

INHALT

0	Einleitung	15
1	Zum Kontext der Arbeit	18
1.1	Informationsflut und Forschungsförderung	18
1.2	Informationslinguistik und Linguistische Informationswissenschaft	22
1.3	Information Retrieval	24
2	Rechtsinformationssysteme	26
2.1	Zur Praxis der Rechtsinformationssysteme	26
2.2	Möglichkeiten für den Einsatz von Rechtsinformationssystemen	30
3	Entwicklung und Stand der Forschung auf dem Gebiet des Information Retrieval	34
3.1	Die frühe Phase	34
3.2	Stand der Forschung	35
3.2.1	Praxis und Forschungsentwicklung	35
3.2.2	Substantielle Ergebnisse und Tendenzen	36
3.2.3	Problemfelder	40
3.3	Zur Ablehnung des linguistischen Ansatzes	41
3.3.1	SALTON 1981	42

3.3.2	Linguistische Verfahren in SMART	42
3.3.2.1	Problemstellung	42
3.3.2.2	Darstellung der getesteten Verfahren	43
3.3.3	Kritik des SALTON-Ansatzes	45
3.3.4	Der linguistische Ansatz und der Aspekt der Komplexität	46
3.4	Fazit	47
4	Syntax und Information Retrieval	49
4.1	Zur Komplexität des Information Retrieval	49
4.2	Syntax im Rahmen des Information Retrieval	50
4.2.1	Verwendung von Syntaxanalyse bei der Aufbereitung der Dokumenttexte	51
4.2.2	Syntaktische Aspekte der Dokumentbeschreibung	52
4.2.3	Syntaktische Aspekte der Anfragekomponente	53
4.3	Leistungssteigerung durch linguistische Teilalgorithmen	53
4.3.1	Reduktion auf Grundformen	55
4.3.2	Homographendifferenzierung	57
4.3.3	Die Erschließung syntaktischer Strukturen	60
5	Zielsetzungen syntaktisch orientierter Ansätze	64
5.1	Zur Einsetzbarkeit linguistischer Verfahren	64
5.2	Zur Allgemeinsprachlichkeit	66
5.2.1	Allgemeinsprachlichkeit und Syntax	66
5.2.2	Allgemeinsprachlichkeit und Lexikon	68
5.3	Zur Homographendifferenzierung	71
5.4	'Vollständige Syntaxanalyse' vs. 'partiell parsing'	72
5.5	Fazit	74

6	Zur Methodik auf dem Gebiet des Information Retrieval	76
6.1	Klassische Verfahren	76
6.1.1	Information Retrieval Experiment als Forschungsinstrument	77
6.1.2	System und Modell	79
6.1.3	Bewertung	80
6.1.3.1	Die Relevanzproblematik	82
6.1.3.2	Relevanzunabhängige Bewertung	84
6.1.3.3	Weitere Aspekte zur Relevanzproblematik	86
6.1.4	Zum Wissenschaftsverständnis	87
6.2	Gedankenexperiment	88
6.2.1	COOPER 1981	88
6.2.2	Kritik	90
6.3	Zur Verbindung quantitativer und qualitativer Methoden	92
6.4	Qualitative Bewertung als methodische Alternative	95
7	Zur Legitimation linguistischer Teilalgorithmen für Information-Retrieval-Systeme	97
7.1	Zur Syntaxanalyse	98
7.1.1	Die Mächtigkeitgrammatik	98
7.1.2	Zur Syntaxanalyse in JUDO	100
7.2	Evaluierung als Forschungsinstrument	101
7.2.1	Möglichkeiten der Argumentationssteuerung	101
7.2.2	Untersuchungsaufbau	105
7.3	Evaluierung und ihre Rahmenbedingungen	106

8	Eine Untersuchung zu syntaktischen Teilstrukturen im Rahmen der automatischen Indexierung juristischer Texte	108
8.0	Vorbemerkungen	108
8.0.1	Zur Textgrundlage	109
8.0.2	Zum Stand von SATAN	109
8.0.3	Zur Retrievalkomponente TELDOK (JUDOT)	111
8.0.4	Allgemeine Festlegungen bei der Auswertung	112
8.1	Bewertung in JUDO 1980	113
8.1.1	Einzellemmatisierung	114
8.1.2	Komplexe Deskriptoren	116
8.2	Bewertung der Analyseergebnisse	118
8.2.1	Überprüfung der maschinell ermittelten ADJ-SUB-Verbindungen	118
8.2.2	Zusätzliche Dokumentzuordnungen und ADJ-SUB-Verbindungen	119
8.2.3	Vergleich der Ergebnisse zum BDSG und zu den Ländergesetzen	120
8.3	Distanz der ADJ-Elemente und der SUB-Elemente im Text	122
8.3.1	Die Retrievalmöglichkeiten in JUDO im Vergleich zur Booleschen Suchlogik mit Abstandsfunktionen	123
8.3.2	Abstände innerhalb des Satzes	125
8.3.2.1	Abstand und Analyseerfolg	128
8.3.2.2	Abstand und inhaltlicher Zusammenhang der Elemente	129

