares Dramen</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| <i>Shakespeares + Dramen</i> | 4 | 3 | 0 | 0 | 0 | 0 | 9 | 7 | 0,75 | | | |
| <i>Shakespeare + Dramen</i> | 0 | 0 | 12 | 10 | 0 | 0 | 12 | 10 | | | | |
| <i>Shakespeare + Drama</i> | 0 | 0 | 15 | 13 | 10 | 10 | 15 | 13 | | 0,87 | 1,00 | 0,87 |
| <i>Energiewirtschaft</i> | 16 | 10 | 79 | 41 | 27 | 27 | 79 | 41 | 0,63 | 0,52 | 1,00 | 0,52 |
| <i>Sprachwandel</i> | 15 | 15 | 45 | 43 | 39 | 38 | 45 | 43 | 1,00 | 0,96 | 0,97 | 0,96 |
| <i>Betriebspsychologie</i> | 5 | 5 | 9 | 9 | 8 | 8 | 9 | 9 | 1,00 | 1,00 | 1,00 | 1,00 |
| <i>Eherecht</i> | 8 | 5 | 64 | 22 | 20 | 20 | 64 | 22 | 0,63 | 0,34 | 1,00 | 0,34 |
| <i>Bundestagswahl</i> | 11 | 8 | 21 | 19 | 15 | 15 | 21 | 19 | 0,73 | 0,90 | 1,00 | 0,90 |
| <i>Scheidung + Kinder</i> | 6 | 4 | 5 | 2 | 0 | 0 | 20 | 13 | | | | 0,65 |
| <i>Scheidung + Kind</i> | 2 | 2 | 27 | 8 | 0 | 0 | 27 | 8 | 1,00 | 0,30 | | |
| <i>Kindesmissbrauch / Kindesmisshandlungen</i> | 6 | 6 | 6 | 6 | 0 | 0 | 34 | 34 | 1,00 | | | 1,00 |
| <i>Kindesmissbrauch / Kindesmisshandlung</i> | 8 | 8 | 34 | 34 | 26 | 26 | 34 | 34 | | 1,00 | 1,00 | |
| <i>Marktanalyse</i> | 6 | 0 | 14 | 12 | 6 | 6 | 14 | 12 | 0,00 | 0,86 | 1,00 | 0,86 |
| <i>Philosophie + Gegenwart</i> | 17 | 3 | 37 | 7 | 0 | 0 | 37 | 7 | 0,18 | 0,19 | | 0,19 |
| <i>Altersversorgung</i> | 44 | 41 | 166 | 103 | 78 | 78 | 166 | 103 | 0,93 | 0,62 | 1,00 | 0,62 |
| <i>Ergonomie</i> | 26 | 25 | 57 | 54 | 28 | 28 | 57 | 54 | 0,96 | 0,95 | 1,00 | 0,95 |
| <i>„Alternative Energiequellen“</i> | 0 | 0 | 54 | 0 | 0 | 0 | 0 | 0 | | | | |
| <i>Alternative Energiequelle</i> | 0 | 0 | 54 | 0 | 35 | 35 | 0 | 0 | | | 1,00 | |
| <i>Alternative + Energiequellen</i> | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 7 | | | | 1,00 |
| <i>Alternative + Energiequelle</i> | 0 | 0 | 54 | 0 | 0 | 0 | 88 | 88 | | 0,00 | | |
| <i>Wirtschaftsethik</i> | 30 | 30 | 140 | 95 | 92 | 92 | 141 | 95 | 1,00 | 0,68 | 1,00 | 0,67 |
| <i>Strahlenschutz + Röntgen</i> | 1 | 1 | 12 | 11 | 0 | 0 | 12 | 11 | 1,00 | 0,92 | | 0,92 |
| <i>Vergesslichkeit / Gedächtnistraining</i> | 19 | 19 | 54 | 54 | 0 | 0 | 54 | 54 | 1,00 | 1,00 | | 1,00 |
| <i>Risk-Management</i> | 3 | 3 | 2 | 1 | 0 | 0 | 4 | 4 | 1,00 | 0,50 | | 1,00 |
| <i>Risk + Management</i> | 6 | 6 | 69 | 69 | 0 | 0 | 70 | 70 | | 1,00 | | |
| <i>Hyperaktivität + Kindern</i> | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 4 | | | | 1,00 |
| <i>Hyperaktivität + Kinder</i> | 2 | 1 | 4 | 4 | 0 | 0 | 18 | 18 | | | | |
| <i>Hyperaktivität + Kind</i> | 2 | 2 | 32 | 32 | 0 | 0 | 32 | 32 | 1,00 | 1,00 | | |
| <i>„Zeugen Jehovas“</i> | 0 | 0 | 0 | 0 | 14 | 14 | 0 | 0 | | | 1,00 | |
| <i>Zeugen + Jehovas</i> | 15 | 15 | 0 | 0 | 14 | 14 | 19 | 19 | 1,00 | | | 1,00 |
| <i>Zeuge+ Jehova</i> | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | | 1,00 | | |
| <i>Aromatherapie</i> | 48 | 46 | 68 | 66 | 66 | 65 | 69 | 66 | 0,96 | 0,97 | 0,98 | 0,96 |
| <i>Direktmarketing</i> | 16 | 15 | 55 | 31 | 28 | 27 | 55 | 31 | 0,94 | 0,56 | 0,96 | 0,56 |
| <i>Behandlung + Schlaganfälle</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| <i>Behandlung + Schlaganfall</i> | 1 | 1 | 7 | 5 | 1 | 1 | 8 | 8 | 1,00 | 0,71 | 1,00 | 1,00 |

| Suchformulierung | StW-Register | | IDX-Register | | SW-Register | | Basic-Index | | Precision | | | |
|---|--------------|------------|--------------|------------|-------------|------------|-------------|------------|-----------|-----------|----------|-------------|
| | gef. Titel. | rel. Titel | gef. Titel | rel. Titel | gef. Titel | rel. Titel | gef. Titel | rel. Titel | STW-Index | IDX-Index | SW-Index | Basic-Index |
| <i>Einwanderer + Amerika</i> | 0 | 0 | 4 | 4 | 1 | 1 | 4 | 4 | | 1,00 | 1,00 | 1,00 |
| <i>Zimmerpflanzen</i> | 43 | 37 | 0 | 0 | 70 | 61 | 81 | 81 | | | 0,87 | 1,00 |
| <i>Zimmerpflanze</i> | 1 | 1 | 87 | 65 | 0 | 0 | 87 | 65 | 1,00 | 0,75 | | |
| <i>Nomaden</i> | 15 | 11 | 0 | 0 | 0 | 0 | 15 | 11 | 0,73 | | | 0,73 |
| <i>Nomade</i> | 1 | 0 | 29 | 21 | 7 | 6 | 29 | 21 | 0,00 | 0,72 | 0,86 | |
| <i>Reiseführer + Toskana</i> | 1 | 0 | 4 | 4 | 0 | 0 | 5 | 4 | 0,00 | 1,00 | | |
| <i>Reiseführer + Toscana</i> | 0 | 0 | 2 | 2 | 0 | 0 | 2 | 2 | | | | 1,00 |
| <i>Baby-Massage</i> | 4 | 4 | 4 | 4 | 0 | 0 | 4 | 4 | 1,00 | | | 1,00 |
| <i>Baby+Massage</i> | 2 | 2 | 11 | 11 | 0 | 0 | 11 | 11 | | 1,00 | | |
| <i>„Lateinamerikanische Tänze“</i> | 0 | 0 | 4 | 3 | 0 | 0 | 0 | 0 | | 0,75 | | |
| <i>Lateinamerikanische + Tänze</i> | 3 | 3 | 4 | 3 | 0 | 0 | 6 | 3 | 1,00 | | | |
| <i>Lateinamerikanisch + Tanz</i> | 0 | 0 | 3 | 3 | 0 | 0 | 3 | 3 | | | | 1,00 |
| <i>Lateinamerika +Tanz</i> | 0 | 0 | 2 | 1 | 0 | 0 | 2 | 1 | | | | |
| <i>Homebanking</i> | 3 | 3 | 3 | 3 | 0 | 0 | 3 | 3 | 1,00 | 1,00 | | 1,00 |
| <i>Qumranrollen</i> | 4 | 4 | 0 | 0 | 0 | 0 | 4 | 0 | | | | |
| <i>Qumran-Rollen</i> | 4 | 4 | 0 | 0 | 0 | 0 | 4 | 0 | 1,00 | | | |
| <i>Qumranrolle</i> | 0 | 0 | 4 | 4 | 0 | 0 | 4 | 4 | | 1,00 | | 1,00 |
| <i>Gesprächsführung</i> | 28 | 20 | 115 | 89 | 93 | 93 | 115 | 89 | 0,71 | 0,77 | 1,00 | 0,77 |
| <i>Italienische + Kochbücher</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| <i>Italienisch + Kochbuch</i> | 0 | 0 | 26 | 26 | 0 | 0 | 26 | 26 | | 1,00 | | 1,00 |
| <i>Italien + Kochbuch</i> | 0 | 0 | 57 | 55 | 52 | 52 | 57 | 55 | | 0,96 | 1,00 | |
| <i>Antriebstechnik</i> | 15 | 15 | 63 | 62 | 0 | 0 | 64 | 63 | 1,00 | 0,98 | | 0,98 |
| <i>„Islamischer Fundamentalismus“</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| <i>Islamischer + Fundamentalismus</i> | 3 | 3 | 0 | 0 | 0 | 0 | 3 | 3 | 1,00 | | | 1,00 |
| <i>Islamisch + Fundamentalismus</i> | 0 | 0 | 34 | 26 | 0 | 0 | 34 | 26 | | | | |
| <i>Islam + Fundamentalismus</i> | 5 | 5 | 34 | 26 | 29 | 22 | 34 | 26 | | 0,76 | 0,76 | |
| <i>Geschichte + Israels</i> | 8 | 7 | 0 | 0 | 0 | 0 | 17 | 11 | | | | 0,65 |
| <i>Geschichte + Israel</i> | 11 | 6 | 76 | 34 | 57 | 27 | 77 | 34 | | 0,45 | 0,47 | |
| <i>Lichtbogenschweißen</i> | 4 | 4 | 15 | 13 | 11 | 11 | 16 | 13 | 1,00 | 0,87 | 1,00 | 0,81 |
| <i>Einführungen + Volkswirtschaftslehre</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| <i>Einführung + Volkswirtschaftslehre</i> | 44 | 42 | 95 | 56 | 18 | 18 | 95 | 56 | | 0,59 | 1,00 | 0,59 |
| <i>Heilfasten</i> | 12 | 12 | 34 | 34 | 34 | 34 | 34 | 34 | 1,00 | 1,00 | 1,00 | 1,00 |
| <i>„Excel 5.0“</i> | 0 | 0 | 0 | 0 | 74 | 74 | 0 | 0 | | | 1,00 | |
| <i>Excel + 5.0</i> | 40 | 37 | 0 | 0 | 74 | 74 | 79 | 77 | 0,93 | | | 0,97 |
| <i>Elektronik + Kraftfahrzeug</i> | 2 | 2 | 2 | 2 | 8 | 6 | 8 | 6 | 1,00 | 1,00 | 0,75 | 0,75 |
| <i>Homöopathische + Mittel</i> | 5 | 3 | 0 | 0 | 0 | 0 | 43 | 43 | | | | 1,00 |
| <i>Homöopathisch + Mittel</i> | 0 | 0 | 145 | 145 | 0 | 0 | 145 | 145 | | 1,00 | | |
| <i>Homöopathie + Mittel</i> | 1 | 1 | 64 | 64 | 0 | 0 | 65 | 65 | 1,00 | | | |
| <i>„Künstliche Intelligenz“</i> | 0 | 0 | 0 | 0 | 90 | 90 | 0 | 0 | | | 1,00 | |
| <i>Künstliche + Intelligenz</i> | 35 | 35 | 0 | 0 | 93 | 93 | 101 | 101 | 1,00 | | | 1,00 |
| <i>Künstlich + Intelligenz</i> | 0 | 0 | 119 | 119 | 0 | 0 | 119 | 119 | | 1,00 | | |
| <i>Karate</i> | 29 | 28 | 38 | 33 | 20 | 20 | 39 | 33 | 0,97 | 0,87 | 1,00 | 0,85 |

