



Die GIRT-Testdatenbank als Gegenstand informationswissenschaftlicher Evaluation

Michael Kluck

Informationszentrum Sozialwissenschaften (IZ)
Lennéstr. 30
D-53113 Bonn
kluck@bonn.iz-soz.de

Zusammenfassung:

Die Motive für die Einrichtung der GIRT-Testdatenbank sowie ein Überblick die einzelnen Versionen des GIRT-Korpus und ihre Verwendungszusammenhänge werden beschrieben. Die Nutzung der GIRT-Daten in verschiedenen informationswissenschaftlichen Kontexten von 1997 bis 2003 wird dargestellt und die erzielten Ergebnisse skizziert. Schließlich wird eine Zusammenfassung der Trends in den mit GIRT durchgeführten Evaluationskampagnen gegeben.

Abstract:

The motivations behind the creation of the GIRT test database are described and an overview of the structure of the different versions of GIRT and their use is given. The way in which GIRT has been employed in various information science contexts from 1997 to 2003 is then illustrated with a short description of methods and procedures used. The paper concludes with a summary of the trends in the GIRT tracks of these evaluation campaigns.

1 Die GIRT-Daten im Kontext des Cross-Language Information Retrieval (CLIR)

Ausgangspunkt für die Entwicklung des GIRT-Testkorpus (GIRT = German Indexing and Retrieval Testdatabase) waren die folgenden Überlegungen: GIRT sollte den Rahmen für einen aussagefähigen Vergleich moderner Retrievalsysteme schaffen, auf dessen Basis die Leistungsfähigkeit dieser Systeme gegenüber herkömmlichen Standardsystemen, wie sie bisher vom IZ



bzw. den kommerziellen Anbietern der IZ-Datenbanken eingesetzt werden, beurteilt werden kann [Knorz 98].

Die bisher umfangreichste Serie von Systemevaluationen wurde von der Initiative TREC¹ vorgelegt und seit 1992 jährlich wiederholt. Deren Ergebnisse ließen sich jedoch in zweierlei Hinsicht nur bedingt auf die Situation der deutschsprachigen Fachinformation (und damit des IZ Sozialwissenschaften) übertragen: Zum einen wurde in TREC als Testdatenbestand eine englischsprachige Pressedatenbank zur Verfügung gestellt, die andere Anforderungen an die Recherche stellt als das Referenzretrieval in einer Literaturnachweisdatenbank. Vergleichbare Testreihen an einem Datenbestand aus dem Bereich der Fachinformation mit seiner Problematik einer domänenspezifischen Terminologie lagen bis dahin nicht vor. Zum anderen waren Studien an deutschsprachigem Material ebenfalls noch ein Desiderat, so dass die an englischen Texten gewonnenen Ergebnisse von TREC hinsichtlich der Leistungsfähigkeit der Morphologie-Komponenten maschineller Systeme noch an deutschsprachigem Material überprüft werden mussten. Ferner fehlten Erfahrungen beim Umstieg von monolingualen (meist englischsprachigen) Retrievalsystemen auf mehrsprachige Retrievalsysteme (CLIR-Systeme) [vgl. Womser-Hacker 96, 19].

Die Erstellung der GIRT-Testdatenbank sollte einen Beitrag zur Behebung dieser angesprochenen Defizite leisten und damit eine valide Grundlage für einen Vergleich verschiedener Retrievalsysteme und -techniken liefern. Die Notwendigkeit, die eigenen Ressourcen optimal einzusetzen und sich dabei am Stand der Technik und der Forschung zu orientieren, ließ es sinnvoll erscheinen, dass sich Fachinformationseinrichtungen, und damit auch das Informationszentrum Sozialwissenschaften (IZ) auf empirischer Basis über die gegenwärtigen Möglichkeiten informieren und die Weiterentwicklung solcher Systeme stimulieren. Es galt die Vorteile und Problembereiche der verschiedenen Systeme im praktischen Test kennen zu lernen und Entscheidungskriterien für die Auswahl und Kombination verschiedener Ansätze bzw. Module zu entwickeln [vgl. Womser-Hacker 96, 319 ff].

Bei dem Angebot des IZ an die Softwareproduzenten und -entwickler, im Rahmen des CLIR-Task von TREC und später des Cross Language Evaluation Forum (CLEF) anhand der Datenbank GIRT Testreihen

¹ Die Text REtrieval Conference (TREC) ist eine Initiative des National Institute for Standards and Technology (NIST) in Gaithersburg (MD, USA), siehe: trec.nist.gov [eingesehen: 05.07.2004].

durchzuführen, wurden den zu testenden Retrievalsystemen Dokumente mit inhaltstragenden Informationsfeldern zur Verfügung gestellt. Damit konnte hinsichtlich der Indexierung von Dokumenten getestet werden, ob die Einbeziehung der Indexierungsbegriffe bei bestimmten Testsystemen zu einer nennenswerten Steigerung der Leistungsfähigkeit führt². Und zwar sollte verglichen werden, welche Auswirkungen die reine automatische Indexierung, die von einigen der Testsystemen nach unterschiedlichen Algorithmen vorgenommen wird, oder andererseits die Einbeziehung der vorab erfolgten intellektuell durchgeführten Indexierung auf die Resultate haben.

Ferner war zu berücksichtigen, dass die besondere Bedeutung von Fachsprachen entsprechende Anforderungen an die Indexierungs- und Retrievalsysteme stellt. Detaillierte Forschungen zeigen die besondere Schwierigkeit, Fachbegriffe der Soziologie von allgemeinsprachlichen Begriffen zu differenzieren, deutlich auf: „its [sociology] words are common words that are in general use such as community and immigrant“ [Haas 97, S. 78]. Es gibt neben der hohen Überschneidung der Fachbegriffe der Sozialwissenschaften mit Wörtern der Allgemeinsprache in vielen Fällen eine deutliche Abweichung der Bedeutung der Fachbegriffe von der allgemeinsprachlichen Verwendung und starke Bedeutungsunterschiede der Fachbegriffe an sich aufgrund inhaltlicher Konnotationen wie „Schulen“, Theorien, politische Implikationen, ethische Grundüberzeugungen usw., die eine Verwendung automatisch generierter Schlagwörter für die Indexierung und die Recherche erschweren³.

2 Die Versionen des GIRT-Korpus – Struktur und Inhalt

Generell bestehen die GIRT-Daten aus Dokumenten, die aus den Datenbanken des IZ entnommen worden sind, d.h. es werden Daten zur Verfügung gestellt, die für fachwissenschaftliche Zwecke gesammelt,

² Entsprechende Hinweise auf eine Verbesserung der Leistung durch Einbeziehung der intellektuellen Indexierung finden sich bereits bei TREC-4: „... the manually indexed fields were included in the test documents, though not in the learning documents. As can be seen, there is a 5% to 6% improvement ..., which is reasonably substantial.“ [Buckley et al. 96, 34]

³ siehe [Haas 97, S. 74]: „T tests between discipline pairs showed that physics, electrical engineering, and biology had significantly more domain terms in sequences than history, psychology, and sociology (...) the domains with more term sequences are those which may be considered the hard sciences, while those with more isolated domain terms tend to be the social sciences and humanities.“

aufbereitet und öffentlich gegen Entgelt angeboten werden. Die GIRT-Daten stammen also aus dem Bereich der Fachinformation und sind im Original deutschsprachig⁴; allerdings wurden die Titel, die Abstracts bzw. Inhaltsangaben und die Deskriptoren der meisten neueren Dokumente ins Englische übersetzt, um den internationalen Zugriff zu erleichtern. Diese Dokumentenauszüge aus den Datenbanken FORIS (Forschungsinformationssystem Sozialwissenschaften)⁵ und SOLIS (Sozialwissenschaftliches Literaturinformationssystem)⁶ des IZ wurden in einer integrierten Testdatenbank für den Test intelligenter Retrievalsysteme zur Verfügung gestellt. Für alle GIRT-Versionen wurden mindestens die folgenden Auszüge aus den Originaldokumenten der Ursprungsdatenbanken bereit gestellt: Autor, deutscher Titel, Sprache des Dokuments, Erscheinungsjahr. Zusätzlich liegen für alle Dokumente intellektuelle Indexierungen (Schlagwörter) und Klassifizierungen (Klassifikationstexte) sowie deutschsprachige Kurzreferate oder Inhaltsangaben vor. Detailliertere Angaben zu einzelnen Varianten des GIRT-Korpus sind in einem IZ-Arbeitsbericht beschrieben [Kluck 2003].

