

HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKS- UND
INFORMATIONSWISSENSCHAFT

HEFT 284

**AUTOMATISCHE INDEXIERUNG VON WIRTSCHAFTSWISSEN-
SCHAFTLICHEN DOKUMENTEN**

**IMPLEMENTIERUNG UND EVALUIERUNG AM BEISPIEL DER
DEUTSCHEN ZENTRALBIBLIOTHEK FÜR WIRTSCHAFTS-
WISSENSCHAFTEN**

VON
THOMAS GROß

**AUTOMATISCHE INDEXIERUNG VON WIRTSCHAFTSWISSEN-
SCHAFTLICHEN DOKUMENTEN**

**IMPLEMENTIERUNG UND EVALUIERUNG AM BEISPIEL DER
DEUTSCHEN ZENTRALBIBLIOTHEK FÜR WIRTSCHAFTS-
WISSENSCHAFTEN**

**VON
THOMAS GROß**

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Konrad Umlauf
Humboldt-Universität zu Berlin

Heft 284

Groß, Thomas

Automatische Indexierung von wirtschaftswissenschaftlichen Dokumenten: Implementierung und Evaluierung am Beispiel der Deutschen Zentralbibliothek für Wirtschaftswissenschaften / von Thomas Groß. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2010. – 98 S. : graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 284)

ISSN 14 38-76 62

Abstract:

Die Bewertung der Indexierungsqualität bzw. –güte ist ein grundlegendes Problem von manuellen und automatischen Indexierungsverfahren. Letztere werden aber gerade im digitalen Zeitalter als einzige Möglichkeit angesehen, den zunehmenden Schwierigkeiten bibliothekarischer Informationsstrukturierung gerecht zu werden.

Diese Arbeit befasst sich mit der Funktionsweise, Implementierung und Evaluierung der Sacherschließungssoftware MindServer Categorizer, der Firma Recommind, an der Deutschen Zentralbibliothek für Wirtschaftswissenschaften (ZBW). Grundlage der maschinellen Sacherschließung und anschließenden quantitativen und qualitativen Auswertung bilden rund 39.000 wirtschaftswissenschaftliche Dokumente aus den Datenbanken Econis und EconStor. Unter Zuhilfenahme des rund 6.000 Deskriptoren umfassenden Standard-Thesaurus Wirtschaft (STW) wird der ursprünglich rein statistische Indexierungsansatz des MindServer Categorizer zu einem begriffsorientierten Verfahren weiterentwickelt und zur Inhaltserschließung digitaler Informationsressourcen eingesetzt. Der zentrale Fokus dieser Arbeit liegt vor allem auf der Evaluierung der maschinell beschlagworteten Titel, in Anlehnung und entsprechender Anpassung der von Stock (2008) und Lancaster (2003) hierzu vorgeschlagenen Kriterien: Indexierungskonsistenz, -tiefe, -breite, -spezifität, -effektivität. Zusätzlich wird die Belegungsbilanz des STW evaluiert und es erfolgt ferner eine stichprobenartige, qualitative Bewertung der Ergebnisse seitens der zuständigen Fachreferenten und -referentinnen.

Diese Veröffentlichung geht, in leicht geänderter Fassung, zurück auf eine Masterarbeit im postgradualen Fernstudiengang Master of Arts (Library and Information Science) an der Humboldt-Universität zu Berlin.

Online-Version: <http://edoc.hu-berlin.de/series/berliner-handreichungen/2010-284>

Inhaltsverzeichnis

Abbildungsverzeichnis	II
Tabellenverzeichnis.....	III
Abkürzungsverzeichnis	IV
1 Einleitung	1
2 Forschungsstand	11
2.1 Überblicks- und theoretische Arbeiten.....	11
2.2 Evaluierungsverfahren I: Retrievaltests	14
2.3 Evaluierungsverfahren II: Indexierungskonsistenz.....	16
2.3.1 Manuelle Sacherschließung.....	17
2.3.2 Manuelle vs. automatische Sacherschließung.....	21
2.4 Evaluierungsverfahren III: Sonstige.....	23
2.5 Evaluierung des MindServer	24
3 Methode, Daten, Messverfahren	27
3.1 Der MindServer als statistisches Verfahren	27
3.2 Datenlage und Stichprobenumfang	31
3.3 Zum Messverfahren: Die Evaluierung der Indexierungsqualität	32
4 Ergebnisse	39
4.1 Indexierungskonsistenz	39
4.2 Indexierungstiefe	42
4.2.1 Indexierungsbreite.....	43
4.2.2 Indexierungsspezifität	48
4.3 Indexierungseffektivität	50
4.4 Belegungsbilanz	53
4.5 Referentenauswertung.....	56
5 Fazit	61
Literaturverzeichnis.....	67
Anhang	73

Abbildungsverzeichnis

Abbildung 1: Indexierungsgütekriterien	33
Abbildung 2: Analysedimensionen	37
Abbildung 3: Anzahl der Dokumente pro Indexierungskonsistenzintervall	41
Abbildung 4: Indexierungsbreite: maschinelles Verfahren (Trainingsset)	44
Abbildung 5: Indexierungsbreite: manuelles Verfahren (Trainingsset).....	45
Abbildung 6: Indexierungsbreite: maschinelles Verfahren (Testset).....	46
Abbildung 7: Korrelation zwischen Indexierungsbreite (automatisches Verfahren) und Indexierungskonsistenz.....	47
Abbildung 8: Vergabehäufigkeit: manuelles vs. automatisches Verfahren	55
Abbildung 9: Zentrale intervenierende Variablen: qualitative Bewertung (Trainingsset).....	56
Abbildung 10: Zentrale intervenierende Variablen: qualitative Bewertung (Testset).....	59
Abbildung 11: Referentenauswertung, Trainingsset (Auszug): Kategorisierung der intervenierenden Variablen	87
Abbildung 12: Referentenauswertung, Testset (Auszug): Kategorisierung der intervenierenden Variablen	89

Tabellenverzeichnis

Tabelle 1: Indexierungsspezifität Trainingsset	48
Tabelle 2: Indexierungseffektivität der Gesamtstichprobe	50
Tabelle 3: Vergabehäufigkeit absolut und in %	53
Tabelle 4: Evaluierung automatischer Indexierungsverfahren – eine Forschungsstand- übersicht	74
Tabelle 5: Berechnung der Indexierungskonsistenz anhand der für ein Dokument vergebenen manuellen und automatischen Schlagwörter (Indexierungsbreite) und deren Übereinstimmung, Trainingsset (Auszug)	78
Tabelle 6: Dokumentenanzahl (absolut) pro Indexierungskonsistentwert (0-100%).....	79
Tabelle 7: Indexierungsbreite Testset (Auszug).....	80
Tabelle 8: Indexierungsspezifität Trainingsset (Auszug).....	81
Tabelle 9: Indexierungsspezifität Testset (Auszug).....	82
Tabelle 10: Indexierungseffektivität: automatisches vs. manuelles Verfahren, Trainings- und Testset (Auszug).....	83
Tabelle 11: Vergabehäufigkeit: automatisches Verfahren (Auszug)	85
Tabelle 12: Vergabehäufigkeit: manuelles Verfahren (Auszug).....	86

Abkürzungsverzeichnis

AIR/PHYS	Automatisches Indexierungsverfahren in Anwendung auf die Datenbank des Fachinformationszentrums Karlsruhe
AUTINDEX	Automated Indexing
BWL	Betriebswirtschaftslehre
DDC	Dewey-Dezimalklassifikation
DIN	Deutsche Industrienorm
ECONIS	Economic Information System
GBI	German Business Information, GBI-Genios Deutsche Wirtschaftsdatenbank GmbH
GBV	Gemeinsamer Bibliotheksverbund
GESIS	Leibniz-Institut für Sozialwissenschaften (ehemals Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V.)
HE	Hierarchieebene
HWWA	Hamburgisches Welt-Wirtschafts-Archiv
IAI	Institut der Gesellschaft zur Förderung der angewandten Informationsforschung e. V. an der Universität des Saarlandes
IB	Indexierungsbreite
IDF	Inverse Dokumenthäufigkeit
IE	Indexierungseffektivität
IFO	Institut für Wirtschaftsforschung
IK	Indexierungskonsistenz
IR/PR	Retrievaltest (Precision/Recall)
IS	Indexierungsspezifität
IS/A	Indexierungsspezifität automatisches Verfahren
IS/M	Indexierungsspezifität manuelles Verfahren
IS/Ü	Indexierungsspezifität Übereinstimmung
IT	Indexierungstiefe
JEL	Journal of Economic Literature Classification System
jnF	je nach Formel
k. A.	keine Angabe
KASCADE	Kataloganreicherung durch Scanning und automatische Dokumenterschließung
LINGO	Linguistisches Lego
M/A	Intraindexierungskonsistenz im Vergleich manuelles vs. automatisches Indexat
MILOS	Maschinelle Indexierung zur Literaturschließung in Online-Systemen
M/M	Intraindexierungskonsistenz im Vergleich manueller Indexate
OPAC	Online Public Access Catalogue
OSIRIS	Osnabrück Intelligent Research Information System
PASSAT	Programm zur Automatischen Selektion von Stichwörtern aus Texten
PLSA	Probabilistic Latent Semantic Analysis
PPN	Pica-Produktionsnummer
Q/S	Qualitative Einschätzung bzw. Sonstiges
SWD	Schlagwortnormdatei
STW	Standard-Thesaurus-Wirtschaft
TIB	Technische Informationsbibliothek Hannover
VWL	Volkswirtschaftslehre
WWW	World Wide Web
ZBW	Deutsche Zentralbibliothek für Wirtschaftswissenschaften – Leibniz Informationszentrum Wirtschaft

1 Einleitung

Where is the information we have lost in data? (Stefan Gradmann)¹

Die immer stärker wahrnehmbare Transformation von einer fordistisch geprägten Gesellschaftsordnung hin zu einer postindustriellen Informations- bzw. Wissensgesellschaft, verursacht durch einen immer rasanteren technologischen Fortschritt, stellt den Produktionsfaktor Information, welcher originär ein öffentliches Gut darstellt (Kuhlen 2005: 1), in den Mittelpunkt sozioökonomischer Entwicklungsprozesse (vgl. Nohr 2005: 11). Informationen sind eine entscheidende Basis zur Wissensgenerierung, weil sie dessen Konstituierung erst ermöglichen. Doch selbst der generelle Zugriff auf Informationen sichert weder einen strukturierten Zugang zu dieser noch das Verstehen derselben (vgl. Gradmann 2008). Die Ursachen hierfür liegen in der anwachsenden digitalen Datenflut begründet, in der die Informationen zunehmend für immer zu verschwinden scheinen, quasi „ertrinken“ (Puzicha 2009: 1, vgl. Schneider 2008: 9). Niemand ist nach heutigem Sachstand in der Lage, alle digitalen Daten bzw. Informationen gedanklich zu erfassen, zu verstehen und zueinander in Beziehung zu setzen (Gradmann 2008: 5). Eine strukturierte und normierte Erschließung digitaler Dokumente² ist deshalb ein entscheidender Aspekt der Zugänglichkeit und Handhabbarkeit von Informationen. Dieser auch in der Zukunft sicherlich noch zunehmende Problemdruck erfordert immer dringlicher eine qualitative Informationsstrukturierung.

Eine zentrale Rolle in diesem Informationsfindungsprozess kann dem aktuell im Entstehen befindlichen Semantic Web³ zugesprochen werden. Aufbauend auf ontologiebasierten Metadaten, werden semantische Relationen zwischen den verschiedenen WWW-Inhalten und den darin enthaltenen Informationen hergestellt, um eine bessere Wiederauffindbarkeit durch entsprechende Suchmaschinen (Information Retrieval) zu gewährleisten. Dies setzt vor allem ein kontrolliertes Vokabular voraus (Weller 2009). Mit dem Standard-Thesaurus-Wirtschaft (STW) der Deutschen Zentralbibliothek für Wirtschaftswissenschaften – Leibniz Informationszentrum Wirtschaft (ZBW) steht zur Inhaltserschließung von wirtschaftswissen-

1 Erste Konsultation zum Seminar „Wissensmanagement-Systeme und Data-Mining“, 25.10.2008, Humboldt-Universität Berlin. In Anlehnung an T. S. Eliot's Gedicht „The Rock“ (1934): [http://en.wikiquote.org/wiki/T. S. Eliot](http://en.wikiquote.org/wiki/T._S._Eliot) (Zugriff: 27.10.10).

2 Zur Kontroverse über den Dokumentenbegriff siehe z. B. Buckland (1998) oder Pédauque (2003). In dieser Arbeit wird im Folgenden von einem weiten Dokumentenbegriff ausgegangen.

3 Die heutzutage oft zu findende Kennzeichnung des Semantic Web als „Web 3.0“ wird in dieser Arbeit nicht verwendet, weil „Web 3.0“ eher einen Marketingausdruck darstellt (ähnlich „Bibliothek 2.0“, „E-Learning 2.0“ usw.) und zudem eine Linearität in der Webentwicklung unterstreicht, die in der Form nicht konstatiert werden kann (vgl. Gradmann 2009).

schaftlichen Webinhalten ein entsprechendes Begriffssystem zur Verfügung.⁴ Seit April 2009 ist der STW auch als Semantic-Web-Anwendung – „STW-Online“⁵ – im Netz zu finden und kann dort von jedermann heruntergeladen und zur Erschließung wirtschaftswissenschaftlicher Informationen genutzt werden (vgl. Neubert 2009).⁶ Durch diese semantische Einbindung eines domänenspezifischen, kontrollierten Vokabulars wird die wirtschaftswissenschaftliche Sacherschließung in einen „webbasierten Informationsraum“ (Gradmann 2009) integriert. Automatische Indexierungsverfahren eröffnen hier der traditionellen bibliothekarischen Sacherschließung in der ZBW neue Anwendungsmöglichkeiten. Aufbauend auf den Erfahrungen und Ergebnissen im Rahmen der automatischen Indexierung von Onlinedokumenten, ist es denkbar, dass über eine entsprechende Deskriptorenzuweisung mittelfristig auch digitale Inhalte aus anderen, ZBW-fremden Datenbanken unter einer einheitlichen Suchoberfläche miteinander semantisch verbunden werden. Heterogene Webinhalte werden somit einer homogenen, strukturierten Suchmöglichkeit zugeführt (Information Retrieval).

Mit Bezug auf die DIN 31 623-1 werden unter der Indexierung, die auch als Verschlagwortung oder Sacherschließung bezeichnet wird (vgl. Bertram 2005: 67), alle Methoden und Verfahren subsumiert, die genormte Metadaten (Schlagwörter, Notationen, Deskriptoren) einem Dokument zuweisen, um dessen inhaltliche Erschließung und Wiederauffindbarkeit zu gewährleisten (Nohr 2005: 24, vgl. Knorz 2004: 181). Dieser Sacherschließungsprozess verläuft immer zweistufig: Zunächst geht es in einem ersten Schritt um das Verstehen des Textes (Inhaltsanalyse), anschließend wird der erkannte Inhalt durch Zuweisung sprachlicher Elemente aus einem normierten Konzeptraum (Thesaurus, Klassifikation) entsprechend abgebildet (vgl. Nohr 2004: 216).

4 Unter einem Thesaurus wie dem STW wird ein terminologisch kontrolliertes und genormtes Begriffssystem, bestehend aus konsistent hierarchischen Bezeichnungen (oder Deskriptoren), verstanden, das umfassend ein Wissenschaftsgebiet, in diesem Falle die Betriebswirtschafts- und Volkswirtschaftslehre, abdecken bzw. beschreiben kann (vgl. Nohr 2005: 26, Brühl 2005: 16, Burkhart 2004: 141). Dies wird durch eine möglichst vollständige Erfassung der Synonyme, eine besondere Kennzeichnung der Homonyme und Polyseme und durch die Verwendung von Vorzugsbezeichnungen erreicht, die den jeweiligen Deskriptor definieren (Panyr 2006: 142). Im Gegensatz zur Schlagwortnormdatei ist durch die Anwendung des STW eine stringendere Tiefenerschließung eines Fachgebietes möglich. Weiterhin schaffen Thesauri, im Gegensatz zu entsprechenden Klassifikationssystemen, einen natürlichsprachigen Zugang zu Informationen (vgl. Stock 2000: 76).

5 Siehe: <http://zbw.eu/stw/versions/latest/about> (Version 8.06, Zugriff: 27.10.10).

6 Der ursprüngliche STW ist 1995-1997 für die Anwendung in bestimmten wissenschaftlichen Einrichtungen entwickelt worden (HWWA, ZBW, IFO, GBI) (vgl. Stock 2000: 84). Die Online-Version zielt auf eine breitere Nutzung ab. Erste Nutzerstudien zeigen eine internationale Anwendung (Schwerpunkt: Deutschland, USA, Großbritannien), hauptsächlich im Hochschulbereich (18 %), in IT-Firmen (15 %) und im Informationssektor (9 %).

Die Indexierung kann in zwei Verfahren unterschieden werden: einerseits in die manuelle (oder intellektuelle) Indexierung, die als klassische Variante der Sacherschließung heutzutage noch in vielen Bibliotheken Anwendung findet (vgl. Nohr 2004: 25). Auf der anderen Seite findet sich die automatische (oder maschinelle) Indexierung. Bei dieser Variante der Sacherschließung erfolgt die Inhaltsanalyse in einem automatischen Verfahren. Das Resultat sind entweder den Inhalt repräsentierende, reine Stichwörter – dies wird als Extraktionsverfahren bezeichnet – oder entsprechende Schlagwörter, die aus einem kontrollierten Vokabular, z. B. Thesaurus oder Klassifikation, entnommen werden – darunter werden Additionsverfahren subsumiert (Oberhauser/Labner 2003: 306, Nohr 2005: 216). Automatische Extraktionsverfahren weisen generell eine höhere Indexierungskonsistenz auf, während Additionsverfahren ein umfassendes „Verständnis“ der Indexierungssoftware für das jeweilige Fachgebiete erfordern (Luckhardt 2010, vgl. Knorz 2004: 183). Wie im weiteren Verlauf dieser Arbeit noch ausführlich dargestellt, findet in der ZBW bei der Indexierung von digitalen Dokumenten ein Additionsverfahren Anwendung.⁷

Automatische Indexierungsverfahren, die seit den 1990er Jahren im deutschsprachigen Raum verstärkt konzipiert und mit unterschiedlichem Erfolg in der Praxis eingesetzt worden sind (u. a. MILOS I und II, AUTINDEX, KASCADE, OSIRIS), haben sich in den letzten Jahren entsprechend weiterentwickelt und werden heutzutage als sinnvolle, unter dem Gesichtspunkt des zukünftig zu erwartenden digitalen Publikationsaufkommens zudem als einzig gangbare Alternative zur manuellen Inhalterschließung angesehen (u. a. Oberhauser/Labner 2003: 3, Siegmüller 2007: 8).

In den Jahren 2005-2009 wurden in der ZWB jedes Jahr im Mittel rund 13.300 ausschließlich als Onlineausgabe vorliegende Arbeitspapiere katalogisiert. Zusätzlich existieren zahlreiche Parallelausgaben, gerade im Periodikabereich (Zeitschriften, Working Paper, Jahrbücher), die einer maschinellen Indexierung zugeführt werden können. Weiterhin ist der Trend zu reinen Onlineausgaben aus den Erfahrungen der täglichen Bibliotheksarbeit heraus ungebrochen. Die zu verzeichnende, stetige und überproportionale Zunahme an (ausschließlich) elektronisch vorhandenen Dokumenten, bei gleichzeitig stagnierenden bzw. knapper werdenden personellen Ressourcen, erschwert eine flächendeckende, vor allem aber zeitnahe und qualitativ hinreichende Sacherschließung der eingehenden Literatur (vgl. Hübner/Groth

⁷ Während automatische Sacherschließungsverfahren ohne manuellen Eingriff ablaufen können, findet in der ZBW ein semiautomatisches Verfahren Anwendung. Hierbei werden maschinell erstellte Schlagwörter anschließend manuell überprüft und gegebenenfalls angepasst (vgl. Bertram 2005: 98).

2004: 2, Stock 2000: 172). Damit einher erodiert die für die Zukunft immer wichtiger werdende inhaltliche Strukturierung von Information(en), weil es weder für den Einzelnen noch eine Gruppe, sei es innerhalb der ZBW oder im Bibliotheksverbund (hier: GBV), möglich ist, diese Vielzahl an Informationen gedanklich zu erfassen und entsprechend zu strukturieren (vgl. Gradmann 2008). Zudem ist die manuelle Sacherschließung nicht skalierbar (Puzicha 2009: 10), denn mit steigenden Eingabemengen (= zunehmender Bestand an Onlinedokumenten) steigt *proportional* auch der personelle Ressourcenbedarf an Indexierern, wenn der Qualitätsstandard beim Indexieren nicht gesenkt werden soll.

Die Anwendung automatisierter Verfahren ergibt sich nicht nur aus der gerade beschriebenen quantitativen Notwendigkeit, die Unumgänglichkeit resultiert zudem aus der Anwendungsdiskrepanz des STW. Die Erschließung von digitalen und analogen Dokumenten auf Basis eines kontrollierten Vokabulars und unter Anwendung von Sacherschließungsregeln (s. ZBW 2007) wird momentan ausschließlich manuell von entsprechend ausgebildeten FachreferentInnen⁸ nach Lektüre der eingehenden Literatur vorgenommen. Aufgrund der Tatsache, dass Indexierung ein genuin subjektiver Prozess ist (Xu 2007: 18), der von zahlreichen persönlichen Faktoren abhängig ist (u. a. Qualifikation, Tagesform, Medium), ergeben sich zwischen der Anwendungstheorie (Sacherschließungsregeln) des STW und der Indexierungspraxis selbstverständlich Heterogenitäten. Zum einen zwischen den einzelnen Indexaten eines Sacherschließers und zum anderen auch auf der Metaebene in Bezug auf die Sacherschließung des Bestandes allgemein. Hier kann eine maschinelle Indexierung helfen, die bestehenden Unterschiede zu nivellieren, und zu einer homogeneren Erschließung des Bestandes beitragen (vgl. Lingelbach-Hupfauer/Laute 2009: 50).

In dieser Masterarbeit wird die Implementierung eines automatischen Indexierungsverfahrens am Beispiel der ZBW dargestellt. Zentraler Aspekt der folgenden Kapitel und deren Ausführungen ist die qualitative Bewertung bzw. Evaluierung der maschinellen Sacherschließungsergebnisse.⁹ Es geht zentral darum, die automatisch generierten Ergebnisse mit Hilfe eines kontrollierten Vokabulars in ein ontologiebasiertes Begriffssystem, hier den STW, einzuordnen. Dieses soll helfen, digitale Dokumente nicht nur sprachoberflächlich – im Sinne einer Kataloganreicherung mit Hilfe ermittelter Keywords – zu beschreiben, sondern auf die

8 Werden keine Personenbezeichnungen verwendet, die beide Geschlechter mit einschließen, z. B. aus Gründen der besseren Lesbarkeit, so schließt dies selbstverständlich das andere Geschlecht jeweils mit ein.

9 Technische Fragen (Anforderungen und Anpassungsprobleme auf Basis der EDV-Infrastruktur) oder ökonomische Aspekte (Effizienzbetrachtungen i. S. v. Kosten-Nutzen-Analysen) sind nicht Gegenstand dieser Arbeit.

Bedeutungsebene von Inhalten abzielen (vgl. Nohr 2005: 25).¹⁰ Es werden demnach mit den zugewiesenen Deskriptoren qualitative Metadaten geschaffen, die den digitalen Inhalten der jeweiligen Dokumente einen informatorischen Mehrwert zuweisen (= Additionsverfahren) (vgl. Stock/Stock 2008: 349). Diese Fokussierung auf die Bedeutungsebene führt dazu, dass das vorhandene Information-Retrieval (OPAC) nicht nur eine Datenanreicherung erfährt, sondern dass darüber hinaus auch die informationsstrukturierenden Vorteile eines Thesaurus ausgenutzt werden.¹¹ Erst die Verbindung von Inhalts- und Bedeutungsebene der digitalen Dokumente durch eine automatische Indexierung unter Zugrundelegung des STW ermöglicht eine Einbindung dieser ins Semantic Web. Durch den Einsatz maschineller Verfahren kann die ZBW auch in Zukunft sicherstellen, dass die stetig zunehmende Anzahl digitaler Dokumente einer zeitnahen, flächendeckenden, qualitativ hochwertigen und konsistenten inhaltlichen Erschließung zugeführt werden.

Mit der verstärkten Erprobung maschineller Sacherschließungsverfahren Mitte der 1990iger Jahre wuchs auch die Zahl der darauf aufbauenden Untersuchungen und Veröffentlichungen. Zahlreiche Fallstudien aus dem Bibliotheks- und Informationssektor verdeutlichen die Aktualität dieser Thematik (u. a. Baum 2005, Leder 2002, Mittelbach/Probst 2006, Schek 2005, Sachse 1998, Schiffer 2007, Schwarz 2004). In den vorliegenden Arbeiten sind qualitative Aspekte von maschineller Indexierung immer, in unterschiedlicher Art und Weise, Analysegegenstand gewesen, sei es im Rahmen eines Vor- und Nachteilvergleiches (siehe Hübner/Groth 2004) oder um die Möglichkeiten und Grenzen dieser Verfahren deutlicher herauszuarbeiten (Mittelbach/Probst 2006).

Diese Masterarbeit ist gleichfalls als Fallstudie angelegt, deren zentrales Merkmal eine umfassende Beschreibung eines spezifischen Gegenstandes darstellt. Hierbei werden die Teile eines Gegenstands durch die starke personalisierte Fokussierung auf denselbigen umfassend dargestellt und zueinander in Beziehung gesetzt (vgl. Nohlen/Schultze 2005: 239ff.). Es gilt demnach, im Rahmen dieser Arbeit, das Typische am Einsatz eines automatischen Sacherschließungsverfahrens am Beispiel der ZBW als weltweit größter Spezialbibliothek für

10 Während eine Klassifikation (siehe bspw. JEL) eher auf einer allgemeineren Ebene verbleibt (z. B. Spieltheorie, Industriepolitik), versucht ein Thesaurus differenzierter zu erschließen (dynamische Spieltheorie, Bergbaupolitik).

11 Beispielsweise kann das automatische Verfahren zwar die Wörter „Müller“ oder „Essen“ generieren, aber erst die Kontextualisierung dieser Information in semantischen Beziehungssystemen wie dem STW ermöglicht eine zielgerichtete Suche, z. B. im Falle von „Essen“ nach der Stadt oder den Nahrungsmitteln. Während momentan der Nutzer diesen Gedankengang bei der Suchanfrage berücksichtigen muss, kann das Semantic Web dies erkennen und automatisch umsetzen.

wirtschaftswissenschaftliche Literatur herauszuarbeiten. Der Schwerpunkt liegt hierbei auf der Evaluierung des automatischen Indexierungsverfahrens unter qualitativen Gesichtspunkten.

In diesem Zusammenhang wird dieser Masterarbeit analytisch folgende forschungsleitende Hypothese zugrunde gelegt: Die automatische Indexierung digitaler Dokumente wird sich in Bezug auf die Indexierungsqualität den manuellen Sacherschließungsergebnissen annähern, ohne ihnen völlig zu entsprechen.¹² Hierbei stellen die im späteren Verlauf dieser Arbeit dargestellten Indexierungsergebnisse des automatischen Sacherschließungsverfahrens die abhängige Variable dar. Als unabhängige Variablen werden folgende Erklärungsfaktoren herangezogen (in Anlehnung an Lancaster 2003, Stock/Stock 2008: 355): die Anzahl der übereinstimmenden Deskriptoren auf Dokumentenebene (= Indexierungskonsistenz), die Anzahl der verwendeten Deskriptoren (= Indexierungsbreite), der Umfang und die Hierarchie des STW-Vokabulars (= Indexierungsspezifität), die Trennschärfe der Deskriptoren im zugrunde gelegten Datensatz (= Indexierungseffektivität), die Erfassung von Haupt- vs. Randthemen (goal/sub-goal-Problem), die Fokussierung auf sinntragende Dokumentelemente (Abstract, Titel, Keywords) sowie domänenspezifische Aspekte.¹³

Die zentrale, diese Masterarbeit begleitende Fragestellung lautet: Führt der Einsatz eines automatischen Indexierungsverfahrens einerseits zu einer qualitativ hochwertigen Sacherschließung und erfolgt in Bezug auf den gesamten Bestand andererseits eine konsistente Informationserschließung und -strukturierung der digitalen Dokumente? Für den empirischen Teil dieser Masterarbeit lassen sich daraus folgende, erkenntnisleitende Unterfragen ableiten: Ist eine maschinelle Indexierung unter Zuhilfenahme eines kontrollierten, polyhierarchischen Vokabulars, wie des STW, generell möglich, oder können mit diesem Verfahren nur klassifikatorische Ergebnisse i. S. flacher Begriffshierarchien (vgl. SWD, JEL, DDC) erzielt werden? Wie viele Deskriptoren werden von der Indexierungssoftware pro Dokument generiert und welche hierarchische Position haben diese im Thesaurusbaum?

12 Die Indexierungssoftware soll durch die Vorgabe bereits erschlossener Dokumente so trainiert werden, dass sie dem durchschnittlichen Indexierungsverhalten einer heterogenen Indexierergruppe (Anzahl, Qualifikation) entspricht. Es ist zu erwarten, dass Indexierungsfehler Einzelner durch die Gruppe eher ausgeglichen werden. Gleichzeitig können systematische Verwerfungen, d. h. Fehler, die von vielen Personen gemacht werden, die Maschine natürlich auch fehltrainieren.

13 Nicht in Betracht kommen folgende Erklärungsfaktoren, die Lancaster (2003: 70ff.) zusätzlich aufführt: kontrolliertes vs. freies Vokabular (mit dem STW kommt ein kontrolliertes Vokabular zum Einsatz), der Objektcharakter (hier: überwiegend Working Paper), die Terminologie (hier: Wirtschaftswissenschaften), die Indexierungshilfsmittel und der Umfang des zu indexierenden Dokuments. Stellschrauben der eingesetzten Indexierungssoftware (z. B. Titel-/Abstractgewichtung, Lerndokumentenanzahl) bleiben wegen des momentanen Projektcharakters des automatischen Verfahrens an der ZBW gleichfalls unberücksichtigt. Hier bieten sich entsprechende Anschlussstudien an.

Sind die vom automatischen Verfahren durchschnittlich vergebenen Deskriptoren, auch wenn sie nur einen Ausschnitt aus dem gesamten Thesaurusbaum (mit seinen ca. 6.000 Begriffen) darstellen, nicht ausreichend, um ein Dokument umfassend zu beschreiben, weil sie dem manuellen Indexierungsverhalten nahekommen, das gleichfalls nur einen Teil der möglichen Begriffe benutzt (vgl. Zipf's law¹⁴)? Wie gestaltet sich darauf aufbauend dann die Trennschärfe der einzelnen Deskriptoren im Hinblick auf ihr Vorkommen im gesamten Datensatz? Werden die Hauptthemen als zentrale Inhalte der einzelnen Dokumente durch die Indexierungssoftware erfasst und inwieweit werden Randthemen auch abgebildet (goal-/sub-goal-Problem)? Welche Rolle spielen sinntragende Dokumententeile (Titel, Abstract, Keywords) für das Indexierungsergebnis? Können domänenspezifische Muster innerhalb und zwischen den Indexaten erkannt werden? Inwieweit kann die im Rahmen eines Additionsverfahrens in der ZBW eingesetzte automatische Sacherschließung den scheinbar unüberbrückbaren Widerspruch zwischen intellektueller und automatischer Sacherschließung, der sich aus den jeweiligen Vorgehensweisen und qualitativen Ansprüchen ergibt, ein Stück weit auflösen?¹⁵

Gemäß dem Charakter einer Fallstudie und der Tatsache, dass die Bibliotheks- und Informationswissenschaft eine praxisnahe und problemorientierte Wissenschaft ist, basiert diese Masterarbeit, aufbauend auf einer umfassenden Auswertung der Sekundärliteratur (theoretischer und empirischer Forschungsstand), vor allem auf der qualitativen Bewertung der maschinell erzeugten Indexierungsergebnisse anhand der zugrunde gelegten und im Methodenkapitel noch ausführlich dargestellten Evaluierungsmaße.

Die Ergebnisevaluierung automatischer Indexierungsverfahren erfolgt üblicherweise mit einem Retrievaltest, bei dem Recall¹⁶ und Precision¹⁷ bestimmt werden (Rapke 2001: 321, vgl. Salton/Mc Gill 1983). Der mit dieser Methode verbundene Aufwand (beispielhaft Sachse et al. 1998) und die dabei auftretenden methodischen Probleme¹⁸ haben, in Bezug auf die

14 Diese Gesetzmäßigkeit geht auf G.K. Zipf zurück. Hiernach existiert „eine konstante Beziehung zwischen dem Rang eines Wortes in einer Häufigkeitsliste und der Frequenz, mit der dieses Wort in einem Text vorkommt“ (Nohr 2004: 217). Eine ausführliche Beschreibung gibt Nohr (2005: 5f.).

15 Siehe hierzu u. a. Nohr (2004: 216).

16 Der Recall bestimmt die Wahrscheinlichkeit, dass ein Dokument aus einer Grundgesamtheit (Bestand) bei einer Suchanfrage auch tatsächlich gefunden wird (Oberhauser 2005: 32). Nach Stock (2007: 556): „Quotient aus der Anzahl der gefundenen relevanten Dokumentationseinheiten und der Gesamtzahl der relevanten Dokumente in einer Datenbank.“

17 Die Precision bestimmt die Genauigkeit des Ergebnisses, d. h., wie viele der gefundenen Dokumente stimmen mit der Intention der Suchanfrage überein (Oberhauser 2005: 32). Nach Stock (2007: 556): „Quotient aus der Anzahl der gefundenen relevanten Dokumentationseinheiten und aller gefundener Datensätze (also einschließlich des Ballastes).“

18 Ein zentrales Problem für die Berechnung des „Recall“ (Vollzähligerate) ist die genaue Bestimmung der Grundgesamtheit. Im Falle des ZBW-OPAC's müsste aus den vorliegenden 4,4 Mio. Medieneinheiten für eine Suchanfrage die Anzahl aller relevanten Dokumente für einen Themenbereich (z. B. Arbeitslosigkeit)

Evaluierung automatischer Verfahren, „über lange Zeit hinweg spekulative Antworten auf der Basis lokaler Kriterien zugelassen“ (Knorz 2004: 186). Beispielhaft hierfür sind die Ergebnisse des AUTINDEX-Projektes am damaligen HWWA (siehe Haller et al. 2001) und die Studie von Helmbrecht-Schaar (2007) zur Klassifizierung von Textdokumenten.

