

Inhaltsverzeichnis

1 Einleitung	1
1.1 Eine Kurzbeschreibung des World Wide Web	1
1.2 Probleme, die die (Weiter-)Entwicklung des World Wide Web betreffen	2
1.3 Motivation und Aufgabenbeschreibung	4
1.4 Ausblick	6
2 Statische Erfassung von Informationen im WWW	8
2.1 Erfassen der Datenobjekte im Web	8
2.1.1 Eine kurze Klassifizierung der Informationen im Web und Methoden zur Veröffentlichung	8
2.1.2 Erfassen der Informationen, Aufbau eines Crawlersystems	9
2.1.3 Nicht zugängliche Informationen	11
2.2 Anfragesprachen für Web-Strukturen	11
2.3 Empirische Ergebnisse zum Aufbau des Web	13
2.4 Informationsextraktion aus Web-Daten	14
2.4.1 Screen-Scraping	14
2.4.2 Wrapper für die Informations-Extraktion	16
2.4.3 Informationsextraktion mittels automatischer Strukturschätzung	18
2.5 Zusammenfassung	26
3 Methoden zur Erfassung des Dynamic Web	27
3.1 Das Web als dynamische Informationsquelle	27
3.2 Die Bedeutung der Dynamik des Web für die Informationserfassung	29
3.3 Crawler Optimierung	31
3.3.1 Stochastische Methoden zur Crawler Optimierung . .	31

3.3.2	Crawler Optimierung bei regulärem Änderungsverhalten	33
3.3.3	Zusammenfassung und Ausblick	40
4	Zerlegungsbasierte Erfassung von Web Information	42
4.1	Identifikation dynamischer Informationskomponenten	43
4.2	Beschreibungen und Definitionen für Änderungen von Web-Information	46
4.3	Strukturbasierte Erfassung des Änderungsverhaltens von Web Information	47
4.4	Tracken dynamischer Informationskomponenten	51
4.5	Pruning dynamischer Web-Information	55
4.6	Zusammenfassung und Ausblick	57
5	Anfrageverfahren für das Dynamic Web	58
5.1	Statische Anfragen	59
5.1.1	Der Aufbau moderner Suchmaschinen	59
5.1.2	Weiterführende Ansätze	63
5.2	Dynamische Anfragen	63
5.2.1	Motivation	63
5.2.2	Architekturen für dynamische Anfragen	66
5.2.3	Relevanzschätzung bei inhalts-basierten Publish/Subscribe Systemen	70
5.2.4	Beschränkte kontinuierliche Suchanfragen	73
5.2.5	Eine Anfragesprache für Beschränkte Kontinuierliche Suchanfragen	74
5.2.6	Realisierung Beschränkter Kontinuierlicher Suchanfragen	77
5.2.7	Qualitätsfaktoren Beschränkter Kontinuierlicher Suchanfragen	78
5.2.8	Probleme bei der Realisierung Beschränkter Kontinuierlicher Suchanfragen	81
5.2.9	Optimierung beschränkter kontinuierlicher Suchanfragen mittels 'Optimal Stopping'	83
5.2.10	Optimierung beschränkter kontinuierlicher Suchanfragen durch Verteilungsschätzung	91
5.2.11	Evaluation von Verfahren für beschränkte kontinuierliche Anfragen	97
5.3	Zusammenfassung und Ausblick	97

6 Organisieren von Informationen aus dem Dynamic Web	102
6.1 Das Web als Informations-Raum	103
6.1.1 Methoden zur Darstellung der Informationsstruktur des Web	103
6.1.2 Personalisierte Darstellung von Web-Informationsräumen	104
6.1.3 Ein Werkzeug für die Darstellung personalisierter Site-maps	105
6.2 Personal Information Spaces	108
6.2.1 Ansätze zur Organisation von Web-Referenzen	109
6.2.2 Ein Werkzeug für die räumliche Organisation von Informationsquellen	111
6.3 Kollaborative Infomationsräume	112
6.4 Zusammenfassung	116
7 Zusammenfassung und Ausblick	118