2.1 GIRT1

Insgesamt enthielt die erste Version der Testdatenbank GIRT ca. 13.000 Dokumente aus den Erscheinungsjahrgänge 1990 bis 1996. Thematisch stammten die Dokumente von GIRT1 aus den Dokumentmengen, die für den gedruckten sozialwissenschaftlichen Fachinformationsdienst (soFid) mit den Themen „Industrie- und Betriebssoziologie“, „Frauenforschung“ und „Migration und ethnische Minderheiten“ ausgewählt worden waren, und aus den Aufsätzen in den Heften der Zeitschriften „Kölner Zeitschrift für

⁴ Zu den besonderen Probleme der deutschen Sprache in diesem Zusammenhang siehe [Womser-Hacker 96, S. 104].

⁵ FORIS enthält umfangreiche Beschreibungen laufender, geplanter und abgeschlossener Forschungsprojekte aus allen Bereichen der Sozialwissenschaften, die in Deutschland, Österreich, der Schweiz und Liechtenstein durchgeführt werden. Die Gesamtzahl der Dokumente betrug Ende 2003 ca. 40.000, wovon ca. 6.000 aufgrund der Fortschreibung der Forschungsprojekte jährlich geändert werden. Siehe auch <http://www.gesis.org/Information/FORIS/Recherche/index.htm>.

⁶ SOLIS enthält Beschreibungen der sozialwissenschaftlichen Literatur aus den deutschsprachigen Ländern bzw. deutschsprachiger Autoren: Monographien und Sammelwerke, Zeitschriftenaufsätze und Beiträge in Sammelwerken, Forschungsberichte, Graue Literatur, Dissertationen und Habilitationen, digitale Dokumente aus dem Internet. Die Gesamtzahl der Dokumente betrug Ende 2003 ca. 285.000 bei einem jährlichen Zuwachs von ca. 12.000 Dokumenten. Siehe auch <http://www.gesis.org/Information/SOLIS/index.htm>.

Soziologie und Sozialpsychologie“ sowie „Soziale Welt“ der Jahrgänge 1990-1996. Bei einer kleinen Menge lagen zusätzlich die Volltexte zu den Dokumenten vor, die jedoch nicht in die Testdatenbank aufgenommen wurden, sondern separat zur Verfügung standen.

Vor den weiterführenden Tests wurde vorab im IZ ein Pretest mit den Retrievalsystemen Messenger (Boolesche Suche) und freeWAISsf (statistisches Verfahren nach dem Vektorraum-Modell) durchgeführt, der vornehmlich der Erprobung der Versuchsanordnung und der Mess- und Analyseverfahren diente [Frisch/Kluck 97]. Im Gegensatz zu Pressedaten, die für Testkollektionen wie TREC und für die „core tracks“ von CLEF (bilinguales und multilinguales Retrieval) herangezogen werden und die einen sehr allgemeinen, jedenfalls nicht fachwissenschaftlichen Inhalt haben⁷, handelt es sich bei GIRT um sozialwissenschaftliche Dokumente mit einem entsprechenden fachbezogenen Vokabular in den Textpassagen. Dementsprechend wurden fachbezogene Fragen an die Testdatenbank formuliert und eine fachbezogene Relevanzbeurteilung der Ergebnismengen durchgeführt. Diese Relevanzbeurteilung wurde von Fachwissenschaftlern⁸ des IZ vorgenommen. Die Dokumentmenge war bewusst klein und fachlich überschaubar gehalten worden, um die Chance zu haben, zumindest für einige der Testanfragen die gesamten Dokumentbestand bewerten zu können und somit tatsächlich alle relevanten Dokumente zu der jeweiligen Frage zu kennen, aber selbst dafür reichten leider die zeitlichen Ressourcen nicht aus.

Die Sichtung der einzelnen Testergebnisse zeigte, dass die Recherchen, die mit Messenger durchgeführt wurden, im Durchschnitt sowohl beim Recall als auch bei der Präzision bessere Ergebnisse zeitigten als die mit freeWAISsf durchgeführten. Ferner zeigte die detaillierte Durchsicht der einzelnen Recherchen, dass die Möglichkeiten der Recherche mit Messenger hinsichtlich der Nutzung des Thesaurus von den externen Versuchspersonen nicht oder nur begrenzt genutzt worden sind. Für freeWAISsf wiederum war eigentlich mit den Kurzreferaten der Dokumente eine gute Textgrundlage für die automatische Indexierung geschaffen, so dass ein höherer Recall zu erwarten gewesen wäre. Bei freeWAISsf wurde allerdings von den Versuchspersonen die spezifische Funktion des Ranking nicht strategisch

⁷. Deren Dokumente sind dementsprechend in einer mehr allgemein sprachlichen Form formuliert.

⁸ Die Recherchen und die Relevanzbeurteilung werden nicht von Studenten oder fachfremden Personen mit guter Allgemeinbildung durchgeführt, wie dies bei vielen Studien - auch aufgrund ihres allgemeinen und nicht fachbezogenen Scopes - geschieht, sondern von Fachwissenschaftlern.

genutzt. Dasselbe Verhalten der Versuchspersonen wurde auch bei den Tests an der Universität Konstanz beobachtet [Womser-Hacker 98]. Dort wurde festgestellt, dass aufgrund der Komplexität der Retrievalsysteme diese suboptimal und nicht strategisch, sondern eher „auf gut Glück“ eingesetzt wurden [vgl. auch Erfahrungen bei TREC: Vogt et al. 97].

Ein weiteres wichtiges Ergebnis war die von der Mehrzahl der Versuchspersonen explizit und massiv geäußerte Meinung, dass sie sich bei der Handhabung von Messenger sicherer gefühlt haben als bei der Verwendung von freeWAISsf. Dies wurde sowohl während des Test als auch in Nachgesprächen als erster und prägender Eindruck geäußert. Diese subjektive Einschätzung entspricht nicht dem objektiven Befund der mit freeWAISsf erzielten beachtlichen Ergebnisse, beruht aber wohl darauf, dass man sich bei einem System mit strenger Logik sicherer und - bei oder gerade wegen aller allgemeinen Unsicherheit und Unkenntnis - in einer Position der Beherrschung des Systems fühlt. Ein statistisch-mathematisches Verfahren vermittelt offensichtlich ein Gefühl der Vagheit und Unsicherheit und des Misstrauens gegenüber einem Ranking, dessen Zustandekommen man nicht durchschaut. Diese Befindlichkeit beruhte vor allem bei den Laien unter den Versuchspersonen nur auf einem subjektiven Eindruck; denn lediglich ein Retrieval-Profi kann sagen, dass er die Wirkungsweise der Booleschen Logik tatsächlich versteht und ihren Einsatz beherrscht.

