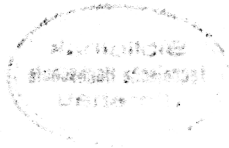


Gerhard Lustig (Hg.)

Automatische Indexierung zwischen Forschung und Anwendung

Technische Universität Braunschweig
Institut für Sprachwissenschaft und Linguistik
Linguistisches Institut



1986

Georg Olms Verlag Hildesheim · Zürich · New York
Gesellschaft für Linguistische Datenverarbeitung e. V.

Inhalt

1. Eine anwendungsorientierte Konzeption der automatischen Indexierung <i>(Gerhard Lustig)</i>	1
1.1 Pragmatische Voraussetzungen	1
1.1.1 Konkretisierung der Aufgabenstellung	1
1.1.2 Forderungen an das Indexierungswörterbuch	2
1.1.3 Fehlertoleranz	3
1.2 Ein Modell für fehlertolerierende Verfahren der automatischen Indexierung	4
1.2.1 Allgemeiner Rahmen	4
1.2.2 Formen des Vorkommens eines Terms	6
1.2.3 Hinweise und Relevanzbeschreibungen	8
1.2.4 Indexierungsfunktionen	9
1.2.5 Probabilistische Indexierungsfunktionen	10
2. Entwicklung von Indexierungswörterbüchern	13
2.1 Verfahren zur Gewinnung von Wörterbuchdaten <i>(Gerhard Lustig)</i>	13
2.1.1 Erfassung von Terms	13
2.1.2 Gewinnung von Relationen	14
2.2 Bestimmung von Mehrwortgruppen mithilfe des Begrenzerverfahrens <i>(Barbara Kienitz-Vollmer, Johannes Reichardt)</i>	18
2.2.1 Das Begrenzerverfahren	19
2.2.2 Zusatzkriterien des Begrenzerverfahrens	21
2.2.3 Weiterentwicklung des Begrenzerverfahrens zur Anwendung auf physikalische Texte	22

2.2.3.1	Modifikation der Begrenzmeng	für Zweiwortgruppen	23
2.2.3.2	Einführung neuer Zusatzkriterien		23
2.2.3.3	Ausdehnung des Verfahrens auf drei- und viergliedrige Mehrwortgruppen		25
2.2.4	Produktionsergebnisse		26
2.2.5	Reduktion der Mehrwortgruppen im Rahmen der Relationengewinnung		27
2.2.6	Möglichkeiten zur Erweiterung des Begrenzerverfahrens		29
2.3	Die Verarbeitung von Formeln in physikalischen Texten (<i>Ingeborg Muhler, Jürgen Repp</i>)		31
2.3.1	Analyse quantitativer physikalischer Angaben und Umsetzung in Bereichsbezeichner		32
2.3.2	Analyse von physikalischen und chemischen Formeln		33
2.3.3	Transformation von Formeln in Formelbezeichner		37
2.3.4	Statistische Ergebnisse der Formelverarbeitung		38
2.3.5	Beurteilung der mit dem System AUFIDET durchgeführten Formelverarbeitung		41
2.4	Die Gewinnung von statistischen Relationen zwischen Terms und Deskriptoren (<i>Norbert Fuhr, Roswitha Jäger-Beck, Michael Schwantner</i>)		43
2.4.1	Die Relation Z		43
2.4.2	Die Berechnung von Werten $z(t, s)$		46
2.4.3	Der Aufbau des Systems ZWERG		49
2.5	Indexierungswörterbücher für das Fachgebiet Physik (<i>Peter Biebricher, Norbert Fuhr, Gerhard Knorz, Michael Schwantner</i>)		52
2.5.1	Struktur des Wörterbuchs		53
2.5.2	Globale Beschreibung der Wörterbuchdaten und ihrer Herkunft		55
2.5.3	Anwendung der Verfahren für den WB-Aufbau		56

2.5.3.1	Gesamtplan	56
2.5.3.2	Quantitative Angaben	61
2.5.4	Die Indexierungswörterbücher	67
2.5.4.1	Beschreibung von PHYS-2	67
2.5.4.2	Beschreibung von PHYS/PILOT	70
3.	Verfahren der automatischen Indexierung	73
3.1	Indexieren mit dem System DAISY <i>(Ute Beinke-Geiser, Gerhard Lustig, Gudrun Putze-Meier)</i>	73
3.1.1	Untersuchung potentieller Beschreibungselemente	73
3.1.2	Bestimmung von Indexierungsfunktionen	78
3.1.2.1	Bestimmung einer probabilistischen Indexierungs- funktion	78
3.1.2.2	Der Fall der ungewichteten Indexierung	81
3.1.2.3	Heuristische Vereinfachung des Verfahrens	83
3.1.2.4	Bestimmung einer Indexierung für den AIR- Retrievaltest	87
3.1.3	Das System DAISY	90
3.1.3.1	TEXTAUFBEREITUNG	92
3.1.3.2	ERSTELLUNG DER RELEVANZBESCHREIBUN- GEN	93
3.1.3.3	DESKRIPTORENZUTEILUNG UND AUSWERTUNG	95
3.2	Die Anwendung von Polynomklassifikatoren für die automati- sche Indexierung <i>(Gerhard Knorz)</i>	98
3.2.1	Konzeption	98
3.2.2	Das Verfahren	102
3.2.3	Das System ALIBABA	105
3.2.4	Experimentelle Ergebnisse	109

3.2.4.1	Zum Aufbau von Beschreibungsvektoren	109
3.2.4.2	Zur Herleitung und Wahl von Indexierungspoly- nomen	114
3.2.4.3	Zur zweistufigen Indexierung	118
3.2.5	Definition der eingesetzten Indexierungsverfahren	122
4.	Automatische Indexierung in der Erprobung	127
4.1	Der AIR-Retrievaltest <i>(Peter Biebricher, Norbert Fuhr, Bernd Niewelt)</i>	127
4.1.1	Ziele des Retrievaltests	127
4.1.2	Vorbereitung und Durchführung des Tests	128
4.1.3	Statistische Auswertungen	131
4.1.4	Globale Ergebnisse	134
4.1.5	Der Einfluß von Testparametern	139
4.1.6	Schlußfolgerungen	142
4.2	Rankingexperimente mit gewichteter Indexierung <i>(Norbert Fuhr)</i>	144
4.2.1	Ranking mit gewichteter Indexierung	144
4.2.2	Das Testmaterial	145
4.2.3	Bewertung der Retrievalergebnisse	146
4.2.4	Retrievalfunktionen mit Berücksichtigung der Booleschen Suchfragenlogik	150
4.2.5	Retrievalfunktionen auf der Basis von Ähnlichkeitsmaßen	152
4.2.6	Zusammenfassende Bewertung	157
4.2.7	Weitere Untersuchungen	161
4.3	Fragen und Probleme der Anwendung der automatischen Inde- xierung <i>(Gerhard Knorz)</i>	163
4.3.1	Bedingungen der Anwendbarkeit der automatischen Inde- xierung	163

4.3.2 Die Pilotanwendung AIR/PHYS	164
4.3.3 Weitere Problemfelder bei Einführung und Betrieb einer automatischen Indexierung	168
4.3.4 Automatische Indexierung als ein erster Schritt in die Zukunft	170
Literatur	174
Stichwortverzeichnis	180