Gleichwohl existiert für die Bewertung der Indexierungsgüte bzw. -qualität ein besser handhabbares Kriterienset (vgl. Stock 2000: 82, Bürkholz 1997: 17): die Indexierungskonsistenz, die Indexierungstiefe, bestehend aus den Maßen Indexierungsbreite und -spezifität, sowie die Indexierungseffektivität.¹⁹ Gerade die Indexierungskonsistenz ist ein starkes Messinstrument (Medelyan 2005: 10), wenn es, wie im Rahmen dieser Arbeit, zu evaluieren gilt, inwieweit eine automatische Indexierungssoftware auf Basis eines Thesaurus qualitativ hinreichend funktionieren kann. Die Indexierungskonsistenz stellt hierbei einen kombinierten Recall-Precision-Wert dar, der bestimmt, wie viele der manuell vorgegebenen Deskriptoren von der Indexierungssoftware eigentlich hätten gefunden werden müssen (Recall) und wie präzise bzw. genau diese letztendlich sind (Precision).

Wie im Folgenden gezeigt werden soll, erreicht die maschinelle Sacherschließung in Bezug auf die Güte des Indexierungsergebnisses einen durchschnittlichen Indexierungskonsistenzwert von 36 %. Hinsichtlich der Indexierungsbreite vergibt das automatische Verfahren im Vergleich zum manuellen Verfahren mehr Deskriptoren pro Dokument, diese sind jedoch unspezifischer und somit auch weniger trennscharf.

Basierend auf den in der Einleitung getätigten grundlegenden Annahmen und der geschilderten Herangehensweise, ergibt sich für den weiteren Verlauf dieser Arbeit folgende Gliederung:

Im 2. Kapitel erfolgt eine kritische Würdigung des Forschungsstandes. Hierbei werden vor allem die bisherigen Evaluierungsansätze automatischer Indexierungsverfahren dargestellt und deren Ergebnisse herausgestellt. Zuerst für diejenigen Untersuchungen, die Retrievaltests ihren Analysen zugrunde gelegt haben. Anschließend werden die Ergebnisse von zentralen Indexierungskonsistenzstudien aufgeführt, einerseits im Vergleich von manuellen Sacherschließungsverfahren untereinander und andererseits in der Gegenüberstellung von manueller und automatischer Indexierung. Weiterhin wird auf den Forschungsstand zur Evaluierung der in der ZBW zum Einsatz kommenden Indexierungssoftware (MindServer) eingegangen.

bekannt sein, was nicht zu ermitteln ist (ähnlich Holl 2009: 60). Weitere Anwendungsprobleme zeigen Sachse et al. (1998), Knorz (2004: 187) oder Lewandowski (2005: 145-148).

¹⁹ Zur genauen Berechnung siehe Kapitel 3.3.

Im 3. Kapitel wird zunächst die Funktionsweise des MindServer-Ansatzes kurz erläutert und in den theoretischen Kontext automatischer Sacherschließungsverfahren eingeordnet. Weiterhin werden die Datenlage und der Stichprobenumfang dargestellt und erläutert. Im Anschluss daran werden das Messverfahren und dessen Evaluierungsmaße zur Bestimmung der Indexierungsqualität vorgestellt und die vorgenommenen Anpassungen an den vorliegenden Untersuchungsgegenstand begründet.

Im 4. Kapitel erfolgt die Darstellung der empirischen Ergebnisse auf Basis der im vorangegangenen Kapitel aufgeführten und definierten Variablen. Zuerst für die berechneten Indexierungskonsistenzwerte und im Anschluss daran für die Maße Indexierungstiefe, als Kombination aus Indexierungsbreite und -spezifität sowie die Indexierungseffektivität. Abschluss der Ergebnisdarstellung bilden die Resultate zur Belegungsbilanz des STW und diejenigen Aussagen, die einer qualitativen Analyse seitens der FachreferentInnen zugrunde liegen.

Im 5. Kapitel werden die Analyseergebnisse anhand der aufgestellten Hypothese und forschungsleitenden Fragestellung abschließend diskutiert und in den allgemeinen Kontext – die Implementierung automatischer Indexierungsverfahren in bibliotheksbezogenen Anwendungen und deren Bewertung unter qualitativen Gesichtspunkten – in Beziehung gesetzt. Weiterhin werden die Grenzen und Schwächen der getroffenen Annahmen und Ergebnisse diskutiert. Den Abschluss bildet ein Ausblick, welche Forschungsmöglichkeiten aus dem Ergebniskontext heraus sich für weiterführende Studien ergeben.

2 Forschungsstand

Mit dem Aufkommen und dem breiten Einsatz automatischer Verfahren im Rahmen der Sacherschließung von Dokumenten stehen diese im Wettstreit mit der manuellen Indexierung um das bessere Indexierungsergebnis (vgl. Nohr 2004: 216). Für die maschinellen Ansätze gilt es zu beweisen, dass sie gleiche oder bessere Resultate erzielen, wohingegen die intellektuellen Ansätze wiederum bemüht sind, ihre Position im Rahmen einer strukturierten Sacherschließung zu verteidigen, indem sie gleichfalls diese Beweise zu erbringen versuchen.

Dieses Kapitel arbeitet den umfangreichen Forschungsstand in Bezug auf den eben beschriebenen Vergleich auf. Dabei gilt es aufzuzeigen, welchen Evaluierungsansatz die einzelnen Arbeiten wählen, welche Stärken und Schwächen die unterschiedlichen Herangehensweisen aufweisen und zu welchem Ergebnis sie letztendlich kommen.²⁰ Zuerst werden überblicksbezogene und theoretische Analysen dargestellt, danach folgt eine Darlegung und kritische Bewertung von Resultaten entsprechender Retrievaltests, anschließend werden Analysen, die die Indexierungskonsistenz berechnen, aufgeführt und deren Ergebnisse miteinander verglichen, um zuletzt Untersuchungen zu beleuchten, die den MindServer als automatisches Sacherschließungsverfahren zu bewerten versuchen.

2.1 Überblicks- und theoretische Arbeiten

Neben theoretisch angelegten Ausführungen finden sich im deutschsprachigen Raum zahlreiche deskriptiv ausgerichtete Überblicksarbeiten, die auf Stärken-Schwächen-Analysen beruhen und keine eigenen Forschungsdaten erheben und auswerten.

Einen pointierten Überblick über die Vor- und Nachteile automatischer Indexierungsverfahren liefert Siegmüller (2007). Anhand der Projekte MILOS/KASCADE²¹, OSIRIS²²,

20 Die jeweilige Datenbasis und der zugrunde gelegte Gegenstand der Untersuchung (z. B. Bibliothek, Pressedokumentation) sind in der Tabelle 4 (im Anhang) zusammenfassend dargestellt. Gleiches gilt für den Evaluierungsansatz, die konkreten Ergebnisse und die Erklärungsfaktoren (= unabhängige Variablen).

21 Beides DFG-geförderte Projekte der Universitäts- und Landesbibliothek Düsseldorf in den 1990iger Jahren. MILOS I und das Nachfolgeprojekt MILOS II waren linguistische Verfahren, während das Anschlussprojekt KASCADE einen statistischen Ansatz verfolgte (vgl. Siegmüller 2007: 68ff.).

22 Gleichfalls ein von der DFG gefördertes Projekt an der Universitätsbibliothek Osnabrück im Zeitraum 1996-1999. Es handelt sich um ein statistisches Verfahren (vgl. Siegmüller 2007: 75ff.).

IntelligentCAPTURE/AUTINDEX²³ und FAST Data Search²⁴ stellt sie die technische Plattform, die Erweiterbarkeit (Skalierbarkeit), die Modularität, den Support, den Workflow, die Datenhaltung und die Softwarebasis der aufgeführten Anwendungen im Bibliotheksbereich vergleichend dar. Die Quintessenz des dargestellten Vergleiches mündet in der Aussage, dass automatische Verfahren nicht die Qualität der manuellen Sacherschließung erreichen, diese aber sehr wohl das vorhandene Datenmaterial für ein Information-Retrieval entsprechend aufbereiten können, um das Suchergebnis zu verbessern (ebd.: 97).

Mit einer kleineren Fallzahl, welche gleichfalls die Anwendungen MILOS/KASCADE, OSIRIS und AUTINDEX als Analysebasis zugrunde legt, arbeitet Pusch (2005). Durch die Fokussierung auf die Beweggründe zum Einsatz automatischer Verfahren, die Datenbasis, den Workflow, die technischen Rahmenbedingungen, die Implementierungsprobleme, die jeweiligen Retrievalergebnisse und den aktuellen Projektstand, werden umfassend die Stärken und Schwächen der untersuchten Verfahren dargestellt. Die zentralen bewertenden Aussagen zu den Ergebnissen des vorgenommenen Vergleiches gleichen denen Siegmüllers (2007). Der Einsatz maschineller Indexierungsverfahren wird grundsätzlich empfohlen, letztlich aber aus quantitativen Aspekten heraus (Datenflut) und weniger aus qualitativen Gründen (vgl. Pusch 2005: 108-122).

Mit einer qualitativen Einschätzung von MILOS und IntelligentCAPTURE stellen auch Mittelbach und Probst (2006) zwei bekannte automatische Indexierungsverfahren ins Zentrum ihrer Untersuchung. In einer Stärken-Schwächen-Analyse zielt der vorgenommene Vergleich vor allem auf die Bestimmung und Bewertung der Diskrepanz zwischen dem Informationsbedürfnis des Nutzers einerseits, welches sich in einer formulierten Suchanfrage konkretisiert, und andererseits dem Informationsgehalt des formell und intellektuell erschlossenen Titels. Basierend darauf wird der Einsatz im Rahmen der Freien Universität Berlin diskutiert (ebd.: 67ff.). Zudem wird eine Kosten-Nutzen-Analyse (Ressourceneinsatz) dargestellt. Letztlich bleiben die qualitativen Bewertungen und Empfehlungen vage und münden in der Aussage, dass „nur bei kombinierter Anwendung intellektueller Erschließungsmethoden und automatischer Indexierungsverfahren optimale Suchergebnisse zu erzielen sind“ (ebd.: 73).²⁵

23 Dieses Verfahren wurde am IAI in Saarbrücken im Zeitraum 1999-2002 entwickelt, vereint statistische und linguistische Komponenten und wurde anschließend in zahlreichen Informationseinrichtungen getestet (vgl. Siegmüller 2007: 84, siehe auch Pusch 2005: 97ff. oder Nübel/Schmidt 2003).

24 An der Universitätsbibliothek Bielefeld im Jahre 2004 gestartetes Projekt. Neben einem statistischen Ansatz kann auch eine linguistische Komponente in Form von Wörterbüchern zugeschaltet werden (vgl. Siegmüller 2007: 90ff.).

25 Dieser Vorschlag entspricht einem semi-automatischen Sacherschließungsverfahren.

Einen Vergleich verschiedener maschineller Indexierungsverfahren unter software- und retrievalbezogenen Gesichtspunkten mit einer anschließenden kurzen qualitativen Einschätzung liefert auch Schwarz (2004). In der Gegenüberstellung dreier Softwareprodukte werden zentrale Implementierungsaspekte von automatischen Sacherschließungsverfahren im Rahmen einer Fallstudie (Lenzing AG, Österreich) analysiert. Die Vorteile eines Einsatzes dieser Verfahren werden in der Verbesserung des Verhältnisses zwischen formulierter Suchanfrage und bestehendem Dokumenteninhalt gesehen („vocabulary problem“)²⁶. Weitere Vorteile liegen nach Schwarz gleichfalls in der Skalierbarkeit (Erweiterungsfähigkeit), in der Einsparung personeller Ressourcen (Kosten-Nutzen-Analyse) sowie in der Einbindung heterogener digitaler Inhalte unter einer Suchoberfläche.

In einer älteren Studie vergleicht Bürkholz (1997) den Entwicklungsstand der Anwendungen PASSAT²⁷, MILOS und AIR/PHYS²⁸. Im Einzelnen werden der Funktionsumfang, die Systembesonderheiten, der Indexierungsablauf, die Erfahrungen der Anwender dargestellt und eine abschließende Systembewertung vorgenommen. Im Fazit wird die Anwendung im bibliothekarischen Kontext als „sinnvolle Lösung“ (ebd.: 42) angesehen, um die manuelle Sacherschließung zu ergänzen (semi-automatischer Ansatz).

Im Fokus von theoretischen Abhandlungen, die sich automatischen Sacherschließungsverfahren zuwenden und keine empirische Überprüfung der getätigten Annahmen vorlegen, stehen vor allem informationstheoretische Aspekte dieser Verfahren und deren Übertragung in softwarespezifische Lösungen. Hauer (2000) zeigt zum Beispiel für das Projekt AUTINDEX, wie das eingesetzte statistische Verfahren die eingespielten Daten bearbeitet und anschließend die Ergebnisse mit Hilfe eines Thesaurus entsprechend abbildet. Lepsky und Vorhauer (2006) legen mit der Open-Source-Software LINGO²⁹ ein linguistisches Verfahren ihrer theoretischen Abhandlung über die automatische Indexierung deutschsprachiger Dokumente zugrunde. Hierbei untersuchen sie vor allem die Softwarefunktionalität (Grundform-, Muster-, Kompositaerkennung) hinsichtlich der Anforderungen an Datenbasis und Einsatzmöglichkeiten. Eine informatikbezogene und ausführliche Darstellung der Funktionsweise dieses

26 Vgl. Holls (2009: 65) Aussagen zur Suchterweiterung und der Vagheit zwischen Suchanfrage einerseits und dem kontrollierten Vokabular andererseits.

27 Dieses automatische Indexierungsverfahren wurde bereits 1968 von der Firma Siemens entwickelt und stellt einen wörterbuchbasierten Ansatz dar (vgl. Bürkholz 1997: 18ff.).

28 Ein in den 1970iger Jahren an der TH Darmstadt entwickeltes statistisches Verfahren, das in den 1980iger Jahren im Fachinformationszentrum Karlsruhe getestet worden ist (vgl. Bürkholz 1997: 34ff.).

29 Eine ausführliche Darstellung findet sich in Gödert und Lepsky (2008) bzw. in Schiffer (2007). Die Software kann unter <http://www.lex-lingo.de/> (Zugriff: 27.10.10) abgerufen werden.

automatischen Sacherschließungsverfahrens und deren software- und retrievalspezifischer Anforderungen findet sich in Manning et al. (2008).

2.2 Evaluierungsverfahren I: Retrievaltests

Retrievaltests, in denen die Maße Recall (Vollzähligkeitsrate) und Precision (Präzisionsrate) bestimmt werden, nehmen eine herausragende Rolle ein, wenn es darum geht, die Performance automatischer Sacherschließungsverfahren zu evaluieren. Gerade bei umfangreichen Datensätzen bzw. Titelmengen ist diese Art der Qualitätsmessung sinnvoll.³⁰ In diesem Unterkapitel sollen exemplarisch die Ergebnisse von durchgeführten, die Qualität automatischer Indexierungsansätze bewertenden Retrievaltests dargestellt und deren Einflussfaktoren kontrastiert werden.

Im Rahmen einer Diplomarbeit untersucht Rapke (2001) die Funktionsweise von zwei grundsätzlich unterschiedlich ausgerichteten automatischen Verfahren – einerseits einem statistischen Ansatz und andererseits einem Mustererkennungsverfahren –, für die jeweils Recall und Precision im Vergleich zur manuellen Sacherschließung ermittelt werden. Während das statistische Verfahren einen Recall-Wert von 74 % und einen Precision-Wert von 57 % erreicht, liegt die Performance des Mustererkennungsverfahrens bei 56 % bzw. 32 %. Im Vergleich hierzu erreicht die manuelle Erschließung einen Recall-Wert von 75 % und einen Precision-Wert von 73 % (ebd.: 332). Die Gründe in dieser Diskrepanz sieht Rapke (ebd.: 335-337) u. a. in der bestehenden Vagheit zwischen der Suchanfrage einerseits und dem Dokumenteninhalte andererseits sowie in der Schwierigkeit maschineller Verfahren hinsichtlich der Kompositazerlegung und Eigennamenerkennung.

Für die Indexierungssoftware AUTINDEX vergleichen Nübel/Schmidt (2003) automatische und maschinelle Indexierungsergebnisse. Mit Hilfe eines Retrievaltests wird die Abbildung der automatisch generierten Indexierungsergebnisse in Form von Thesaurusbegriffen evaluiert.³¹ Die Recallwerte liegen zwischen 26 % (bilinguale Indexierung, deutsch-englisch) und 37 % (monolinguale Indexierung, englisch). Im Bereich der Precision werden Werte zwischen 61 % (monolinguale Indexierung, deutsch) und 37 % (monolinguale Indexierung, englisch) erreicht. Leider finden sich für die zu erklärenden Variablen keine

30 Siehe Tabelle 4 im Anhang.

31 Gleichzeitig erfolgt auch eine Berechnung von Recall und Precision für aus einem flacheren Thesaurus generierte Klassifikationsergebnisse (siehe Nübel/Schmidt 2003: 96-97).

Angaben hinsichtlich der intervenierenden Variablen (Erklärungsfaktoren). Ähnlich oberflächlich verbleibt auch die Analyse von Schek (2005), in der zwar die ermittelten Recall-Werte (75 % bzw. >90 %) für das eingesetzte automatische Indexierungsverfahren aufgeführt werden, die gleichfalls interessanten Precision-Werte aber nicht. Zudem finden sich auch hier keine Hinweise zu möglichen Erklärungsfaktoren.

Die Arbeit von Oberhauser und Labner (2003) ist ein weiteres typisches Beispiel für einen Evaluierungsansatz, der versucht, die Vagheit/Lücke zwischen einerseits der Suchanfrage und andererseits den formell und intellektuell erschlossenen Datensätzen zu bestimmen. In drei Retrievaltests, denen unterschiedliche Annahmen bezüglich dieser Lücke zugrunde liegen, nimmt die Anzahl der gefundenen Titel zu. Der Recall steigt im ersten Test um 32 % (Precision 91 %), im zweiten Test um 55 % (Precision 56 %) und im dritten Test um 32 % (Precision 84 %) (ebd.: 11-16). Ebenso wie bei Nübel/Schmidt (2003) und Schek (2005) werden keine Erklärungsfaktoren aufgezeigt.

Einen anderen Evaluierungsansatz, gleichwohl unter Anwendung eines Retrievaltests, zeigt die Diplomarbeit von Leder (2002). Auf Basis eines linguistischen Verfahrens (MILOS) werden die Indexierungsergebnisse manueller und automatischer Sacherschließungsansätze gegenübergestellt. Der Einsatz der Automatik liefert einen Anstieg des mittleren Recall von 54 % auf 93 %, im Gegenzug sinkt die Precision von 78 % auf 69 % (ebd.: 36). Erklärungsfaktoren werden in der Wörterbuchpflege (neue Wörter, Synonyme etc.) aufgeführt und in den Stellschrauben des automatischen Verfahrens gesehen (z. B. Fehler in der Dekomposition von Wörtern) (ebd.: 35). Als Fazit wird eine gewinnbringende Verbindung von intellektuellen und automatischen Sacherschließungsverfahren in semi-automatischer Form vorgeschlagen.

Sachse et al. (1998) zeigen in einer älteren Arbeit, dass die Retrievalergebnisse durch den Einsatz eines automatischen Indexierungsverfahrens (hier MILOS) zu einem erhöhten Recall-Wert von 92 % und mehr führen, unter Inkaufnahme eines Rückgangs der Precision (79 % bzw. 74 %) (ebd.: 37). Die aufgezeigten Erklärungsfaktoren liegen einerseits in den Stellschrauben des eingesetzten automatischen Systems, z. B. in der Grundformermittlung, der Kompositazerlegung und der Derivation, begründet sowie andererseits in der Wörterbuchpflege des linguistischen Verfahrens (ebd.).³²

Die dargestellten Retrievaltests als Messverfahren zur Bestimmung der Indexierungsgüte haben gezeigt, dass sich bei automatischen Sacherschließungsverfahren grundsätzlich die Zahl der gefundenen Titel bei einer Suchanfrage erhöht (Recall-Anstieg), während die

32 Zu dem gleichen Ergebnis kommt auch Baum (2005). Z. B. steigt durch eine Kompositazerlegung der Recall für einen Suchbegriff um 400 %, unter Anstieg der Precision (von 72 % auf 84 %) (ebd.: 33). Ähnliches wird für die Dekomposition und die Anreicherung mit Synonymen festgestellt.

Precision demgegenüber immer abnimmt, aber in vertretbarem Ausmaße, wie vielfach betont worden ist. Die aufgeführten Arbeiten haben weiterhin, allerdings nur teilweise, mögliche Erklärungsfaktoren dargestellt, die vor allem im Bereich der software- bzw. verfahrensspezifischen Stellschrauben (z. B. Kompositazerlegung, Wörterbuchpflege) zu verorten sind.

2.3 Evaluierungsverfahren II: Indexierungskonsistenz

Die Berechnung der Indexierungskonsistenz als Gütekriterium maschineller Indexierungsergebnisse nimmt, im Vergleich zu den bekannten Evaluierungsmaßen von Retrievaltests, gleichfalls eine prominente Stellung ein, wenn manuelle und automatische Sacherschließungsverfahren gegenübergestellt werden. Gerade bei Dokumentensets mit mittelgroßen Titelzahlen³³ kommt dieses Qualitätsmaß zum Einsatz.

Die Indexierungskonsistenz kann grundsätzlich in folgende Kategorien eingeteilt werden: erstens die Inter-Indexier-Konsistenz. Hierbei werden die Ergebnisse verschiedener Indexierer – hierzu zählen auch Gegenüberstellungen von manuellen vs. automatischen Sacherschließungsergebnissen – miteinander verglichen (siehe Stock 2000: 83, Lancaster 2003: 68). Zweitens die Intra-Indexer-Konsistenz, bei der die Ergebnisse eines Indexierers zur selben Vorlage, aber zu unterschiedlichen Zeitpunkten gegenübergestellt werden. Drittens die Indexer-Anfrager-Konsistenz. Darunter werden Analysen subsumiert, die sich der Lücke zwischen der formell und inhaltlich erschlossenen Vorlage einerseits und der gestellten Suchanfrage des Nutzers als konkretisiertes Informationsbedürfnis andererseits zuwenden und diese entsprechend untersuchen (vgl. Stock 1993: 150).³⁴

Die folgenden Ausführungen geben den Forschungsstand von wichtigen und aktuellen Indexierungskonsistenzanalysen wieder. Neben den reinen Konsistenzwerten sind hierbei vor allem die aufgezeigten Erklärungsfaktoren für das Zustandekommen der jeweiligen Ergebnisse von Interesse und werden entsprechend herausgestellt. Zunächst wird auf die Einflussfaktoren der Indexierungskonsistenz zwischen manuellen Indexierern eingegangen (= manuelle Inter-Indexer-Konsistenz), um anschließend auf konsistenzbezogene Vergleiche

33 Siehe Tabelle 4 im Anhang.

34 Eine ausführliche Darstellung findet sich in Cooper (1969) oder Fugman (1992). Ein aktuelles Beispiel für eine Analyse zur Indexer-Anfrager-Konsistenz stellt die Arbeit von Holl (2009) dar.

zwischen manuellen und automatischen Sacherschließungsverfahren zu fokussieren (manuell/automatische Inter-Indexer-Konsistenz).

Es überrascht nicht, dass, wenn der eigentliche Vorgang des Indexierens als Kunst charakterisiert wird (Stock/Stock 2008: 345), hinsichtlich der Qualitäten der Ergebnisse heftige Diskurse entbrennen können, so dass Studien zu dem Ergebnis kommen, dass „indexers do not always agree on the choice of keywords representating the contents of the same document“. (Bertrand/Cellier 1995: 459). Während dieser Fakt, wie noch zu sehen sein wird, in zahlreichen Arbeiten immer wieder herausgestellt wird und offensichtlich unstrittig ist, so unterschiedlich sind jedoch demgegenüber die aufgezeigten Erklärungsansätze dieses Phänomens.

2.3.1 Manuelle Sacherschließung

Die Ursachen für eine geringe manuelle Inter-Indexer-Konsistenz liegen, um einen ersten Erklärungsansatz zu nehmen, in der Kognition der Indexierer bzw. der Indexiererinnen begründet. Deren Wissen³⁵ bildet eine intellektuelle Basis, um im Rahmen einer sogenannten Indexierungsstrategie, d. h. dem konkreten Arbeitsablauf des Indexierens, zur Anwendung zu gelangen (Bertrand/Cellier 1995: 460). Die Indexierungskonsistenz zwischen manuellen Indexierern variiert diesem Ansatz nach zwischen 62 % für das Fachgebiet Wirtschaftswissenschaften und 55,1 % für das Fachgebiet Psychologie. Eine Aufteilung in Experten vs. Laien ergab Werte von 64,9 % bzw. 53,4 % und somit – wenig überraschend – eine höhere Konsistenz zwischen den erfahrenen Indexierern (ebd.: 464). Dies ist, wenn die Indexierungsstrategie als ein zentraler Erklärungsfaktor angesehen wird, auf eine unterschiedliche Indexierungsbreite der einzelnen Indexate zurückzuführen (im Mittel wurden 4,48 Terme pro Dokument vergeben)³⁶ und auf die vorhandene Struktur des Thesaurus, im konkreten Falle vor allem die Aktualität und Kontrolle des vorhandenen Vokabulars.

Zu ähnlichen Ergebnissen kommen auch David und Giroux (1995), die eine Inter-Indexer-Konsistenz von 45 % für alle Deskriptoren und 66 % für die jeweils zuerst aufgeführten Deskriptoren feststellen (ebd. 7). Der gleichfalls in der kognitiven Psychologie

35 Hierzu zählen nicht nur die generelle Vertrautheit eines Sacherschließers mit dem jeweiligen Fachgebiet, z. B. durch ein Hochschulstudium des betreffenden Faches, sondern auch der Erfahrungsschatz, welcher in jahrelanger Anwendung eines kontrollierten Vokabulars entsteht.

36 Es unterbleibt eine Korrelation zwischen Indexierungsbreite und -konsistenz, um zu eindeutigen Aussagen hinsichtlich deren Zusammenhangs zu gelangen.

zu verortende Erklärungsansatz fokussiert ebenso auf das strategische Verhalten der IndexiererInnen im Rahmen der Sacherschließung als Problem lösender Prozess:

With experts, for example, it can be assumed that with experience comes the assimilation of general rules and norms, knowledge relative to the particular work context, knowledge of the domain, etc., all of which leads to the development of a goal structure and *sub-goals different from that of other indexers*. (ebd.: 3, Hervorhebung durch T. G.).

Gerade die Diskrepanz zwischen Hauptthemen („goal“) und Randthemen („sub-goal“) wird in den nachfolgenden Studien immer wieder als ein zentraler Faktor für die Veränderung der Inter-Indexer-Konsistenz angesehen. So auch bei Saarti (2002). Gerade weiche Themen, „so called soft or abstract sciences, especially for fictional material“ (ebd.: 50), beeinflussen durch die Vielzahl an möglichen Deskriptoren eine konsistente Sacherschließung zwischen manuellen IndexiererInnen. Der Konsistenzwert für die Experten (Bibliothekare) liegt bei 77,1 %, während die Laien (Bibliotheksbenutzer) im Vergleich dazu nur 56,3 % erreichen. Weitere intervenierende Variablen stellen die Granularität des Thesaurus dar (Vergabe Ober-, Unter-, verwandter Begriffe) sowie die Fokussierung der Indexierer auf sinntragende Dokumentelemente, wie Titel, Abstract, Einleitung (ebd.: 61).

Die Gegenüberstellung der Indexierungsergebnisse von Experten im Vergleich zu denen von Laien durch Bildung von entsprechenden Versuchs- und Kontrollgruppen ist ein oft anzutreffendes Untersuchungsdesign, um den Mehrwert normierter Sacherschließung näher zu bestimmen. Während die Ergebnisse von Experten sich tendenziell annähern (vgl. David/Giroux 1995, Saarti 2002), ohne jedoch völlig übereinzustimmen, variiert die Konsistenz zwischen Laien naturgemäß stärker, wie die Arbeit von Kipp und Campbell (2006) aufzeigt. In der Studie, die auf manuelle Sacherschließungsergebnisse durch Laien im Rahmen von Social Tagging³⁷ fokussiert, kommen sie zu dem Ergebnis, dass nutzergenerierte Schlagwörter (oder Tags) ohne die Zuhilfenahme eines kontrollierten Vokabulars naturgemäß so unterschiedlich sind, dass eine Evaluierung der Indexierungskonsistenz in Form eines konkreten Wertes nicht möglich ist (ebd.: 2). Eine stattdessen vorgenommene Kategorisierung (Clusterung) der vorliegenden Tags führt zu dem Ergebnis, dass Web-2.0-Begriffe (u. a. blog, wiki) sowie Zeitangaben am häufigsten vergeben werden. Ohne kontrolliertes Vokabular halten sich Laien eng an den Textinhalt, um häufig vorkommende oder an prominenter Stelle

37 Hierbei werden Inhalte im WWW durch die Nutzer mit entsprechenden Metadaten (= Tags) beschrieben (vgl. Semantic Web Company 2010).

auftretende Wörter zu extrahieren.³⁸ Auch Lancaster (2003: 73) zeigt diesen Aspekt in einem Experiment, in dem die Inter-Indexer-Konsistenz beim freien Indexieren generell höher ist als unter Anwendung eines entsprechenden Thesaurus.

Die Varianz freier Sacherschließung durch Laien verdeutlicht gleichfalls die Studie von Olson und Wolfram (2005, 2007), welche gleichfalls Indexierungskonsistenzen zwischen Laien im Rahmen des Social Tagging untersuchen: „[...] over a reasonably large number of people, consistency in identifying key concepts is small.” (ebd.: 2005: 3). Die Gründe liegen, wie bei Kipp und Campell (2006) bereits dargestellt, in der Natur der Sache:³⁹ Je höher die Anzahl der Indexierer, die einen bestimmten Dokumenteninhalte verschlagworten sollen, desto stärker variieren die Ergebnisse. Wenn selbst die Kernthemen eines Dokuments de facto mit unterschiedlichen Worten beschrieben werden (können),⁴⁰ kommt es verständlicherweise gerade bei den peripheren Aspekten zu einer noch größeren Inkonsistenz (Olson/Wolfram 2005: 1).⁴¹

Einen Lösungsansatz hierzu zeigen Stubbs et al. (1999) auf. Durch eine Erhöhung der Indexierungsbreite kann die Indexierungskonsistenz zwischen manuellen Indexatoren gesteigert werden.⁴² Im ersten Sample lag der Konsistenzwert bei 29 % bei Indexierungsbreiten von 2-30 Begriffen. Im zweiten Sample konnte der Konsistenzwert durch die durchschnittliche Vergabe von 18-24 Deskriptoren mit einem Anstieg auf 60 % erheblich gesteigert werden (ebd.: 65). Dieses Ergebnis ist aber nur auf statistische Effekte zurückzuführen und kann als nicht überzeugend für den praktischen Einsatz im Rahmen bibliothekarischer Sacherschließung angesehen werden. Für den Bereich der Sacherschließung durch Laien, z. B. via Tagging, mag dies ein Lösungsansatz sein, für das Terrain professioneller Indexierung ist dieser nicht geeignet.⁴³

Die Ergebnisse von Tonta (1991), der die Inter-Indexierer-Konsistenz zwischen Bibliothekaren der British Library und der Library of Congress miteinander vergleicht, relativieren ein Stück weit die Ergebnisse von Kipp/Campbell (2006) und Wolfram/Olson (2005, 2007). Während bei einer unterschiedlichen Indexierungsbreite die Inter-Indexierer-Konsistenz

38 In gleicher Art und Weise arbeiten auch statistische Indexierungsverfahren.

39 Olson und Wolfram (2007) sprechen von „[...] nature of consistency or inconsistency.“ (Ebd.: 602).

40 Z. B. der Begriff Indexierung, welcher auch als Sacherschließung oder Verschlagwortung bezeichnet werden kann.

41 Vgl. das goal/sub-goal-Problem bei David/Giroux (1995: 3).

42 Zu diesem Ergebnis gelangen auch Neshat und Horri (2006: 67): „[...] whenever the number of assigned terms increases, the probability of consistency decreases.“

43 Dies würde für eine Annäherung manueller mit automatischer Sacherschließung bedeuten, dass beide Verfahren nur möglichst viele Begriffe vergeben, um somit die mögliche Schnittmenge zu erhöhen.

Werte von 16 % (exakte Übereinstimmung) und 36 % (partielle Übereinstimmung) aufweist, liegen die Zahlen bei einer identischen Indexatbreite bei 14 % bzw. 41 % (Tonta 1991: 4-5).

Auch Watermann (1997) vergleicht die Inter-Indexierer-Konsistenz zwischen Bibliothekaren unter Rückgriff auf zwei Anwender des STW, das ifo-Institut und das damaligen HWWA. Als durchschnittlicher Konsistenzwert wurden 31 % ermittelt, die auf eine unterschiedliche Indexierungspolitik der beiden Institutionen (ebd.: 17ff.) sowie eine divergierende Indexierungsbreite zwischen den einzelnen Indexaten zurückzuführen sind (ebd.: 43). Als abschließende Bewertung wird die Inter-Indexierungs-Konsistenz beider STW-Anwender als „dürftig und desolat“ (ebd.: 43) beschrieben.