Als nächstes wurden mit den GIRT1-Daten ein Vergleich zwischen FreeWAISsf und Fulcrum durchgeführt [Krause/Mutschke 99] sowie ein Vergleichstest Messenger/Fulcrum durchgeführt [Binder/Stahl/Faulbaum 00]. Ferner wurde die Daten von GIRT1 bei Evaluationen bzw. Tests an den Universitäten Regensburg [Zettel 96], Konstanz [Womser-Hacker 98, Käter/Rittberger/Womser-Hacker 99, Käter 99, Griesbaum 00], Hildesheim [Mandl 01] und Düsseldorf [Griesbaum/Rittberger/Bekavac 02] verwendet. Dabei wurden weitere Systeme getestet oder entwickelt und die GIRT1-Daten, die verwendeten Fragestellungen oder die bereits vorliegenden Relevanzbewertungen zu diesen Fragen verwendet. Auch für weitere Untersuchungen im IZ zur Visualisierung des Information Retrieval und der Suchergebnisse [Eibl 00] sowie zum Einsatz neuronaler Netze im Information Retrieval [Mandl 01] wurden die GIRT1-Daten herangezogen; diese Ergebnisse wurden als Dissertationen vorgelegt und in Buchform veröffentlicht.

2.2 GIRT2

GIRT2 umfasste die Dokumente aus GIRT1 und weitere Dokumente mit derselben inhaltlichen Begrenzung, aber erweitert um die Jahrgänge von 1978-1989, so dass nun insgesamt die Jahrgänge 1978-1996 enthalten waren. Die ca. 38.000 Dokumente von GIRT2 wurden bei TREC7 und TREC8 in den Jahren 1998 und 1999 angeboten. Das Ziel der Ausdehnung der Datenbasis war eine leicht realisierbare Vergrößerung der Dokumentmenge.

2.3 GIRT3

Neben den Informationsfeldern Autor, deutscher Titel, Sprache des Dokuments, Erscheinungsjahr, deutsches Abstract, deutsche Deskriptoren, deutsche Klassifikation standen für GIRT3 englische Titelübersetzungen (nur bei SOLIS) und englische Indexierungsbegriffe sowie bei einem Teil der Dokumente englischsprachige Kurzfassungen und freie deutschsprachige Schlagwörter zur Verfügung. GIRT3 umfasste die Dokumente von GIRT2 (was auch diejenigen von GIRT1 einschloss) und wurde inhaltlich auf alle Dokumente der Zeitspanne 1976-1996 erweitert, so dass nun ein repräsentativer Querschnitt der sozialwissenschaftlichen Forschung und Literatur dieses Zeitraums zur Verfügung stand, der nicht mehr inhaltlich beschränkt war. Neben der weiter Erhöhung der Dokumentmenge war die Steigerung der Repräsentativität für die deutschen Sozialwissenschaften das Hauptziel der Erweiterung, außerdem sollte die Nutzung englischsprachiger Informationselemente angeboten werden. Insgesamt umfasste GIRT3 76.128 Dokumente. GIRT3 wurde bei CLEF 2000, 2001 und 2002 verwendet.

Feldname	Anzahl Vorkommen Feldes	der des	in X % der GIRT3 Dokumente	Durchschnittliche Anzahl Einträge in diesem Feld pro Dokument
DOC	76.128		100,00	1,00
DOCNO	76.128		100,00	1,00
LANGUAGE	76.128		100,00	1,00
PUBLICATION YEAR	76.128		100,00	1,00
TITLE	76.128		100,00	1,00
TITLE-ENG	54.275		71,29	-
TEXT	73.291		96,27	-
TEXT-ENG	6.063		7,96	-
CONTROLLED TERM	755.333		-	9,92
FREE TERM	6.588		-	0,09
CLASSIFICATION	169.064		-	2,22
AUTHOR	126.322		-	1,66

Tabelle 1: Auszählung der Feldbelegung bei GIRT3⁹

⁹ aus [Kluck/Gey 01, 52]

2.4 GIRT4

Die Datenstruktur, die Auswahl und der Umfang der Daten wurden für GIRT4 völlig neu gestaltet. GIRT4 wurde nun in zwei nach Sprachen getrennte Korpora aufgeteilt: Deutsch (GIRT4-DE) und Englisch (GIRT4-EN). Die Gesamtzahl der Dokumente in jeder Kollektion beträgt jeweils 151.319 Dokumente, die inhaltlich identisch sind; also insgesamt 302.638 einzelne Datensätze. Auswahlkriterium war für die Dokumente, dass zu jedem der verwendeten deutschsprachigen Dokumente eine entsprechende Übersetzung zumindest des Titels vorliegen musste. Außerdem wurde der zeitliche Rahmen auf die Erscheinungsjahre 1990 – 2000 festgelegt. Zu GIRT3 ergab sich somit nur eine teilweise Überschneidung, nämlich im wesentlichen für die Erscheinungsjahre 1990-1996 (allerdings unter der oben genannten Prämisse, dass parallel zu dem deutschen ein englischer Titel vorhanden war).

GIRT4 wurde bei CLEF 2003 verwendet und wird für CLEF 2004 angeboten. Mit den pseudo-parallelen Kollektionen¹⁰ wurde dem multilingualen Charakter von CLEF verstärkt Rechnung getragen und die durch die Übersetzungen ins Englische bestehenden Voraussetzungen im Datenmaterial des IZ dezidiert genutzt. Für GIRT4-DE und GIRT4-EN ergibt sich folgende Gesamtstatistik der Feldbelegungen.

Informationsfeld	Anzahl der Eintragungen insgesamt pro Korpus	Anzahl der Eintragungen in diesem Feld pro Dokument: GIRT4-DE	Anzahl der Eintragungen in diesem Feld pro Dokument: GIRT4-EN
Dokumentnummer DE = EN	151.319	1	1
Autor DE = EN	237.301	1,75	1,75
Titel DE + EN	302.638	1	1
Deskriptoren DE = EN	1.535.709	10,15	10,15
Klassifikationstexte DE = EN	305.504	2,02	2,02
Methodendeskriptoren DE	354.968	2,35	-
Methodendeskriptoren EN	292.387	-	1,93
Abstract DE	145.941	0,96	-
Abstract EN (HT + MT)	22.058	-	0,15
Freie Schlagwörter DE	38.505	0,25	-
Methodentext DE	10.258	0,07	-

Tabelle 2: Auszählung der Feldbelegung für GIRT4

¹⁰ Die beiden Kollektionen werden hier als pseudo-parallel bezeichnet, weil die englischsprachige Variante im Wesentlichen das Resultat von Übersetzungen und nicht genuin englisch ist, und im englischen Teil wesentlich weniger text enthalten ist, da nicht alle Dokumente eine Übersetzung des deutschsprachigen Abstracts enthalten.

2.4.1 GIRT4-DE

Die folgenden mit Tags versehenen Datenfelder sind in den deutschsprachigen Daten von GIRT4 vorhanden.

DOCNO	ursprüngliche Dokumentnummer in den zugrunde liegenden Datenbanken SOLIS und FORIS
DOCID	eindeutige Identifikationsnummer, entspricht DOCNO
AUTHOR	Verfasser des Dokuments
TITLE-DE	deutschsprachiger Titel, im allgemeinen der Originaltitel, wenn der Dokumenttitel ursprünglich in deutscher Sprache vorlag; ansonsten wurde der Titel von einem Übersetzer ins Deutsche übersetzt; bei allen Dokumenten vorhanden
PUBLICATION-YEAR	Veröffentlichungsjahr
LANGUAGE-CODE	Sprachkode, hier immer DE
COUNTRY CODE	Erscheinungsland
CONTROLLED-TERM-DE	kontrollierter deutscher Deskriptor aus dem Thesaurus Sozialwissenschaften, jedes Dokument ist zumindest mit einem kontrollierten Term versehen, im Durchschnitt sind 10 Deskriptoren pro Dokument vergeben worden
METHOD-TERM-DE	kontrollierter deutscher Deskriptor für die verwendete Methode, soweit zutreffend
METHOD-TEXT-DE	deutscher Text zur verwendeten Methode und zum Forschungsdesign, soweit zutreffend (nur bei FORIS)
CLASSIFICATION-TEXT-DE	deutscher Text der Klassifikation, die dem Dokument zugeordnet worden ist, ein Eintrag ist vorgeschrieben, aber es können auch mehrere sein
FREE-TERM-DE	zusätzliche freie Terme oder Deskriptoren in Deutsch, die nicht kontrollierte Thesaurusbegriffe sind, bei weniger als 10 % der Dokumente vorhanden
TEXT-DE	Beschreibung oder Abstract des Dokumentinhalts, bei 96,4 % der Dokumente vorhanden