| Suchformulierung | StW-Register | | IDX-Register | | SW-Register | | Basic-Index | | Precision | | | |
|---|--------------|------------|--------------|------------|-------------|------------|-------------|------------|-----------|-----------|----------|-------------|
| | gef. Titel. | rel. Titel | gef. Titel | rel. Titel | gef. Titel | rel. Titel | gef. Titel | rel. Titel | STW-Index | IDX-Index | SW-Index | Basic-Index |
| Außenpolitik + USA | 3 | 2 | 41 | 34 | 35 | 35 | 41 | 34 | 0,67 | 0,83 | 1,00 | 0,83 |
| Internet | 21 | 21 | 23 | 23 | 22 | 22 | 23 | 23 | 1,00 | 1,00 | 1,00 | 1,00 |
| Alarmanlagen + Auto | 3 | 3 | 0 | 0 | 0 | 0 | 4 | 4 | 1,00 | | | 1,00 |
| Alarmanlage + Auto | 0 | 0 | 4 | 4 | 0 | 0 | 4 | 4 | | 1,00 | | |
| Herstellung + (Zeitschriften / Zeitungen) | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | | | | 1,00 |
| Herstellung + (Zeitschrift / Zeitung) | 6 | 3 | 9 | 4 | 0 | 0 | 10 | 4 | 0,50 | 0,44 | | |
| Kurden | 7 | 7 | 0 | 0 | 13 | 13 | 14 | 13 | 1,00 | | 1,00 | |
| Kurde | 0 | 0 | 14 | 14 | 0 | 0 | 14 | 14 | | 1,00 | | 1,00 |
| Reaktorunfall / Tschernobyl | 13 | 11 | 19 | 17 | 14 | 12 | 19 | 17 | 0,85 | 0,89 | 0,86 | 0,89 |
| Chaostheorie | 4 | 4 | 81 | 76 | 74 | 74 | 82 | 75 | 1,00 | 0,94 | 1,00 | 0,91 |
| Pflanzen + Aquarium | 16 | 16 | 19 | 19 | 0 | 0 | 19 | 19 | 1,00 | 1,00 | | 1,00 |
| Pflanze + Aquarium | 0 | 0 | 21 | 21 | 0 | 0 | 21 | 21 | | | | |
| „Weiße Rose“ | 0 | 0 | 0 | 0 | 22 | 22 | 0 | 0 | | | 1,00 | |
| Weißer + Rose | 16 | 15 | 0 | 0 | 22 | 22 | 24 | 23 | 0,94 | | | 0,96 |
| Magersucht | 22 | 22 | 20 | 20 | 0 | 0 | 25 | 20 | 1,00 | 1,00 | | 0,80 |
| Geschichte + Mittelmeerraumes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| Geschichte + Mittelmeerraum | 0 | 0 | 9 | 6 | 8 | 8 | 9 | 6 | | 0,67 | 1,00 | 0,67 |
| Ying / Yang | 17 | 8 | 24 | 13 | 10 | 9 | 29 | 17 | 0,47 | 0,54 | 0,90 | 0,59 |
| Selbstbewusstsein + stärken | 7 | 7 | 0 | 0 | 0 | 0 | 7 | 7 | 1,00 | | | 1,00 |
| „Wissenschaftliches Arbeiten“ | 0 | 0 | 0 | 0 | 32 | 32 | 0 | 0 | | | 1,00 | |
| Wissenschaftliches + Arbeiten | 14 | 11 | 0 | 0 | 32 | 32 | 50 | 49 | 0,79 | | | 0,98 |
| wissenschaftlich + Arbeiten | 3 | 2 | 0 | 0 | 0 | 0 | 56 | 12 | | | | |
| wissenschaftlich + Arbeit | 0 | 0 | 126 | 31 | 0 | 0 | 126 | 31 | | 0,25 | | |
| Krankheit + Stress | 0 | 0 | 16 | 13 | 5 | 4 | 16 | 13 | | 0,81 | 0,80 | 0,81 |
| Umweltgifte | 6 | 6 | 0 | 0 | 3 | 3 | 6 | 6 | 1,00 | | 1,00 | 1,00 |
| Umweltgift | 0 | 0 | 10 | 7 | 0 | 0 | 10 | 7 | | | | |
| Sexualerziehung + Schule | 0 | 0 | 13 | 4 | 0 | 0 | 13 | 4 | | 0,31 | | 0,31 |
| „Pluralistische Gesellschaft“ | 0 | 0 | 0 | 0 | 17 | 17 | 0 | 0 | | | | |
| Pluralistische + Gesellschaft | 0 | 0 | 0 | 0 | 17 | 17 | 17 | 17 | | | 1,00 | |
| pluralistisch + Gesellschaft | 0 | 0 | 19 | 19 | 0 | 0 | 19 | 19 | | 1,00 | | |
| Pluralismus + Gesellschaft | 1 | 1 | 7 | 7 | 2 | 2 | 7 | 7 | 1,00 | | | 1,00 |
| Politischer Skandal | 0 | 0 | 0 | 0 | 13 | 13 | 0 | 0 | | | | |
| Politischer + Skandal | 0 | 0 | 1 | 0 | 13 | 13 | 14 | 13 | | | 1,00 | 0,93 |
| Politik + Skandal | 0 | 0 | 14 | 13 | 0 | 0 | 14 | 13 | | 0,93 | | |
| Studium + Ausland | 0 | 0 | 26 | 12 | 0 | 0 | 26 | 12 | | 0,46 | | 0,46 |
| Koedukation + Schule | 4 | 4 | 5 | 5 | 1 | 1 | 5 | 5 | 1,00 | 1,00 | 1,00 | 1,00 |
| Anleitung + Videofilm | 1 | 0 | 8 | 8 | 8 | 8 | 8 | 8 | 0,00 | 1,00 | 1,00 | 1,00 |
| (Entstehung / Entwicklung) + Schrift | 2 | 2 | 34 | 7 | 0 | 0 | 34 | 7 | 1,00 | 0,21 | | 0,21 |
| Denken + Lernen | 15 | 14 | 22 | 15 | 1 | 1 | 38 | 26 | 0,93 | 0,68 | 1,00 | 0,68 |
| Software + Textverarbeitung | 0 | 0 | 19 | 18 | 0 | 0 | 19 | 18 | | 0,95 | | 0,95 |
| Geschichte + Plastik | 1 | 1 | 83 | 70 | 35 | 35 | 83 | 70 | 1,00 | 0,84 | 1,00 | 0,84 |
| Individuum+Gesellschaft | 11 | 7 | 27 | 22 | 12 | 12 | 28 | 22 | 0,64 | 0,81 | 1,00 | 0,79 |

| Suchformulierung | StW-Register | | IDX-Register | | SW-Register | | Basic-Index | | Precision | | | |
|----------------------------------|--------------|------------|--------------|------------|-------------|------------|-------------|------------|-----------|-----------|----------|-------------|
| | gef. Titel. | rel. Titel | gef. Titel | rel. Titel | gef. Titel | rel. Titel | gef. Titel | rel. Titel | STW-Index | IDX-Index | SW-Index | Basic-Index |
| <i>Straßenkinder</i> | 6 | 6 | 0 | 0 | 0 | 0 | 6 | 6 | 1,00 | | | 1,00 |
| <i>Straßenkind</i> | 1 | 0 | 7 | 6 | 0 | 0 | 7 | 6 | | 0,86 | | |
| <i>Kriegsberichterstattung</i> | 0 | 0 | 15 | 13 | 14 | 13 | 15 | 13 | | 0,87 | 0,93 | 0,87 |
| <i>Verein + Gründung</i> | 3 | 3 | 8 | 3 | 0 | 0 | 8 | 3 | 1,00 | 0,38 | | 0,38 |
| <i>Medizin + Drittes Reich</i> | 0 | 0 | 83 | 13 | 10 | 10 | 0 | 0 | | 0,16 | 1,00 | |
| <i>Medizin + Drittes + Reich</i> | 0 | 0 | 34 | 3 | 10 | 10 | 34 | 3 | | | | 0,09 |
| <i>Verkehrsgeographie</i> | 3 | 3 | 5 | 4 | 4 | 3 | 5 | 4 | 1,00 | 0,80 | 0,75 | 0,80 |
| <i>Theatergeschichte</i> | 6 | 6 | 7 | 6 | 1 | 1 | 7 | 6 | 1,00 | 0,86 | 1,00 | 0,86 |
| <i>Interkontinentalrakete</i> | 0 | 0 | 649 | 0 | 0 | 0 | 649 | 0 | | 0,00 | | 0,00 |
| <i>Politik + Massenmedien</i> | 3 | 3 | 42 | 27 | 2 | 2 | 43 | 28 | 1,00 | 0,64 | 1,00 | 0,65 |
| <i>Politik + Massenmedium</i> | 0 | 0 | 19 | 18 | 0 | 0 | 19 | 18 | | | | |

| | |
|------------------|--|
| Suchformulierung | Formulierungen der Suchfragen; ggf. in mehreren Varianten |
| StW-Register | Register der Stichwörter |
| IDX-Register | Register der IDX-Deskriptoren |
| SW-Register | Register der RSWK-Schlagwörter |
| Basic-Index | Register der Stichwörter, IDX-Deskriptoren und RSWK-Schlagwörter |
| Precision | Precisionwert für das jeweilige Register |
| gef. Titel | Gesamtzahl gefundener Titel |
| rel. Titel | Anzahl relevanter Titel unter den gefundenen Titeln |

