

Gerda Rüge

# Wortbedeutung und Termassoziation

Methoden zur automatischen semantischen Klassifikation

1995

Georg Olms Verlag Hildesheim . Zürich. New York

# Inhalt

1. Motivation	11
1.1 Lexika und Wissensbasen	12
1.2 Anwendungsmöglichkeiten einer semantischen Klassifikation	13
1.3 ^Ursprünge und Umgebung dieser Arbeit	16
1.4 Aufgabenstellung	16
2. Sprachphilosophische Grundlagen	19
2.1 Wortbedeutung	19
2.1.1 Modelltheoretische Semantik	20
2.1.2 Strukturelle Semantik oder Merkmalssemantik	25
2.1.3 Der Wittgensteinsche Bedeutungsbegriff	27
2.1.4 Vergleichende Darstellung der Wonbedeutungstheorien	28
2.2 Synonymie und Ähnlichkeit	29
2.2.1 Synonymie- und Ähnlichkeitskriterien	30
2.2.2 Auswahl eines Synonymie- und eines Ähnlichkeitskriteriums	32
2.3 Eigenschaften und Relationen	34
2.4 Die Head/Modifier-Relation in der Dependenzgrammatik	36
3. Struktureller Vergleich von Extensionen	39
3.1 Unterschiede zu herkömmlichen KI-Systemen	39
3.2 Die Strukturhypothese des logischen Empirismus	41
3.3 Voraussetzungen einer Theorie des strukturellen Extensionsvergleichs	42
3.3.1 Hinreichende Komplexität	43
3.3.2 Vollständigkeit der Beschreibung	43
3.3.3 Wahrheit und Sinn der Beschreibung	43
3.3.4 Modellierbarkeit	44
3.3.5 Normalisierung des Modells	45
3.4 Ein Axiomensystem für Extensionen	47
3.4.1 Ausdrückbarkeit	47
3.4.2 Vorbestimmtheit der externen Eigenschaften	48
3.4.3 Homogenität von Extensionen	49
3.4.4 Widerspruchsfreiheit des Axiomensystems	51

3.5	Synonymiekriterium: Übereinstimmung der externen Eigenschaften	54
3.6	Externe Eigenschaften und die Head/Modifier-Relation	65
3.6.1	Denotate von Head/Modifier-Relationen	65
3.6.2	Head/Modifier-Relationen zwischen Nomen	67
3.6.3	Abdeckung der Relation der externen Eigenschaft durch die Head/Modifier-Relation	71
3.7	Theorie und Anwendung	73
3.7.1	Abweichungen von den Voraussetzungen der Theorie	73
3.7.1.1	Hinreichende Komplexität des Zustandes	74
3.7.1.2	Vollständigkeit der Beschreibung	75
3.7.1.3	Wahrheit, Sinn und Relevanz der Aussagen	75
3.7.1.4	Modellierbarkeit	76
3.7.2'	Das Münchhausenproblem	78
3.7.3	Anwendung des Ansatzes auf geeignete Korpora	79
4.	Frühere Ansätze zur Termassoziation	81
4.1	Abgrenzung	81
4.2	Evaluierungsverfahren	83
4.3	Der Korpusansatz oder Anwendung des Wittgensteinschen Bedeutungsbegriffs	87
4.3.1	Klassifikation der rein statistischen Ansätze	92
4.3.2	Kritik der rein statistischen Termassoziation	96
4.3.3	Einsatz linguistischen Wissens	100
4.3.4	Zusammenhang mit der vorliegenden Arbeit	102
4.4	Der Lexikonansatz oder Anwendung der Merkmalstheorie	104
4.4.1	Grundlegende Annahmen	104
4.4.2	Termrelationen auf der Basis von Genustermen	105
4.4.3	Term-Relationen auf der Basis von Genustermen und Differenzia	107
4.4.4	Vor- und Nachteile des Lexikonansatzes	108
4.4.5	Bezug zur vorliegenden Arbeit	109
4.5	Verfahren auf der Basis von intellektuell erstellten Relevanzdaten	111
4.5.1	Interpretation von a priori Relevanzinformation	112
4.5.2	Interpretation von benutzerabhängigen Relevanzaussagen	115
4.5.2.1	Lernen aus Benutzerverhalten	115
4.5.2.2	Relevanz Feedback	116
	<b>4.5.3 Interpretation von intellektuellen Indexierungen</b>	<b>117</b>
4.6	<b>Zusammenfassung</b>	<b>118</b>