Tabelle 3: Informationsfelder mit Tags in GIRT4-DE

2.4.2 GIRT4-EN

Die folgenden mit Tags versehenen Datenfelder sind in den englischsprachigen Daten von GIRT4 vorhanden:

DOCNO	künstlich generierte Dokumentnummer, die nicht der Dokumentnummer des entsprechenden Dokuments in GIRT4-DE entspricht ¹¹
DOCID	eindeutige Identifikationsnummer, entspricht DOCNO
AUTHOR	Verfasser des Dokuments
TITLE-EN	intellektuelle Übersetzung des Titels ins Englische, wenn der Originaltitel nicht in Englisch vorlag, oder anderenfalls der ursprüngliche englischsprachige Titel, bei allen Dokumenten vorhanden
PUBLICATION-YEAR	Veröffentlichungsjahr
LANGUAGE-CODE	Sprachkode, hier immer EN
COUNTRY CODE	Erscheinungsland
CONTROLLED-TERM-EN	kontrollierter englischer Deskriptor aus dem Thesaurus Sozialwissenschaften, jedes Dokument ist zumindest mit einem

¹¹ Damit sollte eine direkte Identifikation zwischen den deutschsprachigen und englischsprachigen Dokumenten erschwert bzw. verhindert werden. Eine Konkordanzliste der Nummern liegt im IZ vor.

	kontrolliertem Term versehen, im Durchschnitt sind 10 Deskriptoren pro Dokument vergeben worden
METHOD-TERM-EN	kontrollierter englischer Deskriptor für die verwendete Methode, soweit zutreffend
CLASSIFICATION-TEXT-EN	englischer Text der Klassifikation, die dem Dokument zugeordnet worden ist, ein Eintrag ist vorgeschrieben, aber es können auch mehrere sein; der Text ist eine intellektuelle Übersetzung des deutschen Eintrags
TEXT-EN-HT	intellektuelle Übersetzung der Beschreibung oder des Abstracts des Dokumentinhalts ins Englische, bei 9,1 % der Dokumente vorhanden
TEXT-EN-MT	maschinelle Übersetzung der Beschreibung oder des Abstracts des Dokumentinhalts ins Englische, bei 5,5 % der Dokumente vorhanden. Diese maschinelle Übersetzung mit Hilfe von SYSTRAN ist manchmal ungenau und kann unübersetzte deutsche Wörter oder Phrasen enthalten, wenn das MT-System nicht in der Lage war, eine angemessene Übersetzung zu liefern, aber sie ist für die Suche zuverlässig genug. (nur bei FORIS)

Tabelle 5: Informationsfelder mit Tags in GIRT4-EN

3 Die für GIRT zur Verfügung gestellten Thesauri

Der für GIRT in CLEF zur Verfügung gestellte maschinenlesbare Deutsch-Englische Thesaurus Sozialwissenschaften [Schott 99a,b] umfasst die folgenden Bestandteile: die deutschen Deskriptoren mit Ober- und Unterbegriffen sowie verwandten Begriffen, die deutschen Nicht-Deskriptoren sowie die jeweiligen englischen Übersetzungen für alle Deskriptoren und einen großen Teil der Nicht-Deskriptoren. Die Nützlichkeit dieser Ressource wird auch von [Gonzalo 01] betont, der sie als mit anderen Quellen wie z.B. Lexika durchaus vergleichbar bezeichnet.

Die maschinenlesbare Deutsch-Russische Wortliste, die aus dem entsprechenden Deutsch-Englisch-Russischen Thesaurus [Basarnova et al. 97] entnommen ist, stellt die russischen Äquivalente für deutsche Deskriptoren zur Verfügung, aber keine strukturellen Thesaurusinformationen. Um den kyrillischen Zeichensatz abzubilden, ist diese Wortliste in UTF-8 kodiert worden.

4 Die Evaluationskampagnen mit den GIRT-Daten

Im Zusammenhang mit der Etablierung einer deutschsprachigen Komponente des CLIR-Tasks von TREC wurden auch die GIRT-Daten in einem Sub-Task angeboten. Nachdem die Organisatoren von TREC ihr Interesse stärker auf nicht-europäische Sprachen gerichtet hatten, wurde die Evaluation von CLIR-Systemen anhand europäischer Sprachen nach Europa verlegt und als CLEF¹² etabliert. Dabei wurde weiterhin eine inhaltliche und methodische Koordination mit TREC und der parallelen japanischen Initiative NTCIR [Kando 03] aufrecht erhalten. Teilweise wurden auch Korpora, Evaluationssoftware und Topics gemeinsam genutzt.

Über die Evaluationskampagnen von TREC und CLEF hinaus wurden die GIRT-Daten auch im Rahmen einer Evaluation des neuentwickelten Fuzzy-Rechercheassistenten der Universitätsbibliothek Bielefeld¹³ eingesetzt.

4.1 TREC

Für die internationale CLIR-Evaluation wurden erstmalig 1998 die GIRT3-Daten mit über 76.000 Dokumenten zur Verfügung gestellt, um CLIR-Systeme in einem spezifischen Fachgebiet und anhand strukturierter Daten zu testen [Harman et al. 01]. Dazu wurden 28 fachspezifische Topics (Informationsanfragen) formuliert und in 3 Sprachen (Deutsch, Englisch, Französisch) angeboten. Ferner stand der Deutsch-Englische Thesaurus Sozialwissenschaften maschinenlesbar zur Verfügung. Allerdings war die Bereitstellung sehr kurzfristig erfolgt und es hat sich im Jahre 1998 noch keine Teilnehmergruppe gefunden, die mit diese Daten gearbeitet hätte. Daher wurden dieselben Daten und Fragen auch 1999 im Rahmen des CLIR-Track von TREC angeboten [Vorhees/Harman 00]. In dieser Kampagne haben zwei Gruppen (Universität von Kalifornien, Berkeley, USA; Eurospider, Schweiz) Evaluationen mit den GIRT3-Daten durchgeführt.

Das generelle Retrievalmodell der Forschungsgruppe aus Berkeley basiert auf einem probabilistischen Verfahren [Cooper/ Chen/Gey 94]. Im GIRT-Kontext hat die Berkeley-Gruppe in starkem Maße von dem zweisprachigen deutsch-englischen Thesaurus Gebrauch gemacht und ihn sowohl zur Übersetzung der englischen Topics als auch zur Expansion der Anfragen durch das

¹² siehe www.clef-campaign.org

¹³ Der Rechercheassistent ist unter <http://www.ub.uni-bielefeld.de/rechercheassistent/> aufrufbar. Der Bericht über die Evaluation ist unveröffentlicht.

Retrievalsystem genutzt [Gey/Jiang 00]. Im Vergleich zum ebenfalls verwendeten Systran (im Internet unter Babelfish frei zugänglich) als maschinellem Übersetzungssystem wurde festgestellt, dass durch die zur Verfügung gestellten Thesaurusbegriffe eine deutliche Verbesserung (d.h. mehr als eine Verdoppelung!) der relevanten Ergebnisse gegenüber der maschinellen Übersetzung erreicht werden konnte. Dagegen hat die zusätzliche Expansion der Anfrage mit den Unterbegriffen des Thesaurus zu einer gefundenen Übersetzung eines Thesaurusbegriffes insgesamt zu einer Verringerung der Präzision des Ergebnisses geführt, wengleich unter den ersten Treffern mehr relevante Dokumente gefunden wurden.