6.2. Tabelle 2: Relationencodes

| | |
|---|--|
| 001 Synonym | 046 Übersetzung (falsch in Quellwort) |
| 002 Teilwort → Kompositum | 047 Nichtstichwort (Singular) → Stichwort (Plural) |
| 003 Kompositum → Teilwort | 048 Stichwort (Plural) → Nichtstichwort (Singular) |
| 004 Derivation | 049 Nichtstichwort (Plural) → Stichwort (Singular) |
| 005 Akronym → Langform | 050 Stichwort (Singular) → Nichtstichwort (Plural) |
| 006 Langform → Akronym | 051 Deskriptor (ohne Relationierung) |
| 007 Antonym (alle Richtungen) | 052 Nichtdeskriptor (ohne Relationierung) |
| 008 Sprachvarianten (Deutsch/Schweiz/Österr.) | 053 Assoziation (alle Richtungen) |
| 009 siehe - auch | 054 Übersetzung (falsch in Zielwort) |
| 010 orthograph. Variante | 055 Genusproblem |
| 011 Stichwort → Nichtstichwort | 056 Schreibproblem (Verwechslung) |
| 012 Nichtstichwort → Stichwort | 057 grammatischer Begriff |
| 013 (Syntakt.) Homograph | 058 chemisches Element → Langform |
| 014 Quasi - Synonym | 059 Langform → chemisches Element |
| 015 Unterbegriff → Oberbegriff | 060 Wort → Beispiel |
| 016 Oberbegriff → Unterbegriff | 061 Regelname → Regel |
| 017 Grundform → Ablaut/Umlaut | 062 Regelbeispielwort → Regel |
| 018 Ablaut/Umlaut → Grundform | 063 Regelnummer → Regeltext |
| 019 Wort (Begriff) → Klassifizierungscode | 064 siehe: |
| 020 Klassifizierungscode → Wort (Begriff) | 065 vergleiche: |
| 021 Wort → Worterläuterung (nur eine Richtung) | 066 Regel → Regelbeispielwort |
| 022 Einzelwort → Mehrwort | 067 Verb als Kompositumteil → Nominalumsetzung |
| 023 Mehrwort → Einzelwort | 068 Translation Memory → Übersetzung (allgemein) |
| 024 Synonym als Stichwort → Synonym als Nicht-Stichwort | 069 Translation Memory Spezial → Übersetzung (spezial) |
| 025 Synonym als Nicht-Stichwort → Synonym als Stichwort | 070 Phrasenstrukturregel |
| 026 Homonym (alle Richtungen; mit Semantik-Differenzierung) | 071 Struktureintrag |
| 027 Grammat. Hinweis | 072 ST-Eintrag |
| 028 Zeichensetzung | 073 GWS-Eintrag |
| 029 Groß/Klein-Hinweis | 074 Expansions-Eintrag |
| 030 Kontextbeispiel | 075 Reduktionseintrag |
| 031 unregelmäßiger Nominativ Plural → Nominativ Singular | 076 E-H-String-Eintrag |
| 032 Nominativ Singular → unregelmäßiger Nominativ Plural | 077 H-E-String-Eintrag |
| 033 Pseudokompositum → Teilwort | 078 Wort-Feinwortklasse |
| 034 Kompositum mit Präfix → Teilwort | 079 Silbentrennung |
| 035 Korrekte Schreibweise → Rechtschreibfehler | 080 Mehrwort → Phrasenstruktur |
| 036 Rechtschreibfehler → Korrekte Schreibweise | 081 Terminalsymbol → Feinwortklasse |
| 037 Zusammenbildung → Element | 082 Struktursymbol → kompaktes Struktursymbol |
| 038 Element → Zusammenbildung | 083 kompaktes Struktursymbol → Struktursymbol |
| 039 chemische Formel → Langform | 084 Feinwortklasse → Homographentyp |
| 040 Langform → chemische Formel | 085 Homographentyp → Feinwortklasse |
| 041 Mehrwort kanonisch → Mehrwort Text | 099 Schrottrelation (z.B.: abdizieren → zieren) |
| 042 Mehrwort Text → Mehrwort kanonisch | 103 formal mögl. Dekomposition |
| 043 Zusammen- oder Getrennschreibung | 104 formal mögl. Derivation |
| 044 Pluralbildung | 124 formal mögl. Rel.: SW → Synonym |
| 045 Übersetzung | 125 formal mögl. Rel.: Synonym → SW |

6.3 Praktische Hinweise zur Durchführung von Retrievaltests

Die folgenden Hinweise zur Durchführung eines Retrievaltests wurden einem Artikel von Jean Tague-Sutcliffe entnommen⁵⁴. Die hier wiedergegebene Darstellung ist eine gekürzte und zusammengefasste Version dieses Artikels.

6.3.1 Allgemeines

Ein Retrievaltest muss einen bestimmten Zweck erfüllen. Daher sollte vorher klar sein, welches Ziel ein Test verfolgt, welches zusätzliches Wissen durch diesen Test erlangt werden kann und ob dieses Wissen nicht schon von anderen untersucht wurde. Für den letztgenannten Aspekt sind gründliche Recherchen in der einschlägigen Literatur unerlässlich, um sich über die wichtigsten bereits durchgeführten Tests zu informieren.

6.3.2 Art des Tests

In einem Experiment oder Test, im Gegensatz zur Forschung, können die unbekannt Variablen klar definiert und variiert werden und ihr Effekt auf das Retrievalergebnis untersucht werden. In einem Labortest können alle Hauptkomponenten (Nutzer, Sucher, Datenbank, Suchverfahren) kontrolliert werden, während in einem Praxistest keine Kontrolle möglich ist. Praxistests liefern realistischere Ergebnisse, erbringen aber weniger spezifische Informationen. Da weniger Komponenten kontrolliert werden können, ist es schwieriger, beobachtete Ergebnisse auf bestimmte Ursachen zurückzuführen.

6.3.3 Definition der Variablen

Vor der Durchführung des Tests müssen alle Komponenten des Tests definiert werden und es muss klar sein, welche Komponenten kontrolliert bzw. nicht kontrolliert sind. Hierbei ist auch zu überlegen, wie diese Komponenten zu messen und zu beurteilen sind. Folgende Komponenten können untersucht werden:

| | |
|--|--|
| Datenbank / Dokumentensammlung | Größe, Art der enthaltenen Dokumente, Sachgebiet, Art der Einträge |
| Form der Informationsrepräsentation: logisches Modell der Datenbank | Boole'sches Modell, Vector Space-Modell, semantisches Modell, Cluster-Modell, Hypergraph-Modell, Production Grammar-Modell |
| physikalische Struktur | Baumstruktur, ungeordnete Struktur (Hashing), Signaturen, Multilists |

⁵⁴ Tague-Sutcliffe, J.: The pragmatics of information retrieval. In: Information Processing and Management. 28 (1992), Nr. 4, S. 467-490.

| | |
|----------------------|---|
| Indexierung | Breite und Tiefe, Maß an Kontrolle, Anzahl von Verweisen im Vokabular, Gebräuchlichkeit des Vokabulars, syntaktische Kontrolle, Prä- oder Postkoordination, Genauigkeit der Indexierung, Indexierungskonsistenz innerhalb mehrerer Indexierer |
| Nutzer | Art von Nutzern, Zweck der Nutzung (für Beruf, Bildung, Freizeit), Art der gesuchten Information, Dringlichkeit des Informationsbedürfnisses |
| Suchfrage | der Praxis entnommen oder extra für den Test formuliert |
| Suchformulierung | Art der Suchelemente, Fragenstruktur, Länge der Suchformulierung |
| Suchprozess | Suche durch Endnutzer oder durch Informationsvermittler Zugangsmodus: Kommando- oder menüorientiert, Hypertext-links Länge der Suche (zeitlich, Anzahl der eingegebenen Suchkommandos) |
| Retrievalperformance | Recall, Precision oder Fallout mit dem System verbrachte Zeit, Nutzerfreundlichkeit. |

6.3.4 Datenbank

Bei der Entscheidung für eine Datenbank muss überlegt werden, ob das Datenbanksystem und die Retrievalsoftware selbst erstellt und eingerichtet oder aus einer externen, kommerziellen Quelle bezogen werden sollen. Mehrere Variationen sind hier möglich. Selbsterstellte Datenbanken sind in der Regel teuer in der Herstellung und deswegen oftmals nur klein, so dass die Ergebnisse nur unter Vorbehalt auf größere Datenmengen übertragen werden können. Für bestimmte Zwecke kann dies aber nötig sein, beispielsweise wenn ein experimentelles Retrievalverfahren getestet werden soll. Bei der Planung eines Praxistests sollten folgende Aspekte bedacht werden:

- Umfang der Datenbank hinsichtlich Sprache, Zeitraum, Sachgebiet
- Quelle der Dokumente
- Erschließungsvokabular, Einsatz von Normdateien
- Form der Dokumente: Volltext, Abstract, Zitierungen
- Wahl der Felder zur Beschreibung der einzelnen Dokumente / Einträge
- Anzeigeformat, Reihenfolge in der Anzeige
- Einsatz unter Windows
- Art der Indices, was wird indexiert.

Wird vor allem Wert auf die Untersuchung von experimentellen Retrieval-Strategien gelegt, können bereits bestehende Testkollektionen genutzt werden (z.B. TREC, GIRT).

Sollen externe Datenbanken eingesetzt werden, können diese entweder als Tape erworben werden oder direkt online genutzt werden. Ersteres ist sinnvoll, wenn neu entwickelte Software für das Retrieval eingesetzt werden soll und letzteres, wenn bestimmte Retrievalstrategien untersucht werden sollen.

6.3.5 Fragensauswahl

Die Wahl der Suchfragen ist sehr wichtig für einen Retrievaltest, da sie der Ausgangspunkt für eine Recherche sind, den Suchprozess und die Brauchbarkeit der Ergebnisse bestimmen.

'Richtige' Nutzer sind die beste Quelle für Suchfragen. Gleichzeitig ist diese Einbeziehung aber auch problematisch. Es ist schwierig, Nutzer zu kontrollieren und sie adäquat in den Suchprozess und die Auswertung einzubeziehen. Zudem können sie möglicherweise vor Beendigung des Tests aussteigen oder die Anweisungen nicht befolgen, so dass die Integrität des Tests beeinträchtigt werden könnte.

Eine Möglichkeit, Fragen eigens für einen Test zu formulieren, besteht darin, den Titel eines Aufsatz als Suchfrage und die darin zitierten Publikationen als relevante Antworten zu wählen. Eine andere Möglichkeit ist, ältere Suchfragen zu nehmen, die oftmals in Bibliotheken oder Informationseinrichtungen aufbewahrt werden. Problematisch ist dann allerdings, dass diese Fragen kein konkretes Informationsbedürfnis darstellen und deshalb keine zusätzlichen Informationen zu diesen Fragen erhältlich sind und die Relevanzbeurteilung schwierig ist, was zu Inkonsistenzen führen kann.

6.3.6 Retrievalsoftware, Durchführung der Suche

Der Einsatz von selbstentwickelter Retrievalsoftware ist sinnvoll, wenn neue Retrievalansätze untersucht werden sollen. Zusätzlich können dadurch detailliertere Informationen über das eingesetzte System gewonnen werden. Die Software sollte gut dokumentiert und strukturiert und absolut fehlerfrei sein. Deswegen sollten bei der Entwicklung Informationsvermittler und Programmierer eng zusammen arbeiten.

Richtet sich das Interesse des Tests mehr auf den Gebrauch eines Retrievalsystems, z.B. Suchverhalten oder Nutzer-Sucher-Interaktion, so kann bereits existierende Software verwendet werden. Kommerzielle Software kann zudem durch vor- oder nachher eingesetzte Zusatzsoftware an die speziellen Anforderungen eines Tests angepasst werden.

Sehr wichtig bei der Durchführung des Tests ist ein standardisiertes Suchverfahren, um Variationen in und Manipulationen an den Testergebnissen auszuschließen. Erreicht wird dies durch gute Testdokumentation und angemessene Einweisungen der an dem Test beteiligten Personen. Informationen müssen gegeben werden hinsichtlich des eingesetzten Suchverfahrens, der Suchhilfen (z.B. Thesauri) und des erforderlichen Datenoutputs.

Während des Suchprozesses sollten mögliche äußere Einflussfaktoren weitestmöglich ausgeschlossen werden. So sollte eine Suche beendet werden, sobald bei einer der Testpersonen Müdigkeit auftritt; Systemausfällen sollte vorgebeugt werden, indem Ausweichrechner zur Verfügung stehen.

6.3.7 Datenerhebung

Während der Durchführung eines Retrievaltests entstehen in den verschiedensten Bereichen Daten. Daher ist es wichtig, die Datenauswertung so zu planen, dass nur die für die Zwecke des Tests relevanten Daten analysiert werden.

Zu den folgenden Bereichen können Daten gesammelt werden:

| | |
|------------------------------|--|
| Datenbank (Größe, Abdeckung) | Datenbankproduzent, statistische Programme |
|------------------------------|--|

Nutzer, Prozesse (Suchverfahren, Beobachtung, Interviews, Fragebogen, Indexierung, Evaluierung), Ergebnisse (Effektivität, Kosten, Wert) Protokolle, Log-Dateien.

6.3.8 Datenanalyse

Die Art der Datenanalyse ist abhängig von der Art der erhobenen Daten. Wurden normierte Daten gesammelt, kann eine statistische Analyse durchgeführt werden. Daten in Form von natürlicher Sprache müssen zunächst kategorisiert und normiert werden, bevor statistische Methoden darauf angewandt werden können.

