

Matthias Dehmer



© 2008 [AGI-Information Management Consultants](#)
May be used for personal purposes only or by
libraries associated to [dandelion.com](#) network.

Strukturelle Analyse Web-basierter Dokumente

Deutscher Universitäts-Verlag

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation der Arbeit	1
1.2	Zielsetzung der Arbeit	3
1.3	Aufbau der Arbeit	5
1.4	Wissenschaftlicher Beitrag der Arbeit	7
2	Strukturelle Aspekte hypertextueller Einheiten	11
2.1	Hypertext und Hypermedia	11
2.2	Problemstellungen des Web Mining	14
2.2.1	Probleme des World Wide Web bezüglich der Informationssuche	14
2.2.2	Bereiche des Web Mining und deren Kernaufgaben	16
2.3	Existierende graphentheoretische Analysemethoden von Hypertextstrukturen	20
2.3.1	Motivation	21
2.3.2	Maße für die strukturelle Analyse von Hypertexten	22
2.3.3	Zusammenfassende Bewertung	30
2.3.4	Fazit	30
2.4	Existierende Clusteringverfahren zur Analyse hypertextueller Daten	31
2.4.1	Interpretation von Clusterlösungen	33
2.4.2	Hierarchische Clusteringverfahren	35
2.4.3	Partitionierende Clusteringverfahren	38

2.4.4	Sonstige Clusteringverfahren	40
2.5	Modellbildung: Polymorphie und funktionale Äquivalenz	43
2.6	Konkreter Modellierungsansatz auf der Basis von GXL	45
2.7	Zusammenfassende Bewertung und Fazit	47
3	Grenzen der inhaltsbasierten Kategorisierung von Hypertextstrukturen	51
3.1	Motivation	51
3.2	Das Testkorpus und die Extraktion web-basierter Hypertexte	54
3.3	Motivation des maschinellen Lernverfahrens	56
3.4	Das Kategorisierungsexperiment	59
3.5	Interpretation der Evaluierungsergebnisse	62
3.6	Fazit	64
4	Graphentheorie und strukturelle Ähnlichkeit: Bekannte Methoden	67
4.1	Erforderliche Grundlagen	67
4.1.1	Überblick und Resultate der Graphentheorie	69
4.1.2	Ähnlichkeit strukturierter Objekte	72
4.1.3	Abstand, Distanz und Metriken	74
4.2	Strukturelle Ähnlichkeit von Graphen	75
4.3	Graph Mining und weitere graphorientierte Ähnlichkeitsmaße	80
4.4	Zusammenfassende Bewertung	89
5	Graphbasierte Analyse und Retrieval: Neuer Ansatz	93
5.1	Motivation	94
5.2	Gradsequenzen von Graphen	98
5.3	Hierarchisierte und gerichtete Graphen	102

5.4	Zentraler Lösungsansatz	105
5.5	Berechnungsgrundlagen	108
5.6	Strukturelle Ähnlichkeit hierarchisierter und gerichteter Graphen .	113
5.7	Ergebnisse	122
5.8	Experimentelle Ergebnisse	124
5.8.1	Experimente mit Website-Strukturen	125
5.8.2	Experimente mit web-basierten Dokumenten	132
5.8.3	Fazit	137
6	Exkurs: Strukturvorhersage	139
6.1	Erkennung struktureller Beziehungen zwischen Graphmengen	139
6.2	Ergebnisse	142
6.3	Fazit	144
7	Zusammenfassung und Ausblick	145
7.1	Zusammenfassung der Ergebnisse	145
7.2	Ausblick	148
7.3	Weiterführende Fragestellungen	151
	Literaturverzeichnis	153