Die Eurospider-Gruppe [Braschler et al. 00] hat den selbst entwickelten Ähnlichkeitsthesaurus [Sheridab/Ballerini 96] für Französisch-Deutsch und Englisch-Deutsch benutzt und für Französisch-Deutsch auch alternativ Systran eingesetzt. Außerdem wurden für Deutsch ein eigener Stemmer und eine Kompositazerlegung angewendet. Die in den Daten enthalten Indexierungsbegriffe aus dem Thesaurus Sozialwissenschaften wurden nicht explizit genutzt, ebenso wenig wurden die Thesaurusbegriffe in die Übersetzung einbezogen.

4.2 CLEF 2000

Im Rahmen von CLEF 2000 wurden die GIRT3-Daten für die Evaluation fachspezifischer CLIR-Systeme angeboten. In diesem Task wurden 25 Topics (Informationsanfragen) zur Verfügung gestellt, die auf Deutsch entwickelt und ins Englische und Russische übersetzt wurden. Bei CLEF 2000 hatten sich 3 Gruppen am GIRT-Task beteiligt: Xerox (Frankreich); Universität von Kalifornien, Berkeley (USA); Universität Dortmund (Deutschland).

Die Berkeley-Gruppe benutzte weiterhin ihr probabilistisches Retrievalsystem und entwickelte eine deutschen Stemmer ähnlich dem Porter-Stemmer, der bereits eine erhebliche Verbesserung der Ergebnisse brachte [Gey et al. 01]. Die Übersetzung der Informationsanfragen (Topics) wurde mit verschiedenen Verfahren durchgeführt:

- Thesaurus-Nachschau: Nutzung des Deutsch-Englischen Thesaurus, um ein Wörterbuch zu erstellen.
- Fuzzy Matching mit dem Thesaurus, um Wörter im Thesaurus zu finden, die andere Schreibweisen haben.
- Entry Vocabulary Module: ein Verfahren zur Abbildung englischer Wörter in den englischsprachigen Informationsfeldern auf deutsche Thesaurusbegriffe.

- maschinelle Übersetzung (MT) mit der Software Power Translator.

Die Kombination aller Verfahren führte zu einer erheblichen Verbesserung des Gesamtergebnisses, vor allem weil die einzelnen Verfahren bei jeweils anderen Topics zusätzliche relevante Dokumente fanden.

Die Forschungsgruppe der Universität Dortmund hatte die im Internet frei zugängliche Übersetzungsmaschine Babelfish (= Systran) für die Übersetzung der Topics eingesetzt [Gövert 01]. Dieses System hat allerdings für GIRT, also im fachspezifischen Bereich, wesentlich schlechter abgeschnitten als für den Bereich der Zeitungsartikel, wo die Performanz im multilingualen Bereich nahe dem monolingualen Resultat lag.

Die Xerox-Gruppe hatte sich auf die monolinguale Suche in deutscher Sprache konzentriert, aber keinen ausführlichen Bericht zu den Tests geliefert.

4.3 CLEF 2001

Für CLEF 2001 wurden wiederum die Daten von GIRT3 zur Verfügung gestellt. Die 25 Topics (Informationsanfragen) wurden auf Deutsch, Englisch und Russisch angeboten. In der CLEF-Kampagne des Jahres 2001 benutzte nur eine Gruppe (Universität von Kalifornien, Berkeley) die GIRT-Daten.

Die Berkeley-Gruppe[Gey/Jiang/Perelman 02] konzentrierte sich diesmal auf die bilinguale Suche in russischer Sprache und verwandte dabei die maschinelle Übersetzung (MT) von Promt und das Fuzzy Matching für den russisch-deutschen Thesaurus, um Variationen und Komposita von russischen Begriffen mit zu erfassen. Hier zeigten sich die erwartete Nützlichkeit der russischen Übersetzung des Thesaurus und der Transliteration der kyrillischen Zeichen in die lateinische Schrift [Gey 01].

4.4 CLEF 2002

Im Jahr 2002 war GIRT ein Bestandteil des Tasks “Mono- and Cross-Language Information Retrieval for Scientific Collections”, der außerdem noch die Kollektion von AMARYLLIS mit 150.000 französischen bibliografischen Dokumenten aus allen wissenschaftlichen Disziplinen umfasste. Die 25 Topics (Informationsanfragen) für GIRT wurden auf Deutsch, Englisch und Russisch angeboten. Als maschinenlesbare Daten standen wieder GIRT3, der englisch-deutsche Thesaurus und die deutsch-russische Wortliste zur Verfügung. In dieser Kampagne nahmen 4 Gruppen

am GIRT-Task teil: Universität von Kalifornien, Berkeley (USA), Universität Amsterdam (Niederlande), Universität Hildesheim (Deutschland), Xerox (Frankreich).

Die Forscher der Universität Amsterdam setzte ein CLIR-System ein, das auf dem Standard-Vektorraum-Modell basiert [Kamps/Monz/de Rijke 02]. Für GIRT wurde die Ähnlichkeit der Vektoren für das Vorkommen der Schlagwörter im Dokument und im Topic ermittelt, wobei die Vektoren auf 10 Dimensionen reduziert wurden. Die Ergebnisse dieses Verfahren erwiesen sich als unbefriedigend und ergaben nur in Kombination mit dem base run eine Verbesserung, ohne jedoch an die Werte im allgemeinen Teil von CLEF heranzureichen. Für die bilinguale Suche Englisch-Deutsch wurde das Ding-Wörterbuch der TU Chemnitz zur Übersetzung der Titel und Abstracts herangezogen.

Die Berkeley-Gruppe führte sowohl monolinguale Suchen (Deutsch) als auch bilinguale Suchen (Englisch-Deutsch und Russisch-Deutsch) durch [Petras/Perelman/Gey 02]. Für Englisch-Deutsch wurden neben der Kompositazerlegung maschinelle Übersetzungen von Systran und Power Translator verglichen, in Kombination verwendet und zusätzlich mit der Thesaurusnutzung verbunden. Für die monolinguale Suche in Deutsch erbrachten die Nutzung aller Felder bei den Dokumenten und bei den Topics deutlich bessere Ergebnisse, allerdings trat dieser Erfolg bei der bilingualen Suche Russisch-Deutsch nicht ein. Für Russisch-Deutsch wurden die maschinellen Übersetzungen von Systran und Prompt zusammen mit der Thesaurusnutzung eingesetzt.

Die Xerox-Gruppe extrahierte aus parallelen bzw. vergleichbaren Korpora bilinguale Lexika und optimierte die Werte der Konstanten in den Formeln zur Berechnung der Ähnlichkeit von Begriffen bzw. Dokumenten [Renders/Déjean/Gaussier 02]. Schließlich zeigte sich, dass eine Kombination der verschiedenen Verfahren zu einer weitere Optimierung des Ergebnisses führte.

Die Forschungsgruppe der Universität Hildesheim testete ihr adaptives System MIMOR anhand der GIRT-Daten mit einer monolingualen Suche in Deutsch [Hackl et al. 02]. Als Basis-Retrievalsystem wurde IRF verwendet, das von NIST im Rahmen von TREC entwickelt wurde und als open source zur Verfügung steht. Durch die Setzung unterschiedlicher Parameter wurde der Einsatz zweier Retrievalsysteme simuliert. Hinsichtlich der Kombination

der Ergebnisse wurde mit verschiedenen Gewichtungen experimentiert, was zu leichten Variationen der Ergebnisse führte.