Statistische Verfahren können entweder deskriptiv oder inferentiell durchgeführt werden. Deskriptive Methoden beschreiben die Daten, ohne sie tiefer gehend zu analysieren. Alle angefallenen Daten sollten zunächst deskriptiv dargestellt werden, in Form von Graphiken oder Tabellen. Bei der inferentiellen Methode wird versucht, Schlussfolgerungen aus den aufgetretenen Sachverhalte zu ziehen. Folgende Verfahren können angewandt werden: Schätzung, Vergleich der Verfahren oder Ergebnisse, Erforschung der Beziehungen, Vorhersagen und Bestimmung der Struktur von mehrfach variierbaren Daten.

Die Berechnung und Analyse der Daten kann mit Statistikprogrammen wie SPSS (Statistical Package for the Social Sciences) oder SAS (Statistical Analysis System) durchgeführt werden.

6.3.9 Darstellung der Ergebnisse

Ein Bericht über einen Retrievaltest sollte folgende Aspekte umfassen:

- Ziel des Tests: warum wurde er durchgeführt, was sollte herausgefunden werden
- Hintergrund des Tests: Hinweise auf vorherige Arbeiten, die sich speziell mit dem Thema des Test auseinandergesetzt haben
- Methodologie: Beschreibung der Testumgebung und -verfahren, so detailliert, dass andere Interessierte dies nachvollziehen können; Schilderung der aufgetretenen methodologischen Probleme
- Präsentation der Ergebnisse: in Worten oder Tabellen und graphischen Darstellungen
- Schlussfolgerungen: Rückblick und Zusammenfassung sowie eine Darstellung der Haupteckkenntnisse und ihre Auswirkungen auf zukünftige Forschung.

Zuletzt muss überlegt werden, wo der Bericht publiziert werden soll. Wurde der Test für ein kommerzielles Unternehmen durchgeführt, wird der Bericht diesem zur Verfügung gestellt. Einer größeren Anzahl an Interessierten können die Ergebnisse auf Fachkonferenzen, beispielsweise auf der International Conference on Research and Development in Information Retrieval von der Association for Computing Machinery Special Interest Group on Information Retrieval (SIGIR), vorgestellt werden. Eine weitere Möglichkeit zur Veröffentlichung bieten Fachzeitschriften.

6.4 Verzeichnis der zitierten Literatur

- Aitchison, T. M.: Comparative evaluation of index languages. Part I, Design. Part II, Results. London: Inspec 1969, 1970.
- Cleverdon, C. W., J. Mills u. E. M. Keen: Factors determining the performance of indexing systems. Vol. 1: Design. Cranfield: Aslib 1966.
- Cleverdon, C. W.: Progress in documentation: evaluation tests of information retrieval systems. In: Journal of Documentation. 26 (1970), Nr. 1, S. 59-75.
- Cleverdon, C. W.: Report on the first stage of an investigation into the comparative efficiency of indexing systems. Cranfield: College of Aeronautics 1960.
- Cleverdon, C. W.: Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems. Cranfield: College of Aeronautics 1962.
- Cooper, W. S.: Expected search length: a single measure of retrieval effectiveness based on weak ordering action of retrieval systems. In: Journal of the American Society for Information Science. 19 (1968), S. 30-41.
- Cuadra, C. A., R. V. Katter: Experimental studies of relevance judgements. Santa Monica, Calif. 1967.
- Cuadra, C. A., R. V. Katter: Opening the black box of relevance. In: Journal of Documentation. 23 (1967), S. 291-303.
- Denning, R., M. Shuttleworth u. P. Smith: Interface design concepts in the development of a web-based information retrieval system. In: Bulletin of the American Society for Information Science. (1998) April/May, S. 17-20.
- Dreis, G.: Benutzerverhalten an einem Online-Publikumskatalog für wissenschaftliche Bibliotheken: Ergebnisse und Erfahrungen aus dem OPAC-Projekt der Universitätsbibliothek Düsseldorf. Frankfurt am Main: Klostermann 1994.
- Ellis, D.: Progress and problems in information retrieval. London: Library Association Publishing 1996.
- Frisch, E., M. Kluck: Pretest zum Projekt German Indexing and Retrieval Testdatabase (GIRT) unter Anwendung der Retrievalsysteme Messenger und freeWAIS-sf. IZ-Arbeitsbericht, (1997), Nr. 10.
- Glöckner-Rist, A.: Suchfragen im Information Retrieval: eine empirische Untersuchung zum Rechercheverhalten von Informationsvermittlern und Endbenutzern. Konstanz: Universitätsverlag 1993.
- Gödert, W., M. Liebig: Maschinelle Indexierung auf dem Prüfstand: Ergebnisse eines Retrievaltests zum MILOS-II-Projekt. In: Bibliotheksdienst. 31 (1997), Nr. 1, S. 59-68.
- Gödert, W.: Sacherschließung in EDV-Katalogen: Probleme der Sacherschließung unter Einsatz von PC-Systemen. In: Mitteilungen der Arbeitsgemeinschaft der Parlaments- und Behördenbibliotheken. (1992), Nr. 73, S. 5-37.
- Gull, C. D.: Seven years of work on the organization of materials in the special library. In: American Documentation. 7 (1956), S. 320-329.
- Harman, D.: The Text Retrieval Conferences (TREC): providing a test-bed for information retrieval systems. In: Bulletin of the American Society for Information Science. (1998) April/May, S. 11-13.
- Harter, S. P.: Variations in relevance assessments and the measurement of retrieval effectiveness. In: Journal of the American Society for Information Science. 47 (1996), Nr. 1, S. 37-49.
- Keen, E. M.: The Aberystwyth index languages test. In: Journal of Documentation. 29 (1973), Nr. 1, S. 1-35.
- Keen, E. M.: Aspects of computer-based indexing languages. In: Computers in libraries international '91. Proceedings of the 5th annual conference on computers in libraries,

- held in London in February 1991. London 1991.
- Keen, E. M., J. A. Digger: Report of an information science index language test. Part 1: Text. Part 2: Tables. Aberystwyth: Department of Information Retrieval Studies, College of Librarianship 1972.
- Gluck, M.: Eine deutschsprachige Testdatenbank für moderne Erschließungs- und Retrievalsysteme (German Indexing and Retrieval Testdatabase - GIRT). In: Die digitale Revolution: Deutscher Dokumentartag 1996. Frankfurt a. M. 1996. S. 427-431.
- Knorz, G.: Testverfahren für intelligente Indexierungs- und Retrievalsysteme anhand deutschsprachiger sozialwissenschaftlicher Fachinformation (GIRT): Bericht über einen Workshop am 12. September 1997 im IZ Sozialwissenschaften, Bonn. In: Nachrichten für Dokumentation. 49 (1998), Nr. 2, S. 111-116.
- Lancaster, F. W.: Evaluation of the MEDLARS demand search service. Washington: U.S. Government Print Office 1968.
- Lancaster, F. W., J. Mills: Testing indexes and index language devices: the ASLIB Cranfield project In: American Documentation. 15 (1964), S. 4-13.
- Lepsky, K.: Auf dem Weg zur automatischen Inhaltserschließung?: Das DFG-Projekt MILOS und seine Ergebnisse. In: Mitteilungen der Gesellschaft für Bibliothekswesen und Dokumentation des Landbaus. (1997), Nr. 53, S. 46-52.
- Lepsky, K.: Automatische Indexierung und bibliothekarische Inhaltserschließung: Ergebnisse des DFG-Projekts MILOS I. In: Zukunft der Sacherschließung im OPAC: Vorträge des 2. Düsseldorfer OPAC-Kolloquiums am 21. Juni 1995. Hrsg.: E. Niggemann u. K. Lepsky. Düsseldorf 1996. (Schriften der Universitäts- und Landesbibliothek Düsseldorf; Bd. 25).
- Lepsky, K.: Maschinelles Indexieren zur Verbesserung der sachlichen Suche im OPAC: DFG-Projekt an der Universitäts- und Landesbibliothek Düsseldorf. In: Bibliotheksdienst. 28 (1994), Nr. 8, S. 1234-1242.
- Lepsky, K.: RSWK - und was noch?: Stellungnahme zum Bericht 'Sacherschließung in Online-Katalogen' der Expertengruppe Online-Kataloge. In: Bibliotheksdienst. 29 (1995), Nr. 3, S. 500-519.
- Lepsky, K.: Sacherschließung ohne RSWK?: Neue Praxis an der Universitäts- und Landesbibliothek Düsseldorf. In: ProLibris. 3 (1998), Nr. 2, S. 112-114.
- Lepsky, K., J. Siepmann u. A. Zimmermann: Automatische Indexierung für Online-Kataloge: Ergebnisse eines Retrievaltests. In: Zeitschrift für Bibliothekswesen und Bibliographie. 43 (1996), Nr. 1, S. 47-56.
- Lesk, M. E., G. Salton: Relevance assessments and retrieval system evaluation. In: Information Storage and Retrieval. 4 (1968), S. 343-359.
- Niggemann, E., K. Lepsky (Hrsg.): Zukunft der Sacherschließung im OPAC: Vorträge des 2. Düsseldorfer OPAC-Kolloquiums am 21. Juni 1995. Düsseldorf: Universitäts- und Landesbibliothek 1996. (Schriften der Universitäts- und Landesbibliothek Düsseldorf; Bd. 25).
- Rowley, J. E.: The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research. In: Journal of Information Science. 20 (1994), Nr. 2, S. 108-119.
- Rowley, J. E., C. M. Turner: The dissemination of information. Worcester: André Deutsch Ltd. 1978.
- Sacherschließung. In: Jahresbericht 1996. Hrsg.: Die Deutsche Bibliothek. Frankfurt: Die Deutsche Bibliothek 1997.
- Salton, G.: The state of retrieval system evaluation. In: Information Processing and Management. 28 (1992), Nr. 4, S. 441-449.
- Salton, G.; M. J. McGill: Information Retrieval - Grundlegendes für Informationswissenschaftler. Hamburg: McGraw-Hill 1987.
- Softex GmbH Saarbrücken (Hrsg.): Universitäts- und Landesbibliothek Düsseldorf (Hrsg.): MILOS: Automatische Indexierung für Bibliotheken: Handbuch.

Stand: Juni 1996. Düsseldorf: Universitäts- und Landesbibliothek 1996.

Tague-Sutcliffe, J.: The pragmatics of information retrieval. In: Information Processing and Management. 28 (1992), Nr. 4, S. 467-490.

Thorne, R. G.: The efficiency of subject catalogues and the cost of information searches. In: Journal of Documentation. 11 (1955), S. 130-148.

6.5 Auswahlbibliographie zu durchgeführten Retrievaltests

1963

Cleverdon, C. W., J. Mills: The testing of index language devices. In: Aslib proceedings 15 (1963), Nr. 4, S. 106-130.

1964

VanOot, J. G.: Links and roles in coordinate indexing and searching: an economy study of their use and an evaluation of their effect on relevance and recall. Chicago, Ill.: American Chemical Society, Division of Chemical Literature 1964.

1966

Blagden, J. F.: How much noise in a role-free and link-free co-ordinate indexing system?. In: Journal of documentation. 22 (1966), S. 203-209.

1967

Cuadra, C. A., R. V. Katter: Experimental studies of relevance judgements: final report: Vol. 2: description of individual studies. Santa

Monica, Calif.: System Development Corp. 1967.

Good, I. J.: The decision-theory approach to the evaluation of information-retrieval systems. In: Information storage review. 3 (1967), S. 31-34.

1968

Brookes, B. C.: The measure of information retrieval effectiveness proposed by Swets. In: Journal of documentation. 24 (1968), S. 41-54.

