

Thomas Groß

Die Implementierung eines automatischen Indexierungsverfahrens am Beispiel der deutschen Zentralbibliothek für Wirtschaftswissenschaften

Bibliografische Information der Deutschen Nationalbibliothek:

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de/> abrufbar.

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlanges. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Copyright © 2010 Diplom.de
ISBN: 9783842805033

Thomas Groß

**Die Implementierung eines automatischen Indexierungs-
verfahrens am Beispiel der deutschen Zentralbibliothek
für Wirtschaftswissenschaften**

Thomas Groß

Die Implementierung eines automatischen Indexierungsverfahrens am Beispiel der deutschen Zentralbibliothek für Wirtschaftswissenschaften

Thomas Groß

Die Implementierung eines automatischen Indexierungsverfahrens am Beispiel der deutschen Zentralbibliothek für Wirtschaftswissenschaften

ISBN: 978-3-8428-0503-3

Herstellung: Diplomica® Verlag GmbH, Hamburg, 2010

Zugl. Humboldt-Universität zu Berlin, Berlin, Deutschland, MA-Thesis / Master, 2010

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtes.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Die Informationen in diesem Werk wurden mit Sorgfalt erarbeitet. Dennoch können Fehler nicht vollständig ausgeschlossen werden und der Verlag, die Autoren oder Übersetzer übernehmen keine juristische Verantwortung oder irgendeine Haftung für evtl. verbliebene fehlerhafte Angaben und deren Folgen.

© Diplomica Verlag GmbH

<http://www.diplomica.de>, Hamburg 2010

Inhaltsverzeichnis

Abbildungsverzeichnis	II
Tabellenverzeichnis.....	III
Abkürzungsverzeichnis	IV
1 Einleitung	1
2 Forschungsstand	10
2.1 Überblicks- und theoretische Arbeiten.....	10
2.2 Evaluierungsverfahren I: Retrievaltests	13
2.3 Evaluierungsverfahren II: Indexierungskonsistenz	15
2.3.1 Manuelle Sacherschließung.....	16
2.3.2 Manuelle vs. automatische Sacherschließung.....	20
2.4 Evaluierungsverfahren III: Sonstige.....	22
2.5 Evaluierung des MindServer	23
3 Methode, Daten, Messverfahren	25
3.1 Der MindServer als statistisches Verfahren	25
3.2 Datenlage und Stichprobenumfang	29
3.3 Zum Messverfahren: Die Evaluierung der Indexierungsqualität	30
4 Ergebnisse	36
4.1 Indexierungskonsistenz	36
4.2 Indexierungstiefe	40
4.2.1 Indexierungsbreite	40
4.2.2 Indexierungsspezifität	44
4.3 Indexierungseffektivität	46
4.4 Belegungsbilanz	49
4.5 Referentenauswertung	52
5 Fazit.....	57
Literaturverzeichnis.....	63
Anhang	69

Abbildungsverzeichnis

Abbildung 1: Indexierungsgütekriterien.....	31
Abbildung 2: Analysedimensionen	35
Abbildung 3: Anzahl der Dokumente pro Indexierungskonsistenzintervall	39
Abbildung 4: Durchschnittliche Indexierungsbreite pro Dokumentenintervall, Trainingsset	41
Abbildung 5: Durchschnittliche Indexierungsbreite pro Dokumentenintervall, Testset.....	42
Abbildung 6: Korrelation zwischen Indexierungsbreite (automatisches Verfahren) und Indexierungskonsistenz	43
Abbildung 7: Vergabehäufigkeit: manuelles vs. automatisches Verfahren	51
Abbildung 8: Zentrale intervenierende Variablen: qualitative Bewertung Trainingsset.....	52
Abbildung 9: Zentrale intervenierende Variablen: qualitative Bewertung Testset	55
Abbildung 10: Referentenauswertung Trainingsset (Auszug): Kategorisierung der intervenierenden Variablen	83
Abbildung 11: Referentenauswertung Testset (Auszug): Kategorisierung der intervenierenden Variablen	85

Tabellenverzeichnis

Tabelle 1: Indexierungsspezifität Trainingsset	45
Tabelle 2: Indexierungseffektivität der Gesamtstichprobe	47
Tabelle 3: Vergabehäufigkeit absolut und in %	49
Tabelle 4: Evaluierung automatischer Indexierungsverfahren – eine Forschungsstandsübersicht.....	70
Tabelle 5: Berechnung der Indexierungskonsistenz anhand der für ein Dokument vergebenen manuellen und automatischen Schlagwörter (Indexierungsbreite) und deren Übereinstimmung, Trainingsset (Auszug)	74
Tabelle 6: Dokumentenanzahl (absolut) pro Indexierungskonsistentwert (0-100%).....	75
Tabelle 7: Indexierungsbreite Testset (Auszug).....	76
Tabelle 8: Indexierungsspezifität Trainingsset (Auszug).....	77
Tabelle 9: Indexierungsspezifität Testset (Auszug).....	77
Tabelle 10: Indexierungseffektivität: automatisches vs. manuelles Verfahren, Trainings- und Testset (Auszug).....	79
Tabelle 11: Vergabehäufigkeit: automatisches Verfahren (Auszug)	81
Tabelle 12 Vergabehäufigkeit: manuelles Verfahren (Auszug).....	82

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 Wirtschaftswissenschaft – Leibniz Informationszentrum Wirtschaft