4.5 CLEF 2003

Im Jahre 2003 basierte der CLEF-Task “Mono- and Cross-Language Information Retrieval for Scientific Collections” auf einer neuen, umfangreicheren Kollektion von GIRT. Dadurch dass GIRT4 in zwei getrennten Sprachversionen angeboten wurde, standen zwei parallele Korpora zur Verfügung: ein deutschsprachiger Korpus (GIRT4-DE) und eine pseudo-paralleler englischsprachiger Korpus (GIRT4-EN), der genau betrachtet eine Übersetzung des deutschen Korpus ins Englische darstellte und nicht ganz so umfangreiche Textinformationen enthielt wie der deutsche Originalkorpus. Dennoch wurden nunmehr zwei getrennte parallele Korpora in verschiedenen Sprachen angeboten, während der frühere GIRT3-Korpus eine Mischung aus Informationsfeldern mit deutschen und englischen Inhalten enthielt. Die Topics standen wieder in drei Sprachen zur Verfügung (Deutsch, Englisch, Russisch). Der Englisch-Deutsche Thesaurus und die Deutsch-Russische Wortliste standen wiederum zur Verfügung. Im Jahr 2003 nahmen 4 Gruppen am GIRT-Task teil: Universität von Kalifornien, Berkeley (USA), Fernuniversität Hagen (Deutschland), ENEA/Universität La Sapienza Rom¹⁴ (Italien), Universität Amsterdam (Niederlande).

Die Gruppe der Universität Amsterdam führte ihre Experimente des Vorjahres unter Verwendung der Retrievalsystems FlexIR weiter und verwendete das eingesetzte Vektorraum-Modell nun mit einem 100-dimensionalen Raum [Kamps et al. 04]. Außerdem wurden für das Re-Ranking ein Stemmer (allerdings noch ohne Kompositazerlegung) bzw. ein 4-gram-Modell angewendet. Mit der n-gram-Methode (hier mit n=4) werden n Zeichen lange Zeichenketten gesucht und gleiche Zeichenketten identifiziert¹⁵. Dadurch kommt diese Methode ohne Kenntnisse der jeweiligen Sprachen aus. Nachdem die Topics in die jeweiligen Sprachen der Dokumentsammlungen übersetzt wurden, konnte die 4-gram-Methode zur Auffindung gleicher Zeichenketten genutzt werden. Beide Verfahren haben sich nun als deutlich erfolgreicher als der Ansatz des Vorjahres erwiesen, insbesondere auch in der Kombination miteinander. Damit konnten

¹⁴ ENEA = Ente per le Nuove tecnologie, l’Energia e l’Ambiente, S. Maria di Galeria (Roma); Università degli Studi di Roma La Sapienza

¹⁵ Zum Beispiel wird die Phrase „information retrieval“ in folgende 4-grams zerlegt, wenn die Wortgrenzen eingehalten und die Wörter selbst einbezogen werden: information info nfor form orma rmat mati atio tion, retrieval retr etri trie riev ieva eval.

Verfahren, die für die allgemeinen Fragestellungen des multilingualen Tasks von CLEF angewendet wurden, ebenso erfolgreich auf fachspezifische Daten verwendet werden. Zusätzliche Verbesserungen der Ergebnisse konnten durch die Nutzung der in den Dokumenten enthaltenen Indexierungsbegriffe erreicht werden.

Die Universität von Kalifornien Berkeley, die seit Anbeginn an allen der TREC und CLEF-Kampagnen beteiligt war, hat alle mit GIRT4 angebotene Möglichkeiten von Sub-Tasks durchgeführt und dabei für alle Kombination nachdrücklich herausgearbeitet, dass die Einbeziehung der Thesauri zu einer deutlichen Verbesserung der Ergebnisse führte, auch wenn die öffentlich verfügbaren maschinellen Übersetzungssysteme (MT) inzwischen eine höhere Qualität als in der Vergangenheit aufwiesen. Die besten Ergebnisse erbrachte eine Kombination zweier MT-Systeme mit der Nutzung des Thesaurus. Dennoch sind auch die Abstracts von großer Bedeutung für die Ergebnisse gewesen. Die Titel der Dokumente sind meist nicht aussagekräftig genug, um den Inhalt eines Dokuments korrekt zu ermitteln, und bedürfen zumindest der Ergänzung durch Thesaurusbegriffe. Die Effekte der Nutzung von Thesaurusbegriffen wird von Berkeley weiter intensiv untersucht. „Documents that have controlled vocabulary terms added to the usual title and abstract information prove advantageous in retrieval because the thesaurus terms add valuable search terms to the index. An index containing titles, abstracts, and thesaurus terms will always outperform an index only containing title and abstract.” [Petras/Perelman/Gey 04, 243]

ENEA und Universität Rom La Sapienza [Alderuccio/Bordoni/Loretto 04] haben einen völlig anderen Ansatz als die üblichen CLIR-Verfahren gewählt, nämlich den der Datenkompression, die es ermöglichen sollte, die syntaktische und die semantische Distanz von Zeichenfolgen festzustellen, ohne überhaupt Kenntnisse über die jeweiligen Sprachen und ihre Besonderheiten vorauszusetzen oder zu verwenden. In den monolingualen Experimenten auf Deutsch und Englisch wurden zur Repräsentation der Dokumentinhalte die intellektuell vergebenen Deskriptoren (und in einem Fall das Abstract) verwendet, für die Repräsentation der Topics wurden die Felder „Title“ bzw. „Title“ und „Description“ genutzt. Die Resultate zeigten zwar einen durchaus hohen Recall, aber eine geringe Präzision. Diese unzureichenden Ergebnisse wurden von der Forschergruppe auf die Kürze der verwendeten Texte und das Fehlen einer Disambiguierung der Wörter zurückgeführt. Die Erweiterung des Verfahrens um eine Kombination der verschiedenen Ausprägungen des Systems scheint sich anzubieten, ebenfalls

wäre eine Expansion der Textbasis für Topics und Dokumente möglicherweise hilfreich.

Auch die Fernuniversität Hagen [Leveling 04] hat einen weiteren Ansatz in die CLIR-Evaluation von CLEF eingebracht, der auf einer natürlichsprachlichen Schnittstelle basiert. Für die Analyse der Texte der Topics und der Dokumente wurden vielfältige lexikalische und morphologische Informationen und Ressourcen genutzt, die insbesondere die Disambiguierung von Bedeutungen einzelner Zeichenketten und die Kompositazerlegung unterstützten. Zur Erstellung einer recherchierbaren Datenbank der GIRT4-Daten wurde die Zebra-Software eingesetzt, die eine Z39.50-Schnittstelle hat und Relevanzoperatoren und Ergebnisranking zulässt. Der Thesaurus Sozialwissenschaften wurde ebenfalls als lexikalische Ressource im MultiNet-Stil aufgebaut und genutzt. Es wurden nur monolinguale Tests in deutscher Sprache durchgeführt. In der Nutzung lexikalischer Informationen und von Hintergrundwissen wurde seitens der Fernuniversität Hagen noch ein großes Potenzial zur Steigerung der Retrievaleffektivität für automatisierte Retrievaltechniken gesehen.

4.6 CLEF 2004 bis 2007

Auch für die Kampagne CLEF 2004 wurden der pseudo-parallele Korpus GIRT4, 25 neue Topics und die deutsch-englischen bzw. deutsch-russischen Thesauri bereitgestellt. Es ist geplant, die GIRT4-Daten auch in den weiteren CLEF-Kampagnen (zunächst bis 2007) anzubieten.

5 Trends der Entwicklung von Retrievalkomponenten für fachspezifische Dokumentmengen

Schon bei den monolingualen, bilingualen und multilingualen CLEF-Tasks mit allgemeinen Datenbeständen haben sich ein robustes Stemming, die Anwendung bekannter und oft erprobter Gewichtungformeln und die automatische Expansion der Fragen (Blind Feedback) als erfolgreichste Verfahren durchgesetzt. Hinzu kommt meistens eine Kombination von Übersetzungen auf der Basis verschiedenster Quellen (maschinenlesbare Wörterbücher oder Thesauri, korpusbasierte Verfahren, maschinelle Übersetzung). Meistgenutztes System für die maschinelle Übersetzung war Systran. Ferner wird fast ausschließlich die Übersetzung der Fragen und nicht der Dokumente verwendet, soweit überhaupt ein Übersetzungsverfahren zur Anwendung kommt [Brasch-ler/Peters 04].