Lancaster, F. W.: On the need for role indicators in postcoordinate retrieval systems. In: American documentation. 19 (1968), S. 42-46.

Salton, G., M. E. Lesk: Computer evaluation of indexing and text processing. In: Journal of the Association for Computing Machinery. 15 (1968), S. 8-36.

Treu, S.: The browser's retrieval game. In: American documentation. 19 (1968), S. 404-410.

1969

Aitchison, T. M.: Comparative evaluation of index languages: Part I, Design. Part II, Results. London: Inspec 1969. (OSTI report 5073)

Lancaster, F. W.: MEDLARS: report on the evaluation of its operating efficiency. In: American documentation. 20 (1969), S. 119-142.

Lesk, M. E., G. Salton: Relevance assessments and retrieval system evaluation. In: Information storage and retrieval. 4 (1969), S. 343-359.

Robertson, S. E.: The parametric description of retrieval tests: Part II: Overall measures. In: Journal of documentation. 25 (1969), Nr. 2, S. 93-106.

Robertson, S. E.: The parametric description of retrieval tests: Part I: The basic parameters. In: Journal of documentation. 25 (1969), Nr. 1, S. 1-27.

1971

Harter, S. P.: The Cranfield II relevance assessments: a critical evaluation. In: Library quarterly. 41 (1971), S. 223-228.

King, D. W., E. C. Bryant: The evaluation of information services and products. Washington: Information resources press 1971. 306 S.

1973

Cooper, W. S. : On selecting a measure of retrieval effectiveness, revisited. In: Journal of the American Society for Information Science. 24 (1973), S. 87-100.

Rijsbergen, C. J. van: A test for the separation of relevant and non-relevant documents in experimental retrieval collections. In: Journal of documentation. 29 (1973), Nr. 3, S. 251-257.

1974

Bhattacharyya, K.: The effectiveness of natural language in science indexing and retrieval. In: Journal of documentation. 30 (1974), S. 235-254.

Rijsbergen, C. J. van: Foundations of evaluation. In: Journal of documentation. 30 (1974), S. 365-373.

1975

Byrne, J. R.: Relative effectiveness of titles, abstracts, and subject headings for machine retrieval from the COMPENDEX services. In: Journal of the American Society for Information Science. 26 (1975), S. 223-229.

1976

Cooper, W. S. : The paradoxical role of unexamined documents in the evaluation of retrieval effectiveness. In: Information processing and management. 12 (1976), S. 367-375.

Sparck Jones, K., C. J. van Rijsbergen: Progress in documentation: Information retrieval test collection. In: Journal of documentation. 32 (1976), Nr. 1, S. 59-75.

1979

McGill, M. J., J. Huitfeldt: Experimental techniques of information retrieval. In: Annual review of information science and technology. 14 (1979), S. 93-128. White Plains: Knowledge Industry Publ. 1979.

Schabas, A. H.: A comparative evaluation of the retrieval effectiveness of titles, Library of Congress Subject Headings and PRECIS strings for computer searching of UK MARC data. London: University of London 1979.

1980

Regazzi, J. J.: Evaluating indexing systems: a review after Cranfield. In: Indexer. 12 (1980), Nr. 1, S. 15-20.

1981

Barraclough, E. D.: Opportunities for testing with online systems. In: Information retrieval experiment. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 128-135.

Belkin, N. J.: Ineffable concepts in information retrieval. In: Information retrieval experiment. Ed.: K. Sparck

Jones. London: Butterworths 1981. S. 44-58.

Cooper, W. S. : Gedanken experimentation: an alternative to traditional system testing?. In: Information retrieval experiment. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 199-210.

Evans, L.: An experiment: search strategy variations in SDI. In: Information retrieval experiments. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 285-315.

Heine, M. D.: Simulation, and simulation experiments. In: Information retrieval experiment. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 179-198.

Keen, E. M.: Laboratory tests of manual systems. In: Information retrieval experiment. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 136-155.

Lancaster, F. W.: Evaluation within the environment of an operating information service. In: Information retrieval experiment. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 105-127.

Oddy, R. N.: Laboratory tests: automatic systems. In: Information retrieval experiment. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 156-178.

Rijsbergen, C. J. van: Retrieval effectiveness. In: Information retrieval experiment. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 32-43.

Robertson, S. E.: The methodology of information retrieval experiment. In: Information retrieval experiment. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 9-31.

Salton, G.: The Smart environment for retrieval system evaluation: advantages and problem areas. In: Information retrieval experiments. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 316-329.

Sparck Jones, K. (Hrsg.): Information retrieval experiment. London: Butterworths 1981. 352 S.

Sparck Jones, K.: Retrieval system tests 1958-1978. In: Information retrieval experiment. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 213-255.

Sparck Jones, K.: The Cranfield tests. In: Information retrieval experiments. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 256-284.

Tague, J. M.: The pragmatics of information retrieval experimentation. In: Information retrieval experiment. Ed.: K. Sparck Jones. London: Butterworths 1981. S. 59-102.

1982

Madelung, H.-O.: Subject searching in the social sciences: a comparison of PRECIS and KWIC indexes indexes to newspaper articles. In: Journal of librarianship. 14 (1982), S. 45-58.

Schabas, A. H.: Postcoordinate retrieval: a comparison of two retrieval languages. In: Journal of the American Society for Information Science. 33 (1982), S. 32-37.

1983

Hodges, P. R.: Keyword in title indexes: effectiveness of retrieval in computer searches. In: Special libraries. 74 (1983), Nr. 1, S. 56-60.

Saracevic, T.: On a method for studying the structure and nature of requests in information retrieval. In: Proceedings of the 46th American Society for Information Science, Annual Meeting. 1983. S. 22-25.

1984

Bernstein, L. M., R. E. Williamson: Testing of a natural language retrieval system for a full text knowledge base. In: Journal of the American Society for Information Science. 35 (1984), S. 235-247.

Fidel, R.: Online searching styles: a case-study-based model of searching behavior. In: Journal of the American

Society for Information Science. 35 (1984), S. 211-221.

Hartley, D.: A 'laboratory' method for the comparison of retrieval effectiveness in manual and online searching. In: 7th International Online Information Meeting, London, 6.-8. 12. 1983. London: 1984. S. 157-166.

Peritz, B. C.: On the informativeness of titles. In: International classification. 11 (1984), Nr. 2, S. 87-89.

1985

Blair, D. C., M. E. Maron: An evaluation of retrieval effectiveness for a full-text document-retrieval system. In: Communications of the Association for Computing Machinery. 28 (1985), S. 280-299.

Donkersloot, H. B.: Zoeken op titelwoorden: een onderzoek met de online publiekscatalogus. In: Open. 17 (1985), Nr. 12, S. 542-546.

Jochum, F., V. Weissmann: Struktur und Elemente des Information Retrieval Experiments. In: Anwendungen in der Klassifikation. I. Proc. 8. Jahrestagung der Gesellschaft für Klassifikation, Hofgeismar, 10.-13. 4. 1984. Hrsg.: R. G. Henzler. Frankfurt: Indeks 1985. S. 76-87. (Studien zur Klassifikation, Bd. 14)

1986

Blair, D. C.: Full text retrieval: Evaluation and implications. In: International classification. 13 (1986), S. 18-23.

1987

Bellardo, T., T. Saracevic: Online searching and search output: relationships between overlap, relevance, recall and precision. In: Proceedings of the 50th Annual Meeting of the American Society for Information Science 1987. Ed.: C. Chen. Medford, NJ: Learned Information 1987. S. 11-13.

MacCain, K. W., H. D. White u. B. C. Griffith: Comparing retrieval performance in online data bases. In: Information processing and management. 23 (1987), S. 539-553.

1988

Salton, G.: Thoughts about modern retrieval technologies. In: Information services and use. 8 (1988), Nr. 2/ 3/4, S. 107-113.

Sievert, M. E., E. J. McKinin u. M. Slough: A comparison of indexing and full-text for the retrieval of clinical medical literature. In: ASIS'88. Information technology: planning for the next fifty years. Proceedings of the 51st annual meeting of the American Society for Information Science, Atlanta, Georgia, 23-27. 10. 1988. Vol. 25. Ed. by C. L. Borgman u. E. Y. H. Pai. Medford, New Jersey: Learned Information 1988. S. 143-146.

Sullivan, M. V., C. L. Borgman: Bibliographic searching by end-users and intermediaries: front-end software vs native DIALOG commands. In: ASIS '88. Information Technology: planning for the next fifty years. Proceedings of the 51st annual meeting of the American Society for Information Science, Atlanta, Georgia, 23-27. 10. 1988. Vol. 25. Ed. by C. L. Borgman u. E. Y. H. Pai. Medford, New Jersey: Learned Information Inc. 1988. S. 120-126.

1989

Croft, W. B., R. H. Thompson: Support for browsing in an intelligent text retrieval system. In: International journal of man-machine studies. 30 (1989), S. 639-668.

Feng, S. : A comparative study of indexing languages in single and multidatabase searching. In: Canadian Journal of Information Science. 14 (1989), Nr. 2, S. 26-46.

Gordon, M., M. Kochen: Recall-precision trade-off: a derivation. In: Journal of the American Society for Information Science. 40 (1989), Nr. 3, S. 145-151.

Kilgour, F. G.: Retrieval on information from computerized book texts. In: Academic librarianship past, present, and future. A festschrift in honor of David Kaser: Ed. by J. Richardson u. J. Y. Davis. Englewood, CO: Libraries Unlimited 1989. S. 31-39.

- Lochbaum, K. E., A. R. Streeter: Comparing and combining the effectiveness of latent semantic indexing and the ordinary vector space model for information retrieval. In: Information processing and management. 25 (1989), Nr. 6, S. 665-676
- MacCain, K. W.: Descriptor and citation retrieval in the medical behavioral sciences literature: retrieval overlaps and novelty distribution. In: Journal of the American Society for Information Science. 40 (1989), S. 110-114.
- Ongering, H., G. J. A. Riesthuis: De ontsluiting in het ADION-systeem: precision en recall. In: Open. 21 (1989) nos. 7/8, S. 261-264.
- Pao, M. L., D. B. Worthen: Retrieval effectiveness by semantic and citation searching. In: Journal of the American Society for Information Science. 40 (1989), S. 226-235.
- Pao, M. L.: Retrieval differences between term and citation indexing. In: Information, knowledge, evolution. Proceedings of the 44th FID congress, Helsinki, 28. 8.-1. 9. 1988. Ed. by S. Koshiala u. R. Lau, Nr. Amsterdam: Elsevier Science Publishers 1989. S. 113-120.
- Prasher, R. G.: Evaluation of indexing system. In: Herald of library science. 28 (1989), Nr. 3, S. 157-65.
- Sievert, M. E., E. J. McKinin: Why full-text misses some relevant documents: an analysis of documents not retrieved by CCML or MEDIS. In: ASIS'89. Managing information and technology. Proceedings of the 52 nd annual meeting of the American Society for Information Science, Washington D. C., 30. 10.-2. 11. 1989. Vol. 26. Ed. by J. Katzer u. G. B. Newby. Medford, New Jersey: Learned Information 1989. S. 34-39.
- 1990**
- Bauer, G., C. Schneider: PADOK-II: Untersuchungen zur Volltextproblematik und zur interpretativen Analyse der Retrievalprotokolle. In: Nachrichten für Dokumentation. 41 (1990) Nr. 1, S. 21-26.
- Blair, D. C., M. E. Maron: Full-text information retrieval: further analysis and clarification. In: Information processing and management. 26 (1990), S. 437-447.
- Bollmann-Sdorra, P.: Probleme der Validität bei Retrievaltests. In: Nachrichten für Dokumentation. 41 (1990) Nr. 1, S. 9-12.
- Dalrymple, P. W.: Retrieval by reformulation in two library catalogs: toward a cognitive model of searching behavior. In: Journal of the American Society for Information Science. 41 (1990), S. 272-281.
- Harter, S. P.: Search term combinations and retrieval overlap: a proposed methodology and case study. In: Journal of the American Society for Information Science. 41 (1990), S. 132-146.
- Huffman, G. D., D. A. Vital u. R. G. Bivins: Generating indices with lexical association methods: term uniqueness. In: Information processing and management. 26 (1990), Nr. 4, S. 549-558.
- Krause, J., C. Womser-Hacker: PADOK-II: Retrievaltests zur Bewertung von Volltextindexierungsvarianten für das deutsche Patentinformationssystem. In: Nachrichten für Dokumentation. 41 (1990) Nr. 1, S. 13-19.
- Logan, E.: Cognitive styles and online behaviour of novice searchers. In: Information processing and management. 26 (1990), S. 503-510.
- Saracevic, T., H. Mokros u. L. Su: Nature of interaction between users and intermediaries in online searching: a qualitative analysis. In: ASIS'90: Information in the year 2000, from research to applications. Proc. of the 53 rd Annual Meeting of the American Society for Information Science, Toronto, Canada, 4.-8. 11. 1990. Ed. by Diana Henderson. Medford, NJ: Learned Information Inc. 1990. S. 47-54.
- Shenouda, W.: Online bibliographic searching: how end-users modify their

search strategies. In: ASIS'90: Information in the year 2000, from research to applications. Proc. of the 53 rd Annual Meeting of the American Society for Information Science, Toronto, Canada, 4.-8. 11. 1990. Ed. by Diana Henderson. Medford, NJ: Learned Information Inc. 1990. S. 117-128.