Demgegenüber ermittelt Leininger (2000) durchschnittlich höhere Inter-Indexer-Konsistenzwerte. Je nach zugrunde gelegter Formel liegt die ermittelte Konsistenz zwischen manuellen Indexaten bei 50,4 % bzw. 60,83 %. Eine differenzierte Analyse der Verschlagwortung von Kern- bzw. Randthemen ergab, dass die Indexierungskonsistenz zurückgeht, wenn nur die ersten Schlagworte miteinander verglichen werden. Hier tritt wieder das goal/sub-goal-Problem manueller Indexierungsverfahren deutlich hervor: „[...] the central concepts addressed can be reliably found in the indexing of a record as a whole.“ (Ebd.: 7).

Einen anderen Erklärungsansatz von geringen Inter-Indexer-Konsistenzen zeigt Iivonen (1990) auf. Der zugrunde gelegte organisationstheoretische Ansatzpunkt legt den Fokus auf die Umwelt des Indexierungsprozesses: „[...] a library as an indexing environment, proves to be a very complex whole, where numerous factors influence each other.“ (Ebd.: 259). Die geringen Konsistenzwerte – je nach Testsituation liegen sie bei 8 %, 22,4 % und 30,6 % (ebd.: 261) – haben ihre Ursache in zahlreichen Faktoren: der Indexierungsbreite, der Gewichtung zentraler/peripherer Themen (goal/sub-goal-Problem), der Indexierungspolitik (Indexierungsregeln), der Aktualität der angewandten Dokumentationsprache, der Thesaurushierarchie (Vergabe von Ober- statt Unterbegriffen) und der Vergabe offensichtlich falscher Terme (ebd.: 262-263). Diese intervenierenden Variablen sehen auch Reich und Bieber (1991) als ursächlich für unterschiedliche Inter-Indexer-Konsistenzen an.⁴⁴ Die ermittelten Konsistenzwerte liegen in ihrer Untersuchung je nach Sample bei 24 % bzw. 45 % (ebd.: 338). Zusätzlich wird der Indexierungstiefe und hier insbesondere der Indexierungsspezifität entscheidender, konsistenzbezogener Einfluss zugeschrieben: „Depth of indexing may play an additional role inasmuch as it may influence the choice of terms“ (ebd.: 340).

44 Es wird nach thesaurusspezifischen Faktoren (Hierarchie/Granularität) und nicht thesaurusspezifischen Faktoren (Indexierungspolitik, Gewichtung zentraler/peripherer Themen) unterschieden.

Hier setzt auch die umfassende Arbeit von Xu (2007, 2008) an, der eine Reihe konsistenzbezogener Erklärungsfaktoren evaluiert. Es werden je nach Formel Inter-Indexer-Konsistenzwerte von 62,2 % und 70,7 % ermittelt (ebd.: 147). Während die Indexierungsbreite, der Wissenshintergrund (Qualifikation) der Sacherschließer und die Aktualität des kontrollierten Vokabulars die Konsistenz erhöhen, kann den Faktoren Thesaurushierarchie, Indexierungsspezifität, Dokumentenlänge kein Einfluss zugerechnet werden (ebd.: 148). Zudem wird die Inter-Indexer-Konsistenz domänenspezifisch beeinflusst: „[...] lower indexing consistency can be observed in humanities and social science, the so called soft science.“ (ebd.).⁴⁵

2.3.2 Manuelle vs. automatische Sacherschließung

Die Inter-Indexer-Konsistenz zwischen manuellen und automatischen Sacherschließungsverfahren stellt vor allem auf die Entsprechung des maschinell generierten Indexates im Vergleich mit den manuellen Vorgaben bzw. Ergebnissen ab. Während manuelle Indexierungskonsistenzmessungen vor allem auf die Qualifikation des Indexierers und dessen Vorgehen beim Prozess der Sacherschließung fokussieren, untersuchen die folgenden Analysen maschineller Indexierungsverfahren vor allem die entsprechenden (vor allem softwarespezifischen) Stellschrauben.

Inter-Indexer-Konsistenzen werden vor allem auch durch die Anzahl der vergebenen Deskriptoren maßgeblich beeinflusst (vgl. Tonta 1991, Stubbs et al. 1999). Dieser Sachverhalt wird auch bei Gerards et al. (2006) deutlich herausgearbeitet.⁴⁶ Als Grundlage ihrer Untersuchung, die ein semi-automatisches Sacherschließungsverfahren analysiert, dient die Indexierungssoftware AUTINDEX. Die durchschnittliche Anzahl an vergebenen Deskriptoren liegt im manuellen Verfahren bei 6-7, im Falle der Automatik werden 8-9 Schlagwörter generiert (ebd.: 13). Die daraus ermittelte Inter-Indexer-Konsistenz liegt im manuellen Bereich bei 42,58 % bzw. 56,99 %.⁴⁷ Im Gegensatz hierzu erreicht AUTINDEX Werte von 25,22 % bzw. 40,24 % und bleibt damit deutlich hinter den Zahlen, die die manuellen IndexiererInnen erreichen, zurück (ebd.: 15). Neben der Indexierungsbreite beeinflussen die

45 Dies stellt auch schon Saarti (2002) fest.

46 Keine grundsätzlich anderen Ergebnisse liefert eine aktuelle Evaluationsstudie von 2010 (vgl. Vortrag von Michael Gerards auf der FIS Bildungstagung, 03.-04.05.2010, Frankfurt am Main: <http://www.dipf.de/de/pdf-dokumente/bildungsinformation/fis-bildung/tagung2010.pdf>, Zugriff: 27.10.10).

47 Je nach Formel: 1. Wert nach Hooper (1965) berechnet, der 2. Wert nach Rolling (1981).

Thesaurusstruktur⁴⁸ sowie die fehlende englischsprachige Komponente die Indexierungsergebnisse entscheidend. Konsistenzmindernden oder -erhöhenden Einfluss haben vor allem auch die Homonyme innerhalb des Thesaurus. Im Falle mehrdeutiger Schlagwörter vermag es die Automatik nicht, diese hinreichend trennscharf den jeweiligen Sachverhalten zuzuordnen (ebd.: 17).⁴⁹

Die Bedeutung, die die sprachbezogene Pflege eingesetzter Thesauri hat, wird in Nübel/Schmidt (2003) einmal mehr deutlich herausgestellt. Eine monolinguale Indexierung im Deutschen erreicht einen Wert von 15 %, im Bereich englischer Dokumente liegt der Konsistenzwert bei 20 %. Die bilinguale Indexierung deutscher Dokumente erreicht 11 % und diejenige englischer Dokumente 10 % (ebd.: 96).

Die Analysen von Haller et al. (2001)⁵⁰ beruhen, wie die Gerards et al. (2006), auf der Anwendung und Bewertung des AUTINDEX-Verfahrens, in diesem Falle aber im Kontext wirtschaftswissenschaftlicher Dokumente bzw. Inhaltserschließung. In einem damaligen DFG-Projekt wurde diese Indexierungssoftware von der ZBW und dem damaligen HWWA eingesetzt und evaluiert. Während die menschlichen Indexierer im Durchschnitt 7,0 bis 7,4 Deskriptoren den Dokumenten zuordneten, generierte die Automatik im Mittel 3,47 bis 3,9 Schlagwörter. Das maschinelle Verfahren erreichte letztendlich Konsistenzwerte gegenüber den FachreferentInnen von 49,3 % bzw. 52,3 % (ebd.: 7). Die Ursachen für diese Werte bzw. Differenzen liegen in der Datenbasis, konkret der fehlenden Strukturiertheit der Dokumente begründet.⁵¹ Weiterhin beeinflussen die Thesaurusstruktur⁵² und dessen bilinguale Ausrichtung⁵³ die Indexierungsgüte der automatisch generierten Indexate. Nicht zuletzt wird das Ergebnis durch eine fehlerbehaftete Kompositazerlegung oder mangelnde Worterkennung von Eigennamen (Länder, Personen) beeinflusst (ebd.: 12-13).

48 Ständige Kontrolle und Pflege der Deskriptoren und vor allem der Nicht-Deskriptoren.

49 Ein gutes Beispiel hierfür sind folgende STW-Begriffe: Bank, Entwicklung, Kooperation.

50 Grundlegende Aspekte und Gedanken zu diesem Projekt finden sich zudem in Gastmeyer (2004).

51 „Die Einbeziehung des gesamten Dokuments in die Inhaltserschließung durch AUTINDEX bringt oft eine erhebliche Verzerrung bei den vergebenen Deskriptoren mit sich, weil u. a. Autorennamen, herausgebende Institutionen und deren Standorte in die Analyse und das Ergebnis mit eingehen.“ (Gastmeyer 2004: 1).

52 Hierzu zählen die Aktualität des kontrollierten Vokabulars und dessen entsprechende Pflege (laufende Ergänzung von Synonymen, Homonymen ...). Probleme bereiten zudem domänenspezifische Allgemeinwörter, z. B. Ablauf, Bewertung oder Messung (Haller et al. 2001: 13).

53 „Eine valide englische Thesaurusübersetzung, die in maschinellen Indexierungsverfahren einsetzbar wäre, kann nur mit einigem Aufwand weitgehend intellektuell erstellt werden, weil es für die Mehrzahl der Fachbegriffe keine 1:1 Beziehung gibt.“ (Gastmeyer 2004: 5).

Medelyan (2005)⁵⁴ legt den Schwerpunkt ihrer Untersuchung auf die Kombination von Schlagwortextraktions- und Schlagwortadditionsverfahren, um zu berechnen, inwieweit die manuell vergebenen Schlagworte nicht nur absolut gleich sind, sondern vor allem welche semantischen Relationen zwischen verwandten/ähnlichen Deskriptoren (= Ober-, Unterbegriffe) bestehen. Im Vergleich zu manuellen Indexierern, deren Inter-Indexer-Konsistenz (je nach Formel) Werte von 38 % bzw. 51 % annehmen kann (bei einer Indexierungsbreite von 5-16 Deskriptoren), erreicht die automatische Sacherschließung mit Werten von 27 % bzw. 38 % generell eine 10-15 % geringere Konsistenz gegenüber der manuellen Indexierung (ebd.: 83).

2.4 Evaluierungsverfahren III: Sonstige

Nachdem die zwei hauptsächlichen Evaluierungsverfahren maschineller Sacherschließungsverfahren – Retrievaltests (Kapitel 2.2) und Indexierungskonsistenzberechnungen (Kapitel 2.3) – dargestellt worden sind, gilt es nun, andere Qualitätsmessungsansätze kurz zu erwähnen, die nicht diesen beiden Forschungsstandskategorien zuzuordnen sind. Kennzeichnendes Merkmal aller dieser Arbeiten ist, dass diese eher allgemeine Aussagen beinhalten und selten konkrete Ergebnisse bestimmter Untersuchungsdesigns darstellen und entsprechend bewerten.

In einer Studie zur Funktionsweise semi-automatischer Verfahren untersucht Helmbrecht-Schaar (2007) die Möglichkeiten und Grenzen maschineller Klassifizierungsverfahren. Die qualitative Analyse mündet in der Feststellung, dass die Ergebnisse des eingesetzten automatischen Sacherschließungsverfahrens „recht gut“ (ebd.: 51) bzw. „akzeptabel“ (ebd.: 58) sind. Die Ursachen hierfür liegen vor allem in den abzubildenden Fachtermini, den entsprechenden Retrieval-Stellschrauben (Dekomposition) sowie den zugrunde gelegten Wörterbüchern und dessen Pflege begründet.

Eine ähnliche qualitative Einschätzung liefern Hübner und Groth (2004), indem sie die Vor- und Nachteile eines automatischen Indexierungsverfahrens in Bezug zur intellektuellen Sacherschließung setzten. Neben einem Vergleich der Indexierungsbreite der untersuchten Titel (ebd.: 65-68) wird jeder Titel hinsichtlich seiner Erschließungsgüte auch kurz verbal eingeschätzt (ebd.: 69-74). Diese Einschätzungen variieren innerhalb der Bereiche „sehr gut“ bis „sehr schlecht“. Ähnlich wie Mittelbach und Probst (2006) und Bürkholz (1997) kommen

54 Eine kurze Zusammenfassung des Ergebnisses aus der Masterarbeit von Medelyan (2005) findet sich in Medelyan/Witten (2006).

auch Hübner und Groth (2004) zu dem Ergebnis, dass semi-automatische Indexierungsverfahren den zukünftigen Anforderungen bibliothekarischer Sacherschließung am ehesten gerecht werden (ebd.: 96).

Einen Stärken-Schwächen-Vergleich nimmt auch Schiffer (2007) vor. In seiner Arbeit über die automatische Indexierung wissenschaftlich-technischer Kongressschriften mit Hilfe des linguistischen Verfahrens LINGO werden vor allem die grundsätzliche Funktionsweise der zum Einsatz gekommenen Indexierungssoftware vorgestellt, deren Stellschrauben (u. a. Mehrworterkennung, Grundformerzeugung, Synonymzuordnung) beleuchtet und mit einer allgemeinen Einschätzung verbunden. Das eingesetzte Verfahren ist „[...] geeignet [...] erfolgreich zu indexieren.“ (Ebd.: 53).

2.5 Evaluierung des MindServers

Im deutschen Anwendungskontext wird die Indexierungssoftware MindServer vor allem im Bereich der Pressedokumentation eingesetzt.⁵⁵

Ein Beispiel hierfür ist das ZDF, welches dieses automatische Sacherschließungsverfahren implementiert hat, um Zeitungsartikel, die vorher ausschließlich von Dokumentaren beschlagwortet worden sind, nun semi-automatisch zu bearbeiten bzw. zu erschließen (vgl. Lingelbach-Hupfauer/Laute 2009). Als zentrales Ergebnis des seit 2001 laufenden Projektes hat sich herauskristallisiert, dass der eingesetzte Thesaurus zum einen nicht trennscharf genug ist, um der Automatik eine eindeutige Kategoriezuordnung zu ermöglichen, und zum anderen hat sich gezeigt, dass die Anzahl der zur Verfügung stehenden Begriffe von 3.500 auf 2.500 verringert werden kann, weil viele der möglichen Schlagwörter in nur sehr geringem Ausmaße Anwendung finden (ebd.: 48). In Bezug auf die Indexierungskonsistenz hat sich im Laufe des Projektes verfestigt, dass die Intra-Indexierungskonsistenz, die sonst ausschließlich abhängig von der Tagesform des Dokumentars gewesen ist, ebenso gesteigert werden kann wie die Inter-Indexierungskonsistenz zwischen den einzelnen Sacherschließern (ebd.: S. 40).

Im Bereich der Informationsinfrastruktureinrichtung hat die GESIS den MindServer zur Suchterweiterung innerhalb des bestehenden Retrievals eingesetzt (vgl. Holl 2009). Der Fokus lag hierbei auf der Frage, ob das entwickelte Modell einer Suchterweiterung (Search Term Recommender Modul) geeignet ist, die Vagheit zwischen einerseits der

⁵⁵ Siehe auch: Aufstellung der deutschen und internationalen Anwender, online unter: http://www.recommind.de/customer_success (Zugriff: 27.10.10). Andere Anwender wurden im Rahmen dieser Arbeit nicht kontaktiert, um deren Evaluierungsansätze zu eruieren. In der TIB in Hannover läuft hierzu momentan ein Pilotprojekt (Sachstand: Februar 2010).

Suchanfrage und andererseits dem kontrollierten Vokabular zu beseitigen und somit zu einer Verbesserung der Suchergebnisse beizutragen. Laut Holl (ebd.: 65) konnte diese Fragestellung „im Rahmen der durchgeführten Tests positiv beantwortet werden“.

3 Methode, Daten, Messverfahren

Im vorangegangenen Kapitel wurde aufgezeigt, mit welchen Herangehensweisen die unterschiedlichen Studien arbeiten, um sich vor allem einer qualitativen Bewertung maschineller Indexierungsverfahren anzunähern. Nun gilt es in diesem Kapitel zunächst darzustellen, welches automatische Verfahren der maschinellen Indexierung im Rahmen der ZBW zugrunde liegt, um anschließend die Datengrundlage der vorgenommenen Untersuchung darzulegen und darauf aufbauend abschließend den Evaluierungsansatz dieser Arbeit in Bezug auf die Indexierungsqualität zu begründen.

3.1 Der MindServer als statistisches Verfahren

Automatische Indexierungsverfahren können in vier Kategorien eingeteilt werden (nach Nohr 2005: 38-39). Statistische Verfahren (1. Kategorie) bestimmen die Häufigkeit von Wörtern bzw. Termen⁵⁶ in einem Text und weisen den signifikanten, weil in der ermittelten Häufigkeitsliste⁵⁷ oben stehenden Wörtern eine sinntragende Bedeutung für den Dokumenteninhalte zu (vgl. Stock 2000: 148). Die informationslinguistischen (bzw. computerlinguistischen) Verfahren (2. Kategorie) fokussieren demgegenüber auf die Sprach- und Schriftsystemerkennung von Texten. Diese Verfahren gehen deshalb folgendermaßen vor, um die sprachliche Ebene des Dokumentes zu analysieren (vgl. Schneider 2008: 17): Entfernung nicht sinntragender Wörter, Rückführung der Flexionsform eines Wortes auf seine Grund- oder Stammform, Erkennung zusammengehöriger Wortterme und Pronomina sowie Kompositazerlegung. Informationslinguistische Verfahren können wiederum unterteilt werden in regelbasierte und wörterbuchbezogene Ansätze, die sich aber auch miteinander kombinieren lassen (vgl. Bertram 2005: 104, Nohr 2004: 217). Pattern-Matching- bzw. Mustererkennungsverfahren (3. Kategorie) bestimmen auf Basis vorher trainierter Muster die in Dokumenten entsprechend relevanten/sinntragenden Wortgruppen. Begriffsorientierte (oder additive) Verfahren (4. Kategorie) abstrahieren vom Dokumenteninhalte auf die Bedeutung des Textes und weisen diesem dann eine oder mehrere Kategorien (Deskriptoren

56 Terme/Wörter werden hierbei als Buchstabenketten und nicht als sinntragende Einheiten angesehen (Haller/Schmidt 2006: 108)

57 Auch als Rangordnung oder relevance ranking bezeichnet.

oder Notationen) aus einem kontrollierten Vokabular zu (Bertram 2005: 108, vgl. Nohr 2005: 93). Letzteres Verfahren beinhaltet auch ordnungstheoretische Aspekte (Stock 2000: 164), gilt es doch, die Vorlage in ein ontologisches Begriffssystem – hierzu zählen domänenspezifische Konzepträume – einzuordnen. Zudem weist Luckhardt (2010, Kap. 2.3) darauf hin, dass ein Additionsverfahren (vgl. S. 5), wie es in dieser Arbeit zur Anwendung kommt, generell zu einer geringeren Indexierungskonsistenz führt als eingesetzte Extraktionsverfahren.

Die in der ZBW zur automatischen Indexierung eingesetzte MindServer-Software bedient sich der PLSA (Probabilistic Latent Semantic Analysis).⁵⁸ Dahinter verbirgt sich eine Maschinenlertechnik, mit deren Hilfe in einer vorhandenen Dokumentensammlung relevante Konzepte oder Themen automatisch identifiziert und entsprechend strukturiert werden (vgl. Puzicha 2009: 7). Die Analyse der sinntragenden Teile eines Dokumentes erfolgt hierbei über ein konsequent statistisches Verfahren, das

„im Gegensatz zu linguistischen Ansätzen ein Kategorieverständnis über das gemeinsame Auftreten von Worten erlangt. Im Kategorisierungsfall werden dem System Kategorien/Taxonomien vorgegeben, die zur Verschlagwortung herangezogen werden sollen. Für jede der Kategorien werden dann Trainingsdokumente, die bereits kategorisiert sind, in das Softwaresystem eingespeist. Über diese vorkategorisierten Trainingsdokumente ist das System in der Lage, eine Konfiguration für jede Kategorie zu extrahieren, die es dem System ermöglicht, Dokumente unter Angabe der Konfidenz in die Kategorien einzusortieren“ (Hartwig Laute in Lingelbach-Hupfauer/Laute 2009: 48).

Der PLSA-Algorithmus benötigt für diese eben beschriebene Kategorisierung generell keinen Input in Form von Lexika, Klassifikationen, Thesauri oder Ontologien. Die Software stellt ein lernendes System dar, welches Informationsstrukturen aus einer Gesamtdokumentenanzahl abstrahiert und Lernmuster generiert. Dies geschieht mit Hilfe eines statistischen Verfahrens und mündet in einer quantitativen und zugleich qualitativen Beschreibung aller Dokumente (Puzicha 2009: 7). Die semantische Verknüpfung von Worttermen oder Inhaltsaspekten erfolgt über die statistische Ermittlung von Häufigkeiten. Diese Vorgehensweise ermöglicht eine sprach- und fachspezifisch unabhängige Analyse der vorliegenden Texte (vgl. Lingelbach-Hupfauer/Laute: 2009: 49).⁵⁹ Zudem ermöglicht dieser Ansatz, latent im Dokument enthaltene Inhalte durch den Abgleich mit der Gesamtdokumentenanzahl zu

58 Auch als PLSI (Probabilistic Latent Semantic Indexing) bezeichnet (vgl. Lingelbach-Hupfauer/Laute 2009: 48).

59 Aufgrund der geringen Lerndokumentenanzahl wird die automatische Indexierung in der ZBW nur im Rahmen deutsch- und englischsprachiger Dokumente sinnvoll einzusetzen sein.

erkennen, die ein manueller Indexierer durch seinen eingeschränkten Blickwinkel normalerweise nicht erkennen kann (ebd. 12).⁶⁰

Die generelle Lernfähigkeit der eingesetzten Indexierungssoftware wird durch deren Einsatz im Rahmen eines semi-automatischen Verfahrens noch erhöht. Mit Hilfe dieses Verfahrens können durch die FachreferentInnen Indexierungsfehler⁶¹ ausgebessert und Wortkombinationen, eine Schwachstelle vieler maschineller Verfahren (vgl. Gödert/Lepsky 2008: 4-6), dem System als Regeldefinition vorgegeben werden, um die Kontexterkenkung zu verbessern. Dadurch werden nicht nur statistisch häufige Muster erkannt, sondern darüber hinaus durch die Trainingsdokumente auch Gesetzmäßigkeiten konstruiert, die bei der Erschließung von neuen Dokumenten („unseen documents“) Berücksichtigung finden (Oberhauser 2005: 22).

Die in der ZBW momentan vorgenommene Implementierung einer semi-automatischen Indexierung läuft folgendermaßen ab. Zuerst wird ein Trainingsset benötigt, auf dessen Basis die Indexierungssoftware die Vergabe der einzelnen Kategorien/Schlagwörter auf Grundlage des manuellen Indexierungsverhaltens trainieren kann. In der Regel braucht das System eine ausreichende Anzahl an Dokumenten pro Kategorie (mindestens 50 Titel), um diese eindeutig zuzuordnen. Im Falle des STW, der zweisprachig ausgestaltet ist, benötigt das System für jeden Deskriptor insgesamt 100 Dokumente: jeweils 50 deutsche und 50 englischsprachige Titel.⁶² Hierbei extrahiert das System nicht nur einzelne, häufig vorkommende Stichworte, sondern Wortmuster, die wiederum für die Entscheidung bezüglich einer Kategoriezuordnung genutzt werden (Lingelbach-Hupfauer/Laute 2009: 49). Dieser Lernvorgang wird mit Hilfe des sog. Taxonomie-Browsers⁶³ durchgeführt. Auf dieser Erfahrungsbasis können jetzt neue Dokumente dem System im Rahmen des alltäglichen Geschäftsprozesses zur Verschlagwortung zur Verfügung gestellt werden. Neue Dokumente stellen damit das jeweilige Testset dar, welches unter Zuhilfenahme des Annotationstools⁶⁴ kategorisiert wird (vgl. Lingelbach-Hupfauer/Laute 2009: 48). Diese Kategorien bzw. Deskriptoren können nun vom menschlichen Indexierer angenommen, geändert oder abgelehnt werden. Das dann vollendet

60 Hier liegt die Stärke des PLSA-Ansatzes, denn gegenüber linguistischen Verfahren, denen oft lexikarisch entsprechende Synonyme und Polyseme vorgegeben werden müssen, erkennt dieser Ansatz potentielle Mehrdeutigkeiten und verwandte Begriffe auf Basis der Gesamtdokumentenanzahl (vgl. Puzicha 2009: 7).

61 Ein Beispiel hierfür ist die Wortkombination „business cycle“, die von einem getesteten linguistischen Verfahren dem Deskriptor „Fahrradverkehr“ zugeordnet worden ist.

62 Die aktuelle Version des MindServer kommt auch mit weniger Dokumenten aus (Stand: Sept. 2010). Neue Algorithmen können auch bei 5-6 Titeln für eine Indexierung ausreichende Muster erkennen.

63 Hier können den einzelnen Kategorien auch Negativbeispiele zugeordnet werden. Zudem kann der statistische Lernprozess über kategoriespezifische Regeln verfeinert werden (Recommind 2007: 13-24).

64 Zur genauen Funktionsweise siehe Holl (2009: 39ff.).

erschlossene Testdokument wird im Anschluss an diesen Arbeitsvorgang publiziert, d. h. durch den Fachreferenten als fertig bearbeiteter Titel angesehen (vgl. Recommind 2007: 38). Gleichzeitig dient dieses Testdokument wiederum als Verbreiterung der Lernbasis innerhalb des Trainingssets. Damit wird erreicht, dass nicht nur die Oberbegriffe innerhalb einer vorgegebenen Thesaurusstruktur ausreichend trainiert werden, sondern auch die entsprechenden Unterbegriffe. Neben dem Taxonomie-Browser und dem Annotationstool steht das Administrationstool zur Verwaltung der eingespielten und wieder auszuwerfenden Daten zur Verfügung.

Während die eingesetzte Indexierungssoftware MindServer auf der Basis von PLSA bei der Kategorisierung von Dokumenten grundsätzlich ohne manuellen Input, d. h. irgendeine Form von vorgegebenen Kategorien, auskommt, weil es auf Basis der vorliegenden Dokumenteninhalte lernt, erfolgt im Rahmen ihres Einsatzes an der ZBW nunmehr eine qualitative, weil semantische Erweiterung des ursprünglich rein statistischen Ansatzes. Die Ergebnisse der statistischen Textanalyse werden anschließend mit entsprechenden Thesaurusbegriffen abgebildet.⁶⁵ Dadurch erfolgt eine Zuordnung in einen domänenspezifischen Konzeptraum (vgl. Pizucha 2009: 8), in diesem Falle die Volks- und Betriebswirtschaftslehre. Durch diese Verbindung erfährt das ursprünglich rein statistische Verfahren eine Erweiterung hin zu einem begriffsorientierten Verfahren. Diese Art automatischer Sacherschließung ermöglicht eine „sprachunabhängige, auf Bedeutung abhebende Analyse“ (Nohr 2005: 93), weil sie inhalts- und nicht stichwortbezogen funktioniert.⁶⁶ Aus diesen Gründen kann das Verfahren, das dieser Arbeit zugrunde liegt, gleichfalls in den Bereich der Additionsverfahren subsumiert werden, denn es erfolgt keine reine Wortextraktion aus einem Text, sondern diese Extraktion wird mit einem informatorischen Mehrwert versehen (vgl. Oberhauser/Labner 2003: 306).

Gerade in Bezug auf die Informationsstrukturierung im Semantic-Web-Kontext, die vor allem auf die Bedeutungsebene von Informationen und dessen Kontextualisierungsfunktion bei der Wissensgenerierung abzielt (Gradmann 2009), wird sich im weiteren Verlauf dieser Arbeit zeigen, ob das anzuwendende und zu evaluierende Verfahren geeignet ist, dies zu gewährleis-

⁶⁵ Diese Zuordnung setzt eine entsprechend umfangreiche Synonymzuordnung bei den jeweiligen Deskriptoren voraus (vgl. Stock 2000: 164f.). Diese sogenannten Nicht-Deskriptoren, die im Falle des STW jeweils in deutscher und englischer Sprache vorliegen, sollen dafür sorgen, dass die in ihrer Wortwahl variierenden Begriffe im Rahmen der statistischen Analyse dieser übersehen werden.

⁶⁶ Die Wörter „business cycle“, boom, Depression, Rezession können durch statistische und informationslinguistische Verfahren, konkreter durch Häufigkeitsanalysen oder die Rückführung auf den Wortstamm, einzeln erkannt werden, dass sie aber allesamt unter dem Begriff „Konjunktur“ subsumiert werden können und damit einer gemeinsamen Bedeutung unterliegen, vermag nur ein begriffsorientiertes Verfahren zu bestimmen (vgl. Nohr 2005: 93).

ten. Grundsätzlich kommen begriffsorientierte Ansätze dem Verhalten eines menschlichen Indexierers näher als statistische oder informationslinguistische Verfahren. Die Imitation eines manuellen Arbeitsprozesses wird aber auch mit diesem Verfahren nicht vollständig gelingen, denn es wird zunächst mit statistischen Methoden und damit „letztlich wiederum über die Sprachoberfläche auf Bedeutungen geschlossen“ (Nohr 2005: 93). Trotzdem kann über die Messung der Indexierungskonsistenz dieser Abbildungsvorgang, bei dem Inhalte über zugewiesene Deskriptoren beschrieben werden, gemessen und entsprechend bewertet werden.

3.2 Datenlage und Stichprobenumfang

Der Bestand der ZBW als weltweit größte wirtschaftswissenschaftliche Spezialbibliothek umfasst momentan 4,4 Millionen Bände (ZBW 2010), die jeweils im Onlinekatalog ECONIS nachgewiesen werden. Entsprechend ihrem Sammelauftrag beschafft, erschließt und vermittelt die ZBW nationale und internationale Literatur der Wirtschaftswissenschaften und der Wirtschaftspraxis. Weiterhin erwirbt sie Publikationen der engen Nachbar- und Hilfswissenschaften, sofern sie einen ökonomischen Schwerpunkt beinhalten (Meyer/Süssenbach 2009: 971). Neben deutschen und englischen Titeln, die grundsätzlich bei der Beschaffungsentscheidung bevorzugt werden und daher einen Schwerpunkt im Bestand darstellen, wird zudem auch empirische bzw. länderbezogene Literatur in den jeweiligen Landessprachen gesammelt (ebd.: 972). Rund zwei Drittel des Bestandes (65 %) können dem Bereich der „grauen Literatur“ zugeordnet werden. Hierzu zählen vor allem die Serien von Arbeits- und Diskussionspapieren – insgesamt ca. 19.000 –, von denen 78 % in elektronischer Form verfügbar sind (ebd. S. 974) und somit für eine automatische Indexierung in Betracht kommen.⁶⁷

Aus der Gesamttitelmenge kamen ca. 120.000 Dokumente für eine automatische Sacherschließung in die engere Auswahl, weil sie grundsätzlich elektronisch vorhanden und damit für eine Software bearbeitbar waren. Diese Teilmenge wurde für die vorliegende Arbeit um diejenigen Titel bereinigt, bei denen ein Zugriff technisch ohne weiteres nicht möglich gewesen ist (u. a. broken links). Als Stichprobe verblieben 38.878 Dokumente, die somit für

⁶⁷ Von den 27.000 laufend abonnierten Zeitschriften haben 62,2 % einen elektronischen Zugang zum Volltext. Aus lizenz- und urheberrechtlichen Gründen ist ein Zugriff auf diese Daten im Rahmen eines automatischen Indexierungsverfahrens momentan an der ZBW noch nicht möglich.

eine automatische Sacherschließung und deren anschließende Evaluierung zur Verfügung standen. Die Dokumente lagen überwiegend im PDF-Format vor und waren unterschiedlich strukturiert.⁶⁸ Ergänzend muss darauf hingewiesen werden, dass eine Doublettenbereinigung nicht durchgeführt worden ist. Zudem wurde eine zeitliche Eingrenzung automatisch zu indexierender Dokumente nicht vorgenommen.

Die verbliebene Teilmenge von rund 39.000 Titeln wurde nun mit Hilfe des in Kapitel 3.1 beschriebenen automatischen Indexierungsverfahrens inhaltlich erschlossen, indem von der Automatik jedem Dokument STW-Deskriptoren zugewiesen worden sind. Dabei konnten 26.645 Dokumente als Trainingsset und somit für eine Lernbasis genutzt werden, weil sie bereits manuell zugewiesene Deskriptoren aufwiesen. Die verbleibende Teilmenge von 12.233 Titeln konnte durch die Indexierungssoftware als Testset auf Basis des beschriebenen Trainings gleichfalls, aber nunmehr ohne manuelle Vorgaben erschlossen werden. Der STW enthält in der neuesten Version (8.04) insgesamt 5.770 Deskriptoren, die für eine manuelle bzw. automatische Indexierung zur Verfügung standen.

3.3 Zum Messverfahren: Die Evaluierung der Indexierungsqualität

Die Ergebnisevaluierung eines automatischen Indexierungsverfahrens kann mit einer Qualitätskontrolle verglichen werden (vgl. Stock 1993: 135). Dabei soll sichergestellt werden, dass eine Informationsdienstleistung – hierzu zählt meiner Ansicht nach die Abbildung der Inhalte digitaler Dokumente mit Hilfe von Schlagwörtern – auch einen informatorischen Mehrwert darstellt. Im Falle der in der ZBW aktuell zum Einsatz kommenden Indexierungssoftware soll mit der vorgenommenen Bewertung der Indexierungsqualität auch versucht werden zu überprüfen, ob sich manuelle und automatische Sacherschließung hinsichtlich des Arbeitsergebnisses im Rahmen eines begriffsorientierten Verfahrens angleichen, wie beispielsweise von Nohr (2005: 93) ausgeführt wird. Demnach verlangt die Erweiterung des ursprünglich statistischen Verfahrens um eine semantische Komponente (kontrolliertes Vokabular) hin zu einem bedeutungsorientierten Ansatz nach einem geeigneten Messinstrument, mit dem die Indexierungsqualität der Ergebnisse entsprechend bewertet werden kann. Stock (2000: 82) fasst das Messproblem pointiert zusammen, wenn er feststellt, dass „die

68 Neben Autorennamen, Titeln, Literaturverzeichnissen, die in allen Dokumenten vorkamen, gab es deutliche Streuungen z. B. hinsichtlich des Vorkommens eines Abstracts oder Keywords.