Auch bei den GIRT-Tasks waren diese Verfahren besonders erfolgreich. Hinzu kommen die Entwicklung spezifischer Stemmer und einer Kompositazerlegung bzw. –behandlung für die deutsche Sprache, die sich als erfolgssteigernd erwiesen haben. Ferner waren diejenigen Varianten der genannten Systeme, besonders erfolgreich, die sich der Indexierungsbegriffe in den Dokumenten bzw. des Thesaurus für die Übersetzung bedienen. Aber auch andere Verfahren, die sprachunabhängig sind, wie die n-gram-Methode oder die Datenkompression kamen zu Anwendung, allerdings mit unterschiedlichem Erfolg. Schließlich wurde auch ein Verfahren der Verarbeitung der natürlichen Sprache verwendet, das vielfältige linguistische Quellen und Verfahren genutzt hat; aber hier sind die Ergebnisse bisher noch vergleichsweise weniger überzeugend ausgefallen.

Für zukünftige Evaluationen erwarten wir uns einen gezielten Vergleich der in den beiden Sprachversionen von GIRT4 erreichten Resultate und eine Analyse der Gründe für differierende Ergebnisse [vgl. Kluck 04]. Darüber hinaus sollten Fragen der multilingualen Auslegung von Übersetzungssystemen, insbesondere im Hinblick auf Erkennung von Eigennamen, geografischen Bezeichnungen und Mehrwortbegriffen stärkere Berücksichtigung finden [vgl. Thurmair 04].

6 Literaturverzeichnis

- [Alderuccio/Bordoni/Loretto 04] Daniela Alderuccio; Luciana Bordoni; Vittorio Loretto: Data Compression Approach to Monolingual GIRT Task: An Agnostic Point of View. In: [Peters et al 04] (im Druck)
- [Basarnova et al. 97] Svetlana Basarnova; Helen Magaj; Robert Mdivani; Hannelore Schott; Dagmar Sucker (Hrsg.): Thesaurus Sozialwissenschaften Bd.1: Deutsch-Englisch-Russisch, Bd. 2: Russisch-Deutsch-Englisch, Bd. 3: Register. Bonn/Moskau: Informationszentrum Sozialwissenschaften/Institut für wissenschaftliche Information in den Gesellschaftswissenschaften (INION RadW) 1997
- [Binder/Stahl/Faulbaum 00] Gisbert Binder; Matthias Stahl; Lothar Faulbaum: Vergleichsuntersuchung MESSENGER – FULCRUM. Bonn: Informationszentrum Sozialwissenschaften 2000 (= IZ-Arbeitsbericht, Nr. 18), http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab18.pdf
- [Braschler et al. 00] Martin Braschler; Peter Schäuble; Min-Yen Kan, Judith L. Klavans: The Eurospider Retrieval System and the TREC-8 Cross-Language Track, In: Ellen M. Voorhees; Donna K. Harman (Hrsg.): The Eighth Text Retrieval Conference (TREC8) [1999]. Gaithersburg: NIST 2000, 367-376
- [Braschler/Peters 04] Martin Braschler; Carol Peters: Cross-Language Evaluation Forum: Objectives, Results, Achievements. In: Information Retrieval, 7 (2004) 1-2, 7-31

- [Buckley et al. 96] Chris Buckley; Amit Singhal; Mandra Mitra; Gerard Salton: New Retrieval Approaches Using SMART. In: Donna Harman (Hrsg.): The Fourth Text Retrieval Conference (TREC-4) [1995]. Gaithersburg 1996, 25-48
- [Cooper/Chen/Gey 94] Cooper, W., Chen, A., Gey, F.: Full Text Retrieval based on Probabilistic Equations with Coefficients fitted by Logistic Regression. In: Harman, D. K. (Hrsg.): The Second Text Retrieval Conference (TREC-2). Gaithersburg: NIST 1994, S. 57-66
- [Frisch/Kluck 97] Elisabeth Frisch; Michael Kluck: Pretest zum Projekt German Indexing and Retrieval Testdatabase (GIRT) unter Anwendung der Retrievalsysteme Messenger und freeWAISsf. Bonn: Informationszentrum Sozialwissenschaften 1997 (2.Aufl.) (IZ-Arbeitsbericht, Nr. 10)
- [Gey 01] Frederic C. Gey: Research to Improve Cross-Language Retrieval – Position Paper for CLEF. In: [Peters 01], 83-88
- [Gey et al. 01] Frederic C. Gey; Hailing Jiang; Vivien Petras; Aitao Chen: Cross-Language Retrieval for the CLEF Collections – Comparing Multiple Methods of Retrieval. In: [Peters 01], 116-128
- [Gey/Jiang 00] Frederic C. Gey; Hailing Jiang: English-German Cross-Language Retrieval for the GIRT Collection - Exploiting a Multilingual Thesaurus, In: Ellen M. Voorhees; Donna K. Harman (Hrsg.): The Eighth Text Retrieval Conference (TREC8) [1999]. Gaithersburg: NIST 2000, 301-306
- [Gey/Jiang/Perelman 02] Frederic C. Gey; Hailing Jiang; Natalia Perelman: Working with Russian Queries for the GIRT, Bilingual, and Multilingual CLEF Tasks. In: Carol Peters, Martin Braschler, Julio Gonzalo, Michael Kluck (Hrsg.): Evaluation of Cross-Language Information Retrieval Systems. Second Workshop of the Cross-Language Evaluation Forum, CLEF 2001, Darmstadt, Germany, September 3-4, 2001, Revised papers. Berlin: Springer 2002, 235-243
- [Gonzalo 01] Julio Gonzalo: Language Resources in Cross-Language Text Retrieval: A CLEF Perspective. In: [Peters 01], 36-47
- [Gövert 01] Norbert Gövert: Bilingual Information Retrieval with HyREX and Internet Translations Services. In: [Peters 01], 237-244
- [Griesbaum 00] Joachim Griesbaum: Evaluierung hybrider Suchsysteme im WWW, Diplomarbeit Informationswissenschaft Universität Konstanz 2000
- [Griesbaum/Rittberger/Bekavac 02] Joachim Griesbaum; Marc Rittberger; Bernard Bekavac: Deutsche Suchmaschinen im Vergleich: AltaVista.de, Fireball.de, Google.de und Lycos.de. In: Rainer Hammwöhner, Christian Wolff, Christa Womser-Hacker (Hg), Information und Mobilität, Optimierung und Vermeidung von Mobilität durch Information, Proceedings des 8. Internationalen Symposiums für Informationswissenschaft, Konstanz 2002, S.201-223. http://www.inf.uni-konstanz.de/~griesbau/files/griesbaum_rittberger_bekavac.pdf
- [Haas 97] Stephanie W. Haas: Disciplinary Variation in Automatic Sublanguage Term Identification. In: Journal of the American Society for Information Science, 48, 1997, 67-79
- [Hackl et al. 02] René Hackl, Ralph Kölle, Thomas Mandl; Christa Womser-Hacker: Domain Specific Retrieval Experiments with MIMOR at the University of Hildesheim. In: [Peters et al. 03], 343-348