Wildemuth, B. M.: Measures of success in searching a full-text fact base. In: ASIS'90: Information in the year 2000, from research to applications. Proc. of the 53 rd Annual Meeting of the American Society for Information Science, Toronto, Canada, 4.-8. 11. 1990. Ed. by Diana Henderson. Medford, NJ: Learned Information Inc. 1990. S. 104-109.

1991

Chen, H., V. Dhar: Cognitive process as a basis for intelligent retrieval system design. In: Information processing and management. 27 (1991), Nr. 5, S. 405-432.

Frei, H. P., P. Schäuble: Determining the effectiveness of retrieval algorithms. In: Information processing and management. 27 (1991), Nr. 2/3, S. 153-164.

Frei, H. P., S. Meienberg u. P. Schäuble: The perils of interpreting recall and precision values. In: Information retrieval: GI/GMD-Workshop, Darmstadt, 23.-24. 6. 1991: Proceedings. Ed.: N. Fuhr. Berlin: Springer 1991. S. 1-10. (Informatik-Fachberichte; 289)

Kaltenborn, K.-F.: Endnutzerrecherchen in der CD-ROM-Datenbank Medline: T. 1: Evaluations-und Benutzerforschung über Nutzungscharakteristika, Bewertung der Rechercheergebnisse und künftige Informationsgewinnung; T. 2: Evaluations-und Benutzerforschung über Recherchequalität und Nutzer-Computer/Datenbank-Interaktion. In: Nachrichten für Dokumentation. 42 (1991) Nr. 2, S. 107-114. (T. 1); 42 (1991) Nr. 3, S. 177-190 (T. 2).

Lancaster, F. W., T. H. Connell u. N. Bishop u. a.: Identifying barriers to effective subject access in library

catalogs. In: Library resources and technical services. 35 (1991), S. 377-391.

Peters, T. A., M. Kurth: Controlled and uncontrolled vocabulary subject searching in an academic library online catalog. In: Information technology and libraries. 10 (1991), S. 201-211.

Prabha, C.: The large retrieval phenomenon. In: Advances in library automation and networking. 4 (1991), S. 55-92.

Saracevic, T.: Individual differences in organizing, searching and retrieving information. In: ASIS '91: systems understanding people. Proc. of the 54th Annual Meeting of the ASIS, vol. 28, Washington, DC, 27.-31. 10. 1991. Ed.: J.-M. Griffiths. Medford: Learned Information Inc. 1991. S. 82-86.

Walt, H. E. A. van der, P. A. van Brakel: Method for the evaluation of the retrieval effectiveness of a CD-ROM bibliographic database. In: African journal of library and information science. 59 (1991), Nr. 1, S. 32-42.

Wildemuth, B. M., E. K. Jacob u. A. Fullington u. a.: A detailed analysis of end-user search behaviours. In: ASIS '91: systems understanding people. Proc. of the 54th Annual Meeting of the ASIS, vol. 28, Washington, DC, 27.-31. 10. 1991. Ed.: J.-M. Griffiths. Medford: Learned Information Inc. 1991. S. 302-312.

1992

Dalrymple, P. W., R. Cox: An examination of the effects of non-boolean enhancements to an information retrieval system. In: Proceedings of the 13th National Online Meeting. Ed.: M. E. Williams. New York: Learned Information 1992. S. 75-81.

Ekmekcioglu, F. C., A. M. Robertson u. P. Willett: Effectiveness of query expansion in ranked-output document retrieval systems. In: Journal of information science. 18 (1992), Nr. 2, S. 139-147.

Janes, J. W., R. McKinney: Relevance judgements of actual users and

secondary judges: a comparative study. In: *Library quarterly*. 62 (1992), Nr. 2, S. 150-168.

Keen, E. M.: Some aspects of proximity searching in text retrieval systems. In: *Journal of information science*. 18 (1992), S. 89-98.

Tonta, Y.: Analysis of search failures in document retrieval systems: a review. In: *Public-access computer systems review*. 3 (1992), Nr. 1, S. 4-53.

1993

Allen, B.: Logical reasoning and retrieval performance. In: *Library and information science research*. 15 (1993), Nr. 1, S. 93-105.

Harman, D. (Hrsg.): *The First Text Retrieval Conference (TREC-1)*. Gaithersburgh, MD: National Institute of Standards and Technology 1993. (NIST special publication; 500-207)

Harman, D. K.: The first text retrieval conference: TREC-1, 1992. In: *Information processing and management*. 29 (1993), Nr. 4, S. 411-414.

Harman, D.: Overview of the first Text Retrieval Conference. In: *Proceedings of the 14th National Online Meeting 1993, New York, 4-6 May 1993*. Ed.: M. E. Williams. Medford, NJ: Learned Information 1993. S. 181-187.

Kristensen, J.: Expanding end-users' query statements for free text searching with a search-aid thesaurus. In: *Information processing and management*. 29 (1993), Nr. 6, S. 733-744.

Marchionini, G. u. a.: Information seeking in full-text end-user-oriented search system: the roles of domain and search expertise. In: *Library and information science research*. 15 (1993), Nr. 1, S. 35-69.

Park, T. K.: The nature of relevance in information retrieval: an empirical study. In: *Library quarterly*. 63 (1993), Nr. 3, S. 318-351.

Qiu, L.: Analytical searching vs. browsing in hypertext information retrieval systems. In: *Canadian journal*

of information and library science. 18 (1993), Nr. 4, S. 1-13.

Qiu, L.: Markov models of search state patterns in a hypertext information retrieval system. In: *Journal of the American Society for Information Science*. 44 (1993), Nr. 7, S. 413-427.

TREC-1: The first text retrieval conference: Rockville, MD, USA, 4-6 Nov. 1993. In: *Information processing and management*. 29 (1993), Nr. 4, S. 411-521.

1994

Buckland, M., F. Gey: The relationship between recall and precision. In: *Journal of the American Society for Information Science*. 45 (1994), Nr. 1, S. 12-19.

Drabenstott, K. M., D. Vizine-Goetz: Using subject headings for online retrieval: theory, practice and potential. San Diego, CA: Academic Press 1994. XVIII, 365 S.

Hofstede, M.: Literatuur over onderwerpen zoeken in de OPC. In: *CRI bulletin*. 29 (1994), Sept., S. 14-15.

Keen, E. M., R. J. Hartley: Phrase processing in text retrieval. In: *Journal of document and text management*. 2 (1994), Nr. 1, S. 23-34.

Large, A., J. Beheshti u. A. Breuleux u. a.: A comparison of information retrieval from print and CD-ROM versions of an encyclopedia by elementary school students. In: *Information processing and management*. 30 (1994), Nr. 4, S. 499-513.

Smithson, S.: Information retrieval evaluation in practice: a case study approach. In: *Information processing and management*. 30 (1994), Nr. 2, S. 205-221.

Tibbo, H. R.: The epic struggle: subject retrieval from large bibliographic databases. In: *American archivist*. 57 (1994), Nr. 2, S. 310-326.

Wood, F., N. Ford u. C. Walsh: The effect of postings information on search behaviour. In: *Journal of information science*. 20 (1994), Nr. 1, S. 29-40.

1995

- Barker, A. L.: Non-Boolean searching on commercial online systems: optimising use of Dialog TARGET and ESA/IRS QUESTQUORUM. In: *Online information 95: Proceedings of the 19th International online information meeting, London, 5-7 December 1995*. Ed.: D. I. Raitt u. B. Jeapes. Oxford: Learned Information 1995. S. 71-79.
- Barry, C. I., L. Schamber: User-defined relevance criteria: a comparison of 2 studies. In: *Forging new partnerships in information: converging technologies. Proceedings of the 58th Annual Meeting of the American Society for Information Science, ASIS '95, Chicago, IL, 9-12 October 1995*. Ed.: T. Kinney. Medford, NJ: Learned Information 1995. S. 103-111.
- Brown, M. E.: By any other name: accounting for failure in the naming of subject categories. In: *Library and information science research*. 17 (1995), Nr. 4, S. 347-385.
- Buckland, M. K. u. a.: Partnerships in navigation: an information retrieval research agenda. In: *Forging new partnerships in information: converging technologies. Proceedings of the 58th Annual Meeting of the American Society for Information Science, ASIS'95, Chicago, IL, 9-12 October 1995*. Ed.: T. Kinney. Medford, NJ: Learned Information 1995. S. 84-89.
- Buckley, C., J. Allan u. G. Salton: Automatic routing and retrieval using Smart: TREC-2. In: *Information processing and management*. 31 (1995), Nr. 3, S. 315-326.
- Callan, J., W. B. Croft u. J. Broglio: TREC and TIPSTER experiments with INQUERY. In: *Information processing and management*. 31 (1995), S. 327-332,343.
- Evans, D. A., R. G. Lefferts: CLARIT-TREC experiments. In: *Information processing and management*. 31 (1995), Nr. 3, S. 385-395.
- Gluck, M.: Understanding performance in information systems: blending relevance and competence. In: *Journal of the American Society for Information Science*. 46 (1995), Nr. 6, S. 446-460.
- Grunder, R.: Qualitätskriterien für Datenbanken: Grundsätzliche Überlegungen und Untersuchung ausgewählter Aspekte am Beispiel bibliographischer Online-Datenbanken. Köln: FHBD 1995.
- Harman, D. (Hrsg.): *The Second Text Retrieval Conference: TREC-2*. [Themenheft] In: *Information processing and management*. 31 (1995), Nr. 3, S. 269-448.
- Harman, D. K.: The TREC conferences. In: *Hypertext -Information Retrieval -Multimedia: HIM '95. Synergieeffekte elektronischer Informationssysteme*. Hrsg.: R. Kuhlen u. M. Rittberger. Konstanz: Universitätsverlag 1995. S. 9-28. (Schriften zur Informationswissenschaft; Bd. 20)
- Harman, D.: Overview of the Second Text Retrieval Conference: TREC-2. In: *Information processing and management*. 31 (1995), Nr. 3, S. 271-289.
- Hersh, W. R., D. H. Hickam: An evaluation of interactive Boolean and natural language searching with an online medical textbook. In: *Journal of the American Society for Information Science*. 46 (1995), Nr. 7, S. 478-489.
- Hersh, W. R., J. Pentecost u. D. H. Hickam: A task-oriented approach to retrieval system evaluation. In: *Forging new partnerships in information: converging technologies. Proceedings of the 58th Annual Meeting of the American Society for Information Science, ASIS'95, Chicago, IL, 9-12 October 1995*. Ed.: T. Kinney. Medford, NJ: Learned Information 1995. S. 112-117.
- Hofmann, M.: TREC Konferenzbericht (7. 10. 93). In: <http://www.inf-wiss.uni-konstanz.de/IW/HT-Link/assfalg:15:Feb:95:4:44:26pm.html>.
- livonen, M.: Consistency in the selection of search concepts and search terms. In: *Information processing and management*. 31 (1995), Nr. 2, S. 173-190.