Kunst des Indexierens darin besteht, die *relevanten*⁶⁹ Informationsinhalte herauszufiltern sowie durch die bestpassenden Deskriptoren, Notationen o. ä. auszudrücken.“ (vgl. Stock/Stock 2008: 345). Wie kann also die Indexierungskunst, die ja bekanntlich eine stark subjektive Komponente hat (vgl. Stubbs et al. 1999), gemessen werden?

Neben qualitätsbezogenen Indexierungsaspekten, die durchweg allgemeiner Natur sind, z. B. Aktualität, Schnelligkeit, Vollständigkeit (Bertram 2005: 72), kann die Qualität oder Güte der Sacherschließung, wie in Abbildung 1 dargestellt, anhand von vier Kriterien untersucht werden (nach Stock 2008: 355):

Abbildung 1: Indexierungsgütekriterien

- Indexierungstiefe
 - Indexierungsbreite
 - Indexierungsspezifität
- Indexierungseffektivität
- Indexierungskonsistenz

Die Indexierungstiefe setzt sich aus der Indexierungsbreite und der Indexierungsspezifität zusammen. Die Indexierungsbreite (IB) berechnet in Bezug auf den fachlichen Inhalt den Erschließungsgrad der zu analysierenden Vorlage, indem die Anzahl der vergebenen Schlagwörter (oder Notationen bei Klassifizierungssystemen) pro Vorlage ermittelt wird (vgl. Stock/Stock 2008: 355). Die Indexierungsspezifität (IS) versucht die Hierarchie des Thesaurus zu erfassen, indem die vergebenen Deskriptoren unter Beachtung ihres hierarchischen Niveaus (Ober- und Unterbegriffe) bewertet werden: „Je höher ein Term in der Begriffsleiter steht, desto niedriger ist seine Spezifität.“ (Stock 2000: 82).⁷⁰ Die Berechnung der Indexierungsspezifität erfolgt folgendermaßen (Stock 2008: 355):

$$\text{Indexierungsspezifität (IS)} = \{ \text{ld}[\text{HE}(\text{B}_1)+1] + \dots + \text{ld}[\text{HE}(\text{B}_n)+1] \} / S$$

69 Hervorhebung im Original.

70 Die Sacherschließungsregeln der ZBW (2007) geben u. a. vor, dass der bestpassende Deskriptor dem Inhalt zuzuordnen ist. Statt den Oberbegriff „Branchenentwicklung“ bei Untersuchungen zur Lage der Bauwirtschaft zu geben, muss der Unterbegriff „Baukonjunktur“ zwingend zugeordnet werden.

HE bezeichnet die hierarchischen Ebenen, denen die Deskriptoren jeweils zugeordnet sind. Dabei wird im Falle polyhierarchischer Thesauri, wie dem STW, immer der kürzeste Weg zum Top-Oberbegriff berücksichtigt. Der Berechnung des logarithmus dualis (ld) liegt die Überlegung zugrunde, dass die jeweiligen Unterbegriffe nicht doppelt so spezifisch sind wie der dazugehörige Oberbegriff (vgl. ebd.).⁷¹ Durch die Summierung der errechneten Indexierungsspezifitätswerte und Normierung anhand der jeweiligen Seitenzahlen (S) erfüllt die Formel für die Berechnung der Indexierungstiefe gleichzeitig deren Bedingung hinsichtlich der Berücksichtigung der Indexierungsbreite.

$$\text{Indexierungstiefe (IT)} = \{\text{ld}[\text{HE}(\text{B}_1)+1]+\dots+\text{ld}[\text{HE}(\text{B}_n)+1]\}/\text{S}$$

Die Berechnung der Indexierungstiefe geht davon aus, dass „im konkreten Fall von zwei Indexierungsergebnissen des gleichen Dokuments mit gleicher Anzahl von Deskriptoren oder Notationen dasjenige tiefer ist, das die spezifischeren Bezeichnungen enthält.“ (DIN 31623/1, 4, zit. nach Stock 2000: 82).⁷²

Die Indexierungseffektivität (IE) errechnet, wie oft ein Deskriptor im jeweiligen Datensatz vergeben worden ist, um so die Trennschärfe der verwendeten Schlagwörter zu erfassen (Stock 2008: 357).⁷³ In Anlehnung an Stock (ebd. 357f.) erfolgt die Berechnung folgendermaßen:

$$\text{Indexierungseffektivität (IE)} = \text{IDF}(\text{B}) = [\text{ld}(\text{N}/\text{n})]+1$$

IDF bezeichnet die inverse Dokumenthäufigkeit,⁷⁴ N umfasst die Gesamtzahl an allen Titelsätzen in einer Datenbank und n bezieht sich auf die Teilmenge an Titelsätzen, denen der Deskriptor (B) jeweils zugeordnet ist. Je häufiger ein Schlagwort vergeben worden ist, desto geringer ist die Indexierungseffektivität und umgekehrt. Das Maximum an Effektivität liegt bei einem vergebenen Deskriptor im ganzen Datensatz, wohingegen das Minimum von der Gesamttitelanzahl determiniert wird. In dieser Arbeit wird der Faktor N durch die Gesamt-

71 Weil $\text{ld}1 = 0$, wird der Wert 1 addiert, um zu vermeiden, dass der Top-Term (Hierarchiestufe 1) den Wert Null annimmt.

72 Es kann nicht davon ausgegangen werden, dass die Indexierungstiefe umso besser ist, je größer die jeweiligen Werte sind. Der Kurvenverlauf entspricht einer Gauss'schen Glockenkurve, mit der Einschränkung, dass der Höhepunkt der Kurve nicht a priori festgelegt, sondern nur über Nutzerstudien ermittelt werden kann (nach Stock 2010).

73 Klassisches Beispiel sind oft vergebene Geo-Begriffe wie „Deutschland“ oder „USA“ sowie die in der ZBW vorgenommene Verschlagwortung von theoretischen Abhandlungen mit dem Deskriptor „Theorie“.

74 „Die Inverse Dokumenthäufigkeit (IDF) dient [...] der Trennfähigkeit eines Wortes bzw. Termes für die Indexierung von Dokumenten. Ein Wort, das nur in wenigen Dokumenten oft vorkommt, ist geeigneter als eines, das in fast jedem Dokument [...] auftritt.“ (zit. nach Wikipedia 2010).

stichprobe von 38.878 Dokumenten festgelegt, denn es gilt, die Schlagworttrennschärfen innerhalb der Stichprobe zu bestimmen, um zu schauen, inwieweit sich die Indexierungseffektivitäten zwischen den manuellen und den automatischen Verfahren gestalten. Hierzu ist eine Berücksichtigung aller Titelsätze aus der Datenbank ECONIS ungeeignet.

Die Indexierungskonsistenz (IK) als weiteres wichtiges Qualitätskriterium (vgl. Bertram 2005: 72) misst den Grad der Übereinstimmung unterschiedlicher Indexierungsergebnisse des gleichen Dokuments in derselben Dokumentationssprache. Die Konsistenz wird folgendermaßen berechnet:

$$\text{Indexierungskonsistenz (V1, V2)} = A / B+C-A^{75}$$

A umfasst die Anzahl der übereinstimmenden Deskriptoren der Vorlagen (V1, V2). Im Rahmen dieser Arbeit steht V1 für ein manuell erstelltes Indexat und V2 stellt ein automatisch generiertes Indexat dar. B und C sind die Summe der vergebenen Deskriptoren beider Vorlagen (= Indexierungsbreite).⁷⁶

Die Berechnung der Konsistenz zwischen unterschiedlichen Indexaten der gleichen Vorlage gibt einen ersten qualitätsbezogenen Anhaltspunkt, der zur Evaluierung des momentan in der ZBW zum Einsatz kommenden automatischen Indexierungsverfahrens dienen kann. Die Indexierungskonsistenz kann dabei als kombinierter Recall-Precision-Wert angesehen werden. Recall: Wie viele von den zu findenden Deskriptoren werden gefunden (Vollzähligerkeitsrate)? Precision: Wie viele von den gefundenen Deskriptoren sind richtig, i. S. v. einer präzisen Inhaltsbeschreibung? Beide Maße gehen davon aus, dass die Vorgabe des manuellen Indexierens getroffen werden muss.

Wenn eine Indexierungsmethode so beschaffen ist, dass ein „ideales“ Indexierungsergebnis (im Sinne einer Methode) verbindlich vorgegeben werden kann, kann die Abweichung von diesem vorgegebenen Standard in einer Fehlerstatistik erfasst werden. (Knorz 2004: 186)

75 Entspricht der Formel nach Hooper (1965). Eine andere Berechnungsmöglichkeit ist nach L. Rolling (1981) benannt und lautet $2 \cdot A / B + C$. Nach diesen zwei Formeln werden in der Regel Indexierungskonsistenzen berechnet (vgl. Medelyan/Witten 2006: 1).

76 Beispiel: Vorlage V1 hat zehn Deskriptoren, Vorlage V2 hat fünf Deskriptoren, die Anzahl der übereinstimmenden Schlagwörter ist zwei, also $2 / 10 + 5 - 2 = 15 \%$. (Die Anzahl der übereinstimmenden Deskriptoren muss im Nenner wieder subtrahiert werden, weil sonst die Ergebnisse nie einen Wert größer als 50 % annehmen können.)

Gleichwohl können auf Basis dieser Berechnungen immer nur Tendenzaussage getroffen werden, denn „der blind geführte Schluss von der Indexierungskonsistenz zur Indexierungsgüte ist [...] verfehlt.“ (Stock/Stock 2008: 358, ähnlich Lancaster 2003: 91)⁷⁷. Die Konsistenz kann generell den Wert Null annehmen, wenn davon ausgegangen wird, dass der menschliche Indexierer falsche oder nicht die bestpassenden Begriffe gewählt hat, das automatische Verfahren hingegen mit den maschinell generierten Begriffen die Vorlage richtig erschlossen hat. Es ist demnach grundsätzlich denkbar, dass die vorliegende Stichprobe von ca. 39.000 Dokumenten durch die Indexierungssoftware besser im Sinne der Indexierungsbreite und -spezifität beschlagwortet worden ist als durch die entsprechenden FachreferentInnen. Weiterhin können die automatisch erstellten Indexate das Informationsbedürfnis des Nutzers vielleicht eher treffen als die manuellen Indexierer (= Indexer-Nutzer-Konsistenz, vgl. Copper 1969). Auch wenn der ermittelte Konsistenzwert die Qualität der automatisch generierten Ergebnisse vielleicht unterschätzt, liefert er doch einen Anhaltspunkt zur Evaluierung dieser.

Quality and consistency are not the same: one can be consistently bad as well as consistently good! Nevertheless, one intuitively feels that consistency and quality should be related. (Lancaster 2003: 91)

Die Erweiterung des vorgestellten PLSA-Verfahrens hin zu einem begriffsorientierten Verfahren, welches gerade auf Basis des manuellen Indexierungsverhaltens lernen soll, geht dabei natürlich von der Annahme aus, dass die manuellen Vorgaben richtig sind („Goldener Standard“). Die Absolutheit dieser Annahme wird durch den Umstand gemildert, dass durch das Training der Automatik „systemimmanente“ Indexierungsfehler aller Sacherschließer und Sacherschließerinnen durch die Masse an Dokumenten besser herausgefiltert und entsprechend korrigiert werden können. Weiterhin wurde die Indexierungssoftware dahin konfiguriert, dass ein Ausweichen auf die entsprechenden Oberbegriffe und damit fehlendes Training der Tiefenhierarchie des STW nicht erfolgt, sondern dass versucht werden soll, den Thesaurus in seiner teilweise stark ausgeprägten granularen Gesamtheit zu trainieren.

Gleichwohl stellt in dieser Arbeit die Indexierungskonsistenz nicht das alleinige Evaluierungsmerkmal der Indexierungsgüte dar, wie in Abbildung 2 aufgeführt. Die Auswertung und Darstellung der automatischen Sacherschließungsergebnisse im anschließenden 4. Kapitel erfolgt sowohl unter qualitativen als auch unter quantitativen Aspekten; hierbei jeweils für das Trainingsset (= 26.645 Dokumente) und das Testset (= 12.233 Dokumente). Die Belegungs-

⁷⁷ Für Cooper (1969: 268) ist „interindexer consistency a hobgoblin“, der nicht länger bei der Frage nach der Güte von Indexierungsergebnissen einem im Geiste herumspuken dürfe.

bilanz des Thesaurus erfasst durch eine Abbildung der thematischen Schwerpunkte die konkrete Nutzung der angebotenen Vielfalt des STW in Form der 5.770⁷⁸ kontrollierten Schlagwörter.

Abbildung 2: Analysedimensionen

	Trainingsset	Testset
qualitativ	Indexierungskonsistenz Referentenauswertung*	Referentenauswertung*
quantitativ	Indexierungstiefe* Indexierungsbreite Indexierungsspezifität* Indexierungseffektivität* Belegungsbilanz	Indexierungstiefe* Indexierungsbreite Indexierungsspezifität* Indexierungseffektivität* Belegungsbilanz

Eigene Darstellung: *Unterstichproben,⁷⁹ sonst Gesamtstichprobe.

78 STW-Version 8.04 (Stand: 22.02.2010).

79 Aufgrund der Datenlage war es für einzelne Evaluierungsmaße nicht möglich, die Gesamtstichprobe für eine entsprechende Auswertung heranzuziehen. Beispielsweise kann die Indexierungsspezifität nur unter Berücksichtigung der Seitenzahlen pro Dokument ermittelt werden. Dieser Wert lag auf bibliographischer Ebene nicht immer in verwertbarer Form vor, z. B. durch die Kilobyte- statt Seitenzahlangebe. Für die einzelnen Evaluierungsmaße wurden folgende Dokumentenzahlen jeweils zugrunde gelegt: Referentenauswertung Trainingsset und Testset jeweils 128 sowie Indexierungsspezifität Trainingsset 86/ Testset 99.

4 Ergebnisse

Die Evaluierungsergebnisse des in der ZBW im Rahmen der automatischen Sacherschließung von wirtschaftswissenschaftlicher Literatur eingesetzten maschinellen Indexierungsverfahrens werden in diesem Kapitel dargestellt. Zunächst wird auf die Indexierungskonsistenz eingegangen (Kapitel 4.1). Anschließend werden die Berechnungen zur Indexierungstiefe, insbesondere für deren Unterasspekte, die Indexierungsbreite und -spezifität, dargestellt (Kapitel 4.2). Hierbei werden auch deren Einflüsse auf die Indexierungskonsistenz näher beleuchtet. Darauf folgt eine Darstellung der Resultate zur Indexierungseffektivität (Kapitel 4.3) und zur Belegungsbilanz (Kapitel 4.4). Den Abschluss bilden die Ergebnisse der durchgeführten Referentenauswertung (Kapitel 4.5).

4.1 Indexierungskonsistenz

Die Berechnung der Indexierungskonsistenz zwischen unterschiedlichen Indexaten der gleichen Vorlage, in diesem Falle digitaler Dokumente, geht von der Annahme aus, dass eine Vorgabe, in unserem Falle des manuellen Sacherschließungsverfahrens, durch die automatische Indexierung möglichst genau getroffen werden muss. Die ermittelten Abweichungen vom vorgegebenen Indexierungsergebnis – im Sinne einer Fehlerstatistik – geben den Grad der Indexierungsgüte wieder (vgl. Knorz 2004: 186). Aufgrund der Tatsache, dass nur das Trainingsset (26.645 Dokumente) bereits intellektuell erschlossene Titel aufweist, beschränken sich die nachfolgenden Aussagen auf rund 2/3 der durch das automatische Verfahren erschlossenen Dokumente.

Der Konsistenzwert zwischen den Indexierungsergebnissen der eingesetzten Automatik und denen der manuellen Indexierer variiert zwischen 0 % als Minimum und 100 % als Maximum.⁸⁰ Der Mittelwert der Indexierungskonsistenz liegt bei allen Dokumenten des Trainingssets bei 36 %. Somit schafft es die Indexierungssoftware, mit den manuellen Indexierungsvorgaben zu rund 1/3 übereinzustimmen. Auf Basis dieses Ergebnisses ist eine qualitative, weil konsistente Informationsstrukturierung digitaler Dokumente durch den

⁸⁰ Vgl. Tabelle 5: Hier findet sich ein Auszug aus den vorgenommenen Berechnungen zur Indexierungskonsistenz.

Einsatz eines automatischen Indexierungsverfahrens möglich, aber nur im Ansatz (= 36 %). Der überwiegende Teil der digitalen Titel würde nicht dem manuellen Indexierverhalten entsprechen.⁸¹

Einen Konsistenzwert von 0 % wiesen insgesamt 718 (oder 2,8 % aller) Dokumente auf. Die stark variierenden Indexierungsbreiten der jeweiligen automatisch generierten Indexate zeigen deutlich, dass es der Indexierungssoftware trotz der Tatsache, dass stellenweise bis zu 56 bzw. 65 Schlagwörter den einzelnen Titeln zugeordnet worden sind, oft nicht gelingt, die Vorgaben der manuellen Indexierer überhaupt im Ansatz zu treffen. Ferner ist festzustellen, dass selbst diejenigen Deskriptoren, die Allgemeinwortcharakter haben (bspw. Risiko, Theorie, Messung, Bewertung) und die sonst zuverlässig und regelmäßig durch das automatische Verfahren vergeben worden sind, im Falle der Titel mit 0%-Konsistenz generell nicht zugeordnet wurden. Gleiches kann für die häufig vorkommenden Geodeskriptoren – hierzu zählen „Deutschland“, „USA“, „EU-Staaten“ oder „Welt“ – eruiert werden.

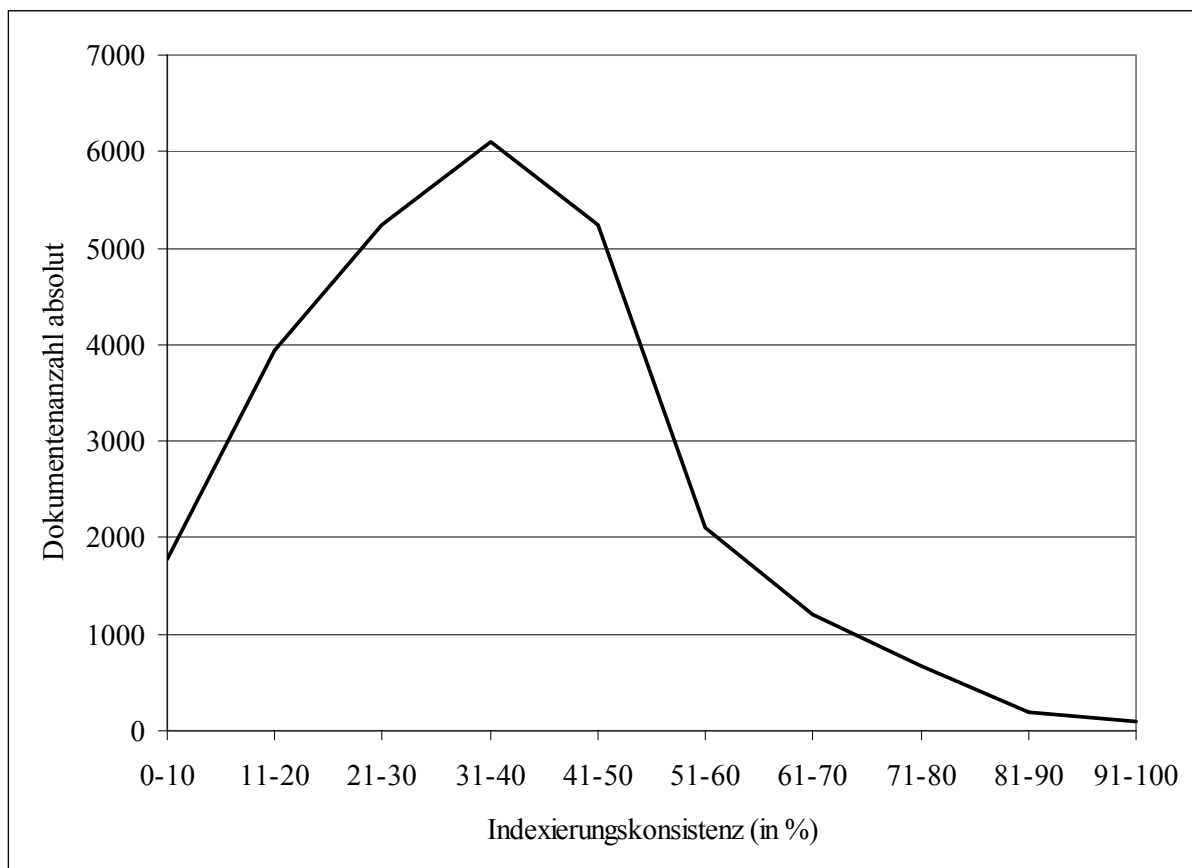
Bei diesen Ergebnissen ist zu beachten, dass eine Abweichung der maschinell generierten Schlagwörter von der manuellen Vorgabe, sei es durch die Vergabe von entsprechenden Ober-, Unter- oder verwandten Begriffen, immer einen ganzen Fehler darstellt. Mitunter können diese Indexierungsabweichungen des automatischen Verfahrens keinen Fehler im Sinne eines qualitativen „Fehlgriffes“ darstellen. Dies ist dann der Fall, wenn sie die Vorlage besser, im Sinne einer Entsprechung mit den Nutzerinteressen bzw. Suchstrategien indexieren, als dies die Fachreferenten tun (vgl. Cooper 1969). Ausgehend von dieser Annahme, können Null-Prozent-Ergebnisse dahingehend interpretiert werden, dass die Automatik deutlich bzw. vollkommen von den manuellen Vorgaben bzw. vom Indexierverhalten abweichen kann. Die Bandbreite dieser Abweichungen bewegt sich innerhalb folgender Pole: von *völlig richtig, nur anders indexiert*, z. B. durch richtige Ober-, Unter- oder verwandte Begriffe, bis hin zu *völlig falsch indexiert*, durch die Vergabe offensichtlich falscher Deskriptoren. Eine vollständige Verifizierung dieser Aussage wäre nur bei einer qualitativen, manuellen Kontrolle aller ca. 26.500 indexierten Trainingsdokumente möglich und ist aus naheliegenden Gründen unrealistisch. In der Tendenz schafft es die Indexierungssoftware aber trotzdem, durchschnittlich 36 % der digitalen Dokumente richtig im Sinne der manuellen Vorgabe zu indexieren und somit konsistent inhaltsbezogen zu strukturieren.

81 Diese Aussagen beruhen auf einer bisher nur rudimentär trainierten Indexierungssoftware, die ihr Potential noch längst nicht ausgeschöpft hat.

Das andere Extremum stellen die 90 (oder 0,3 % aller) Dokumente da, die eine 100%ige Konsistenz zwischen den manuell indexierten und den automatisch vergebenen Deskriptoren aufweisen. Während die Indexierungsbreite im Falle der 0%-Konsistenz-Titel noch stark variierte und sehr hohe Werte annehmen konnte (bis 65 Schlagwörter), liegt die Indexierungsbreite beim automatischen Verfahren im Bereich der 100%-Titel bei einem bis acht Schlagwörtern. Mit Fokus auf die vergebenen Deskriptoren können keine besonderen Muster hinsichtlich deren Vergabe festgestellt werden. Die Indexate umfassen die unterschiedlichsten wirtschaftswissenschaftlichen Teilgebiete (BWL, VWL, Nachbarwissenschaften), thematischen Aspekte sowie Länder(-bereiche).

Abbildung 3 veranschaulicht die Verteilung der Indexierungskonsistenzwerte über das gesamte Trainingsset.⁸²

Abbildung 3: Anzahl der Dokumente pro Indexierungskonsistenzintervall⁸³



⁸² Vgl. Tabelle 6: Dokumentenanzahl (absolut) pro Indexierungskonsistenzwert (0-100 %).

⁸³ Eine graphische Darstellung der einzelnen Ergebnisse für das gesamte Trainingsset ist aus technischen Gründen (Spezifikationen der Tabellenkalkulation) und Gründen der Übersichtlichkeit nicht möglich. Deshalb wurden die Konsistenzwerte in Intervalle zusammengefasst und die absolute Dokumentenanzahl jeweils additiv zugeordnet.

Der größte Teil der Dokumente weist eine Indexierungskonsistenz zwischen 20 und 50 % auf, wobei der Kurvenscheitelpunkt im Bereich des errechneten Mittelwertes von 36 % liegt. Für die Werte ab 50 bis 80 % zeigt sich ein steiler Abfall in der Titelzahl. Im Vergleich zu den rund 26.500 indexierten Trainingsdokumenten weisen nur sehr wenige Titel – rund 300 – höhere Konsistenzwerte als 80 % auf. Im Gegensatz hierzu existiert eine hohe Anzahl an Dokumenten, die nur eine geringe Indexierungskonsistenz (unter 21-30 %) aufweisen.

Zusammenfassend zeigt der Vergleich zwischen den manuellen und den automatisch generierten Sacherschließungsergebnissen im Sinne der Indexierungskonsistenz, dass beide Indexierungsansätze, die oft in einem strengen Gegensatz zueinander gesehen werden, weil sie von unterschiedlichen Annahmen ausgehen (siehe Knorz 2004), über das eingesetzte begriffsorientierte Verfahren entsprechend zusammengeführt werden können. Es ist eine qualitativ hochwertige Sacherschließung in Form einer Entsprechung beider Verfahren im Ergebnisbereich möglich. Dies zwar nicht zu 100 %, aber die Basis ist, wie die Verteilung der Kurve in Abbildung 3 zeigt, vorhanden. Letztlich wird erst der umfassende Einsatz in mittelfristiger Perspektive das wahre Potential des eingesetzten automatischen Verfahrens zeigen können. Das „intuitive Gefühl“ (Lancaster 2003: 91) ist, wie gerade dargestellt wurde, vorhanden; ob es sich verfestigt, werden die folgenden Evaluierungsmaße zeigen.

4.2 Indexierungstiefe

Die Indexierungstiefe setzt sich aus den Maßen der Indexierungsbreite, die den Erschließungsgrad quantitativ anhand der jeweiligen Deskriptorenanzahl misst, und der Indexierungsspezifität, die das hierarchische Niveau der vergebenen Schlagwörter im Thesaurus bestimmt, zusammen. Durch die vorgenommene Berechnung anhand der Formel nach Stock (2003: 355), der die Werte für die Indexierungsspezifität pro Dokument aufsummiert und somit die Indexierungsbreite mit abbildet, entsprechen die nachfolgenden Ergebnisse für die Indexierungsspezifität denen der Indexierungstiefe. Zunächst wird jedoch zuerst die Indexierungsbreite für die beiden Sets (Trainings-, Testset) dargestellt (Kapitel 4.2.1), um anschließend über die Indexierungsspezifität die Indexierungstiefe der manuell bzw. automatisch erschlossenen Dokumente zu bestimmen (Kapitel 4.2.2).

4.2.1 Indexierungsbreite

Der Erschließungsgrad des Trainingssets, gemessen an der Anzahl der vergebenen Schlagwörter pro Dokument, variiert im Vergleich beider Verfahren stark. Die manuellen Indexierer vergeben mindestens immer zwei Deskriptoren pro Dokument, jedoch höchstens 17.⁸⁴ Die Vergabe von mehr als zehn Schlagwörtern ist jedoch in Bezug auf die Gesamtdokumentenzahl des Trainingssets sehr selten.⁸⁵ Im Durchschnitt werden fünf Deskriptoren den zu indexierenden Vorlagen intellektuell zugeordnet.⁸⁶ Im Vergleich hierzu vergibt das automatische Verfahren im Mittel acht Schlagwörter und erschließt somit die einzelnen Titel in Bezug auf die Indexierungsbreite umfassender als das intellektuelle Verfahren.

Die maschinellen Indexierungsergebnisse für das Trainingsset zeigen grundsätzlich zwei Extrempunkte. Einerseits weist das automatische Verfahren in zahlreichen Fällen den Dokumenten überhaupt keine Deskriptoren zu. Die Ursachen hierfür liegen in dem ungenügenden Training einzelner Deskriptoren begründet.⁸⁷ Auf der anderen Seite vergibt das maschinelle Sacherschließungsverfahren, wie bei einer Vorlage geschehen, gleich insgesamt 71 Schlagwörter. Zudem finden sich zahlreiche Titel mit sehr hohen Indexierungsbreiten, wie die Abbildung 4 veranschaulicht. Diese Indexate sind für eine qualitativ gute Informationsstrukturierung nicht zu gebrauchen, weil sie das Hauptthema eines Dokumentes nicht fokussiert wiedergeben, sondern eher Randbereiche erschließen.

Dieses Ergebnis wird noch durch den Umstand verschärft, dass die automatisch zugewiesenen Schlagwörter zudem alle eine unterschiedliche Gewichtung (cut-off-Level) haben. Die Analyse einer Stichprobe von einzelnen Titeln mit sehr hohen Indexierungsbreiten verdeutlicht, dass in diesen Fällen fast alle Schlagwörter eine sehr geringe Gewichtung (= einstellige Prozentwerte) aufweisen und für das Retrieval überwiegend Ballast darstellen. Hier existieren hinsichtlich der automatischen Indexierung gleich zwei Optimierungsmöglichkeiten: zum einen in Bezug auf die Reduzierung der Deskriptorenanzahl pro Vorlage und zum anderen in Bezug auf deren Gewichtung, wie auf der folgenden Seite deutlich wird.

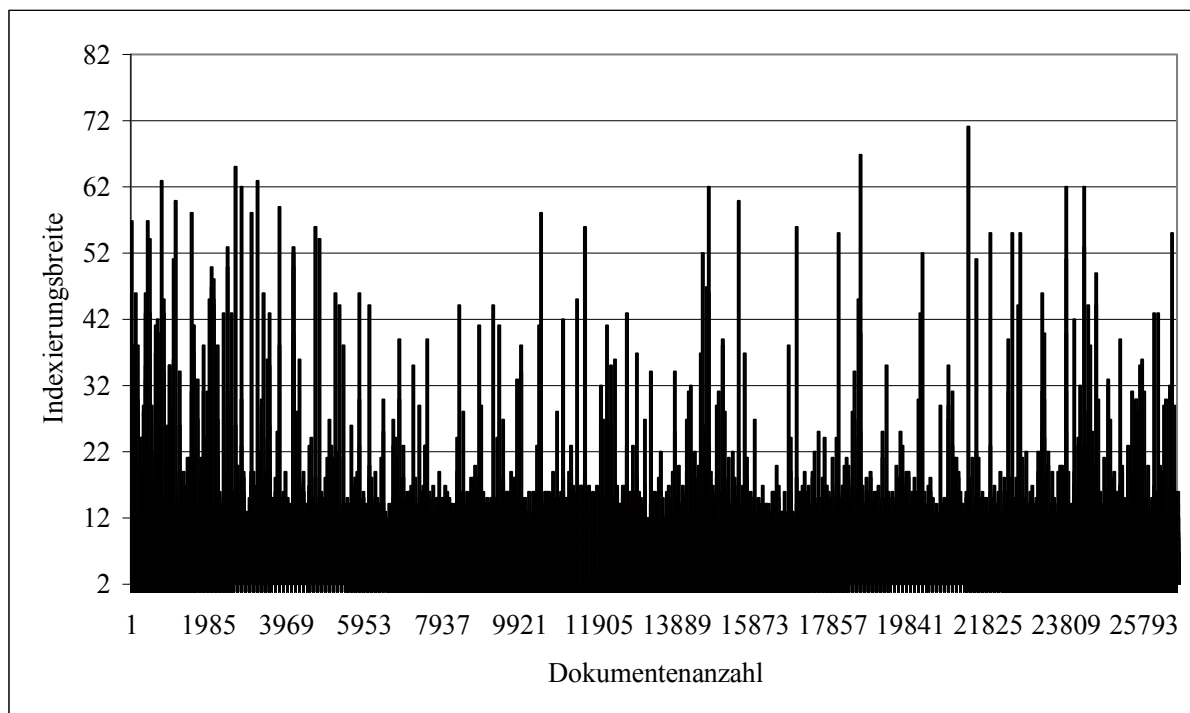
84 Gemäß den Sacherschließungsregeln der ZBW (2007: 6) muss dem Dokument mindestens ein Deskriptor zugeordnet werden. Die Praxis zeigt, dass bei einem Sachschlagwort mindestens immer ein Geodeskriptor oder ein Allgemeinwort (Theorie, Messung) vergeben worden ist.

85 Sie liegt z. B. bei (Schlagwörter/Titelzahl): 13/30, 14/12, 15/2, 16/5, 17/2.

86 Siehe hierzu auch Tabelle 5 im Anhang, die einen Auszug aus den Ergebnissen zur Indexierungsbreite des Trainingssets wiedergibt.

87 Wenn die jeweils für einen Deskriptor berechneten Häufigkeiten einen bestimmten Relevanzwert nicht übersteigen, wird durch die Indexierungssoftware auf die Vergabe der entsprechenden Begriffe verzichtet.