8.4	Zur Legitimation der ADJ-SUB-Struktur	135
8.4.1	Frequenz von ADJ-SUB-Verbindungen	137
8.4.2	Die ADJ-SUB-Verbindung und alternative syntaktische Strukturen	138
8.4.2.1	Problemstellung	138
8.4.2.2	Vorgehen	140
8.4.2.3	Beispiele zur Abgrenzung der ALT-Strukturen	141
8.4.2.4	ALT-Strukturen	143
8.4.2.5	Bewertung der ALT-Strukturen	151
8.4.2.5.1	Zusätzliche Dokumentzuordnungen	151
8.4.2.5.2	Gruppierung der ALT-Strukturen	152
8.4.2.6	Fazit	156
8.5	Inhaltliche Interpretation der ADJ-SUB-Verbindungen	156
8.5.1	Überschriften	157
8.5.2	Namen - Termini	158
8.5.3	Differenzierung nach Elementen	159
8.5.3.1	ADJ Elemente mit Serialisierungsbezug	159
8.5.3.2	ADJ-Elemente mit Kontextbezug	159
8.5.3.3	Problematische Partizipien	163
8.5.3.3.1	Der fachsprachliche Aspekt	165
8.5.3.3.2	Zur Relevanz der ADJ(PTZ)-Elemente	166
8.5.3.3.3	Die ALT-Strukturen der ADJ(PTZ)-SUB-Verbindungen	168
8.5.3.4	SUB-Elemente mit Bezug zum Textaufbau	169
8.5.3.5	Problematische Substantive	169
8.5.3.5.1	Der fachsprachlich-textuelle Aspekt	170
8.5.3.5.2	Anwendung auf die problematischen Substantive	173
8.5.4	ADJ-Ellipsen	175
8.5.4.1	Elliptische attributive Adjektive	175
8.5.4.2	Die generierte ADJ-Relation in JUDO 1980	176
8.5.4.3	Elliptische Adjektive im Text des BDSG	178

8.5.4.4	Distanz der ADJ(ell)-Elemente von den SUB-Elementen	182
8.5.4.5	Analyse von Nominalgruppenanreihungen	184
8.5.5	Fazit	185
8.6	Vollständige Syntaxanalyse vs. partielles Parsing Vergleich anhand von Analyseergebnissen zum BDSG	186
8.6.1	Vergleich der Ergebnisse zu den ADJ-SUB- Verbindungen	187
8.6.1.1	Gesamtvergleich	188
8.6.1.2	Problematische Partizipien	189
8.6.1.3	ADJ-Ellipsen	190
8.6.2	Die Genitivrelation	191
8.6.2.1	Die Genitivrelation in JUDO	191
8.6.2.1.1	Zur Abgrenzung der Genitivrelationen	191
8.6.2.1.1.1	Juristische Verweise	191
8.6.2.1.1.2	SUB-Element oder Präposition	192
8.6.2.1.2	Zur Analysestrategie	195
8.6.2.1.3	Bewertung der Analyseergebnisse	196
8.6.2.1.3.1	Fehler bei den maschinell erstellten Relationen	196
8.6.2.1.3.2	Maschinell nicht analysierte Genitiv- verbindungen	198
8.6.2.2	Genitive in DETECT	198
8.6.3	Überlegungen zu Substantiven und zur Gestaltung einer Retrieval-Komponente	200

9	Zusammenfassung	203
9.1	Methodische Grundlagen	203
9.2	Zum Stellenwert linguistischer Teilalgorithmen in der automatischen Indexierung	203
9.3	Ergebnisse der empirischen Untersuchung	204
9.4	Zur Verbindung der Bereiche Linguistik und Informationswissenschaft	205
	ANMERKUNGEN	207
	LITERATUR	222