- Iivonen, M.: Factors lowering the consistency in online searching. In: *Online information 95: Proceedings of the 19th International online information meeting*, London, 5-7 December 1995. Ed.: D. I. Raitt u. B. Jeapes. Oxford: Learned Information 1995. S. 101-107.
- Kelley, F., A. F. Smeaton: Thresholding the postings lists in information retrieval: experiments on TREC data. In: *New review of document and text management*. 1995, Nr. 1, S. 111-129.
- Losee, R. M.: Determining information retrieval and filtering performance without experimentation. In: *Information processing and management*. 31 (1995), Nr. 4, S. 555-572.
- McJunkin, M. C.: Precision and recall in title keyword searching. In: *Information technology and libraries*. 14 (1995), Nr. 3, S. 161-171.
- Meadows, C. J.: A study of user performance and attitudes with information retrieval interfaces. In: *Journal of the American Society for Information Science*. 46 (1995), Nr. 7, S. 490-505.
- Mokros, H. B., L. S. Mullins u. T. Saracevic: Practice and personhood in professional interaction: social identities and information needs. In: *Library and information science research*. 17 (1995), Nr. 3, S. 237-257.
- Nelson, M. J.: The effect of query characteristics on retrieval results in the TREC retrieval tests. In: *Connectedness: information, systems, people, organizations. Proceedings of CAIS/ACSI 95, the proceedings of the 23rd Annual Conference of the Canadian Association for Information Science*. Ed. by Hope A. Olson u. Denis B. Ward. Alberta: Alberta University, School of Library and Information Studies 1995. S. 156-163.
- Nicholas, D.: Are information professionals really better online searchers than end-users?: (and whose story do you believe?). In: *Online information 95: Proceedings of the 19th International online information meeting*, London, 5-7 December 1995. Ed.: D. I. Raitt u. B. Jeapes. Oxford: Learned Information 1995. S. 383-387.
- Robertson, S. E., S. Walker u. M. M. Hancock-Beaulieu: Large test collection experiments of an operational, interactive system: OKAPI at TREC. In: *Information processing and management*. 31 (1995), Nr. 3, S. 345-360.
- Shafique, M., A. S. Chaudhry: Intelligent agent-based online information retrieval. In: *Online information 95: Proceedings of the 19th International online information meeting*, London, 5-7 December 1995. Ed.: D. I. Raitt u. B. Jeapes. Oxford: Learned Information 1995. S. 259-265.
- Smith, M. P., A. S. Pollitt: Ranking and relevance feedback extensions to a view-based searching system. In: *Online information 95: Proceedings of the 19th International online information meeting*, London, 5-7 December 1995. Ed.: D. I. Raitt u. B. Jeapes. Oxford: Learned Information 1995. S. 231-240.
- Sparck Jones, K.: Reflections on TREC: TREC-2. In: *Information processing and management*. 31 (1995), Nr. 3, S. 291-314.
- Spink, A., A. Goodrum u. D. Robins: Search intermediary elicitations during mediated online searching. In: *Forging new partnerships in information: converging technologies. Proceedings of the 58th Annual Meeting of the American Society for Information Science, ASIS'95, Chicago, IL, 9-12 October 1995*. Ed.: T. Kinney. Medford, NJ: Learned Information 1995. S. 97-102.
- Spink, A.: Term relevance feedback and mediated database searching: implications for information retrieval practice and systems design. In: *Information processing and management*. 31 (1995), Nr. 2, S. 161-171.
- Tillotson, J.: Is keyword searching the answer?. In: *College and research libraries*. 56 (1995), Nr. 3, S. 199-206.
- Turtle, H., J. Flood: Query evaluation: strategies and optimizations. In: *Information processing and management*. 31 (1995), Nr. 3, S. 345-360.

mation processing and management. 31 (1995), Nr. 6, S. 831-850.

Wilkes, A., A. Nelson: Subject searching in two online catalogs: authority control vs. non authority control. In: *Cataloging and classification quarterly*. 20 (1995), Nr. 4, S. 57-79.

1996

Aldous, K. J.: A system for the automatic retrieval of information from a specialist database. In: *Information processing and management*. 32 (1996), Nr. 2, S. 139-154.

Armstrong, C. J., K. Medawar: Investigation into the quality of databases in general use in the UK. London: British Library 1996. 149 S. (Research and innovation report; 11)

Bates, M. J.: Document familiarity, relevance, and Bradford's law: the Getty Online Searching Project report; Nr. 5. In: *Information processing and management*. 32 (1996), Nr. 6, S. 697-707.

Beaulieu, M., S. Robertson u. E. Rasmussen: Evaluating interactive systems in TREC. In: *Journal of the American Society for Information Science*. 47 (1996), Nr. 1, S. 85-94.

Blair, D. C.: STAIRS Redux: thoughts on the STAIRS evaluation, ten years after. In: *Journal of the American Society for Information Science*. 47 (1996), Nr. 1, S. 4-22.

Borgman, C. L.: Why are online catalogs still hard to use?. In: *Journal of the American Society for Information Science*. 47 (1996), Nr. 7, S. 493-503.

Buckley, C., A. Singhal u. M. Mitra u. a.: New retrieval approaches using SMART: TREC 4. In: *The Fourth Text Retrieval Conference (TREC-4)*. Ed.: K. Harman. Gaithersburgh, MD: National Institute of Standards and Technology 1996. S. 25-48. (NIST special publication; 500-236)

Burnett, M., C. Fisher u. R. Jones: InTEXT precision indexing in TREC4. In: *The Fourth Text Retrieval Conference (TREC-4)*. Ed.: K. Harman.

Gaithersburgh, MD: National Institute of Standards and Technology 1996. S. 287-294. (NIST special publication; 500-236)

Crawford, G. A.: Varieties of access: a comparison of databases via Z 39.50, FirstSearch and CD-ROM. In: *Proceedings of the 17th National Online Meeting 1996*, New York, 14-16 May 1996. Ed.: M. E. Williams. Medford, NJ: Information Today 1996. S. 55-64.

Crestani, F., C. J. van Rijsbergen: Information retrieval by imaging. In: *Information retrieval: new systems and current research. Proceedings of the 16th Research Colloquium of the British Computer Society Information Retrieval Specialist Group*, Drymen, Scotland, 22-23 Mar 94. Ed.: R. Leon. London: Taylor Graham 1996. S. 47-76.

Crestani, F., I. Ruthven u. M. Sanderson u. a.: The troubles with using a logical model of IR on a large collection of documents: experimenting retrieval by logical imaging on TREC. In: *The Fourth Text Retrieval Conference (TREC-4)*. Ed.: K. Harman. Gaithersburgh, MD: National Institute of Standards and Technology 1996. S. 509-526. (NIST special publication; 500-236)

Dimitroff, A., D. Wolfram u. A. Volz: Affective response and retrieval performance: analysis of contributing factors. In: *Library and information science research*. 18 (1996), Nr. 2, S. 121-132.

Drabenstott, K. M., M. S. Weller: A comparative approach to system evaluation: delegating control of retrieval tests to an experimental online system. In: *Global complexity: information, chaos and control. Proceedings of the 59th Annual Meeting of the American Society for Information Science, ASIS'96*, Baltimore, Maryland, 21-24 Oct 1996. Ed.: S. Hardin. Medford, NJ: Learned Information 1996. S. 34-40.

Drabenstott, K. M., M. S. Weller: Improving personal-name searching in online catalogs. In: *Information*

- technology and libraries. 15 (1996), Nr. 3, S. 135-155.
- Ellis, D.: Progress and problems in information retrieval. 2nd ed. London: Library association publishing 1996. 220 S.
- Ellis, D.: The dilemma of measurement in information retrieval research. In: Journal of the American Society for Information Science. 47 (1996), Nr. 1, S. 23-36.
- Evans, J. E.: Some external and internal factors affecting users of interactive information systems. In: Herausforderungen an die Informationswirtschaft: Informationsverdichtung, Informationsbewertung und Datenvisualisierung. Proceedings des 5. Internationalen Symposiums für Informationswissenschaft (ISI'96), Humboldt-Universität zu Berlin, 17.-19. Oktober 1996. Hrsg.: J. Krause u. a. Konstanz: Universitätsverlag 1996. S. 239-250. (Schriften zur Informationswissenschaft; Bd. 27)
- Feldman, S. : Testing natural language: comparing DIALOG, TARGET, and DR-LINK. In: Online. 20 (1996), Nr. 6, S. 71-74,76-79.
- Gluck, M.: Exploring the relationship between user satisfaction and relevance in information systems. In: Information processing and management. 32 (1996), Nr. 1, S. 89-104.
- Grossman, D. A., D. O. Holmes u. O. Frieder u. a.: Improving accuracy and run-time performance for TREC-4. In: The Fourth Text Retrieval Conference (TREC-4). Ed.: K. Harman. Gaithersburgh, MD: National Institute of Standards and Technology 1996. S. 433-441. (NIST special publication; 500-236)
- Guglielmo, E. J., N. C. Rowe: Natural language retrieval of images based on descriptive captions. In: ACM transactions on information systems. 14 (1996), Nr. 3, S. 237-267.
- Harman, D. K.: Overview of the fourth Text Retrieval Conference: TREC-4. In: The Fourth Text Retrieval Conference (TREC-4). Ed.: K. Harman. Gaithersburgh, MD: National Institute of Standards and Technology 1996. S. 1-23. (NIST special publication; 500-236)
- Harman, K. (Hrsg.): The Fourth Text Retrieval Conference (TREC-4). Gaithersburgh, MD: National Institute of Standards and Technology 1996. 776 S. (NIST special publication; 500-236)
- Harter, S. P.: Variations in relevance assessments and the measurement of retrieval effectiveness. In: Journal of the American Society for Information Science. 47 (1996), Nr. 1, S. 37-49.
- Hersh, W., J. Pentecost u. D. Hickam: A task-oriented approach to information retrieval evaluation: overview and design for empirical testing. In: Journal of the American Society for Information Science. 47 (1996), Nr. 1, S. 50-56.
- Hull, D. A.: Stemming algorithms: a case study for detailed evaluation. In: Journal of the American Society for Information Science. 47 (1996), Nr. 1, S. 70-84.
- Jacso, P.: CD-ROM commentaries: the speed of the retrieval software. In: Information today. 13 (1996), Nr. 10, S. 24.
- Johnson, K. E.: OPAC missing record retrieval. In: Information technology and libraries. 15 (1996), Nr. 3, S. 169-171.
- Keyes, J. G.: Using conceptual categories of questions to measure differences in retrieval performance. In: Global complexity: information, chaos and control. Proceedings of the 59th Annual Meeting of the American Society for Information Science, ASIS'96, Baltimore, Maryland, 21-24 Oct 1996. Ed.: S. Hardin. Medford, NJ: Learned Information 1996. S. 238-242.
- Kluck, M.: Eine deutschsprachige Testdatenbank für moderne Erschließungs- und Retrievalsysteme: German Indexing and Retrieval Testdatabase - GIRT. In: Die digitale Revolution: Deutscher Dokumentartag 1996, Neue Universität Heidelberg, 24.-26. 9. 1996. Hrsg.: W. Neubauer. Frankfurt a. M.: DGD 1996. S. 427-431. (DGD-Schrift: DOK-9)

