

Gerhard Heyer  
Uwe Quasthoff  
Thomas Wittig

# Text Mining: Wissensrohstoff Text

Konzepte, Algorithmen,  
Ergebnisse

# Inhaltsverzeichnis

1	<b>Wissen und Text</b> *	1
1.1	Text Mining *	1
1.2	Aufbau & Struktur von Text **	7
1.3	Wissensverarbeitung gestern & heute *	12
2	<b>Grundlagen der Bedeutungsanalyse</b> *	19
2.1	Syntagmatische Relationen *	20
2.2	Paradigmatische Relationen *	25
2.3	Semantische Relationen *	30
2.4	Logische Relationen **	39
2.5	Fach- & Allgemeinsprache *	45
3	<b>Textdatenbanken</b> *	51
3.1	Textressourcen *	51
3.2	Aufbau von Textdatenbanken *	56
3.3	Segmentierung von Text **	62
3.4	Datenstrukturen für Wörter **	68
3.5	Abfragemöglichkeiten in Textdatenbanken *	77
4	<b>Sprachstatistik</b> *	85
4.1	Zipfsche Gesetze *	87
4.2	Differenzanalyse *	95
4.3	Probabilistisches Sprachmodell 1 *	100
4.4	Probabilistisches Sprachmodell 2 *	109
4.5	Hidden Markov-Modelle **	115
4.6	Tagging - als Anwendung der HMMs **	126
4.7	Signifikante Kookkurrenzen *	134
4.8	Signifikante Kookkurrenzen - Beispiele & Anwend. *	141
4.9	Visualisierung von signifikanten Kookkurrenzen *	149
4.10	Anwendung auf andere Sprachen *	156
4.11	Kookkurrenzen höherer Ordnung **	161
4.12	Netze von Kookkurrenten *	168
4.13	Small Worlds und skalenfreie Netze **	172
4.14	Disambiguierung *	180
4.15	Communities **	187
5	<b>Clustering</b> *	195
5.1	Clustering-Verfahren *	196
5.2	Dokumentenähnlichkeit *	201
5.3	Clustern von Wortformen *	209

5.4	Merkmalsextraktion *	220
<b>6</b>	<b>Musteranalyse *</b>	<b>227</b>
6.1	Reguläre Ausdrücke *	227
6.2	Syntaktische Muster *	233
6.3	Morphemmuster **	238
<b>7</b>	<b>Hybride Verfahren *</b>	<b>247</b>
7.1	Lexikalische Filter *	248
7.2	Kombination verschiedener Wissensquellen *	254
7.3	Bootstrapping **	260
<b>8</b>	<b>Beispielanwendungen *</b>	<b>267</b>
8.1	Fachterminologie *	268
8.2	Erkennung von Fachterminologie *	272
8.3	Beispielanalysen von zwei Textsammlungen *	278
8.4	Aufbau einer Wissenslandkarte *	287
<b>Glossar</b>		<b>299</b>
<b>Literatur</b>		<b>307</b>
<b>A</b>	<b>Linguistische Grundlagen - Linguistische Ebenen *</b>	<b>318</b>
<b>B</b>	<b>Linguistische Ebenen **</b>	<b>323</b>
<b>C</b>	<b>Zerlegung eines Satzes *</b>	<b>324</b>
<b>D</b>	<b>Formale Rekonstruktion linguist. Ebenen *<sup>e</sup></b>	<b>325</b>
<b>E</b>	<b>Linguistische Grundlagen - Morphologie *</b>	<b>327</b>
<b>F</b>	<b>Das Projekt »Deutscher Wortschatz« *</b>	<b>331</b>
<b>G</b>	<b>Kurzeinführung Statistik *</b>	<b>333</b>
<b>H</b>	<b>Poisson-Verteilung und Signifikanz-Maß **</b>	<b>338</b>
<b>I</b>	<b>Eine Einführung in Grafen *</b>	<b>341</b>
<b>Sachindex</b>		<b>343</b>