- [Harman et al. 01] Donna Harman; Martin Braschler; Michael Hess; Michael Kluck; Carol Peters; Peter Schäuble; Páraic Sheridan: CLIR Evaluation at TREC. In: [Peters 01], 7-23
- [Kamps et al. 04] Jaap Kamps; Christof Monz; Maarten de Rijke; Börkur Sigurbjörnsson: Language-Dependent and Language-Independent Approaches to Cross-Lingual Information Retrieval. In: [Peters et al 04] (im Druck)
- [Kamps/Monz/de Rijke 02] Jaap Kamps; Christof Monz; Maarten de Rijke: Combining Evidence for Cross-language Information Retrieval. In: [Peters et al. 03], 111-126
- [Kando 03] Noriko Kando: CLIR at NTCIR Workshop 3: Cross-Language and Cross-Genre Retrieval. In: [Peters et al. 03], 485-504
- [Käter/Rittberger/Womser-Hacker 99] Thorsten Käter; Marc Rittberger; Christa Womser-Hacker: Evaluierung der Text-Retrievalsysteme Domestic, Intelligent Miner for Text, Lars II und TextExtender. In: Wolfgang Semar; Rainer Kuhlen (Hrsg.): Information Engineering. Proceedings des 4. Konstanzer Informationswissenschaftlichen Kolloquiums (KIK '99). Konstanz: UVK 1999, 63-73
- [Käter 99] Thorsten Käter: Evaluierung des Text-Retrievalsystems "Intelligent Miner for Text" von IBM - Eine Studie im Vergleich zur Evaluierung anderer Systeme, März 1999 Universität Konstanz, <http://www.ub.uni-konstanz.de/kops/volltexte/1999/283/>
- [Kluck 04] Michael Kluck: Evaluation of Cross-Language Information Retrieval Using the Domain-Specific GIRT Data as Parallel German-English Corpus. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, Lisbon – Portugal, 26 May - 28 May 2004, Paris: ELRA - European Language Resources Association 2004, Vol. IV, 1343-1346
- [Kluck 03] Michael Kluck: Die Evaluation von Cross-Language-Retrieval-Systemen mit Hilfe der GIRT-Daten des IZ. Ein Bericht über die Entwicklung im Zeitraum von 1997 bis 2003. Bonn: Informationszentrum Sozialwissenschaften 2003 (IZ-Arbeitsbericht)
- [Kluck 98] Michael Kluck: German Indexing and Retrieval Test Data Base (GIRT): Some Results of the Pre-test. In: Dunlop, Mark D. (Hrsg.): The 20th BCS IRSG Colloquium: Discovering New Worlds of IR (IRSG-98), Grenoble, France, 25-27 March 1998, Grenoble 1998 (electronic workshops in computing) <http://ewic.bcs.org/conferences/1998/20thirsg/papers/paper5.pdf> [eingesehen 05.07.2004]
- [Kluck/Gey 01] Michael Kluck; Frederic C. Gey: The Domain-Specific Task of CLEF – Specific Evaluation Strategies in Cross-Language Information Retrieval. In: [Peters 01] 48-56
- [Knorz 98] Gerhard Knorz: Testverfahren für intelligente Indexierungs- und Retrievalsysteme anhand deutschsprachiger sozialwissenschaftlicher Fachinformation (GIRT). Bericht über einen Workshop am IZ Sozialwissenschaften, Bonn, 12. September 1997. In: LDV-Forum 14 (2), 43-56, und in: nfd - Information Wissenschaft und Praxis, 49, 1998, 2, 111-116; auch: <http://www.inf-wiss.uni-konstanz.de/RIS/confre01/tagber01/03.html>
- [Krause/Mutschke 99] Jürgen Krause, Peter Mutschke: Indexierung und Fulcrum-Evaluierung. Bonn: Informationszentrum Sozialwissenschaften 1999 (IZ-Arbeitsbericht Nr. 17), http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab17.pdf
- [Leveling 04] Johannes Leveling: University of Hagen at CLEF 2003: Natural Language Access to the GIRT4 Data. In: [Peters et al 04] (im Druck)

- [Peters 01] Carol Peters (Hrsg.): Cross-Language Information Retrieval and Evaluation. Workshop of Cross-Language Evaluation Forum, CLEF 2000, Lisbon, Portugal, September 21-22, 2000, Revised Papers. Berlin: Springer 2001
- [Peters et al. 03] Carol Peters, Martin Braschler, Julio Gonzalo, Michael Kluck (Hrsg.): Advances in Cross-Language Information Retrieval. Third Workshop of the Cross-Language Evaluation Forum, CLEF 2002, Rome, Italy, September 19-20, 2002, Revised Papers. Berlin: Springer 2003
- [Peters et al 04] Carol Peters, Martin Braschler, Julio Gonzalo, Michael Kluck (Hrsg.): Comparative Evaluation of Multilingual Information Access Systems. Fourth Workshop of the Cross-Language Evaluation Forum, CLEF 2003, Trondheim, Norway, August 17-22, 2003, Revised Papers. Berlin: Springer 2004 (im Druck]
- [Petras/Rerelman/Gey 03] Vivien Petras; Natalia Perelman; Frederic Gey: Using Thesauri in Cross-Language Retrieval for German and French. In: [Peters et al. 03], 349-362
- [Petras/Rerelman/Gey 04] Vivien Petras; Natalia Perelman; Frederic C. Gey: UC Berkeley at CLEF 2003 – Russian Language Experiments and Domain-Specific Cross-Language Retrieval. In: [Peters et al 04] (im Druck)
- [Renders/Déjean/Gaussier 02] Jean-Michael Renders; Hervé Déjean; Éric Gaussier: Assessing Automatically Extracted Bilingual Lexicons for CLIR in Vertical Domains: XRCE Participation in the GIRT Track of CLEF-2002. In: [Peters et al. 03], 363-371
- [Schott 99a] Hannelore Schott (Hrsg.): Thesaurus Sozialwissenschaften – Thesaurus for the Social Sciences [Ausgabe – Edition] 1999. [Bd. 1:] Deutsch-Englisch – German-English, [Bd. 2] Englisch-Deutsch – English-German. Bonn: Informationszentrum Sozialwissenschaften 1999
- [Schott 99b] Hannelore Schott (Hrsg.): Thesaurus Sozialwissenschaften [Ausgabe] 1999. [Bd. 1:] Alphabetischer Teil, [Bd. 2] Systematischer Teil. Bonn: Informationszentrum Sozialwissenschaften 1999
- [Sheridan/Ballerine 96] Paraic Sheridan; Jean Paul Ballerini: Experiments in Multilingual Information Retrieval Using the SPIDER System. In: Proceeding of the 19th Annual International ACM SIGIR Conference 1996, 58-65
- [Thurmair 04] Gregor Thurmair: Multilingual Content Processing. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, Lisbon – Portugal, 26 May - 28 May 2004, Paris: ELRA - European Language Resources Association 2004, Vol. V, XI-XVI
- [Vogt et al. 97] Christopher C. Vogt, Garrison W. Cottrell, Richard K. Belew, Brian T. Bartell: Using Relevance to Train Mixture of Experts. In: Donna K. Barman (Hrsg.) The Fifth Text Retrieval Conference (TREC-5) 1996. Gaithersburg: NIST 1997, S. 503-515
- [Vorhees/Harman 00] Ellen M. Vorhees; Donna K. Harman (Hrsg.): The Eighth Text Retrieval Conference (TREC8) [1999]. Gaithersburg 2000 (NIST Special Publication 500-246), at: http://trec.nist.gov/pubs/trec8/t8_proceedings.html
- [Womser-Hacker 96] Christa Womser-Hacker: Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitationsschrift. Universität Regensburg 1996
- [Womser-Hacker 98] Christa Womser-Hacker (Hrsg.) u.a.: Projektkurs Informationsmanagement: Durchführung einer Evaluierungsstudie, Vergleich der

Michael Kluck

Information-Retrieval-Systeme (IRS) DOMESTIC - LARS II - TextExtender.
Universität Konstanz 1998

[Zettel 96] W. Zettel: Indexierung auf der Basis formaler Texteingenschaften am Beispiel sozialwissenschaftlicher Volltexte. Magisterarbeit Universität Regensburg 1996