Abbildung 4: Indexierungsbreite: maschinelles Verfahren (Trainingsset)



Im Vergleich zum automatischen Indexierungsverfahren, welches die Schlagwörter mit Gewichtung ausweist, vergeben die manuellen Sacherschließer und Sacherschließerinnen die einzelnen Deskriptoren mit der Wertigkeit „eins“. Weil eine hohe Indexierungsbreite den Ballast für das Retrieval erhöht, vergeben die manuellen Indexierer und Indexiererinnen pro Dokument im Durchschnitt nur fünf Deskriptoren. Im Falle der zum Teil sehr hohen Indexierungsbreiten auf Seiten der automatisch generierten Indexate müssten, u. a. für eine sinnvolle Darstellung im OPAC, deshalb Schlagwörter „abgeschnitten“ werden.⁸⁸ Dies könnte auf zweierlei Wegen erfolgen. Entweder durch die Berücksichtigung eines absoluten cut-off-Levels – hierbei werden nur die ersten paar Deskriptoren berücksichtigt, unabhängig von deren Gewichtung –, oder es finden alle Deskriptoren mit einem bestimmten Gewichtungsniveau, z. B. >50%, Eingang ins Indexat.

Weiterhin ist festzustellen, dass die manuellen Sacherschließer auf eine Postkoordination⁸⁹ der einzelnen Begriffe achten, während das automatische Verfahren dies nicht berücksichtigen kann. Im Falle der Software werden diejenigen Deskriptoren vergeben, die sehr gut

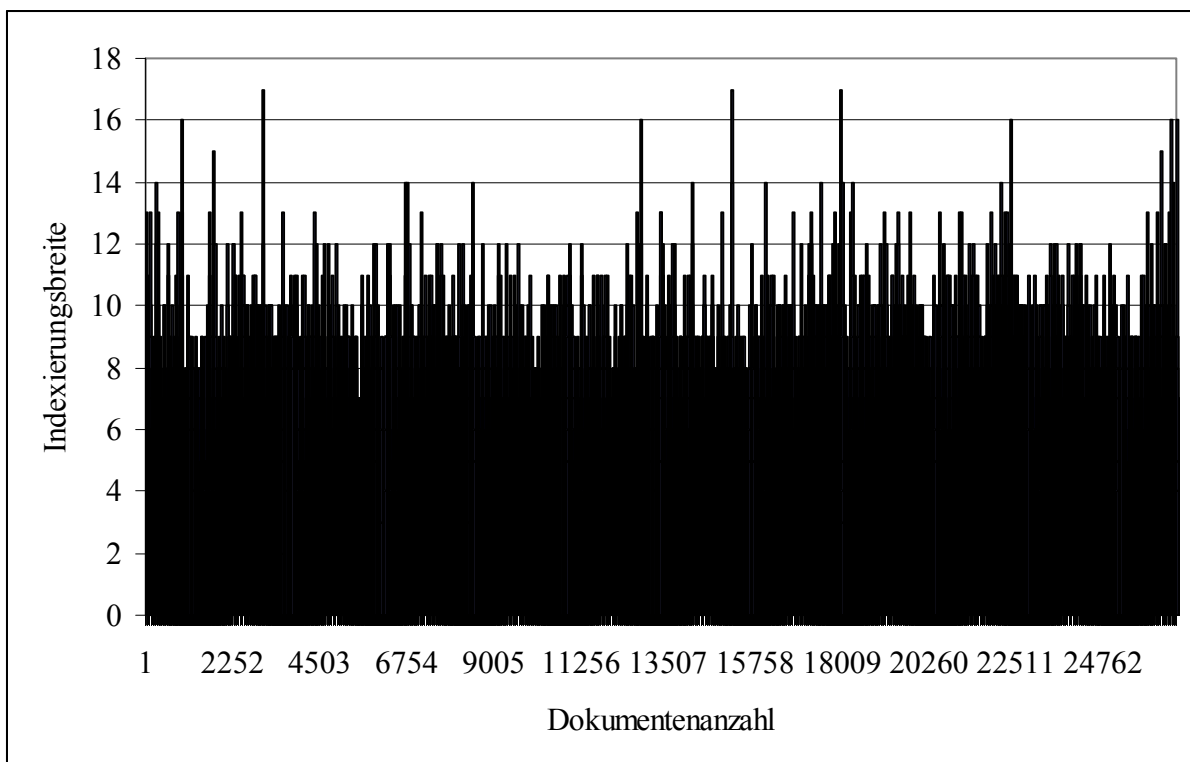
⁸⁸ Für die Analysen im Rahmen dieser Arbeit wurde auf eine Berücksichtigung des Einflusses verschiedener cut-off-Level verzichtet. Im weiteren Verlauf des Einsatzes des MindServers in den nächsten Jahren ergibt sich hier noch Bedarf an Forschung, die beispielsweise Fragen des Trainingsverhaltens und der Veränderung der Gewichtungen umfassen und klären könnte.

⁸⁹ „Mit Postkoordination ist gemeint, dass Begriffe durch die Tätigkeit der Sacherschließung zusammengeführt werden, z. B. Führungskräfte UND Gehalt [...]“ (ZBW 2007: 6).

trainiert sind, d. h. deren Vergabemuster durch den Menschen der Maschine entsprechend vorgeben worden ist. Dies führt dazu, dass Allgemeinbegriffe wie Theorie, Messung, Risiko oder Simulation an erster Stelle stehen, gefolgt von den Geodeskriptoren. Die wichtigen, weil den Inhalt eigentlich strukturierenden Sachschlagworte (z. B. Arbeitslosigkeit, Wirtschaftspolitik), werden erst danach aufgelistet, je nach ermittelter Gewichtung.⁹⁰ Hierbei zeigt sich erneut die unterschiedliche Funktionsweise beider Verfahren. Die fehlende Postkoordinierung innerhalb der automatisch generierten Indexate kann aus diesem Grunde nicht als Fehler oder Mangel der Indexierungssoftware angesehen werden.

Die Abbildung 5 veranschaulicht, ebenfalls für das Trainingsset, die unterschiedliche Indexierungsbreite der manuell verschlagworteten Titel. Dabei wird, im Vergleich zum automatischen Verfahren (siehe Abbildung 4), die generell wesentlich geringere Anzahl an Schlagwörtern pro Vorlage deutlich. Es finden sich keine Titel, mit mehr als 17 Schlagwörtern. Gleichfalls ist die Vergabe von mehr als 10 Deskriptoren selten und Indexate mit >10 Schlagwörtern die Regel.

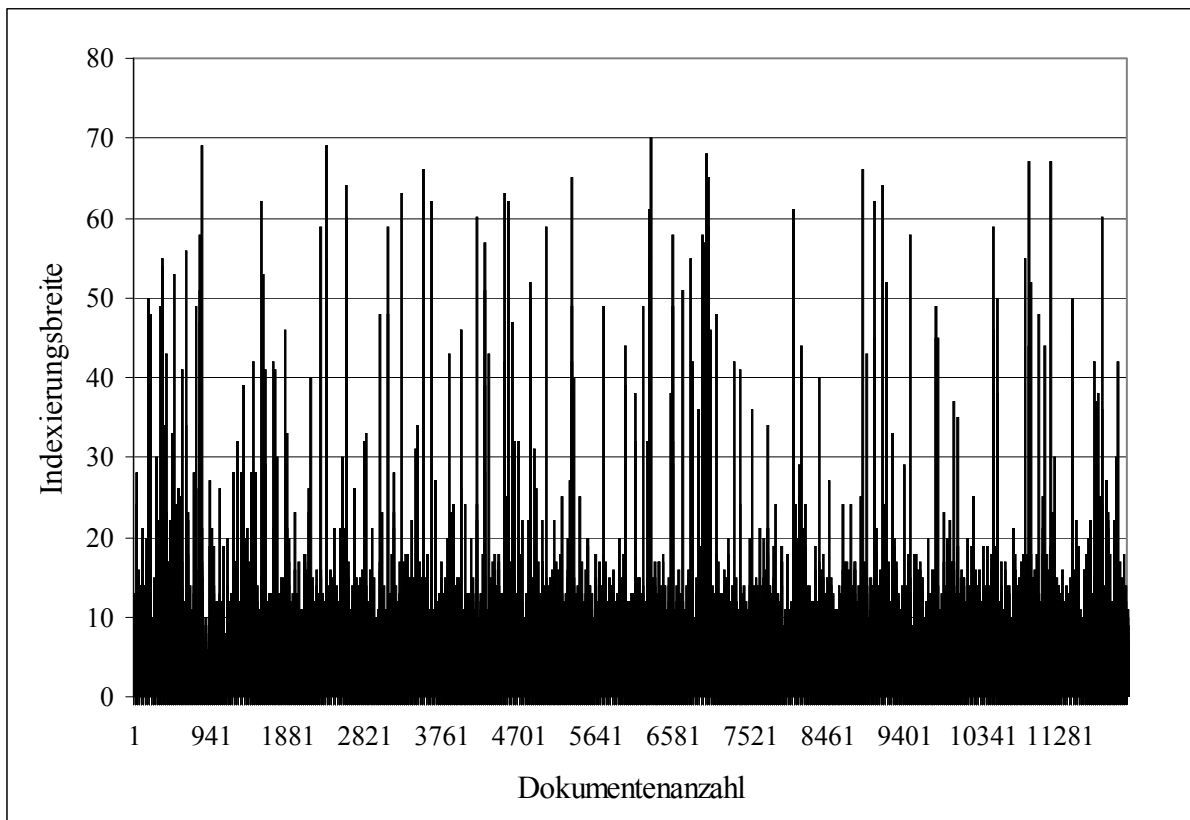
Abbildung 5: Indexierungsbreite: manuelles Verfahren (Trainingsset)



90 Eine Nichtberücksichtigung aller automatisch zugeordneten Schlagwörter ab einem bestimmten – absoluten oder prozentualen – cut-off-Level würde nach jetzigem Stand zahlreiche Sachschlagwörter wegfallen lassen.

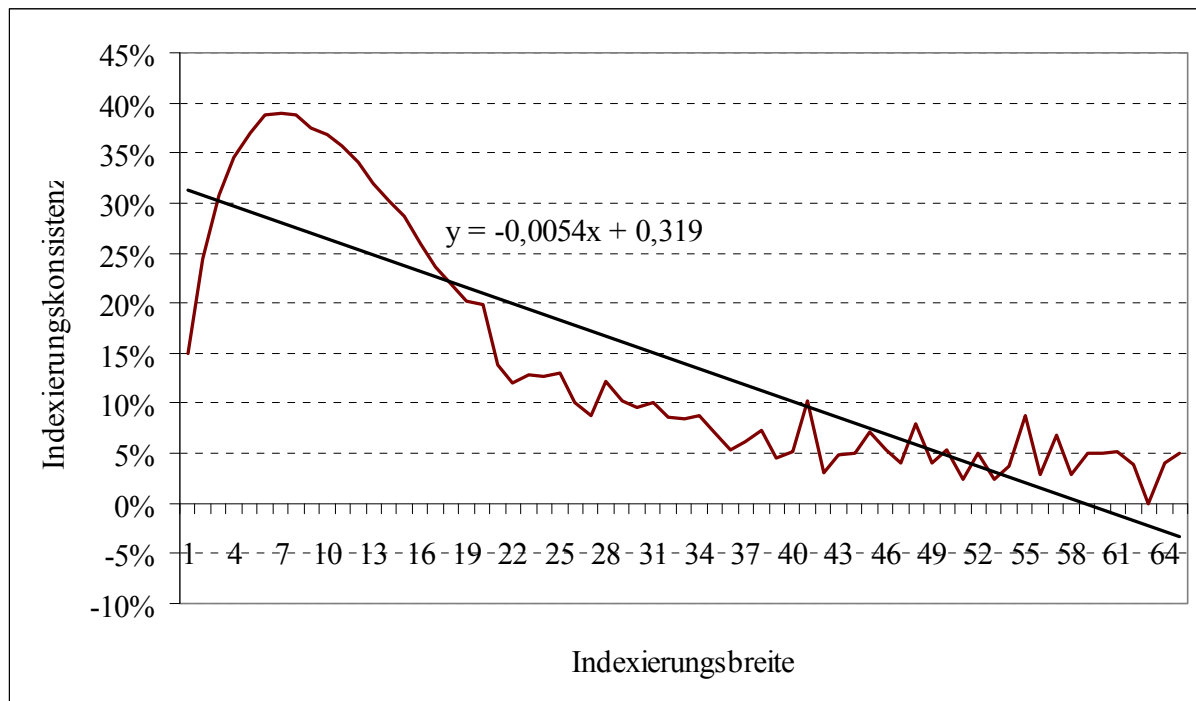
Ein Vergleich der automatisch generierten Indexierungsbreiten des Testsets mit denen des Trainingssets (Abbildung 4) ergibt keine wesentlichen Abweichungen, wie in Abbildung 6 dargestellt. Es finden sich zahlreiche Titel, die sehr hohe Indexierungsbreiten von 40, 50 oder 60 Deskriptoren aufweisen. Es gibt nur einen wesentlichen Unterschied. Während es im Trainingsset vorkam, dass Dokumenten keine Deskriptoren zugewiesen worden sind, wurde durch das automatische Verfahren im Falle des Testsets den zu indexierenden Dokumenten jeweils mindestens immer ein Schlagwort zugeordnet. In einem Falle waren es sogar 70 Schlagwörter. Ein Maximalwert, der gleichfalls im Trainingsset erreicht worden ist. Gleichwohl wurden im Mittel 7,4 Deskriptoren den einzelnen Titeln zugewiesen.

Abbildung 6: Indexierungsbreite: maschinelles Verfahren (Testset)



Die Anzahl der vergebenen Schlagwörter beeinflusst auch die Indexierungskonsistenz. Gemäß Lancasters Annahme: „If enough terms are assigned, a high consistency will eventually be reached” (2003: 71), müsste die Übereinstimmung zwischen beiden Verfahren mit Erhöhung der jeweiligen Indexierungsbreiten zunehmen. Abbildung 7 stellt den Zusammenhang zwischen der Anzahl der Schlagwörter, die die automatische Indexierung generiert, und deren Konsistenz mit den Ergebnissen der manuellen Sacherschließer dar und widerlegt deutlich Lancasters Annahme.

Abbildung 7: Korrelation zwischen Indexierungsbreite (automatisches Verfahren) und Indexierungskonsistenz⁹¹



Bis zu einer Indexierungsbreite von sieben bis neun Deskriptoren ist die Korrelation beider Werte positiv. Die Indexierungskonsistenz steigt von 15 % bei einem vergebenen Schlagwort bis auf ca. 38 %, wenn sieben bis neun Deskriptoren zugewiesen werden. Ab einer Indexierungsbreite über dem Wert 10 sinkt der Konsistenzwert kontinuierlich, um sich ab einer Breite von mehr als 35 Schlagwörtern um den Wert von 5 % einzupendeln.⁹²

Die Kausalität beider Maße ist für eine qualitative Bewertung der Indexierungsergebnisse von entscheidender Bedeutung. Es macht keinen Sinn, dies wurde bereits bei der Bewertung der Indexierungsbreite des automatischen Verfahrens deutlich, den Dokumenten „unendlich“ viele Schlagwörter, die zudem noch eine geringe Gewichtung haben, zuzuweisen, weil dies zu viel Retrieval-Ballast erzeugen würde. Zudem führt im Hinblick auf die Indexierungskonsistenz dieses Vorgehen nicht zu der angestrebten Angleichung beider

91 Die Regressionsgleichung $y = -0,0054x + 0,319$ zeigt an, dass mit jedem durch das automatische Verfahren zusätzlich vergebenen Deskriptor die Indexierungskonsistenz um den Wert 0,0054 abnimmt. Der Schnittpunkt der y-Geraden liegt bei 0,319 bzw. 31,9 %.

92 Dieses Ergebnis wurde bereits in einem Pretest, der im November 2009 an 23 Dokumenten durchgeführt worden ist, deutlich. Eine Herabsetzung des cut-off-Niveaus von 50 % auf 30 % führte beim automatischen Verfahren zu einer Zunahme an Indexierungsbreite von vier auf sechs Deskriptoren, verbunden mit einer Verringerung der Indexierungskonsistenz von 48 % auf 46 %. Dieser Effekt ist natürlich teilweise statistisch bedingt, weil die Anzahl der manuell vergebenen Deskriptoren generell niedriger ist als beim automatischen Verfahren.

Sacherschließungsverfahren, wenn nicht gleichzeitig den manuellen Indexaten mehr Deskriptoren zugewiesen werden als momentan.

4.2.2 Indexierungsspezifität

Der Vergleich beider zur Anwendung gelangten Verfahren im Rahmen der Auswertung des Trainingssets verdeutlicht, dass die Indexierungsspezifität bzw. -tiefe⁹³ im Minimalfall den Wert Null annehmen kann, jeweils für beide Indexierungsverfahren (vgl. Tabelle 1). Dies geschieht, wenn entweder durchgehend nur Oberbegriffe (mit der Hierarchiestufe Null) vergeben werden oder der Umfang der zu indexierenden Dokumente die Spezifität der auf den hierarchischen Stufen weit unten anzusiedelnden Schlagwörter (= Unterbegriffe) beeinflusst.⁹⁴ Im Maximalfall erreicht die Indexierungssoftware beim Trainingsset eine Indexierungsspezifität von 1,72, wohingegen das manuelle Verfahren nur einen Wert von 1,33 erreicht. Im Durchschnitt liegt die Indexierungstiefe bei 0,28 (Automatik) bzw. 0,23 (Mensch). Damit erreicht das automatische Indexierungsverfahren im Falle des Trainingssets maximal 29 % und im Mittel 22 % spezifischere Resultate, als dies dem menschlichen Indexierer gelingt.

Tabelle 1: Indexierungsspezifität Trainingsset

	automatisches Verfahren (A)	manuelles Verfahren (M)	Differenz A → M	bereinigt nach Indexierungskonsistenz (B)	Differenz A → B
minimal	0,00	0,00	0 %	0,00	0 %
maximal	1,72	1,33	+29 %	0,67	-50 %
Mittelwert	0,28	0,23	+22 %	0,14	-39 %

Diese Aussagen sind auf die unbereinigte Indexierungsspezifität bzw. -tiefe bezogen. Eine Bereinigung dieser Werte durch den Wegfall der automatisch falsch zugeordneten Schlagwörter und Berücksichtigung der verbleibenden übereinstimmenden Deskriptoren (= Indexierungskonsistenz) zeigt ein umgekehrtes Bild. Die Indexierungssoftware erreicht für das Trainingsset nur noch einen Spezifitätswert von maximal 0,67 und im Durchschnitt von 0,14. Es werden demnach durch das automatische Verfahren Indexierungsergebnisse erreicht,

93 Beide Maße können durch die den Berechnungen zugrunde gelegte Formel (vgl. Kapitel 3.3) im weiteren Verlauf dieser Arbeit als Synonym angesehen werden.

94 Die Indexierungsspezifität ist ein Quotient aus der bestimmten Hierarchieebene und der Seitenzahl des Dokumentes.

die hinsichtlich der Indexierungsspezifität hinter denjenigen der manuellen Indexierung deutlich zurückbleiben, im Mittel -39 % und maximal -50 %. Obwohl das automatische Sacherschließungsverfahren auf die Vergabe des bestpassenden Deskriptors trainiert worden ist⁹⁵ und ein Ausweichen auf die jeweiligen Oberbegriffe nicht stattfinden sollte, wählt die Software auf Basis der vorliegenden Ergebnisse eher Oberbegriffe bzw. Allgemeinwörter, die innerhalb des STW die Hierarchiestufe Null haben, wie z. B. Theorie, Risiko oder Messung.

Die Berechnung der Indexierungsspezifität bzw. -tiefe für das Testset führt zu folgenden Ergebnissen. Im Minimum liegt die Spezifität bei Null. Die vergebenen Schlagwörter sind demnach innerhalb der STW-Hierarchie sehr weit oben anzusiedeln. Der Maximalwert befindet sich bei 2,25. In diesen Fällen ist eine sehr tiefe bzw. spezifische Indexierung der zu indexierenden Titel erfolgt. Im Durchschnitt erschließt das automatische Verfahren die Dokumente des Testsets mit einer Indexierungsspezifität von 0,26. Dieser Wert liegt deutlich näher dem Minimalwert (0) als dem Maximum (2,25). Weiterhin entsprechen, bis auf den errechneten Höchstwert, die Ergebnisse zur Indexierungstiefe des automatischen Sacherschließungsverfahrens für das Testset grundlegend denjenigen des Trainingssets.

Die Indexierungsspezifität bzw. -tiefe fokussiert, wie im Kapitel 3.3 ausführlich dargestellt, auf die jeweiligen hierarchischen Ebenen, denen die Deskriptoren innerhalb des Thesaurusbaumes zugeordnet sind. Zwar liefert ein Vergleich beider Verfahren erste Anhaltspunkte für eine qualitative Bewertung der Indexierungsergebnisse. Dabei sollte folgender Aspekt aber nicht aus den Augen verloren werden. Die Indexierungstiefe stellt einen *relativen* Wert dar, der nicht umso besser ist, je höher die ermittelten Werte für die einzelnen Verfahren sind (und umgekehrt). Letztlich kann nur über eine Nutzerstudie bestimmt werden, welche Indexierungstiefe als optimal für eine Recherche angesehen wird, um dem Informationsbedürfnis des Suchenden zu entsprechen.⁹⁶ Insofern könnte die Indexierungstiefe, die durch die automatische Indexierung generiert wird, durch die Nutzersicht als ausreichend angesehen werden.

95 Entsprechend den ZBW-Sacherschließungsregeln (2007).

96 Vgl. Stock (2010).

4.3 Indexierungseffektivität

Durch die Bestimmung der Indexierungseffektivität⁹⁷ wird die Trennschärfe der vergebenen Schlagwörter für die Gesamtstichprobe ermittelt. Je häufiger ein Deskriptor innerhalb einer Gesamttitelmengende durch ein Indexierungsverfahren vergeben worden ist, desto geringer wird seine semantische Funktion für eine entsprechende Inhaltsstrukturierung (vgl. Stock 2008: 357).

Die von der automatischen Sacherschließung am häufigsten vergebenen Deskriptoren stellen Allgemeinwörter dar: Vergleich (18.006 Mal indexiert), Theorie (16.985), Risiko (6.292), Messung (4.420), Kritik (4.223) oder Bewertung (2.405).⁹⁸ Diesen Schlagwörtern kann durch ihr häufiges Auftreten im Datensatz nur noch eine minimale Informationsstrukturierungsfunktion zugeschrieben werden. Der Begriff „Vergleich“ wurde durch die Indexierungssoftware am häufigsten vergeben (18.006 Mal) und stellt mit 2,110 das Minimum an Indexierungseffektivität dar (vgl. Tabelle 2). Für das Retrieval führt die inflationäre Vergabe von Deskriptoren zu einem hohen Ballast. Verstärkt wird dieser Fakt durch die Tatsache, dass diese Schlagwörter oft an erster oder nachfolgender und somit an prominenter Stelle innerhalb der automatisch generierten Indexate auftreten. Indexierungsergebnisse, die an den ersten drei Stellen jeweils Vergleich, Risiko und Deutschland aufweisen, führen zu keiner sinnvollen Inhaltsstrukturierung, zumal die eigentlichen Sachdeskriptoren (z. B. Wirtschaftspolitik, Baukonjunktur) stellenweise einfach in der Masse von Allgemeinbegriffen untergehen.

Tabelle 2: Indexierungseffektivität der Gesamtstichprobe

automatisches Verfahren		manuelles Verfahren	
Deskriptor- verwendung	Inverse Dokument- häufigkeit	Deskriptor- verwendung	Inverse Dokument- häufigkeit
18.006 (max.)	2,110 (min.)	6.285 (max.)	3,629 (min.)
14 (min.)	12,439 (max.)	1 (min.)	16,247 (max.)
302 (mittel)	9,062 (mittel)	26 (mittel)	13,678 (mittel)

Eine weitere, oft vergebene Gruppe an automatisch generierten Schlagwörtern stellen die sogenannten Geodeskriptoren dar: USA (12.582), Deutschland (4.736), Welt (4.051), EU-Staaten (2.720), Großbritannien (2.002) (vgl. Tabelle 10). Das Auftreten dieser Deskriptoren

97 Über die Berechnung der Inversen Dokumenthäufigkeit (IDF).

98 Vgl. Tabelle 10 (Anhang), in der ein Auszug aus den Gesamtergebnissen abgebildet ist.

ist aufgrund der regionalen und sprachbezogenen Sammelschwerpunkte der ZBW wenig überraschend. Gleichwohl kann im Gegensatz zu den eben beschriebenen Allgemeinwörtern den Geodeskriptoren eine wichtige, ja zentrale Strukturierungsfunktion von Sachverhalten zugeschrieben werden. Eine Nichtbeachtung dieser aufgrund ihres häufigen Auftretens würde zu einem hohen Informationsverlust führen.

Die Kumulierung und Auswertung der Deskriptoren-Häufigkeiten liefert zudem Anhaltspunkte für eine Verbesserung des eingesetzten automatischen Indexierungsverfahrens. Unerklärlich oft anzutreffende Begriffe, wie beispielsweise Kohlendioxid (2.322 Mal vergeben), Pharmazeutisches Produkt (2.242) oder Computergestütztes Verfahren (953), die vom manuellen Sacherschließer demgegenüber wesentlich sparsamer indexiert worden sind (zum Vergleich: Kohlendioxid = 85, Pharmazeutisches Produkt = 48, Computergestütztes Verfahren = 39 Mal), zeigen wesentliche Ansatzpunkte für ein im weiteren Verlauf des Softwareeinsatzes dringend notwendiges Training mit entsprechender Regelspezifikation.

Am wenigsten wurden durch das automatische Sacherschließungsverfahren Schlagwörter wie Gesundheitspolitik (14), Angebot (14), Entscheidung bei Risiko (15), Kostenfunktion (18) oder Informationskosten (19) den vorliegenden Dokumenten zugeordnet. Diese Begriffe sind qua Definition dieses Evaluierungsmaßes somit am trennschärfsten. Die Schlagwörter Gesundheitspolitik und Angebot wurden 14 Mal vergeben und stellen damit das erreichte Indexierungseffektivitätsmaximum dar, welches den Wert 12,439 annimmt (vgl. Tabelle 2).

Die manuellen IndexiererInnen sind bei der Vergabe von Allgemeinwörtern wesentlich zurückhaltender. Beispielsweise wurde das Schlagwort „Kritik“ insgesamt nur 71 Mal vergeben, im Vergleich zu 4.223 Mal im Falle der eingesetzten Indexierungssoftware. Die am häufigsten manuell zugeordneten Schlagwörter sind Theorie (6.285), gefolgt von USA (3.338), Schätzung (2.999), Deutschland (2.084), EU-Staaten (1.567) und Welt (1.316) (vgl. Tabelle 10). Diese Schlagwörter weisen somit nur eine geringe Trennschärfe auf. Abgesehen von den Geodeskriptoren, deren zentrale Funktion für eine Inhaltsstrukturierung der vorliegenden Dokumente bereits betont worden ist, transformiert das Sachschlagwort „Theorie“ durch seine häufige Verwendung zu einem reinen Formalschlagwort. Das Minimum an Indexierungseffektivität liegt im Falle der intellektuellen Indexate – für den Begriff „Theorie“ – bei 3,629.

Dahingegen kann das Maximum bei 16,247 verortet werden (s. Tabelle 2). Während das Minimum nach oben hin offen ist – einerseits abhängig davon, wie oft ein Schlagwort indexiert worden ist, sowie andererseits von der Gesamttitelmenge –, steht das Maximum der

Trennschärfe generell fest, denn es befindet sich bei genau denjenigen Begriffen, die nur einmal vergeben worden sind (vgl. Stock 2008: 357). Dieses Maximum weist im Falle der manuellen Sacherschließung eine ganze Reihe von Deskriptoren – insgesamt ca. 1.400 – auf, exemplarisch können hier folgende Deskriptoren genannt werden: Baskisch, Containerhafen, Eisenbahnpersonenverkehr, Geldfälschung oder Hausmüll.

Im Vergleich beider Verfahren kann zusammenfassend eruiert werden, dass die manuelle Sacherschließung trennschärfer indexiert als das automatische Verfahren. Im Durchschnitt wird bei der intellektuellen Indexierung jeder Begriff 26 Mal vergeben, dies entspricht einer Indexierungseffektivität von 13,678. Im Gegensatz dazu indexiert das automatische Verfahren jeden Begriff 302 Mal, was einer Indexierungseffektivität von 9,062 entspricht (vgl. Tabelle 2). Deutliche Abweichungen können bei dem Effektivitätsminimum festgestellt werden, das im Falle der manuellen Indexierung bei 3,629 liegt und sich beim automatischen Verfahren bei 2,110 befindet. Weiterhin werden unterschiedliche Maximalwerte von 16,247 für die intellektuelle Sacherschließung und 12,439 im Falle des automatischen Verfahrens erreicht.

4.4 Belegungsbilanz

Eng im Zusammenhang mit der Indexierungseffektivität steht die Belegungsbilanz⁹⁹ des Thesaurus. Während Ersteres auf die *Trennschärfe* der einzelnen Deskriptoren in Bezug auf ein bestimmtes Dokumentenset abzielt, fokussiert Letzteres auf die *tatsächliche Nutzung* des STW-Vokabulars im Rahmen der manuellen oder automatischen Sacherschließung.

Mit dem STW stehen für die Sacherschließung wirtschaftswissenschaftlicher Dokumente insgesamt 5.770 Deskriptoren aus sieben Taxonomien¹⁰⁰ zur Verfügung. Von den manuellen SacherschließernInnen wurden innerhalb des Trainingssets 4.106 bzw. 71 % der verfügbaren Schlagwörter tatsächlich auch genutzt (vgl. Tabelle 3). Die Summe aller vergebenen Deskriptoren liegt bei 130.407.¹⁰¹ Keine Beachtung bei der intellektuellen Sacherschließung fanden hingegen 1.664 bzw. 29 % der zur Verfügung stehenden Schlagwörter. Der Mittelwert der Vergabehäufigkeit liegt bei 31,76, d. h., 21 % (absolut 862) der manuell benutzten Deskriptoren wurden 32 Mal oder öfter vergeben. Dieser Befund entspricht der Annahme, die dem Zipf'schen Gesetz¹⁰² zugrunde liegt. Demnach reicht letztlich nur ein Bruchteil an möglichen (Schlag-)Wörtern aus, um die Dokumente zu beschreiben, weil sie der Häufigkeitsverteilung innerhalb der natürlichen Sprache entsprechen und daher überproportional zur Anwendung gelangen.

Tabelle 3: Vergabehäufigkeit absolut und in %

		absolut	in %
Gesamtdeskriptorenzahl ¹⁰³		5.770	100 %
Manuelle Vergabe	Benutzte Deskriptoren	4.106	71 %
	Unbenutzte Deskriptoren	1.664	29 %
	insgesamt vergeben	130.407	-
	Mittelwert Vergabehäufigkeit	31,8	21 %
Automatische Vergabe	Benutzte Deskriptoren	1.002	17 %
	Unbenutzte Deskriptoren	4.768	83 %
	insgesamt vergeben	303.280	-
	Mittelwert Vergabehäufigkeit	302	20 %

99 Auch als Besetzungsbilanz bezeichnet.

100 BWL, VWL, Nachbarwissenschaften, Allgemeinwörter, Produkte, Wirtschaftssektoren und geographische Begriffe.

101 Dieser Wert stellt die auf die Gesamttitelzahl kumulierte Indexierungsbreite dar.

102 Siehe hierzu die Ausführungen von Nohr (2005f.), der diese Gesetzmäßigkeit in einer ausführlichen Darstellung abhandelt.