5. Experimente	121
5.1 Zugrundeliegende Daten	123
5.1.1 Textkorpus	123
5.1.2 Syntaktische Analyse	124
5.1.3 Head/Modifier-Relationen im Korpus	127
5.1.4 Feldstudien	128
5.1.5 Head/Modifier-Relationen in den Experimenten	131
5.2 Methodik	132
5.2.1 A-posteriori-Evaluierung auf der Basis der gesamten Datenmenge	134
5.2.2 Evaluierung anhand von a-priori-Relevanzinformation	136
5.2.3 Signifikanztests	137
5.2.4 Der mathematische Rahmen für den Head/Modifier-Vergleich	139
5.3 ^Evaluierung der Arbeitshypothese	141
5.3.1 Intellektuelle Bewertung von Termassoziationen mit verschiedenen Ähnlichkeitsvierten	141
5.3.1.1 Auswahl der Stichproben	142
5.3.1.2 Anordnung der Paare bei der Bewertung	143
5.3.1.3 Ergebnisse	144
5.3.2 Gegenprobe: Automatischer Vergleich von a priori ausgewählten Synonymen und Zufallspaaren	145
5.3.3 Interpretation der Ergebnisse	146
5.4 Vergleichende Untersuchung verschiedener Ähnlichkeitsmaße	147
5.4.1 Parameter in der Untersuchung	147
5.4.1.1 Ähnlichkeitsmaße und ihre Anschauung	147
5.4.1.2 Räumliche Ähnlichkeitsanschauungen	150
5.4.1.3 Mengenüberlappung als Ähnlichkeitsanschauung	151
5.4.1.4 Gewichte	152
5.4.1.5 Gewichtung nach der Häufigkeit	153
5.4.1.6 Gewichtung nach der Wichtigkeit	155
5.4.1.7 Definition der untersuchten Ähnlichkeitsmaße	158
5.4.2 Versuchsaufbau	162
5.4.2.1 Intellektuelle Term-Precision-Untersuchung	162
5.4.2.2 Automatische Term-Recall-Untersuchung	163
5.4.2.3 Zusammenhang zwischen Term-Recall und Term-Precision	164
5.4.3 Ergebnisse	165
5.4.4 Interpretation der Ergebnisse	166
5.4.4.1 Der Einfluß der Head- bzw. Modifier-Wichtigkeit	166
5.4.4.2 Der Einfluß der Ähnlichkeitsmaßeschemata	168
5.4.4.3 Einfluß der Link-Häufigkeit	169
5.5 Ein Ähnlichkeitsmaß für Retrieval-Anwendungen	171
5.6 Ergebnisqualität Adjektive vs. Nomen	174
5.7 Zusammenfassung	175

6. Exkurs: Spreading-Activation-Netze oder Gedächtnismodelle	177
6.1 Frühere Anwendungen für Spreading-Activation-Netze	177
6.1.1 Spreading-Activation-Netze in der Computerlinguistik	178
6.1.2 Spreading-Activation-Netze im Information Retrieval	181
6.2 Gedächtnismodelle	183
6.2.1 Gedächtnismodelle zur Erklärung von Verstehen, Erkennen und Entscheiden	183
6.2.2 Gedächtnismodelle zur Erklärung des Erinnerungsvermögens	186
6.2.3 Assoziationismus und Termassoziation	188
6.3 Ein Spreading-Activation-Netz aus Heads und Modifiers	191
6.3.1 Indirekte Aktivierung von semantisch verwandten Wörtern	192
6.3.2' Berücksichtigung von synonymen Heads und Modifiers beim # Head/Modifier-Vergleich	193
6.4 Experimente	195
6.4.1 Datenbasis	195
6.4.2 Netzparameter	196
6.4.3 Ähnlichkeitsmaß vs. Spreading-Activation	198
6.5 Bewertung des Netzansatzes	200
7. Ausblick	203
7.1 Technische Möglichkeiten zur Qualitätssteigerung	203
7.1.1 Erweiterung der Datenbasis	203
7.1.2 Cluster-Analyse	204
7.1.3 Iterieren der Ähnlichkeitsberechnung	205
7.1.4 Verbesserungsmöglichkeiten beim Spreading-Activation	205
7.2 Anpassung der linguistischen Analyse	208
7.2.1 Besser geeignete Grundformen	208
7.2.2 Weitere Sprachen	209
7.3 Inhaltliche Differenzierung der Assoziationsergebnisse	210
7.3.1 Differenzierung der Ähnlichkeitsaussagen	210
7.3.2 Disambiguierung von Polysemen	212
7.4 Mehrwortbegriffe und Abkürzungen	212
Danksagungen	215
Literatur	217
Anhänge	
Heads und Modifiers von car.	233
Glossar	239
Sachregister	241