

Inhaltsverzeichnis

1	Überblick Suchdienste im World Wide Web	1
1.1	Webkataloge	2
1.2	Suchmaschinen	4
1.3	Metasuchmaschinen	8
1.4	Payed Listing Suchmaschinen	9
2	Systemumfeld der Suchmaschinen im WWW	13
2.1	Client-Server-Prinzip bei Suchmaschinen	13
2.2	Das HTTP-Protokoll	15
2.2.1	HTTP-Protokollinformationen	15
2.2.2	Server-Statusinformationen	17
2.3	Funktion und Aufbau eines URL	19
2.4	Hypertextdokumente	20
2.4.1	HTML-Dokumententitel	21
2.4.2	HTML-Metainformationen	22
2.4.3	HTML-Dokumentenkörper	23
3	Funktionsweisen von Suchmaschinen	25
3.1	Webrobots und die Erfassung des WWW	25
3.1.1	Arbeitsweisen von Webrobots im Überblick	27
3.1.2	Erfassung des WWW durch Webrobots	30
3.1.3	Loader und URL-Datenbank	33
3.1.4	Der Checker	35
3.2	Datenaufbereitung und Analyse	39
3.2.1	Information Retrieval Systeme	39
3.2.2	Verfahren der Datenaufbereitung und Analyse im Überblick	44
3.2.3	Datennormalisierung	46
3.2.4	Wortidentifikation	48
3.2.5	Sprachidentifikation	49
3.2.6	Word Stemming	51
3.2.7	Deskriptorengewinnung	53
3.2.8	Stoppwortliste und Black List	56
3.3	Datenstrukturen der Information Retrieval Systeme	57
3.3.1	Besonderheit der Datenstrukturen von IR-Systemen	57
3.3.2	Direkte Dateien	60
3.3.3	Invertierte Dateien	61
3.3.4	Indexdatei	63

4 Relevanz und Gewichtungsmodelle	67
4.1 Zusammenhang von Relevanz und Rangbildung	67
4.2 Effektivität von Suchmaschinen	70
4.3 Statistische Gewichtungsmodelle	73
4.3.1 Das Vektorraummodell	73
4.3.2 Die relative Worthäufigkeit (TF-Algorithmus)	76
4.3.3 Die inverse Dokumentenhäufigkeit (ITF-Algorithmus)	77
4.3.4 Bedeutung der Lage eines Keywords	79
4.4 Hypermedia basierte Gewichtungsmodelle	80
4.4.1 PageRank von Google	81
4.4.2 Systematik der Click Popularity	87
4.5 Cluster-Verfahren	90
5 Suchprozess und Suchformen	95
5.1 Der Query Processor – Suchtool der Suchmaschine	95
5.2. Suchoperatoren der Suchmaschinen	98
5.3 Expertensuche bei den Suchmaschinen	101
5.4 Ergebnislisten der Suchmaschinen	106
5.4.1 Suchergebnisliste von Fireball.de	107
5.4.2 Suchergebnisliste von Lycos.de	109
5.4.3 Suchergebnisliste von Altavista.de	110
5.4.4 Suchergebnisliste von Google.de	112
6 On the Page Methoden der Optimierung	115
6.1 Priorität der Website Optimierung und Erfordernis eines Konzepts	116
6.2 Wahl des Dokumententyps	118
6.3 Framesets	122
6.4 Keywords und Keyword-Strategie	125
6.5 Groß- und Kleinschreibung, Pluralform und Sonderzeichen	128
6.6 Der Dokumententitel	131
6.7 Die Bedeutung der Meta-Tags	137
6.8 Einfluss von Keyword-Dichte	144
6.9 Lage von Keywords	146
6.10 Keywords in ALT-Tags und Kommentaren	149
6.12 Textauszeichnung, Textgröße und Überschriften	151
6.12 Link-Strukturen und Verzeichnistiefe	153
6.13 Sprachen und Länderindexe	156
6.14 Valides HTML	160
6.15 Vermeidbare Fehler	163
7 Off the Page Methoden der Optimierung	169
7.1 Domain-Name und Bezeichnung der Verzeichnisse	169
7.2 Aktualität und Änderungsfrequenz	171
7.3 Serververfügbarkeit, URL-Anzahl und IP-Sperrung	175
7.4 Link Popularity und Link Analyse	178
7.5 Click Popularity und Dokumententitel	184

8 Spam – unzulässige Maßnahmen	187
8.1 Keyword Stuffing.....	189
8.2 Blindtexte und Schriftgröße.....	191
8.3 Meta-Informationen ohne inhaltlichen Bezug.....	194
8.4 Doorway Pages.....	195
8.5 IP-Delivering und Cloaking.....	197
8.6 Unsichtbare Hyperlinks.....	200
8.7 Erkennung von Dubletten.....	202
8.8 Spam durch Link Farms.....	204
9 Aufnahme in den Datenbestand	207
9.1 Relevante Suchmaschinen für die Indexierung.....	207
9.2 Suchmaschinen und Kooperationen.....	212
9.3 Anmeldung bei Suchmaschinen.....	216
9.4 Aufnahme gegen Bezahlung.....	220
9.5 Filterketten der Webrobot Systeme.....	229
9.6 Überprüfung der Indexierung.....	232
10 Payed Placement – Ranking gegen Bezahlung	235
10.1 Payed Placement Programme bei Google.....	235
10.2 Trusted Feed-Programm von Altavista.....	238
10.3 Index Connect von Inktomi.....	240
10.4 Overture – Payed Placement.....	242
10.5 Espotting – Payed Placement.....	248
10.6 QualiGo – Payed Placement.....	251
11 Monitoring und Controlling	257
11.1 Website-Monitoring.....	257
11.2 Log File Analyse.....	261
11.3 Rank Monitoring.....	269
Anhang	273
Literaturverzeichnis.....	273
Sachwortverzeichnis.....	275