103 STW-Version 8.04.

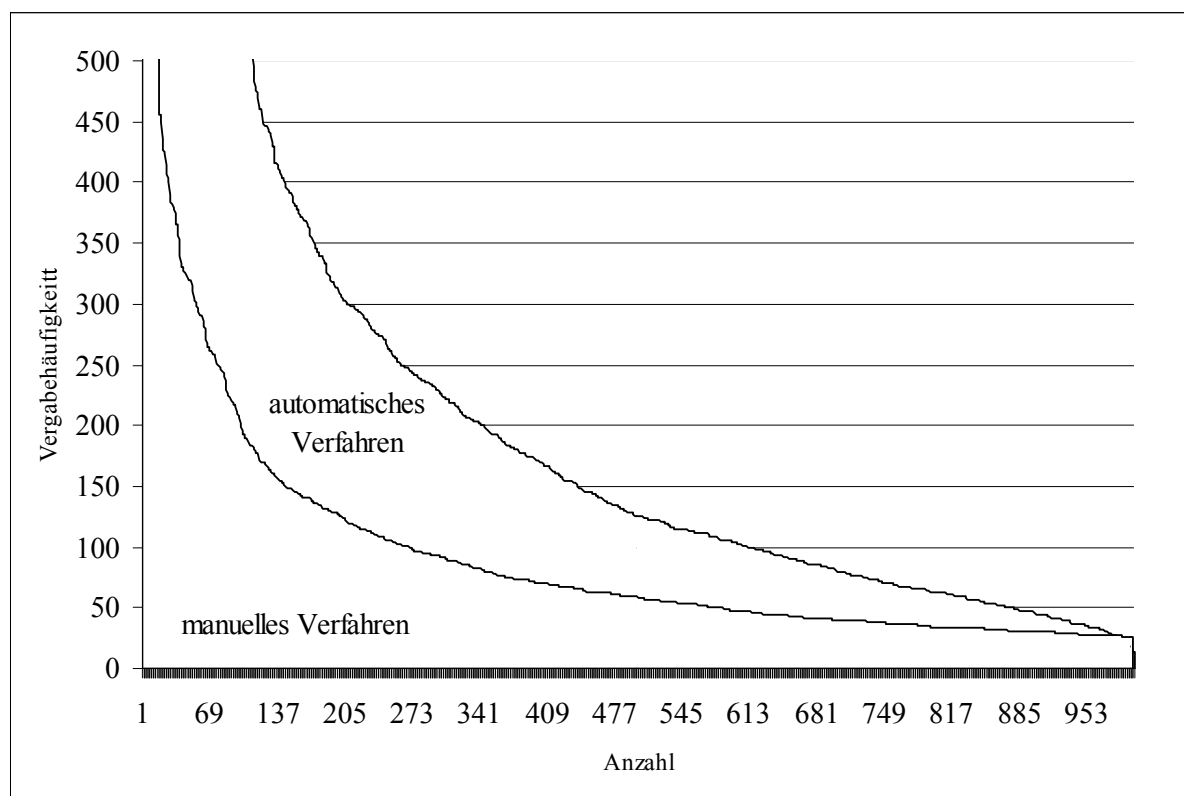
Die automatische Sacherschließung benutzte nur 1.002 bzw. 17 % der zur Indexierung bereitstehenden STW-Deskriptoren (vgl. Tabelle 3). Mit diesem Ausschnitt aus dem STW lag die auf die Gesamttitelzahl kumulierte Indexierungsbreite bei 303.280. Das automatische Verfahren hat demnach quantitativ mehr Deskriptoren vergeben als die manuellen Indexierer. Ein Befund, der den Resultaten zur Indexierungsbreite (s. Kap. 4.2.1) entspricht. Die überwiegende Masse aller möglichen Schlagwörter, nämlich 4.768 bzw. 83 %, wurde nicht zur Sacherschließung der vorliegenden Dokumente eingesetzt. Dies ist aller Wahrscheinlichkeit nach auf die fehlende Trainingsbasis zurückzuführen. Die breite Masse derjenigen Schlagwörter, die von den manuellen Sacherschließern tatsächlich auch zur Indexierung benutzt worden ist, taugt wegen ihres geringen Auftretens im Trainingsset nicht für ein Training der Indexierungssoftware.¹⁰⁴ Der Mittelwert der Vergabehäufigkeit liegt bei 302, d. h., 20 % (206 absolut) der manuell benutzten Deskriptoren wurden 302 Mal oder öfter vergeben.¹⁰⁵

Die Ergebnisse des automatischen Verfahrens entsprechen, wie für die Resultate der manuellen Indexierung bereits festgestellt, den Annahmen des Zipf'schen Gesetzes (vgl. Abbildung 8). Es ist für ein statistisches Indexierungsverfahren, wie es in der ZBW zum Einsatz kommt, typisch, dass im Dokument häufig vorkommende Begriffe als sinntragend für den Inhalt angesehen werden. Das automatische Verfahren orientiert sich daher stark an der jeweiligen Wortwahl der Autoren und versucht diese zu erfassen. Die Abbildung des Dokumenteninhaltes über deren Zuordnung zu den jeweiligen STW-Deskriptoren kann letztlich auch nur in einer entsprechenden Häufung des benutzten Vokabulars münden. Der damit einhergehende Verlust an Trennschärfe für das Retrieval kann nicht vermieden werden. Möglicherweise entspricht diese Art der Sacherschließung, die sich eng an der Dokumentenwortwahl orientiert, eher dem möglichen Suchverhalten der Nutzer.

104 Es werden rund 50 Dokumente pro Schlagwort benötigt, um dieses ausreichend zu trainieren.

105 Von den 1.002 von der Indexierungssoftware verwendeten Deskriptoren sind 875 (= 87 %) entsprechend den Vorgaben – 50 Dokumente pro Kategorie (Deskriptor) als Lerneinheit – ausreichend trainiert worden.

Abbildung 8: Vergabehäufigkeit: manuelles vs. automatisches Verfahren¹⁰⁶



Die Häufigkeitsverteilung im Falle des manuellen Sacherschließungsverfahrens entspricht, wenn auch in flacherer Form, derjenigen der automatischen Indexierung. Anscheinend schaffen es aber die manuellen Indexierer, dies deutet der entsprechende Kurvenverlauf >1.000 an, eine feingliedrigere Sacherschließung vorzunehmen, indem sie stärker von der eigentlichen Wortwahl der AutorInnen abstrahieren.¹⁰⁷ Die Fachbegriffe werden in ihrer Vielfalt umfassender genutzt (nämlich zu 71 %, s. Tabelle 3), was auch dazu führt, dass viele Schlagwörter nur sehr selten zur Anwendung kommen. Dies führt zu einer genaueren Inhaltsbeschreibung, weil die sinntragenden Sachverhalte besser abgebildet werden und damit den möglichen Suchraum für das Retrieval entsprechend zielführender einschränken. Weiterhin hat ein Thesaurus eine normierende Funktion, die, unabhängig von der Wortwahl der jeweiligen Verfasser, ein Wissensgebiet strukturieren soll. Insofern sind beide Kurven das Ergebnis unterschiedlicher Herangehensweisen der miteinander verglichenen Sacherschließungsverfahren.

¹⁰⁶ Siehe hierzu die Tabellen 11 und 12 im Anhang. Zur besseren graphischen Darstellung wurden die Werte oberhalb 500 (Vergabehäufigkeit) und 1.000 (Anzahl) abgeschnitten.

¹⁰⁷ Dieser Annahme Rechnung tragend, würden die automatisch erzeugten Indexierungsergebnisse eher als Zuführungsvokabular für eine entsprechende Notation (JEL, flacher STW) taugen denn für eine Abbildung innerhalb eines teilweise sehr granularen konstruierten polyhierarchischen Thesaurus, wie der STW einer ist.

4.5 Referentenauswertung

Nachdem die in den vorherigen Unterkapiteln dargestellten und diskutierten Evaluierungsmaße überwiegend quantitativen Charakter hatten, befasst sich die nachfolgende Referentenauswertung mit den automatisch generierten Indexierungsergebnissen unter qualitativen Gesichtspunkten. Dies geschieht zuerst für das Trainings- und anschließend für das Testset. Neben bereits bekannten Erklärungsfaktoren zur Indexierungsgüte, wie beispielsweise die Indexierungstiefe, die Trennschärfe der verwendeten Deskriptoren oder die Abbildung von Inhalten in einer bestimmten Thesaurushierarchie, treten in den nachfolgenden Abschnitten neue, aber bereits aus der Darstellung des Forschungsstandes (vgl. Kapitel 2) bekannte, das Indexierungsergebnis intervenierende Variablen auf. Die zentralen Erklärungsfaktoren der Indexierungsergebnisse für das Trainingsset sind in der Abbildung 9 aufgeführt.

Abbildung 9: Zentrale intervenierende Variablen: qualitative Bewertung (Trainingsset)

- goal-/sub-goal-Problem (Haupt- vs. Randthemen)
- sinntragende Dokumentenelemente (Titel, Abstract, Keywords)
- Thesaurushierarchie (Ober-, Unter-, verwandte Begriffe)
- Indexierungstiefe (Indexierungsbreite, -spezifität)
- Indexierungseffektivität (Schlagworttrennschärfe)
- Falsche Deskriptoren

Der erste Erklärungsfaktor für die Bewertung der automatisch generierten Indexierungsergebnisse ist das „goal-/sub-goal-Problem“. Dieses Problem wird durch das Spannungsfeld Haupt- vs. Randthemen determiniert. Auf der einen Seite steht das relevante Hauptthema, das auf jeden Fall via Deskriptor(en) abgebildet werden muss, und auf der anderen Seite sind die Randthemen zu finden, die eher nicht beschlagwortet werden, um den Ballast für das Retrieval zu minimieren.¹⁰⁸ Wenn davon ausgegangen werden kann, dass die Trennung in Haupt- und Nebenaspekte bereits kritisch ist, ist die Einteilung in wichtige und nicht relevante Randthemen noch schwieriger. Die gesammelten Referentenaussagen zu den automatisch

108 Laut den Sacherschließungsregeln der ZBW (2007) werden Nebenaspekte dann indexiert, „wenn das Dokument für eine Recherche nach diesem Aspekt relevant ist.“ (Ebd.: 5).

generierten Indexaten des Trainingssets verdeutlichen diese typische Gradwanderung¹⁰⁹: „[...] zentraler Deskriptor [...] zu geringe Gewichtung.“; „[...] außerdem vergibt Recommind die USA, was nicht falsch ist, aber eigentlich überflüssig (kommt nur am Rande vor).“, oder „[...] einige wichtige Deskriptoren zu geringe Gewichtung“. Hierbei wird nochmals deutlich, wie die Indexierungssoftware arbeitet. Das automatische Indexierungsverfahren versucht, über die Bestimmung der Häufigkeit einzelner Wörter oder Wortgruppen auf deren Relevanz zur Beschreibung des Dokumenteninhaltes zu schließen. Dieser Rückschluss wird anschließend über eine Gewichtung der vorgeschlagenen Deskriptoren abgebildet.¹¹⁰

Die Gewichtung sinntragender Dokumententeile, beispielsweise des Titels, des Abstracts oder der Keywords, stellt den zweiten Erklärungsfaktor der Indexierungsqualität dar. Dieser Gewichtung liegt die Annahme zugrunde, dass die genannten Dokumentenabschnitte, ähnlich den ersten Worten innerhalb einer Häufigkeitsverteilung, komprimiert den zentralen Inhalt eines Dokumentes wiedergeben. Die eingesetzte Indexierungssoftware ermöglicht daher eine entsprechende Gewichtung, beispielsweise des Titels, um den dort aufgeführten Worten und Begriffen eine höhere Relevanz zuzuweisen.¹¹¹ Dass dies nicht immer sauber gelingt, zeigen folgende Aussagen: „[...] wichtige Sachbegriffe fehlen, und nicht einmal das Land wurde vergeben, obwohl es sogar im Titel steht“ bzw. „bei [...] fehlt der Deskriptor Berufsausbildung – ein großer Fehler, da im Abstract 2 Mal erwähnt“.

Die dritte intervenierende Variable stellt die Thesaurusstruktur dar. Dem automatischen Verfahren gelingt es nicht, die Granularität des STW immer zufriedenstellend abzubilden, obwohl das durchgeführte Training desselben auf die bestpassenden Begriffe, entsprechend den Vorgaben des manuellen Indexierens (s. ZBW 2007: 6), stattgefunden hat. Es finden sich Aussagen, wie „[...] statt Sozialreform hat [...] nur Reform gegeben“; „Sparkasse wurde [...] nicht vergeben, obwohl im Text explizit savings bank mehrfach vorkommt – dafür wurde Bank genommen, was auch nicht direkt falsch ist“; und „es fehlt zwar Kalifornien [...], stattdessen wurde [...] USA vergeben“. Offensichtlich besteht trotz der getroffenen Einstellungen und Annahmen eine Tendenz des automatischen Verfahrens, auf die jeweiligen

109 Vgl. Abbildung 10 (Anhang). Die schriftlich fixierten Aussagen der das Trainingsset auszuwertenden FachreferentInnen wurden anhand der in Abbildung 8 dargestellten Kategorien eingeordnet und werden hier nur exemplarisch aufgeführt.

110 Wobei natürlich Schlagwörter, die zentrale Aspekte wiedergeben, aber noch nicht ausreichend trainiert worden sind, geringere Gewichtung erhalten als beispielsweise Deskriptoren, die nur Randbereiche abdecken, dafür aber ausreichend trainiert worden sind.

111 Die Gewichtung anderer sinntragender Dokumentenabschnitte, wie der Abstracts oder Keywords, setzt natürlich deren Existenz voraus. Ein kurzer Blick in die Daten zeigt, dass längst nicht alle Dokumente diese vorweisen können.

Oberbegriffe auszuweichen. Dadurch werden zwar richtige Ergebnisse erzeugt, diese sind aber dann zu unspezifisch, belasten das Retrieval, wie beispielsweise der Deskriptor „Bank“, und stellen keine qualitativ hochwertigen Resultate dar. Weiterhin bilden die *verwandten Begriffe* ein Problemfeld, wie folgende Aussage eines Referenten unterstreicht: „es wurden andere Begriffe gewählt, die auch tolerabel sind“. Hierbei spielt auch die weiter oben beschriebene goal/sub-goal-Problematik mit herein. Werden die zentralen Textaussagen nicht möglichst genau getroffen, sondern nur die Ober- oder verwandten Begriffen vergeben, verschwimmt zugleich die durch eine Sacherschließung eigentlich gewünschte inhaltsbezogene Trennung und Strukturierung von Haupt- und Randthemen.

Die Indexierungstiefe, die sich aus den Maßen der Indexierungsbreite und -spezifität zusammensetzt, hängt als vierter Erklärungsfaktor eng mit der gerade diskutierten Thesaurushierarchie zusammen. Die zahlreichen Kommentare zu den automatisch generierten Indexaten (s. Abbildung 11) verdeutlichen die Schwierigkeiten des automatischen Verfahrens hinsichtlich der Anzahl und der Spezifität der vergebenen Deskriptoren. Oft sind die Indexate unvollständig, wie die nachfolgenden Aussagen belegen: „es fehlen die wichtigen Deskriptoren: Wissenstransfer, Innovation und Öffentlichkeitsarbeit“; „folgende Deskriptoren fehlen: Deutschland, Migranten, Sprachpolitik“ oder „ganz schlechtes Ergebnis: der wichtigste Deskriptor fehlt, dafür eine Fülle von überflüssigen Deskriptoren, deren Relevanz sich nicht erschließt“. Stellenweise trug die Indexierungsbreite nicht zur Treffsicherheit der automatischen Sacherschließung bei: „allerdings etliche unnötige Deskriptoren“; „[...] dafür eine Fülle von überflüssigen Deskriptoren, deren Relevanz sich nicht erschließt“ bzw. „ganz gut, allerdings wieder viele unnötige Deskriptoren“. Weiterhin kann eine mangelhafte Indexierungsspezifität der ermittelten Schlagwörter festgestellt werden: „Welt statt EU-Staaten ist zu unspezifisch“ oder „ordentlich, etwas unspezifisch (New Keynesian P... fehlt), aber zu gebrauchen“.

Als fünfte Variable beeinflusste die Indexierungseffektivität die Qualität der automatisch generierten Indexate. Es ist festzustellen, dass einige Deskriptoren durch die Indexierungssoftware inflationär vergeben werden und somit ihre Trennschärfe für eine Inhaltsstrukturierung fast vollkommen einbüßen: „in mehreren Fällen wurde Deskriptor ‚Vergleich‘ vergeben [...], in einigen Fällen war das auch richtig, in anderen aber überflüssig“; oder „einige

Allgemeinwörter wie Vergleich, Kritik, Risiko kamen ziemlich oft vor – nicht immer zu Recht“.¹¹²

Die Vergabe offensichtlich falscher Deskriptoren stellt den sechsten und letzten Erklärungsfaktor für die Bewertung der Indexierungsergebnisse des Trainingssets dar. Es ist einerseits eine häufige Zuweisung von falschen Deskriptoren festzustellen: „es gab auch Ausreißer: so tauchte in einem Text über Zeitbudgetforschung nicht nachvollziehbar Tansania und Elektrizität auf oder in einem Banken-Kontext der Deskriptor Kohlendioxid“; bzw. „ganz falsch und unverständlich ist, wie ‚Kohlendioxid‘ [...] hineingeraten ist“. Andererseits häuft sich die falsche Vergabe für einige Schlagwörter: „mehrmals ist absolut falsch Deskriptor ‚Pharmazeutisches Produkt‘ vergeben, auch ‚Kohlendioxid‘ mehrmals falsch vergeben“; hier finden sich grobe Ausreißer: „es ist vollkommen unklar, wie [...] auf Tansania und Elektrizität kommt; diese Begriffe kommen im Text überhaupt nicht vor“. Insofern nährt sich die Indexierungssoftware auch in diesem Punkt den manuellen Indexierern an, die gleichfalls falsche Deskriptoren einem Dokument zuordnen können.

Für die qualitative Evaluierung der Indexierungsgüte des Testsets können neben den bereits bekannten sechs Erklärungsfaktoren (vgl. Abbildung 9) zwei neue intervenierende Variablen ausgemacht werden¹¹³: die domänenspezifische Abbildung der jeweiligen Teildisziplinen (BWL, VWL, Ökonometrie) sowie die „Indexierungspolitik“ (siehe Abbildung 10).

Abbildung 10: Zentrale intervenierende Variablen: qualitative Bewertung (Testset)

- goal-/sub-goal-Problem (Haupt- vs. Randthemen)
- sinntragende Dokumentenelemente (Titel, Abstract, Keywords)
- Thesaurushierarchie (Ober-, Unter-, verwandte Begriffe)
- Indexierungstiefe (Indexierungsbreite, -spezifität)
- Indexierungseffektivität (Schlagworttrennschärfe)
- Falsche Deskriptoren
- Fachgebiet (BWL, VWL, Nachbarwissenschaften)
- „Indexierungspolitik“

112 Bei der Durchsicht der einzelnen Titel lassen sich viele Indexate finden, die mit folgenden Begriffen indexiert worden sind: Theorie UND Deutschland bzw. Simulation UND Deutschland.

113 Die gemäß der Referentenauswertung in einzelne Kategorien zusammengefassten Aussagen zu den Indexaten des Testsets sind in der Abbildung 11 (im Anhang) aufgeführt.

Während für die ersten sechs intervenierenden Variablen sich die Problemlagen im Abgleich zwischen Trainings- und Testset gleichen,¹¹⁴ zeigen die beiden neuen Erklärungsfaktoren deutlich die Unterschiede zwischen dem manuellen und dem momentan an der ZBW eingesetzten automatischen Sacherschließungsverfahren auf.

Es ist festzustellen, dass das zu Anwendung kommende automatische Verfahren es oft nicht schafft, die beiden zentralen wirtschaftswissenschaftlichen Fachdisziplinen durch Vergabe der jeweiligen BWL- bzw. VWL-Deskriptoren sauber voneinander zu trennen, wie folgende Aussage belegt: „zu unspezifisch, betriebswirtschaftliche Literatur zu Auslandsinvestitionen wird mit VWL-Begriffen wiedergegeben“ oder „Ansatz gut, es fehlt der (wichtige) kulturelle Aspekt bei Vertragsverhandlungen mit chinesischen Geschäftspartnern (eher BWL, hier VWL-Begriffe)“. Häufig werden zudem VWL- und BWL-Schlagwörter zusammengewürfelt: „ziemlich verquere Mischung aus BWL- und VWL-Deskriptoren: Ehe; Gerechtigkeit; Verteilungsgerechtigkeit; Allokation; Public Choice; Portfolio-Management“. Weiterhin werden theoriebezogene Sachverhalte in Dokumenten, die einen nicht unwesentlichen theoretischen Schwerpunkt haben, seitens der Automatik häufig nicht wiedergegeben, wie folgende Aussagen belegen: „befriedigend, theoretischer Charakter dieser Arbeit nicht erkannt; der Rest geht in die richtige Richtung“ oder „zu viele Begriffe; theoretische Stoßrichtung des Themas nicht erkannt“.

Die Abbildung der durch eine Häufigkeitsverteilung ermittelten zentralen Begriffe eines Textes mit Hilfe eines kontrollierten Vokabulars ist zentrales Merkmal des in der ZBW eingesetzten begriffsorientierten automatischen Verfahrens. Die Unterschiede zu den manuellen Indexierungsergebnissen wurden in den bisherigen Ausführungen sehr deutlich herausgestellt. Doch gerade der Erklärungsfaktor „Indexierungspolitik“ stellt die Unterschiede zwischen einer manuellen und einer automatischen Sacherschließung nochmals pointiert heraus. Das automatische Verfahren arbeitet verständlicherweise anders, als dies ein Mensch tun würde, und deshalb überraschen folgende Aussagen der Referentenauswertung auch nicht: „anders indexiert, als ein manueller Indexierer dies tun würde, aber doch ok“; „sehr gut, einzig die Auflistung von Vergleich, Theorie und Welt an erster Stelle ist ungünstig“ oder „geht in die richtige Richtung, es werden zwar andere Begriffe verwendet, als sie ein manueller Indexierer vergeben würde, ansonsten in Ordnung“.

114 Auf eine erneute Diskussion wird hierbei verzichtet, um stärker auf die zwei zusätzlichen Variablen einzugehen.

5 Fazit

Die Implementierung eines automatischen Indexierungsverfahrens in die Sacherschließung von digitalen Dokumenten am Beispiel der ZBW war Gegenstand der vorliegenden Fallstudie. Im Zentrum der vorhergehenden Ausführungen stand dabei die qualitative Bewertung bzw. Evaluierung der automatisch erzeugten Sacherschließungsergebnisse im Vergleich zu den manuell generierten Indexaten. Hierbei wurde zur automatischen Indexierung von digitalen Dokumenten mit dem MindServer der Firma Reconnind ein statistischer Ansatz eingesetzt, welcher basierend auf einem begriffsorientierten Verfahren den Inhalt der vorliegenden Dokumente mit Hilfe von entsprechenden STW-Schlagwörtern abgebildet hat. Dadurch wurden den einzelnen Titeln Metadaten in Form von Deskriptoren zugewiesen (= Additionsverfahren), die die Dokumente nicht nur sprachoberflächlich beschreiben (= Extraktionsverfahren), sondern durch die Einordnung in ein ontologiebasiertes Begriffssystem auf die Bedeutungsebene von Informationen abzielen. Dadurch wird ein informatorischer Mehrwert erzeugt, der über eine reine Kataloganreicherung hinausgeht. Auf Basis des Forschungsstandes und des zugrunde gelegten automatischen Sacherschließungsverfahrens können die für diese Arbeit leitende Fragestellung und die aufgestellte Arbeitshypothese folgendermaßen beantwortet werden.

Die in der Einleitung aufgestellte Hypothese kann auf Grundlage der herausgearbeiteten Befunde verifiziert werden. Es wurde im Ergebniskapitel deutlich herausgearbeitet, dass die automatische Indexierung digitaler Dokumente sich unter dem Aspekt der Indexierungsqualität anhand der verschiedenen intervenierenden Variablen den manuellen Indexaten in unterschiedlicher Ausprägung annähert. Dieses Ergebnis vermag einerseits nicht zu überraschen, wurde doch mit dem begriffsorientierten Verfahren gerade ein automatischer Indexierungsansatz getestet, der die manuelle und die maschinelle Sacherschließung zusammenführen soll. Andererseits betonen die Resultate gleichzeitig je nach durchdekliniertem Erklärungsfaktor vor allem auch die Unterschiede beider Sacherschließungsverfahren. Eine Annäherung beider Ansätze ist zu konstatieren. Eine mit der manuellen Indexierung völlige Entsprechung der automatisch generierten Indexierungsergebnisse wird mit der eingesetzten Indexierungssoftware momentan nicht erreicht¹¹⁵. Da beide Indexierungsverfah-

115 Die eingesetzte Indexierungssoftware steht gerade erst am Anfang ihres Einsatzes. Somit können noch keine Aussagen zu deren Trainingspotential getroffen werden.

ren aber grundsätzlich von unterschiedlichen Prämissen ausgehen, wird diese völlige Übereinstimmung auch nicht zu erreichen sein. Das Ziel des Einsatzes eines automatischen Indexierungsverfahrens muss sein, einen möglichst hohen Annäherungsgrad zu erreichen, um über den gesamten, heterogenen Bestand hinweg ein gewisses Maß an Homogenität in der bibliothekarischen Sacherschließung zu gewährleisten.

Basierend auf den vorliegenden Daten zu den jeweiligen Erklärungsfaktoren, ist eine Reihe von aussagekräftigen Analyseergebnissen zu verzeichnen, die fokussiert folgendermaßen zusammengefasst werden können.

In Bezug auf die Indexierungskonsistenz ist die erste zentrale Erkenntnis der vorliegenden Arbeit, dass die an der ZBW eingesetzte Indexierungssoftware es schafft, durchschnittlich 36 % aller zu erschließenden Dokumente richtig, im Sinne der manuellen Vorgabe, zu indexieren. Dieses Ergebnis entspricht der Indexierungskonsistenz, die Gerards et al. (2006) mit 25/40 % oder Medelyan (2005) mit 27/38 % angeben. Der durchschnittliche Konsistenzwert dieser Arbeit ist ferner deutlich besser als derjenige Wert, den Nübel/Schmidt (2003) mit 15-20 % errechnen, aber zugleich deutlich schlechter als der Wert von Haller et al. (2001), die 49-52 % angeben.

Die zweite zentrale Aussage dieser Arbeit bezieht sich auf die Befunde zur Indexierungsbreite. Die Resultate zu dieser Variable verdeutlichen, dass das automatische Sacherschließungsverfahren im Gegensatz zu den manuellen Indexierern, die im Mittel fünf Schlagwörter vergeben, den Vorlagen generell eine höhere Anzahl an Deskriptoren zuweist, nämlich durchschnittlich acht (Trainings- und Testset). Zudem ist zwischen den einzelnen Indexaten eine hohe Varianz – zum Teil werden bis 71 Deskriptoren vergeben – feststellbar. Weiterhin wurde der zentrale Einfluss unterschiedlicher Indexierungsbreiten auf die Indexierungskonsistenz untersucht und herausgearbeitet, dass bis zu einer Anzahl von bis zu neun Deskriptoren die Korrelation positiv ist, während der errechnete Zusammenhang ab zehn Schlagwörtern stetig negativ ist. Diese Aussage belegen z. B., wenn auch nicht durch eine entsprechende Korrelation, die Arbeiten von Bertrand und Cellier (1995), Gerards et al. (2006), Haller et al. (2001), Tonta (1991), Iivonen (1990) und Watermann (1997). Es macht, das wurde hier zudem deutlich gezeigt, für die Sacherschließung von digitalen Dokumenten keinen Sinn, so wie einige Studien direkt bzw. indirekt vorschlagen (u. a. Stubbs et al. 1999), den Titeln möglichst viele Schlagwörter zuzuweisen, um die Konsistenz zwischen den manuell und den automatisch erzeugten Indexaten zu erhöhen.

Weiterhin beeinflusst die Indexierungstiefe, dies haben die Resultate veranschaulicht, die Konsistenz und damit auch die Qualität der automatisch erzeugten Indexate. Zu diesem Ergebnis gelangen ebenfalls die Untersuchungen von Saarti (2002), Lancaster (2003), Xu (2007, 2008) sowie Reich und Bieber (1991). Ferner hat diese Masterarbeit aufgezeigt, dass eine maschinelle Indexierung mit Hilfe eines kontrollierten, polyhierarchischen Vokabulars vom Ansatz her möglich ist. Es ist aber offenkundig, dass mit diesem Verfahren in der Tendenz eher Ergebnisse mit flachen Begriffshierarchien erzielt werden. Den manuellen Indexieren gelingt demgegenüber eine tiefere und damit auch für eine Inhaltsstrukturierung qualitativ bessere Sacherschließung von digitalen Dokumenten.

Diese Annahmen wurden zudem mit den dargelegten Befunden zu den Variablen Indexierungseffektivität und Belegungsbilanz untermauert. In Bezug auf die Indexierungseffektivität konnte festgestellt werden, dass die intellektuelle Indexierung die Dokumente trennschärfer erschließt¹¹⁶ als das automatische Verfahren. Gleiches wurde bereits von Gerards et al. (2006) hervorgehoben und ist in ähnlicher Art und Weise auch von Nübel/Schmidt (2003) bestätigt worden. Zudem unterstreicht die Belegungsbilanz, dass die FachreferentInnen das vorhandene STW-Vokabular nicht nur feingliederiger, sondern zudem tatsächlich umfassender nutzen, als dies das maschinelle Verfahren tut.

In der Referentenauswertung für das Trainings- sowie das Testset wurden weitere qualitätsspezifische, intervenierende Variablen aufgezeigt. Zunächst beeinflusst das goal-/subgoal-Problem die automatisch erzeugten Indexierungsergebnisse, und zwar dahingehend, dass nicht nur die Hauptthemen unterschiedlich erschlossen bzw. beschrieben werden, sondern dass diese Varianz bei den Randaspekten noch stärker ausgeprägt ist. Dies entspricht den Ergebnissen, die Leininger (2000), Iivonen (1990), Saarti (2002), Tonta (1991) sowie Olson und Wolfram (2005, 2007) in ihren Arbeiten herausstellen. Der in dieser Arbeit aufgezeigte qualitätsspezifische Einfluss sinntragender Dokumententeile (Titel, Abstract, Keywords) auf die Indexierungsgüte findet sich auch in der Studie von Saarti (2002) und etwas unspezifischer gleichfalls in der Arbeit von Lancaster (2003). Weiterhin konnte der Thesaurusstruktur eine qualitätsspezifische Bedeutung zugemessen werden, denn gerade die Polyhierarchie und Granularität des STW wird durch die Indexierungssoftware oft nicht richtig und umfassend erfasst. Dies deckt sich mit den Resultaten in Xu (2007, 2008), Haller et al. (2001) und Iivonen (1990). Darüber hinaus zeigen die Resultate, dass es das eingesetzte automatische Indexierungsverfahren oft nicht schafft, eine domänenspezifische Trennung in VWL und BWL vorzunehmen. Diesen fachspezifischen Einfluss auf die Indexierungsergebnisse hat

116 Vor allem durch den sparsamen Einsatz von Allgemeinwörtern.

bereits Xu (2007, 2008) konstatiert. Weiterhin verdeutlicht gerade der letzte Erklärungsfaktor innerhalb der Referentenauswertung, die „Indexierungspolitik“, die zentralen Unterschiede beider Verfahren. Die Indexierungssoftware indexiert, obwohl es sich um ein begriffsorientiertes Verfahren handelt, anders, als dies ein Mensch tun würde. Dieser Aspekt wird in den Arbeiten zur manuell-automatischen Indexierungskonsistenz immer wieder prominent herausgestellt (u. a. Gerards et al. 2006, Nüber/Schmidt 2003, Haller et al. 2001).

Basierend auf diesen Befunden, kann die aufgestellte zentrale Fragestellung dieser Arbeit folgendermaßen beantwortet werden. Mit dem Einsatz eines automatischen Indexierungsverfahrens können alle digitalen Dokumente innerhalb des ZBW-Bestandes grundsätzlich indexiert werden. Deren Qualität bemisst sich anhand der dargestellten Erklärungsfaktoren in ihren jeweiligen Ausprägungen. Der konkrete Nutzen der automatisch erzeugten Indexate für die Recherche hängt aber wiederum auch vom Nutzerinteresse und deren Suchverhalten ab. Hier bieten sich Anhaltspunkte für weitere Forschungsmöglichkeiten, zum Beispiel in Form von entsprechenden Nutzerstudien.

Letztlich verbleibt, um auf den zweiten Teil der Fragestellung zu schauen, nur der Versuch einer Tendenzaussage. Eine konsistente Informationserschließung bzw. -strukturierung digitaler Dokumente ist zwar einerseits ansatzweise realisiert worden: Die Indexierungskonsistenz zwischen manueller und automatischer Sacherschließung weist 36 % auf.¹¹⁷ Andererseits ist mit einem weiteren Training der eingesetzten Software zu erwarten, dass sich die Indexate bzw. das Indexierungsverhalten beider Ansätze weiter angleichen werden, ohne jedoch mittel- bis langfristig völlig deckungsgleich zu sein. Letztlich simuliert bzw. reproduziert die Indexierungssoftware durch deren Einsatz als begriffsorientiertes Verfahren das Indexierungsverhalten von momentan 25 hierfür eingesetzten IndexierInnen. Es liegt in der Natur der Sache, dass mit zunehmender Anzahl an Sacherschließern, die einen Thesaurus anwenden, die Ergebnisse, gerade im Spannungsfeld Haupt- vs. Randthemen (goal-/sub-goal-Problem), immer größere Inkonsistenzen festzustellen sind, wie schon Olson und Wolfram (2005, 2007) in ihren Untersuchungen herausgefunden haben.

Mit diesem Befund kann gleichfalls eine jener Unterfragen beantwortet werden, auf deren Fragestellung im Ergebnisteil noch nicht ausreichend eingegangen worden ist. Die in der ZBW eingesetzte automatische Sacherschließung in Form eines begriffsorientierten Verfahrens schafft es, den scheinbar unüberbrückbaren Widerspruch zwischen den qualitativen Ansprüchen des intellektuellen und des automatischen Sacherschließungsverfahrens

117 Die Automatik ist quasi ein Laie bzw. ein Berufsanfänger, der zunächst eingearbeitet werden muss (vgl. Bertrand/Cellier 1995: 471).

rens (vgl. Nohr 2004: 216) ein Stück weit aufzulösen, und zwar dahingehend, dass dieses Vorgehen eine Möglichkeit darstellt, beide Verfahren aneinander anzunähern. Diese Annahme spiegeln einerseits die aufgeführten Ergebnisse wider, andererseits deckt sie sich mit den Aussagen, die zur Arbeitshypothese getroffen worden sind.

Bevor weitere mögliche Forschungsfelder aufgezeigt werden, gilt es, die Grenzen und Schwächen der vorliegenden Arbeit kritisch zu betrachten.

Die durchgeführten Analysen anhand der entsprechenden Gütekriterien Indexierungstiefe (Indexierungsbreite und -spezifität), Indexierungseffektivität, Indexierungskonsistenz und der klassifizierten Aussagen der Referentenauswertung – u. a. goal-/sub-goal-Problem, Domänenspezifität, Indexierungspolitik – stellen den Versuch dar, die Qualität automatischer Sacherschließung nicht über einen Retrievaltest zu bewerten, sondern anhand des von Stock (2008) bzw. Lancaster (2003) vorgeschlagenen Kriteriensets. Dabei stellten die rund 39.000 automatisch verschlagworteten Dokumente aus dem ZBW-Bestand ein sehr umfangreiches Datenset dar, welches es zu analysieren galt. Fallstudien mit dieser Dokumentenanzahl wurden hinsichtlich einer Bewertung der automatisch generierten Indexierungsergebnisse bisher generell nur mit Retrievaltests untersucht. Insofern hat diese Arbeit gezeigt, dass auch andere Ansätze, wie die von Stock (2008) oder Lancaster (2003), geeignet sind, umfangreiche Dokumentensets einer qualitativen Bewertung zu unterziehen. Gleichzeitig wurden die methodischen und inhaltlichen Grenzen des eingesetzten Variablensets deutlich. Beispielsweise haben die Annahmen zur Indexierungsspezifität offenbart, dass es aufgrund fehlender oder im falschen Format vorliegender Daten oft nicht möglich ist, das Gesamtdatenset umfassend zu analysieren.¹¹⁸ Weiterhin ist eine qualitative Inaugenscheinnahme aller automatisch generierten Indexate im Rahmen einer Referentenauswertung zwar grundsätzlich wünschenswert, aber aus Zeit- und Kostengründen nicht realisierbar.