- Leppanen, E.: Homografiongelma tekstihaussa ja homografien disambiguoinnin vaikutukset. In: Informaatiotutkimus. 15 (1996), Nr. 4, S. 133-144.
- Lepsky, K., J. Siepmann u. A. Zimmermann: Automatische Indexierung für Online-Kataloge: Ergebnisse eines Retrievaltests. In: Zeitschrift für Bibliothekswesen und Bibliographie. 43 (1996) Nr. 1, S. 47-56.
- Losee, R. M.: Evaluating retrieval performance given database and query characteristics: analytic determination of performance surfaces. In: Journal of the American Society for Information Science. 47 (1996), Nr. 1, S. 95-105.
- Lu, X. A., J. D. Holt u. D. J. Miller: Boolean system revisited: its performance and its behavior. In: The Fourth Text Retrieval Conference (TREC-4). Ed.: K. Harman. Gaithersburgh, MD: National Institute of Standards and Technology 1996. S. 459-473. (NIST special publication; 500-236)
- Pfeifer, U., T. Poersch u. N. Fuhr: Retrieval effectiveness of proper name search methods. In: Information processing and management. 32 (1996), Nr. 6, S. 667-679.
- Pirkola, A., K. Jarvelin: The effect of anaphor and ellipsis resolution on proximity searching in a text database. In: Information processing and management. 32 (1996), Nr. 2, S. 199-216.
- Pollitt, A. S. , M. Treglown u. A. P. Smith u. a.: Empowering users for improved database access and analysis through the application of knowledge structure views, progressive refinement techniques and a design approach driven by usability. In: Knowledge organization and change: Proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Library of Congress, Washington, DC. Ed.: R. Green. Frankfurt: INDEKS 1996. S. 231-241. (Advances in knowledge organization; vol. 5)
- Ribeiro, F.: Subject indexing and authority control in archives: the need for subject indexing in archives and for an indexing policy using controlled language. In: Journal of the Society of Archivists. 17 (1996), Nr. 1, S. 27-54.
- Sanderson, M. :The Reuters test collection. In: Information retrieval: new systems and current research. Proceedings of the 16th Research Colloquium of the British Computer Society Information Retrieval Specialist Group, Drymen, Scotland, 22-23 Mar 94. Ed.: R. Leon. London: Taylor Graham 1996. S. 219-227.
- Schamber, L., J. Bateman: User criteria in relevance evaluation: toward development of a measurement scale. In: Global complexity: information, chaos and control. Proceedings of the 59th Annual Meeting of the American Society for Information Science, ASIS'96, Baltimore, Maryland, 21-24 Oct 1996. Ed.: S. Hardin. Medford, NJ: Learned Information 1996. S. 218-225.
- Sen, B. K.: An inquiry into the information retrieval efficiency of LISA PLUS database. In: Malaysian journal of library and information science. 1 (1996), Nr. 1, S. 67-84.
- Shakir, H. S. , M. Nagao: Context-sensitive processing of semantic queries in an image database system. In: Information processing and management. 32 (1996), Nr. 5, S. 573-600.
- Singhal, A. u. a.: Document length normalization. In: Information processing and management. 32 (1996), Nr. 5, S. 619-633.
- Smeaton, A. F., L. Kelledy u. R. O'Donnell: TREC-4 experiments at Dublin City University: thresholding posting lists, query expansion with WordNet and POS tagging of Spanish. In: The Fourth Text Retrieval Conference (TREC-4). Ed.: K. Harman. Gaithersburgh, MD: National Institute of Standards and Technology 1996. S. 373-389. (NIST special publication; 500-236)

Spink, A., A. Goodrum: A study of search intermediary working notes: implications for IR system design. In: Information processing and management. 32 (1996), Nr. 6, S. 681-695

Srinivasan, P.: Optimal document-indexing vocabulary for MEDLINE. In: Information processing and management. 32 (1996), Nr. 5, S. 503-514.

Strzalkowski, T., J. P. Carballo: Natural language information retrieval: TREC-4 report. In: The Fourth Text Retrieval Conference (TREC-4). Ed.: K. Harman. Gaithersburgh, MD: National Institute of Standards and Technology 1996. S. 245-258. (NIST special publication; 500-236)

Su, L. T.: Value of search results as a whole as a measure of information retrieval performance. In: Global complexity: information, chaos and control. Proceedings of the 59th Annual Meeting of the American Society for Information Science, ASIS'96, Baltimore, Maryland, 21-24 Oct 1996. Ed.: S. Hardin. Medford, NJ: Learned Information 1996. S. 226-237.

Tague-Sutcliffe, J. M. (Hrsg.): Evaluation of information retrieval systems: special topic issue. [Themenheft] In: Journal of the American Society for Information Science. 47 (1996), Nr. 1, S. 1-105.

Tague-Sutcliffe, J. M.: Some perspectives on the evaluation of information retrieval systems. In: Journal of the American Society for Information Science. 47 (1996), Nr. 1, S. 1-3.

Wilbur, W. J.: Human subjectivity and performance limits in document retrieval. In: Information processing and management. 32 (1996), Nr. 5, S. 515-527.

Wolfram, D., A. Volz u. A. Dimitroff: The effect of linkage structure on retrieval performance in a hypertext-based bibliographic retrieval system. In: Information processing and management. 32 (1996), Nr. 5, S. 529-541.

Wood, F., N. Ford u. D. Miller u. a.: Information skills, searching behaviour and cognitive styles for student-centred learning: a computer-assisted learning approach. In: Journal of information science. 22 (1996), Nr. 2, S. 79-92.

1997

Cavanagh, A. K.: A comparison of the retrieval performance of multi-disciplinary table-of-contents databases with conventional specialised databases. In: Australian academic and research libraries. 28 (1997), Nr. 2, S. 142-154.

Clarke, S. J., P. Willett: Estimating the recall performance of Web search engines. In: Aslib proceedings. 49 (1997), Nr. 7, S. 184-189.

Cleverdon, C. W., J. Mills: The testing of index language devices. In: From classification to 'knowledge organization': Dorking revisited or 'past is prelude'. A collection of reprints to commemorate the forty year span between the Dorking Conference (First International Study Conference on Classification Research 1957) and the Sixth International Study Conference on Classification Research (London 1997). Ed.: A. Gilchrist. The Hague: International Federation for Information and Documentation (FID) 1997. S. 65-83. (FID publication; Nr. 714) (FIC occasional paper; Nr. 14)

Davis, C. H.: From document retrieval to Web browsing: some universal concerns. In: Journal of information; communication; and library science. 3 (1997), Nr. 3, S. 3-10.

Frisch, E., M. Kluck: Pretest zum Projekt German Indexing and Retrieval Testdatabase (GIRT) unter Anwendung der Retrievalsysteme Messenger und freeWAIS-sf. Bonn: IZ Sozialwissenschaften 1997. 66 S. (IZ-Arbeitsbericht; Nr. 10)

Gödert, W., M. Liebig: Maschinelle Indexierung auf dem Prüfstand: Ergebnisse eines Retrievaltests zum MILOS II Projekt. In: Bibliotheksdienst. 31 (1997) Nr. 1, S. 59-68.

- Harter, S. P., C. A. Hert: Evaluation of information retrieval systems: approaches, issues, methods. In: Annual review of information science and technology. 32 (1997), S. 3-94.
- Huang, M.-H.: The evaluation of information retrieval systems. In: Bulletin of the Library Association of China. 1997, Nr. 59, Dec., S. 109-126.
- Lespinasse, K.: TREC: une conference pour l'evaluation des systemes de recherche d'information. In: Documentaliste. 34 (1997), Nr. 2, S. 74-81.
- Robertson, S. E., S. Walker u. M. Beaulieu: Laboratory experiments with Okapi: participation in the TREC programme. In: Journal of documentation. 53 (1997), Nr. 1, S. 20-34.
- Shaw, W. M., R. Burgin u. P. Howell: Performance standards and evaluations in IR test collections: vector-space and other retrieval models. In: Information processing and management. 33 (1997), Nr. 1, S. 15-36.
- Smeaton, A. F., D. Harman: The TREC experiments and their impact on Europe. In: Journal of information science. 23 (1997), Nr. 2, S. 169-174.
- Sparck Jones, K.: Reflections on TREC. In: From classification to 'knowledge organization': Dorking revisited or 'past is prelude'. A collection of reprints to commemorate the forty year span between the Dorking Conference (First International Study Conference on Classification Research 1957) and the Sixth International Study Conference on Classification Research (London 1997). Ed.: A. Gilchrist. The Hague: International Federation for Information and Documentation (FID) 1997. S. 101-125. (FID publication; Nr. 714) (FIC occasional paper; Nr. 14)
- Spink, A., H. Greisdorf: Users' partial relevance judgements during online searching. In: Online and CD-ROM review. 21 (1997), Nr. 5, S. 271-280.
- Spink, A., T. Saracevic: Interaction in information retrieval: selection and effectiveness of search terms. In: Journal of the American Society for Information Science. 48 (1997), Nr. 8, S. 741-761.
- Voorbij, H.: Een goede titel behoeft geen trefwoord, of toch wel?: een vergelijkend onderzoek titelwoorden - trefwoorden. In: Informatie professional. 1 (1997), Nr. 11, S. 23-27.
- Voorbij, H.: Titelwoorden -trefwoorden: een vergelijkend onderzoek. Den Haag: Koninklijke Bibliotheek 1997. 46 Bl.
- Wan, T.-L., M. Evens u. Y.-W. Wan u. a.: Experiments with automatic indexing and a relational thesaurus in a Chinese information retrieval system. In: Journal of the American Society for Information Science. 48 (1997), Nr. 12, S. 1086-1096.
- Wolfram, D., A. Dimitroff: Preliminary findings on searcher performance and perceptions of performance in a hypertext bibliographic retrieval system. In: Journal of the American Society for Information Science. 48 (1997), Nr. 12, S. 1142-1145.
- 1998**
- Chen, H., J. Martinez u. A. Kirchhoff u. a.: Alleviating search uncertainty through concept associations: automatic indexing, co-occurrence analysis, and parallel computing. In: Journal of the American Society for Information Science. 49 (1998), Nr. 3, S. 206-216.
- Gilchrist, A.: Research and consultancy. In: Library and information work worldwide 1998. Ed.: M. B. Line et al. London: Bowker-Saur 1998. S. 201-231.
- Gillman, P.: Text retrieval. In: IT Link. 10 (1998), Nr. 6, S. 10-12.
- Harman, D.: The Text REtrieval Conferences (TREC's): providing a test-bed for information retrieval systems. In: Bulletin of the American Society for Information Science. 24 (1998), April/May, S. 11-13.
- Khan, K., C. Locatis: Searching through cyberspace: the effects of link display and link density on information retrieval

from hypertext on the World Wide Web.
In: Journal of the American Society for
Information Science. 49 (1998), Nr. 2,
S. 176-182.

Knorz, G.: Testverfahren für intelligente
Indexierungs-und Retrievalsysteme an-
hand deutsch-sprachiger sozialwissen-
schaftlicher Fachinformation (GIRT):
Bericht über einen Workshop am 12.

September 1997 im IZ Sozialwissen-
schaften, Bonn. In: nfd Information -
Wissenschaft und Praxis. 49 (1998) Nr.
2, S. 111-116.

Ronthaler, M., H. Zillmann: Literatur-
recherche mit OSIRIS: ein Test der
OSIRIS-Retrievalkomponente. In:
Bibliotheksdienst. 32 (1998) Nr. 7,
S. 1203-1212.