Weitergehende Forschungsmöglichkeiten ergeben sich nicht nur ausgehend von den geschilderten methodischen und datenspezifischen Grenzen der durchgeführten Analysen, sondern auch in weiteren Bereichen. Bei den bisherigen Untersuchungen wurde die unterschiedliche Gewichtung der automatisch vergebenen Deskriptoren nicht berücksichtigt. Unterschiedliche cut-off-Level, absolut oder relativ, haben gemeinhin starken Einfluss auf die einzelnen Variablen und können die Ergebnisse entsprechend beeinflussen. Weiterhin wurde durch das an der ZBW noch im Anfangsstadium befindliche automatische Indexierungsver-

118 In diesem Falle lagen nicht für alle Dokumente die notwendigen Angaben zur Seitenzahl vor oder es war nur der Datenumfang in Kilobyte angegeben.

fahren das Trainingsverhalten des MindServers im Zeitverlauf in einer qualitativen Analyse der Indexierungsergebnisse noch nicht berücksichtigt. Hier bietet sich mittelfristig noch erheblicher Forschungsbedarf, gerade auch bei der Frage, ob die Indexierungssoftware weiterhin semiautomatisch eingesetzt wird oder ob in Zukunft ein vollautomatischer, d. h. ohne manuelle Korrekturen, Einsatz unter den dargestellten Indexierungsqualitätskriterien möglich ist. In einem nächsten Schritt könnte der Indexierungssoftware die Möglichkeit gegeben werden, basierend auf den vorliegenden Dokumenten eine eigene Taxonomie aufzubauen, um zu testen, ob diese, da sie sich eng an der Wortwahl der einzelnen AutorInnen orientiert, nicht eher mit dem Nutzersuchverhalten zusammenfällt, als dies der STW in seiner jetzigen Form es tut. Ein weiteres Forschungsfeld liegt im Bereich der Aktualität des STW und deren Einfluss auf die automatisch erzeugten Indexierungsergebnisse (vgl. Bertrand/Cellier 1995, Haller et al. 2001, Xu 2007/2008). Eventuell ließe sich der Thesaurus zudem mit weiteren Nicht-Deskriptoren so unterfüttern, dass eine bessere maschinelle Indexierung möglich ist.

Die vorliegenden Ergebnisse haben deutlich werden lassen, dass das momentan im Einsatz befindliche automatische Sacherschließungsverfahren noch nicht in der Lage ist, die bisherige Form der manuellen Indexierung in qualitativer Perspektive vollständig zu ersetzen. Gleichwohl zeigen die Ergebnisse auch die Vorteile dieser Verfahren, und dies nicht nur aus der Sichtweise des hinlänglich bekannten Datenflut-Problems. Automatische Indexierung kann – besonders durch die Verbindung mit dem STW bzw. der manuellen Sacherschließung – dazu beitragen, eine Basis aufzubauen, auf deren Grundlage in Zukunft ZBW-fremde Datenbanken mit entsprechenden wirtschaftswissenschaftlichen Inhalten einer konsistenten, hochwertigen inhaltlichen Erschließung zugeführt werden können. Dies würde eine homogene Suche über heterogene Web-Inhalte ermöglichen. Zudem, und hier schließt sich der Kreis zu der in der Einleitung aufgeworfenen Informationsstrukturierungsproblematik, eröffnet sich dadurch eine informations- und damit auch wissensbezogene Chance, nämlich über eine automatische Indexierung von digitalen Dokumenten die Nutzer bei ihren Recherchen zunehmend von den Möglichkeiten des im Entstehen befindlichen Semantic Web profitieren zu lassen.

Literaturverzeichnis

- Baum, Katrin (2005): Integration automatischer Indexierung in die Inhaltserschließung – Entwicklung eines Erschließungskonzepts für den FrauenMediaTurm. Diplomarbeit, Fachhochschule Köln. Köln.
- Bertram, Jutta (2005): Einführung in die inhaltliche Erschließung – Grundlagen, Methoden, Instrumente. Würzburg: Ergon-Verlag.
- Bertrand, Annick und Jean-Marie Cellier (1995): Psychological approach to indexing – effects of the operator's expertise upon indexing behaviour. In: Journal of Information Science, Jg. 21, Heft 6, S. 459-472.
- Brühl, Bettina (2005): Thesauri und Klassifikationen – Naturwissenschaft, Technik, Wirtschaft. Potsdam: Verlag für Berlin und Brandenburg.
- Buckland, Michael (1998): What is a "digital document"? Preprint of article published in Document Numérique (Paris), Heft-Nummer 2, S. 221-230.
<http://people.ischool.berkeley.edu/~buckland/digdoc.html> (Zugriff: 27.10.10).
- Burkart, Margarete (2004): Thesaurus. In: Rainer Kuhlen et al. (Hg.): Grundlagen der praktischen Information und Dokumentation. 5. Auflage, München: Saur. S. 141-154.
- Bürkholz, Elke (1997): Automatische Indexierung als Möglichkeit zur Sacherschließung – Vergleichende Analyse praktizierter Konzepte. Hausarbeit zur Prüfung für den gehobenen Dienst an wissenschaftlichen Bibliotheken und Dokumentationseinrichtungen des Landes Nordrhein-Westfalen. Fachhochschule Köln. Köln.
- Cooper, William S. (1969): Is Interindexer consistency a hobgoblin? In: American Documentation, Jg. 20, Heft 3, S. 268-278.
- David, C. und L. Giroux (1995): Indexing as a problem solving – a cognitive approach to consistency. In: Proceedings of the American Society for Information Science. Jg. 32, S. 49-55. http://www.cais-acsi.ca/proceedings/1995/david_1995.pdf (Zugriff: 27.10.10).
- Fugman, R. (1992): Indexing quality – Predictability versus consistency. In: International Classification, Jg. 19, Heft 1, S. 20-21.
- Gastmeyer, Manuela (2004): HWWA/ZBW-Beitrag zum AUTINDEX-Projekt. Unveröffentlichtes Manuskript.
- Gerards, Michael, Andreas Gerards und Peter Weiland (2006): Der Einsatz der automatischen Indexierungssoftware AUTINDEX im Zentrum für Psychologische Information und Dokumentation (ZPID).
<http://www.zpid.de/download/PSYNDEXmaterial/autindex.pdf> (Zugriff: 27.10.10)
- Gödert, Winfried und Klaus Lepsky (2008): Theorie und Praxis der automatischen Indexierung. <http://www.iws.fh-koeln.de/institut/personen/lepsy/skript-lp-ai-04.pdf> (Zugriff: 27.10.10).
- Gradmann, Stefan (2008): Signal. Information. Zeichen – Zu den Bedingungen des Verstehens in semantischen Netzen. Antrittsvorlesung an der Humboldt-Universität zu Berlin, am 28.10.08.
<http://edoc.hu-berlin.de/libreas/14/gradmann-stefan-44/PDF/gradmann.pdf> (Zugriff: 27.10.10).
- Gradmann, Stefan (2009): Warum sollten sich kulturelle Einrichtungen mit dem Semantic Web beschäftigen? Vortrag auf der Fachtagung Semantic Web, vom 24. bis 25. November 2009. Köln. http://www.swib09.de/vortraege/20091124_gradmann.pdf (Zugriff: 27.10.10).

- Haller, Johann, Bärbel Ripplinger, Dieter Maas und Manuela Gastmeyer (2001): Automatische Indexierung von wirtschaftswissenschaftlichen Texten – Ein Experiment. <http://www.hwwa.de/Publikationen/Dokumentation/docs/0012-gastmeyer.pdf> (Zugriff: 27.10.10).
- Haller, Johann und Paul Schmidt (2006): AUTINDEX – Automatische Indexierung. In: Daniel Lülfig und Irmgard Siebert (Hg.): Geld ist rund und rollt weg, aber Bildung bleibt. Kongressschrift zum 94. Bibliothekartag in Düsseldorf, vom 15. bis 18. März 2005. Frankfurt a. Main: Vittorio Klostermann. S. 104-114.
- Hauer, Manfred (2000): Automatische Indexierung. In: Ralph Schmidt (Hg.): Wissen in Aktion – Wege des Knowledge Managements. 22. Online-Tagung der DGI in Frankfurt a. Main vom 2. bis 4. Mai. Ober-Ramstadt: Garbs-Druck. S. 203-212.
- Helmbrecht-Schaar, Anja (2007): Entwicklung eines Verfahrens zur automatischen Klassifizierung für Textdokumente aus dem Fachbereich Informatik mithilfe eines fachspezifischen Klassifikationssystems. In: Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft, Heft 200. Berlin. <http://www.ib.hu-berlin.de/~kumlau/handreichungen/h200/h200.pdf> (Zugriff: 27.10.10).
- Holl, Daniela (2009): Search Term Recommender auf Basis der Software MindServer. Diplomarbeit im Studiengang Computervisualistik, Universität Koblenz-Landau. Koblenz.
- Hooper, R. S. (1965): Indexer consistency tests – Origin, measurements, results and utilization. Bethesda: IBM.
- Hübner, Juliane und Elke Groth (2004): Darstellung der Vor- und Nachteile ausgewählter automatischer Indexierungsverfahren im Vergleich zum intellektuellen Indexieren. Magisterarbeit, Humboldt-Universität Berlin: Institut für Bibliotheks- und Informationswissenschaft. Berlin.
- Iivonen, Mirja (1990): The Impact of the Indexing Environment on Interindexer Consistency. In: Tools for Knowledge Organisation and the Human Interface. Proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990. Frankfurt/Main: Indeks Verlag. S. 259-266.
- Kipp, Margaret E. I. und D. Grant Campbell (2006): Patterns and Inconsistencies in Collaborative Tagging Systems – An Examination of Tagging Practices. In: Proceedings of the American Society for Information Science and Technology, Jg. 43, Heft 1, S. 1-18. <http://eprints.rclis.org/8315/1/KippCampbellASIST.pdf> (Zugriff: 27.10.10).
- Knorz, Gerhard (2004): Informationsaufbereitung II: Indexieren. In: Kuhlen, Rainer u. a. (Hg.): Grundlagen der praktischen Information und Dokumentation. 5. Auflage, München: Saur. S. 179-188.
- Kuhlen, Rainer (2005): Information ein öffentliches oder privates Gut? – eine globale Perspektive, auch für Bibliotheken. Seminar: Informationsgesellschaft – ein neues Paradigma für Bibliotheken, Goethe-Institut Brasilien, 16.-17. Mai 2005. http://www.kuhlen.name/MATERIALIEN/Vortraege_05Web/RK-vortrag_sao-paulo+rio0505.pdf (Zugriff: 27.10.10).
- Lancaster, Frederick Wilfried (2003): Indexing and abstracting in theory and practice. London: Facet.
- Leder, Beate (2002): Maschinelles Indexieren und die Erschließung von Internet-Ressourcen – eine Untersuchung zur Onlineressourcen-Datenbank des Deutschen Bildungsservers. Diplomarbeit, Fachhochschule Köln, Köln.
- Leininger, Kurt (2000): Interindexer consistency in PsycINFO. In: Journal of Librarianship and Information Science, Jg. 32, Heft 1 (März), S. 4-8.

- Lepsky, Klaus und Johan Vorhauer (2006): Lingo – ein open source System für die automatische Indexierung deutschsprachiger Dokumente. In: ABI-Technik, Jg. 26, Heft 1, S. 18-28.
- Lewandowski, Dirk (2005): Web Information Retrieval – Technologien zur Informationssuche im Internet. Reihe Informationswissenschaft der DGI, Band 7. Wiesbaden: Dinges&Frick GmbH.
- Lingelbach-Hupfauer, Carmen und Hartwig Laute (2009): Die semiautomatische Indexierung von Zeitungsartikeln. In: Info 7, Jg. 24, Heft 2, S. 48-50.
- Luckhardt, Heinz-Dirk (2010): Automatische und intellektuelle Indexierung – Virtuelles Handbuch Informationswissenschaft.
http://is.uni-sb.de/studium/handbuch/exkurs_ind (Zugriff: 27.10.10).
- Manning, Christopher D., Prabhakar Raghavan und Hinrich Schütze (2008): Introduction to information retrieval. Cambridge: University Press.
- Medelyan, Olena (2005): Automatic Keyphrase Indexing with a Domain-Specific Thesaurus. Masterarbeit, Albert-Ludwigs-Universität Freiburg im Breisgau.
http://www.cs.waikato.ac.nz/~olena/master_thesis.pdf (Zugriff: 27.10.10).
- Medelyan, Olena und Ian H. Witten (2006): Measuring inter-indexer consistency using a thesaurus. In: Proceedings of the JCDL, Chapel Hill, NC, USA.
http://www.cs.waikato.ac.nz/~olena/publications/jcdl06_consistency.pdf (Zugriff: 27.10.10).
- Meyer, Thorsten und Christina Süssenbach (2009): Bestandsentwicklung in der Deutschen Zentralbibliothek für Wirtschaftswissenschaften – Leibniz-Informationszentrum Wirtschaft (ZBW). In: Bibliotheksdienst, Jg. 43, Heft 11, S.967-977.
- Mittelbach, Jens und Michaela Probst (2006): Möglichkeit und Grenzen maschineller Indexierung in der Sacherschließung – Strategien für das Bibliothekssystem der Freien Universität Berlin. In: Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft, Heft 183. Berlin. <http://www.ib.hu-berlin.de/~kumlau/handreichungen/h183/h183.pdf> (Zugriff: 27.10.10).
- Neshat, Narges und Abbas Horri (2006): A study of subject indexing consistency between the National Library of Iran and Humanities Libraries in the area of Iranian Studies. In: Cataloguing & Classification Quarterly, Jg. 43, Heft 1, S. 67-76.
- Neubert, Joachim (2009): Bringing the “Thesaurus for Economics” on to the Web of Linked Data. Linked Data on the Web Workshop, am 20. April 2009 in Madrid, Spanien.
http://events.linkeddata.org/ldow2009/papers/ldow2009_paper7.pdf (Zugriff: 27.10.10)
- Nohlen, Dieter und Rainer-Olaf Schultze (Hrsg.) (2005): Lexikon der Politikwissenschaft – Theorien, Methoden, Begriffe. Band 1. München: Beck.
- Nohr, Holger (2005): Grundlagen der automatischen Indexierung – Ein Lehrbuch. 3. überarbeitete Auflage. Berlin: Logos-Verlag.
- Nohr, Holger (2004): Wissensmanagement. In: Kuhlen, Rainer u. a. (Hg.): Grundlagen der praktischen Information und Dokumentation. 5. Auflage, München: Saur. S. 258-270.
- Nübel, Rita und Paul Schmidt (2003): Automatische Indexierung mit dem AUTINDEX System. In: Ralph Schmidt (Hg.): Competence in Content. 25. Online-Tagung des DGI, Frankfurt am Main, 3. bis 5. Juni. Wiesbaden: Dinges & Frick GmbH, S. 88-99.
- Oberhauser, Otto und Josef Labner (2003): OPAC-Erweiterung durch automatische Indexierung: Eine empirische Untersuchung mit Daten aus dem Österreichischen Verbundkatalog. In: ABI-Technik, Jg. 23, Heft 4, S. 305-314.
- Oberhauser, Otto (2005): Automatisches Klassifizieren – Entwicklungsstand, Methodik, Anwendungsbereiche. Frankfurt/Main u. a.: Peter Lang Verlag.

- Olson, Hope A. und Dietmar Wolfram (2005): Indexing consistency and its applications for information architecture – A pilot study. Paper presented at the Information Architecture Summit, Vancouver, Canada. Online abrufbar unter:
http://www.iasummit.org/2006/files/175_Presentation_Desc.pdf (Zugriff: 27.10.10).
- Olson, Hope A. und Dietmar Wolfram (2007): Syntagmatic relationships and indexing consistency on a larger scale. In: Journal of Documentation, Jg. 64, Heft 4, S. 602-615.
<http://www.emeraldinsight.com/Insight/ViewContentServlet?Filename=/published/emeraldfulltextarticle/pdf/2780640406.pdf> (Zugriff: 27.10.10).
- Panyr, Jiri (2006): Thesauri, Semantische Netze, Frames, Topic Maps, Taxonomien, Ontologien – begriffliche Verwirrungen oder konzeptionelle Vielfalt? In: Ilse Harms, Heinz-Dirk Luckhardt und Hans W. Giessen (Hg.): Information und Sprache. Beiträge zu Informationswissenschaft, Computerlinguistik, Bibliothekswesen und verwandten Fächern. Festschrift für Harald H. Zimmermann. München: K. G. Saur. S. 139-150.
- Pédaque, Roger T. (1998): Document: form, sign and medium, as reformulated for electronic documents. http://archivesic.ccsd.cnrs.fr/docs/00/06/22/28/PDF/sic_00000594.pdf (Zugriff: 27.10.10).
- Pusl, Florian (2005): Automatische Indexierung in Bibliotheken – Systeme, Projekte und Einsatzmöglichkeiten. Diplomarbeit, Fachhochschule Köln. Köln.
- Puzicha, Jan (2009): Informationen finden – Intelligente Suchmaschinentechnologie & automatische Kategorisierung. Technical White Paper. Zusendung durch die Firma Recommind auf Anfrage.
http://www.recommind.de/?id=white_papers (Stand 01.01.2009).
- Rapke, Kristin (2001): Automatische Indexierung von Volltexten für die Gruner+Jahr Pressedatenbank. In: Ralph Schmidt (Hg.): Information Research & Content Management. 23. Online-Tagung des DGI, Frankfurt am Main, 8. bis 10. Mai. Ober-Ramstadt: Garbs-Druck: 321-342.
- Recommind (2007): MindServer Categorization – Anwenderhandbuch (technische Dokumentation). Zusendung durch die Firma Recommind auf Anfrage.
http://pages.recommind.com/request_information.html (Zugriff: 27.10.2010).
- Reich, P., und E. J. Bieber (1991): Indexing consistency: The input, output function of thesauri. In: College & Research Libraries, Jg. 52, Heft 4, S. 336-342.
- Rolling, L. (1981): Indexing consistency, quality and efficiency. In: Information Processing & Management, 17. Jg., S. 69-76.
- Sachse, Elisabeth, Martina Liebig und Winfried Gödert (1998): Automatische Indexierung unter Einbeziehung semantischer Relationen – Ergebnisse des Retrievaltests zum MILOS II-Projekt. In: Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft, Band 14. Köln.
- Saarti, Jarmo (2002): Consistency of subject indexing of novels by public library professionals and patrons. In: Journal of Documentation, Jg. 58, Heft 1, S. 49-65.
- Salton, Gerard und Michael J. McGill (1983): Information Retrieval – Grundlegendes für Informationswissenschaftler. Hamburg [u. a.]: McGraw-Hill. S. 167-210.
- Schek, Markus (2005): Automatische Klassifizierung in Erschließung und Recherche eines Pressearchivs. In: Brauer, Margit (Hg.): Spezialbibliotheken zwischen Auftrag und Ressource. Kongressschrift. Hannover. S. 313-324.
- Schiffer, Ralf (2007): Automatisches Indexieren technischer Kongressschriften. Diplomarbeit, Fakultät für Informations- und Kommunikationswissenschaften Fachhochschule Köln. Köln.
<http://klaus.lepsky.googlepages.com/DiplomarbeitRalfSchiffer.pdf> (Zugriff: 27.10.10).

- Schneider, Alexandra (2008): Moderne Retrievalverfahren in klassischen bibliotheksbezogenen Anwendungen – Projekte und Anwendungen. Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft, Heft 238. Berlin.
<http://www.ib.hu-berlin.de/~kumlau/handreichungen/h238/h238.pdf> (Zugriff: 27.10.10).
- Schwarz, Elisabeth (2004): Automatische Inhaltserschließung von Textdokumenten – Eine Analyse von Softwaretools für maschinelle Indexierung und „intelligentes“ Retrieval am Beispiel der Lenzing AG F&E. In: Wissen & Management, Jg. 2004, Nr. 2, University of Applied Sciences Burgenland, Österreich.
- Semantic Web Company (2010): Social Tagging. Eine Definition.
<http://www.semantic-web.at/1.32.catchword.196.social-tagging.htm> (Zugriff: 27.10.10).
- Siegmüller, Renate (2007): Verfahren der automatischen Indexierung in bibliotheksbezogenen Anwendungen. In: Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft, Heft 214. Berlin.
<http://www.ib.hu-berlin.de/~kumlau/handreichungen/h214/h214.pdf> (Zugriff: 27.10.10).
- Stock, Wolfgang (1993): Qualität von elektronischen Informationsdienstleistungen: Wissenschaftstheoretische Grundprobleme. In: Wolfram Neubauer (Hg.): Qualität und Information. Kongressschrift: Deutscher Dokumentartag in Jena, 28. bis 30.09.1993. S. 135-152.
- Stock, Wolfgang (2000): Informationswirtschaft – Managementwissen für Studium und Praxis. München: Oldenbourg-Verlag.
- Stock, Wolfgang (2007): Information Retrieval – Informationen suchen und finden. München: Oldenbourg-Verlag.
- Stock, Wolfgang und Mechthild Stock (2008): Wissensrepräsentation – Informationen auswerten und bereitstellen. München: Oldenbourg-Verlag.
- Stock, Wolfgang (2010): E-Mail-Korrespondenz vom 25.02.2010 zur Frage der Messung von Indexierungsqualität anhand des Beispiels in Stock/Stock (2008: 355).
- Stubbs, Edgardo, Norma E. Mangiaterra, and Anna M. Martínez (1999): Internal quality audit of indexing: A new application of interindexer consistency. Cataloguing & Classification Quarterly, Jg. 28, Heft 4: 53-69.
- Tonta, Yasar (1991): A study of indexing consistency between Library of Congress and British Library cataloguers. In: Library Resources & Technical Services, Jg. 35, Heft 2, 177-185.
- Watermann, Sascha (1997): Thesaurus Wirtschaft – Die Sacherschließungsregeln der Anwender im Vergleich. Hausarbeit zur Prüfung für den gehobenen Dienst an wissenschaftlichen Bibliotheken und Dokumentationseinrichtungen des Landes Nordrhein-Westfalen. Fachhochschule Köln. Köln.
- Weller, Katrin (2009): Ontologien: Stand und Entwicklung der Semantik für das World Wide Web. In: LIBREAS: Library Ideas, Jg. 5, Heft 2, S. 15ff.
<http://www.libreas.eu/ausgabe15/texte/001.htm> (Zugriff: 27.10.2010).
- Wikipedia (2010): Inverse Dokumenthäufigkeit.
http://de.wikipedia.org/wiki/Inverse_Dokumenth%C3%A4ufigkeit (Zugriff: 27.10.10, 08:50 Uhr).
- Xu, Chen (2007): Indexing Consistency between Online Catalogues. Dissertation, Humboldt Universität, Institut für Bibliotheks- und Informationswissenschaft. Berlin.
<http://edoc.hu-berlin.de/dissertationen/chen-xu-2008-05-14/PDF/chen.pdf> (Zugriff: 27.10.10).

- Xu, Chen (2008): The influence of existing consistency measures on the relationship between Indexing consistency and exhaustivity. In: Clément Arsenault und Joseph T. Tennis (Hg.): Culture and Identity in Knowledge Organization. Proceedings of the Tenth International ISKO Conference, Montreal, Canada. Würzburg: Ergon. S. 144-150.
- ZBW (2007): Sacherschließungsregeln. Interne Arbeitsunterlage. Stand: März 2007.
- ZBW (2010): Homepage. <http://www.zbw.eu/kataloge/econis.htm> (Zugriff: 27.10.10).

Anhang

Tabelle 4: Evaluierung automatischer Indexierungsverfahren – eine Forschungsstandübersicht¹¹⁹

Titel/Quelle	Fachgebiet	Messverfahren	Titelzahl	Ergebnis	Erklärungsansätze (intervenierende Variablen)
Baum (2005)	Geschlechterforschung	IR/PR	40.000	Recall-Anstieg um 400 %, Precision-Anstieg von 72 % auf 84 % (Auszug)	Thesaurusstruktur Retrieval-Stellschrauben (u. a. Grundformreduktion)
Bertrand/Cellier (1995)	Wirtschaft Psychologie	IK (M/M)	32	IK: 0.58 (allg.), 0.62 (Wirtschaft), 0.55 (Psychologie) IB: 4.48 (allg.), 4.8 (Wirtschaft), 4.95 (Psychologie)	M/M (Experten vs. Experten; Experte vs. Laie) IB Thesaurusstruktur (Aktualität, Verweise)
David/Giroux (1995)	-	IK (M/M) Q/S	8	IK: 0.44, 0.66 qualitative Einschätzung	M/M (Experte vs. Experte) kognitive Psychologie
Gerards et al. (2006)	Psychologie	IK (M/A)	63	IK: (M/M): 42.58 bzw. 56.99 (jnF) IK (M/A): 25.22 bzw. 40.24 (jnF) IB: 6-7 (M), 8-9 (A)	M/A IB
Haller et al. (2001), Gastmeyer (2004)	Wirtschaft	IK Q/S	89	IK: 0.55 bzw. 0.49 IB: 7.4 bzw. 7.0 (M); 3.9 bzw. 3.47 (A)	Dokumentenstruktur Eigennamen, Fremdwörter Thesaurus (Synonyme ...) Allgemein- vs. Fachbegriffe Deskriptorengewichtung, Kompositazerlegung IB
Helmbrecht-Schaar (2007)	Informatik	Q/S	30	"recht gut" (S. 51), "akzeptabel" (S. 58)	Abbildung von Fachtermini via Automatik Retrieval-Stellschrauben (Dekompositionen ...) Wörterbuchpflege
Hübner/ Groth (2004)	alles	Q/S	98	IB (98 Titel): 3-6 (M), 3 (A) Qualitativ (Stichprobe: 13 Titel): Was ist gut, was schlecht?	Thesaurusstruktur Eigennamen IB

Legende: A = Automatik (automatische Indexierung), IB = Indexierungsbreite (in Anzahl der Schlagworte bzw. Kategorien), IK = Indexierungskonsistenz, IR/PR = Retrievaltest (Precision/Recall), IS = Indexierungsspezifität, IT = Indexierungstiefe, jnF = je nach der verwendeten Formel, k. A. = keine Angabe, M = Mensch (manuelle Indexierung), M/A = Intraindexierungskonsistenz im Vergleich manuelles vs. automatisch generiertes Indexat, M/M = Intraindexierungskonsistenz im Vergleich manueller Indexate, Q/S = Qualitative Einschätzung bzw. Sonstiges.

Titel/Quelle ¹²⁰	Fachgebiet	Messverfahren	Titelzahl	Ergebnis	Erklärungsansätze (intervenierende Variablen)
Iivonen (1990)	alles	IK (M/M)	10	IK: 0.08, 0.224, 0.306 (je nach Testsituation)	M/M Organisationstheorie, Umfeld des Indexierungsprozesses: Indexierungspolitik, Thesaurusstruktur Fehlindexate, Themenrelevanz
Kipp/Campbell (2006)	alles	IK (M/M)	64	IK: schwammige Erklärung IB: min. 53, max. 5.172, mittel: 917	Natur der Sache (je mehr Indexierer, desto mehr heterogenere Indexate), IB
Lancaster (2003)	Medizin	IK (M/M)	ein paar	freie Indexate höhere Konsistenz als unter Benutzung eines kontrollierten Vokabulars	M/M (Laien), kontrolliertes Vokabular
Leder (2002)	-	IR/PR	5000	Recall: 0.11-0.93, Precision: 0.69-0.95	Abbildung von Fachtermini via Automatik (M/A)
Leininger (2000)	Psychologie	IK (M/M)	60	IK: 0.50 bzw. 0.60 (jnF)	deskriptiver Vergleich Formalschlagwörter, Hauptschlagwörter vs. alle Schlagwörter, Hauptschlagwörter vs. Hauptschlagwörter Klassifikationszuordnung
Medelyan (2005)/ Medelyan/Witten (2006)	Landwirtschaft	IK (M/M) IK (M/A)	10	IK (M/M): 0.38 bzw. 0.51 (jnF), IK (M/A): 0.27 bzw. 0.38 (jnF) IB: 5-16	M/M M/A (Stellschrauben) IB
Mittelbach/Probst (2007)	alles	Q/S	k. A.	-	rein deskriptiv (Stärken-Schwächen-Analyse)
Neshat/Horri (2006)	alles	IK	4.893?	IK: max. 0.82, min. 0.51, mittel: 0.70	keine, rein deskriptiv
Nübel/Schmidt (2003)	Maschinenbau	IR/PR IK	500	Recall: 0.26-0.37, Precision: 0.13-0.30 IK: 0.10-0.20	A (Stellschrauben)

Legende: A = Automatik (automatische Indexierung), IB = Indexierungsbreite (in Anzahl der Schlagworte bzw. Kategorien), IK = Indexierungskonsistenz, IR/PR = Retrievaltest (Precision/Recall), IS = Indexierungsspezifität, IT = Indexierungstiefe, jnF = je nach der verwendeten Formel, k. A. = keine Angabe, M = Mensch (manuelle Indexierung), M/A = Intraindexierungskonsistenz im Vergleich manuelles vs. automatisch generiertes Indexat, M/M = Intraindexierungskonsistenz im Vergleich manueller Indexate, Q/S = Qualitative Einschätzung bzw. Sonstiges.

Titel/Quelle ¹²¹	Fachgebiet	Messverfahren	Titelzahl	Ergebnis	Erklärungsansätze (intervenierende Variablen)
Oberhauser/Labner (2003)	alles	IR/PR	100.000	Verfahren 1: Recall-Anstieg: 0.32, Precision: 0.91 Verfahren 2: Recall-Anstieg: 0.55, Precision: 0.56 Verfahren 3: Recall-Anstieg: 0.32, Precision: 0.84	A: drei Verfahren und deren Werte im Überblick
Olson/Wolfram (2005, 2007)	Bibliotheks- und Informationswissenschaft	IK (M/M)	38	allgemeine Einschätzung: "over a reasonably large number of people, consistency in identifying key concepts is small", S. 3.	Natur der Sache (je mehr Indexierer, desto mehr heterogene Interessen), M/M (Laie vs. Laie)
Rapke (2001)	alles	IR/PR	75.000	Verfahren 1, A: Recall: 0.74, Precision: 0.57 Verfahren 2, A: Recall: 0.56, Precision: 0.32 M/A: Recall: 0.74, Precision: 0.73	A: zwei Verfahren und deren Werte im Überblick A/M Thesaurusstruktur Eigennamen, Kompositazerlegung Vagheit der Suchanfrage (natürliche Sprache)
Reich/Biever (1991)	Landwirtschaft	IK (M/M)	13	IK: 0.24 (Sample 1), 0.45 (Sample 2) IB: 8-9 (Sample 1), 5-6 (Sample 2) IT	M/M, IB, IT Indexierungspolitik Thesaurusdesign
Saarti (2002)	Germanistik	IK (M/M)	5	IK: 0.771 (Bibliothekare), 0.563 (Laien)	M/M: Experte (Bibliothekare) vs. Laie (Nutzer) IB, sinntragende Einheiten vs. ganzes Dok. Thesaurushierarchie, Thema
Sachse et al. (1998)	alles	IR/PR	377.336	Recall: 0.92, Precision: 0.74 bzw. 0.79	A: zwei Verfahren und deren Werte im Überblick
Schek (2005)	alles	IR/PR	7.8 Mio.	Recall: 0.75 bzw. 0.90, Precision: k. A.	Cluster (ähnliche Themen)
Schiffer (2007)	Luft-/Raumfahrt	-	412	allgemeine Einschätzung	A (Stellschrauben)
Schwarz (2004)	Chemie	Fallstudie	k. A.	allgemeine Einschätzung	A: drei Verfahren und deren Werte im Überblick
Stubbs et al. (1999)	Bibliotheks- und Informationswissenschaft	IK (M/M)	2	IK: 0.29 (Sample 1), 0.60 (Sample 2)	M/M, zweimaliges Indexieren IB

Legende: A = Automatik (automatische Indexierung), IB = Indexierungsbreite (in Anzahl der Schlagworte bzw. Kategorien), IK = Indexierungskonsistenz, IR/PR = Retrievaltest (Precision/Recall), IS = Indexierungsspezifität, IT = Indexierungstiefe, jnF = je nach der verwendeten Formel, k. A. = keine Angabe, M = Mensch (manuelle Indexierung), M/A = Intraindexierungskonsistenz im Vergleich manuelles vs. automatisch generiertes Indexat, M/M = Intraindexierungskonsistenz im Vergleich manueller Indexate, Q/S = Qualitative Einschätzung bzw. Sonstiges.

Titel/Quelle ¹²²	Fachgebiet	Messverfahren	Titelzahl	Ergebnis	Erklärungsansätze (intervenierende Variablen)
Tonta (1991)	Bibliotheks- und Informationswissenschaft	IK (M/M)	82	IK: 0.36 bzw. 0.16 (unterschiedliche IB) IK: 0.41 bzw. 0.14 (gleiche IB) IB: 3.4 (Library of Congress), 1.5 (British Library)	M/M IB auch deskriptiv (qualitative Beschreibung)
Watermann (1997)	Wirtschaft	IK (M/M)	6	IK: 0.31	rein deskriptiv
Xu (2007, 2008)	alles	IK (M/M)	3307	IK: 62.2 bzw. 70.7 (jnF)	Wissenshintergrund Indexierer, Vertrautheit mit Indexierungssprache/-regeln, terminologische Kontrolle, Aktualität des Thesaurus, IB, IS, Dokumentenlänge, Fachgebiet, Erscheinungsjahr

Legende: A = Automatik (automatische Indexierung), IB = Indexierungsbreite (in Anzahl der Schlagworte bzw. Kategorien), IK = Indexierungskonsistenz, IR/PR = Retrievaltest (Precision/Recall), IS = Indexierungsspezifität, IT = Indexierungstiefe, jnF = je nach der verwendeten Formel, k. A. = keine Angabe, M = Mensch (manuelle Indexierung), M/A = Intraindexierungskonsistenz im Vergleich manuelles vs. automatisch generiertes Indexat, M/M = Intraindexierungskonsistenz im Vergleich manueller Indexate, Q/S = Qualitative Einschätzung bzw. Sonstiges.

Tabelle 5: Berechnung der Indexierungskonsistenz anhand der Indexierungsbreite manueller automatischer Indexierung und deren Übereinstimmung, Trainingsset (Auszug)

PPN ¹²³	übereinstimmende Schlagwörter	Indexierungsbreite automatisch	Indexierungsbreite manuell	Indexierungskonsistenz
091565065	6	8	7	67 %
122515013	5	8	5	63 %
133256154	4	9	5	40 %
148462804	4	8	6	40 %
188458980	1	12	3	7 %
188998675	1	3	5	14 %
254528112	3	4	4	60 %
254570682	2	5	2	40 %
258176849	4	7	4	57 %
258300299	1	8	4	9 %
270698612	5	12	5	42 %
300301758	3	15	3	20 %
300315775	3	11	4	25 %
300447604	3	9	4	30 %
303871822	3	4	5	50 %
305271490	2	9	3	20 %
306492555	8	9	13	57 %
306884828	1	7	7	8 %
306887231	3	6	7	30 %
307319547	5	57	7	8 %
307323102	3	6	6	33 %
308706013	2	8	4	20 %
308900669	4	5	5	67 %
308901037	5	7	6	63 %
309368014	5	12	7	36 %
309368642	4	5	11	33 %
309368820	2	5	6	22 %
309369037	2	5	4	29 %
309373492	4	13	6	27 %
309587883	1	2	2	33 %
309705053	3	4	5	50 %
309705274	3	10	3	30 %
309705495	1	4	3	17 %
311691633	2	9	5	17 %
311692117	2	7	5	20 %
311832989	2	4	3	40 %
312057458	2	4	8	20 %
312411987	2	4	6	25 %

123 Über die PPN (Identifikationsnummer) wurden die manuellen und die automatisch beschlagworteten Datensätze für die Berechnung der Konsistenz miteinander verknüpft.

Tabelle 6: Dokumentenanzahl (absolut) pro Indexierungskonsistenzwert (0-100 %) ¹²⁴

Indexierungs- konsistenz	Dokumenten- anzahl	Indexierungs- konsistenz	Dokumenten- anzahl	Indexierungs- konsistenz	Dokumenten- anzahl
0 %	718	27 %	697	57 %	575
2 %	17	28 %	21	58 %	103
3 %	19	29 %	996	59 %	2
4 %	42	30 %	787	60 %	555
5 %	47	31 %	387	62 %	31
6 %	81	32 %	12	63 %	380
7 %	119	33 %	2230	64 %	103
8 %	232	35 %	44	67 %	606
9 %	224	36 %	860	69 %	13
10 %	279	37 %	5	70 %	76
11 %	350	38 %	1208	71 %	231
12 %	44	39 %	10	73 %	43
13 %	538	40 %	1344	75 %	221
14 %	554	41 %	19	77 %	4
15 %	199	42 %	355	78 %	1
16 %	28	43 %	918	79 %	148
17 %	731	44 %	784	80 %	17
18 %	472	45 %	501	82 %	81
19 %	65	46 %	167	86 %	56
20 %	951	47 %	39	88 %	33
21 %	202	48 %	1	89 %	12
22 %	694	50 %	2463	90 %	5
23 %	297	53 %	29	91 %	2
24 %	53	54 %	95	92 %	2
25 %	1476	55 %	247	100 %	90
26 %	17	56 %	500		

124 Fehlende Prozentwerte (z. B. 72 oder 81 %) bedeuten, dass diesem Wert keine Dokumente zugeordnet werden konnten.

Tabelle 7: Indexierungsbreite Testset (Auszug)

PPN	Indexierungs- breite	PPN	Indexierungs- breite	PPN	Indexierungs- breite
577005030	10	577030957	9	577117351	13
577005294	3	577049399	6	577226053	7
577005472	7	577049496	11	577226266	4
577005561	8	577069357	12	577226622	16
577005693	9	577069381	12	577226762	8
577005863	6	577069608	5	577226886	6
577005952	8	577069756	7	577227025	3
577006096	20	577070452	8	577227254	1
577006193	8	577070517	6	577227343	1
577006320	5	577070614	3	577227564	5
577015524	8	577079654	8	577227653	2
577016768	4	577079751	4	577230492	10
577017721	11	577079867	6	577230646	13
577018639	7	577079980	2	577230719	2
577018949	5	577080083	4	577231049	13
577019201	5	577080148	10	577231391	7
577019252	11	577080210	5	577231898	2
577019589	8	577080342	5	577236814	6
577019953	4	577084232	8	577240412	6
577020293	12	577085719	8	577240560	9
577020919	4	577085999	5	577240951	9
577021435	5	577086677	10	577250507	8
577021494	13	577086855	7	577253549	4
577022199	5	577087231	8	577291793	4
577024205	4	577088785	4	577292064	7
577024639	6	577088963	11	577292307	5
577024949	5	577089161	5	577292447	11
577025201	9	577089293	12	577292749	5
577028804	9	577089404	16	577293028	49
577028952	9	577089889	4	577293079	3
577029401	5	577090720	8	577293370	10
577116576	6	577090909	7	577293567	9
577116800	10	577092952	2	577294563	9
577116940	6	577093940	11	577299107	10
577117173	7	577103784	2	577299778	6

Tabelle 8: Indexierungsspezifität Trainingsset (Auszug)¹²⁵

PPN	Seitenzahl		Schlagwort	H	Schlagwort	H	Schlagwort	H	Schlagwort	H	IS/A	IS/M	IS/Ü
148462804	130	A	USA	3	Betriebliche Liquidität	1	Regulierung	4	Lebensversicherung	1	0,05		
		M	Schadensversicherung	1	Lebensversicherung	1	Versicherungsaufsicht	3	Betriebliche Liquidität	1		0,04	
		Ü	Risiko	0	Reform	0	Bewertung	0	USA	3			0,02
254570682	208	A	USA	3	Bildungsverhalten	1	Geschichte	0	Institutionalismus	3	0,02		
		M	Studienfinanzierung	3	USA	3						0,02	
		Ü	USA	3	Studienfinanzierung	3							0,02
258176849	9	A	Vergleich	0	USA	3	Wirtschaftspolitik	1	Inflation	0	0,33		
		M	Wirtschaftswachstum	3	Wirtschaftslage	0	Wirtschaftspolitik	1	USA	3		0,56	
		Ü	Wirtschaftspolitik	1	Wirtschaftswachstum	3	Wirtschaftslage	1	Studienfinanzierung	3			0,67
258300299	14	A	Vergleich	0	Messung	0	USA	3	Bevölkerungsstatistik	2	0,26		
		M	Wirtschaftsstatistik	1	Statistische Methode	2	Statistisches Amt	2	USA	3		0,44	
		Ü	USA	3	Geldpolitik	2	Wirtschaftswachstum	3	Wirtschaftslage	0			0,4
588120464	26	A	Theorie	0	Spieltheorie	3	Soziale Norm	0	Vertragstheorie	5	0,18		
		M	Vertragstheorie	5	Strafe	0	Pareto-Optimum	5	Soziale Norm	0		0,2	
		Ü	Soziale Norm	0	Vertragstheorie	5							0,1
373883382	289	A	Deutschland	2	E-Business	1	Internet	3			0,02		
		M	Unternehmensberatung	0	E-Business	1	Internet	3	PC-Software	1		0,01	
		Ü	Unternehmensberatung	0	E-Business	1	Internet	3					0,01

124 Legende: H = Hierarchie des Schlagwortes im Thesaurus, IS/A = Indexierungsspezifität automatisches Verfahren, IS/M = Indexierungsspezifität manuelles Verfahren, IS/Ü = Indexierungsspezifität Übereinstimmung (automatisches vs. manuelles Verfahren).

Tabelle 9: Indexierungsspezifität Testset (Auszug)¹²⁶

PPN	Seitenanzahl	Schlagwort	H	Schlagwort	H	Schlagwort	H	Schlagwort	H	IS/A
373615671	32	Vergleich	0	Betriebliche Forschung	1	Pharmazeutisches Produkt	1	Pharmazeutische Industrie	3	0,13
373616007	48	Kritik	0	Indien	2	Globalisierung	2	Internationale Wirtschaftsbez.	1	
		Außenwirtschaftspolitik	1	Fruchtbarkeit	0	Familienplanung	3	Außenhandel	1	
		Freihandelszone	3	Außenhandelsliberalisierung	4	Außenwirtschaft	0			0,26
373616201	31	Vergleich	0	Entwicklungsländer	0	Indien	2	Innovationsmanagement	2	
		Hochtechnologiesektor	2	Technologie	0	Technischer Wettbewerb	1	Technologietransfer	1	0,22
373616422	41	Vergleich	0	Indien	2	Hochtechnologiesektor	2	Technischer Fortschritt	1	
		Lernende Organisation	1	Export	2	Internationaler Wettbewerb	1	Exportindustrie	1	
		Verarbeitendes Gewerbe	1							0,24
373616694	50	Vergleich	0	Kritik	0	Welt	0	Internationale Wirtschaftsbez.	1	
		Wirtschaftsintegration	2	Handelsregionalismus	3	Freihandelszone	3	Außenhandelsliberal	4	
		Außenhandel	1							0,20
373616961	34	International	0	Welt	0	Internationale Zusammenarb.	1	Internationales Währungssystem	1	
		Internationale Kapitalmob.	2							0,11
373617216	32	Netzwerk	0	Kritik	0	Asien	0	Europa	0	
		Kommunikation	0	Internationale Wirtschaftsbez.	1	Handelsregionalismus	3	Außenwirtschaft	0	
		Technologietransfer	1							0,13
373762992	37	Vergleich	0	Betriebliche Standortwahl	1	Multinationales Unternehmen	2	Lohnstruktur	1	
		Einkommensverteilung	0	Verarbeitendes Gewerbe	1					0,12
373763360	41	Vergleich	0	Welt	0	Direktinvestition	1	Kapitalimport	3	
		Internationale Kapitalmob.	2							0,11
373763611	43	Vergleich	0	Welt	0	Panel	2	Regionales Wachstum	4	
		Entwicklungskonvergenz	0							0,09
373763921	43	Vergleich	0	Messung	0	Außenwirtschaftspolitik	1	Export	2	
		Außenhandelsliberal	4	Außenhandelsbeschränkung	4	Außenwirtschaft	0			0,17
373764200	29	Vergleich	0	Kritik	0	Entwicklungsländer	0	Theorie	0	
		Entwicklungshilfe	2	Wirkungsanalyse	0					0,05

125 Legende: H = Hierarchie des Schlagwortes im Thesaurus, IS/A = Indexierungsspezifität automatisches Verfahren.

Tabelle 10: Indexierungseffektivität: automatisches vs. manuelles Verfahren, Trainings- und Testset (Auszug)

Schlagwort	automatisches Verfahren		manuelles Verfahren	
	insgesamt vergeben	IDF ¹²⁷	insgesamt vergeben	IDF
Vergleich	18.006	2,11	633	6,941
Theorie	16.985	2,195	6.285	3,629
USA	12.582	2,628	3.338	4,542
Risiko	6.292	3,627	250	8,281
Deutschland	4.736	4,037	2.084	5,222
Messung	4.420	4,137	224	8,439
Kritik	4.223	4,203	71	10,097
Welt	4.051	4,263	1.316	5,885
Simulation	3.007	4,693	184	8,723
EU-Staaten	2.720	4,837	1.567	5,633
International	2.522	4,946	61	10,316
Bewertung	2.405	5,015	103	9,56
Kohlendioxid	2.322	5,066	85	9,837
Geldpolitik	2.292	5,084	924	6,395
Pharmazeutisches Produkt	2.242	5,116	48	10,662
Wirtschaftswachstum	2.100	5,21	957	6,344
Großbritannien	2.002	5,279	1.061	6,195
Elektrizität	1.792	5,439	31	11,292
Prognose	1.618	5,587	130	9,224
Reform	1.592	5,61	132	9,202
Vorprodukt	1.582	5,619	28	11,439
Konsumentenverhalten	1.562	5,637	257	8,241
Entwicklungsländer	1.453	5,742	436	7,478
Wettbewerb	1.453	5,742	319	7,929
Verarbeitendes Gewerbe	1.402	5,793	334	7,863
Spieltheorie	1.392	5,804	456	7,414
Preis	1.378	5,818	113	9,426
Technischer Fortschritt	1.344	5,854	379	7,681
Informationsverhalten	1.247	5,962	70	10,117
Direktinvestition	1.182	6,04	534	7,186
Produktivität	1.052	6,208	663	6,874
Kapitalertrag	1.025	6,245	280	8,117
Lieferanten-Kunden-Beziehung	1.022	6,249	83	9,872
Public Choice	978	6,313	294	8,047
Einkommen	973	6,32	190	8,677
Computergestütztes Verfahren	953	6,35	39	10,961
Schweden	950	6,355	549	7,146

Legende: IDF = Inverse Dokumenthäufigkeit.

Schlagwort	automatisches Verfahren		manuelles Verfahren	
	insgesamt vergeben	IDF¹²⁷	insgesamt vergeben	IDF
Umweltschutz	949	6,356	63	10,269
Investition	946	6,361	311	7,966
Portfolio-Management	909	6,419	286	8,087
Humankapital	904	6,426	325	7,902
Globalisierung	897	6,438	297	8,032
OECD-Staaten	888	6,452	566	7,102
Arbeitsmarkt	878	6,469	365	7,735
Entscheidung	873	6,477	93	9,708
Australien	868	6,485	384	7,662
Lohn	838	6,536	262	8,213
Japan	822	6,564	450	7,433
Internat. Wirtschaftsbeziehungen	817	6,572	246	8,304
Dauer	808	6,588	117	9,376
Kanada	803	6,597	374	7,7
Abzinsung	798	6,606	41	10,889
Börsenkurs	794	6,614	291	8,062
Regenerative Energie	786	6,628	30	11,34
Arbeitslosigkeit	768	6,662	510	7,252
Umweltpolitik	768	6,662	153	8,989
Innovation	764	6,669	401	7,599
Monte-Carlo-Methode	751	6,694	95	9,677
Beschäftigungseffekt	746	6,704	330	7,88
Armut	736	6,723	426	7,512
Lernprozess	732	6,731	202	8,588
Konjunktur	730	6,735	407	7,578
Geschlecht	728	6,739	265	8,197
Außenwirtschaft	718	6,759	244	8,316
Brasilien	710	6,775	425	7,515
Lohnstruktur	710	6,775	290	8,067
Qualifikation	696	6,804	320	7,925
Europa	693	6,81	217	8,485
Zeit	689	6,818	58	10,389
Haushaltseinkommen	688	6,82	154	8,98
Kooperation	682	6,833	43	10,82
Einkommensverteilung	671	6,856	584	7,057
Bildungsinvestition	670	6,859	217	8,485
Forschung	667	6,865	82	9,889
China	666	6,867	415	7,55
Risikomanagement	651	6,9	119	9,352
Wechselkurs	649	6,905	258	8,235

Tabelle 11: Vergabehäufigkeit: automatisches Verfahren (Auszug)

Vergleich	18.006	Spieltheorie	1.392	Internat. Wirtschaftsbeziehungen	817	Forschung	667
Theorie	16.985	Preis	1.378	Dauer	808	China	666
USA	12.582	Europäische Wirtschafts- und Währungsunion	1.370	Kanada	803	Risikomanagement	651
Risiko	6.292	Technischer Fortschritt	1.344	Abzinsung	798	Wechselkurs	649
Deutschland	4.736	Informationsverhalten	1.247	Börsenkurs	794	Lateinamerika	645
Messung	4.420	Direktinvestition	1.182	Regenerative Energie	786	Bildungsniveau	642
Kritik	4.223	Produktivität	1.052	Arbeitslosigkeit	768	Soziale Lage	642
Welt	4.051	Kapitalertrag	1.025	Umweltpolitik	768	Aktienmarkt	635
Simulation	3.007	Lieferanten-Kunden-Beziehung	1.022	Innovation	764	Multinationales Unternehmen	630
EU-Staaten	2.720	Public Choice	978	Monte-Carlo-Methode	751	Europäische Integration	620
International	2.522	Einkommen	973	Beschäftigungseffekt	746	Inflation	609
Bewertung	2.405	Computergestütztes Verfahren	953	Armut	736	Arbeitsangebot	608
Kohlendioxid	2.322	Schweden	950	Lernprozess	732	Asymmetrische Information	608
Geldpolitik	2.292	Umweltschutz	949	Konjunktur	730	Frankreich	591
Pharmazeutisches Produkt	2.242	Investition	946	Geschlecht	728	Arbeitsmobilität	589
Wirtschaftswachstum	2.100	Wirtschaftspolitische Wirkungsanalyse	940	Außenwirtschaft	718	Bank	583
Großbritannien	2.002	Portfolio-Management	909	Brasilien	710	Schock	575
Elektrizität	1.792	Humankapital	904	Lohnstruktur	710	Indien	572
Prognose	1.618	Globalisierung	897	Qualifikation	696	Außenhandelsliberalisierung	571
Reform	1.592	OECD-Staaten	888	Europa	693	Volatilität	570
Vorprodukt	1.582	Arbeitsmarkt	878	Zeit	689	Arbeitsproduktivität	563
Konsumentenverhalten	1.562	Entscheidung	873	Haushaltseinkommen	688	Privater Haushalt	563
Entwicklungsländer	1.453	Australien	868	Kooperation	682	Internet	562
Wettbewerb	1.453	Lohn	838	Einkommensverteilung	671	Wohlfahrtseffekt	562
Verarbeitendes Gewerbe	1.402	Japan	822	Bildungsinvestition	670	Entwicklung	561

Tabelle 12: Vergabehäufigkeit: manuelles Verfahren (Auszug)

Theorie	6.285	Konjunktur	407	Multinationales Unternehmen	315	Asymmetrische Information	251
USA	3.338	Volatilität	403	Investition	311	Risiko	250
Schätzung	2.999	Innovation	401	Prognoseverfahren	304	Unternehmer	249
Deutschland	2.084	Allgemeines Gleichg.	390	Bank	302	Finnland	248
EU-Staaten	1.567	Australien	384	Finanzpolitik	301	Internationale Wirtschaftsbeziehungen	246
Welt	1.316	Zeitreihenanalyse	382	Indien	298	Außenwirtschaft	244
Großbritannien	1.061	Technischer Fortschritt	379	Globalisierung	297	Philippinen	244
Wirtschaftswachstum	957	Lateinamerika	378	Public Choice	294	Klein- und Mittelunternehmen	243
Geldpolitik	924	Kanada	374	Börsenkurs	291	Polen	240
Produktivität	663	Arbeitsangebot	365	Italien	290	VAR-Modell	237
Vergleich	633	Arbeitsmarkt	365	Lohnstruktur	290	Agency Theory	230
Einkommensverteilung	584	Europäische Wirtschafts.	357	Portfolio-Management	286	EU-Erweiterung	226
OECD-Staaten	566	Frankreich	353	Panel	281	Arbeitsproduktivität	225
Wirtschaftspolitische Wirkungs.	564	Außenhandelslib.	339	Kapitalertrag	280	Aktienmarkt	224
Schweden	549	Verarbeitendes Gewerbe	334	Bildungsertrag	270	Messung	224
Direktinvestition	534	Beschäftigungseffekt	330	Arbeitsmarktpolitik	268	Dänemark	220
Arbeitslosigkeit	510	Schock	330	Geschlecht	265	Spillover-Effekt	218
Spieltheorie	456	Inflation	327	Spanien	265	Bildungsinvestition	217
Wohlfahrtseffekt	456	Finanzmarkt	326	Lohn	262	Europa	217
Japan	450	Humankapital	325	Osteuropa	261	Unternehmensgründung	215
Entwicklungsländer	436	Russland	323	Schätztheorie	261	Arbeitsmobilität	211
Armut	426	Bildungsniveau	320	Corporate Governance	259	Frauenerwerbstätigkeit	209
Brasilien	425	Qualifikation	320	Schweiz	259	Kinder	208
Experiment	424	Unternehmensentw.	320	Wechselkurs	258	Norwegen	206
China	415	Wettbewerb	319	Konsumentenverhalten	257	Lernprozess	202

Abbildung 11: Referentenauswertung, Trainingsset (Auszug): Kategorisierung der intervenierenden Variablen

<p>goal-/sub-goal-Problem (Haupt- vs. Randthemen)</p>
<ul style="list-style-type: none"> - sehr gut, zentraler Deskriptor (Ältere Arbeitskräfte) zu geringe Gewichtung - sehr gut, einige wichtige Deskriptoren zu geringe Gewichtung - schwach, wichtige Schlagwörter haben eine zu geringe Gewichtung; wichtige Aspekte fehlen - Deutschland als Geo-Begriff ist allerdings nicht ausreichend, die Untersuchung bezieht sich auf die EU-Staaten und die USA - Auswanderung und Deutsche [...] nicht vergeben, aber Brain Drain, was ein Teil der Untersuchung ausmacht; die Studie beinhaltet zudem auch nicht ökonomische Motive der Auswanderung - Elementarschadenversicherung wurde [...] nicht erkannt, stattdessen Naturkatastrophe (immerhin) - [...] außerdem vergibt [...] die USA, was nicht falsch ist, aber eigentlich überflüssig (kommt nur am Rande vor)
<p>Sintragende Dokumentenelemente (Titel, Abstract, Keywords)</p>
<ul style="list-style-type: none"> - hier hat [...] ein schlechtes Ergebnis geliefert; wichtige Sachbegriffe fehlen, und nicht einmal das Land wurde vergeben, obwohl es sogar im Titel steht - bei [...] fehlt allerdings der Deskriptor Berufsausbildung – ein großer Fehler, da im Abstract 2 Mal erwähnt
<p>Thesaurushierarchie (Ober-, Unter-, verwandte Begriffe)</p>
<ul style="list-style-type: none"> - eindeutige Schwächen bei den Geo-Deskriptoren - [...] statt Sozialreform hat [...] nur Reform gegeben - [...] hat außerdem noch Beschäftigungseffekt gegeben, was der besser passende Deskriptor ist - Sparkasse wurde [...] nicht vergeben, obwohl im Text explizit savings bank mehrfach vorkommt – dafür wurde Bank genommen, was auch nicht direkt falsch ist - es wurden andere Begriffe gewählt, die auch tolerabel sind - es fehlt zwar Kalifornien [...]; stattdessen wurde [...] USA vergeben - Welt statt EU-Staaten ist zu unspezifisch
<p>Indexierungstiefe (Indexierungsbreite und -spezifität)</p>
<ul style="list-style-type: none"> - oft wurden sogar wichtige Deskriptoren richtig vergeben, die in ECONIS fehlten - in einigen Fällen fand [...] zusätzliche gute Deskriptoren - es fehlen die wichtigen Deskriptoren: Wissenstransfer, Innovation und Öffentlichkeitsarbeit; Publizitätspflicht ergibt in diesem Zusammenhang keinen Sinn - folgende Deskriptoren fehlen: Deutschland, Migranten , Sprachpolitik - folgende Deskriptoren fehlen: Wirtschaftswachstum, Wirtschaftsförderung - es fehlen: Preisbindung, Kundendienst, Dienstleistungsqualität, Buchhandel [...] - in einigen Fällen fand [...] besser passende [...] oder zusätzliche gute Deskriptoren - wichtige Deskriptoren wie Rentenversicherung u. Umlageverfahren (Pay-as-you-go),

Indexierungstiefe (Indexierungsbreite und -spezifität) (Fortsetzung)

- zusätzliche Deskriptoren [...] sind gute Ergänzung
- die [...] zusätzlich gewählten Deskriptoren sind überflüssig, aber wohl nicht falsch
- nicht zu gebrauchen; zentrale Deskriptoren fehlen; übrig bleibt nur Theorie (hier: nutzloses Allgemeinwort); gut. Regionaleskriptor Dtl. gefunden
- nicht zu gebrauchen; Theorie-Allgemeinwort so nutzlos
- ordentlich, etwas unspezifisch (New Keynesian P... fehlt), aber zu gebrauchen
- sehr gutes Ergebnis; alle wichtigen Deskriptoren gefunden
- nicht zu gebrauchen; bis auf Deutschland wird nichts weiter gefunden
- die wesentlichen Begriffe wurden [...] erkannt
- Welt statt EU-Staaten ist zu unspezifisch
- einige Allgemeinwörter wahrscheinlich überflüssig, aber nicht falsch
- [...] hat sogar mehr passende Deskriptoren gefunden als bei der intell. Indexierung
- [...] einige zusätzliche Begriffe, die nicht unbedingt notwendig, aber doch richtig sind
- bis auf Schätzung [...] alle Deskriptoren vergeben und zusätzlich einige ebenfalls nützliche Begriffe wie Informationsbeschaffung
- nur ein Deskriptor (Lohn) ist ungenau
- ein gutes Ergebnis; der zusätzliche Deskriptor „Glaubwürdigkeit“ fehlt in unserer Erschließung
- allerdings etliche unnötige Deskriptoren
- ganz schlechtes Ergebnis: der wichtigste Deskriptor fehlt, dafür eine Fülle von überflüssigen Deskriptoren, deren Relevanz sich nicht erschließt
- ganz gut, allerdings wieder viele unnötige Deskriptoren

Indexierungseffektivität (Schlagworttrennschärfe)

- in mehreren Fällen wurde Deskriptor „Vergleich“ vergeben [...], in einigen Fällen war das auch richtig, in anderen aber überflüssig; in einigen Dokumenten tauchen solche allgemeine Begriffe wie „Forschung“ und „Computergestütztes Verfahren“ auf
- einige Allgemeinwörter wie Vergleich, Kritik, Risiko kamen ziemlich oft vor – nicht immer zu Recht
- allerdings etliche unnötige Deskriptoren
- „Kritik“ ist unverständlich

Falsche Deskriptoren

- mehrmals ist absolut falsch Deskriptor „Pharmazeutisches Produkt“ vergeben, auch „Kohlendioxid“ mehrmals falsch vergeben
- es gibt ab und zu ein paar Ausreißer, die nicht thematisch passen, aber auch nicht oft
- es gab auch Ausreißer: so tauchte in einem Text über Zeitbudgetforschung nicht nachvollziehbar Tansania und Elektrizität auf, oder in einem Banken-Kontext der Deskriptor Kohlendioxid
- der [...] vergebene Begriff Zentralbank spielt in dem Dokument keine Rolle
- [...] USA, Haushaltsökonomik und Bildungsökonomik sowie Wirtschaftspolitische Wirkungsanalyse scheinen mir falsch zu sein
- hier finden sich grobe Ausreißer: es ist vollkommen unklar, wie [...] auf Tansania und Elektrizität kommt; diese Begriffe kommen im Text überhaupt nicht vor
- ganz falsch und unverständlich ist, wie „Kohlendioxid“ [...] hineingeraten ist

Abbildung 12: Referentenauswertung, Testset (Auszug): Kategorisierung der intervenierenden Variablen

<p>goal-/sub-goal-Problem (Haupt- vs. Randthemen)</p> <ul style="list-style-type: none"> - wenn 900 Unternehmen der Chemischen Industrie untersucht werden, muss auch die Branche vergeben werden - zentrales Thema fehlt: Entstehung (Zusammensetzung) des europäischen Gerichtshofes - zentrales Thema verfehlt, Interessenpolitik falsch - zentrales Thema nicht erfasst: Rechtsprechung, Rechtsordnung, Angleichung in der EU - befriedigend, geht in die richtige Richtung, wichtige Aspekte aber nicht erkannt (Schmuggel, grenzüberschreitender Verkehr) bzw. nur umschrieben - guter Ansatz, weitere wichtige Aspekte fehlen (Lohnniveau, Haushaltseinkommen) - befriedigend, Theorie-Aspekte alles richtig, es fehlt der empirische Bezug (Krieg ...)
<p>Sintragende Dokumentenelemente (Titel, Abstract, Keywords)</p> <ul style="list-style-type: none"> - wie aus dem Untertitel, dem Abstract und der Inhaltsübersicht zu ersehen ist, hätten noch Klubtheorie und Verfassungsökonomik hineingehört - Paper komplett auf Finnisch (bis auf Abstract, der ist englisch), sonst guter Ansatz (Umweltschutzwirkungen fehlen) - anscheinend starke Titelgewichtung, es handelt sich um Essays (4 Stück) mit unterschiedlichen Aspekten zum Titelthema
<p>Thesaurushierarchie (Ober-, Unter-, verwandte Begriffe)</p> <ul style="list-style-type: none"> - richtiger Ansatz (Inflationserwartung), die zentralen Aspekte aber nur umschrieben oder nicht erkannt (Phillips-Kurve) - sehr gut, Geodeskriptoren und Wirtschaftsgeschichte gefunden - zusätzlich zum Schlagwort Kinderarbeit (sehr gut gefunden), wird Kinder vergeben - Mehrfachindexierung, z. B. Mütter, Frauen, Arbeitsangebot, Frauenerwerbstätigkeit - alles richtig, aber viele Redundanzen: Geschlechterdiskriminierung, Geschlecht, Frauen, Frauenerwerbstätigkeit, Arbeitsangebot - trifft den Inhalt, aber nur mit zahlreichen Wiederholungen: Geschlecht, Frauen, Einkommen, Frauenarbeitslohn
<p>Indexierungstiefe (Indexierungsbreite, -spezifität)</p> <ul style="list-style-type: none"> - insgesamt zu viel zu ungenau, Modell (Allgemeines Gleichgewicht) nicht gefunden, Aspekt der Energiesubstitution auch nicht - der Auslagerungsaspekt (spin-off) fehlt völlig, Hochschulaspekt auch, einzig Humankapital Dtl. gut gefunden - zentraler Aspekt (Innovation) erkannt, aber insgesamt zu unspezifisch (es geht hier um die Steigerung des Geschäftswertes durch Innovationsmanagement) - die zentralen Aspekte werden zwar erkannt, diese gehen aber in der Vielzahl der vergebenen Deskriptoren unter - zu oberflächlich, zentraler Aspekte nicht erkannt - insgesamt viel zu unspezifisch, alles richtig, den Punkt aber nicht getroffen - die zentralen Aspekte werden zwar erkannt, wenn auch etwas unspezifisch, diese gehen aber in der Vielzahl der vergebenen Deskriptoren unter

<p>Indexierungstiefe (Indexierungsbreite, -spezifität) (Fortsetzung)</p> <ul style="list-style-type: none"> - gut, oft auftauchende Begriffe im Zusammenhang mit Spieltheorie (Kooperation, Koalition) werden erkannt, Risikomanagement falsch - viel zu ungenau und teilweise einfach falsch (es geht um Wohnungsbauinvestitionen) - befriedigend, zu unspezifisch, man erkennt aber die Stoßrichtung des Artikels - sehr gut, wichtige Begriffe getroffen, umfassend indexiert (Migranten-Aspekt gefunden)
<p>Indexierungseffektivität (Schlagworttrennschärfe)</p> <ul style="list-style-type: none"> - Begriffe wie „Vergleich“, „Kritik“, „Risiko“, „Theorie“, „Messung“ stehen oft an erster Stelle des Indexats und verlieren durch ihre häufige Verwendung jegliche Aussagekraft - Indexierungen mit „Deutschland“ und „Theorie“ (auch „USA“ und „Theorie“) treten gehäuft auf und haben keinerlei Aussagekraft hinsichtlich einer Inhaltsstrukturierung, zudem wenn sie ohne weitere Sachschlagwörter auftreten - nicht zu gebrauchen, was sagt „Theorie“ als allein stehender Begriff letztendlich aus (siehe Gesamthäufigkeit)?
<p>Falsche Deskriptoren</p> <ul style="list-style-type: none"> - Niederlande ist falsch, stattdessen Deutschland; Elektrizität ist grundlegend falsch - ziemlich verqueres [...] Indexat, Geo-Begriffe Asien und Europa nicht nachvollziehbar, Leistungsmotivation, Evolutionsökonomik und Wachstumstheorie sind überflüssig - viel zu unspezifisch, schwammig; Australien völlig falsch - völlig falsch, es geht generell über ethischen Verhalten (Verhaltensökonomik) - kein Deskriptor passt, es geht um Umweltbelastung nicht um Wettbewerbstheorie - geht völlig am Thema vorbei, es geht um Ausbildungscontrolling und nicht um Umweltökonomie - zu viele falsche, zu wenig richtige Begriffe, die aber insgesamt zu unspezifisch sind
<p>Fachgebiet (BWL, VWL, Nachbarwissenschaften)</p> <ul style="list-style-type: none"> - Ansatz gut, es fehlt der (wichtige) kulturelle Aspekt bei Vertragsverhandlungen mit chinesischen Geschäftspartnern (eher BWL, hier VWL-Begriffe) - zu unspezifisch, betriebswirtschaftliche Literatur zu Auslandsinvestitionen wird mit VWL-Begriffen wiedergegeben - befriedigend, theoretischer Charakter dieser Arbeit nicht erkannt; der Rest geht in die richtige Richtung - zu viele Begriffe; theoretische Stoßrichtung des Themas nicht erkannt - ziemlich verquere Mischung aus BWL- und VWL-Deskriptoren: Ehe; Gerechtigkeit; Verteilungsgerechtigkeit; Allokation; Public Choice; Portfolio-Management
<p>„Indexierungspolitik“</p> <ul style="list-style-type: none"> - anders indexiert, als ein manueller Indexierer dies tun würde, aber doch ok - sehr gut, einzig die Auflistung von Vergleich, Theorie und Welt an erster Stelle ist ungünstig - geht in die richtige Richtung, es werden zwar andere Begriffe verwendet, als sie ein manueller Indexierer vergeben würde, ansonsten in Ordnung