

Michael Glögger

Suchmaschinen im Internet

Funktionsweisen, Ranking Methoden,
Top Positionen

Mit 47 Abbildungen und 25 Tabellen



Springer

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	WordStemming	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
6.11 Textauszeichnung, Textgröße und Überschriften.....	151
6.12 Link-Strukturen und Verzeichnistiefe.....	153
6.13 Sprachen und Länderindexe.....	156
6.14 ValidesHTML.....	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

Abbildungsverzeichnis

Abb. 1.1.	Altavista.de Suchergebnisliste - verkürzte Darstellung.....	2
Abb. 2.1.	Client-Server-Prinzip bei den Suchmaschinen.....	14
Abb. 2.2.	Struktur von HTML-Dokumenten.....	21
Abb. 3.1.	Komponenten eines Webrobot-Systems.....	29
Abb. 3.2.	Struktur eines invertierten Dateisystems.....	42
Abb. 3.3.	Prozesse der Datennormalisierung und Dokumentenanalyse.....	45
Abb. 3.4.	Zusammenhang von Worthäufigkeit und Wortrelevanz.....	55
Abb. 3.5.	Invertiertes Dateisystem - schematische Darstellung.....	59
Abb. 3.6.	Beispiel eines b-tree organisierten invertierten Dateisystems.....	65
Abb. 4.1.	Vergleich binärer Vektor - gewichteter Vektor.....	74
Abb. 4.2.	Vektorrepräsentation im Vektorraummodell.....	75
Abb. 4.3.	Dokumentengewichtung bei Page Rank-Ausgangssituation.....	84
Abb. 4.4.	Dokumentengewichtung PageRank - erste Kalkulation.....	85
Abb. 4.5.	Dokumentengewichtung PageRank - zweite Kalkulation.....	86
Abb. 4.6.	Rückzahlung bei Fireball.de - Stand 1.2.2002.....	89
Abb. 4.7.	Rückzahlung bei Alltheweb.com - Stand 1.11.2002.....	89
Abb. 4.8.	Cluster-Suche bei Google.....	91
Abb. 4.9.	Cluster-Suche bei Teoma.....	92
Abb. 5.1.	Erfassung von Booleschen Suchanfragen durch Serverprotokollierung.....	99
Abb. 5.2.	Expertensuche bei Alltheweb.com - Teil 1, Stand 1.11.2002.....	101
Abb. 5.3.	Expertensuche bei Alltheweb.com - Teil 2, Stand 1.11.2002.....	103
Abb. 5.4.	Suchergebnisliste Fireball.de - verkürzte Darstellung.....	108
Abb. 5.5.	Suchergebnisliste Lycos.de - verkürzte Darstellung.....	110
Abb. 5.6.	Suchergebnisliste Altavista.de - verkürzte Darstellung.....	111
Abb. 5.7.	Suchergebnisliste Google.de - verkürzte Darstellung.....	112
Abb. 6.1.	Darstellung von Framesets ohne DESCRIPTION-Tag bei Altavista	124
Abb. 6.2.	Darstellung des Title-Tags im Browser.....	131
Abb. 6.3.	Darstellung eines Titels in der Favoritenliste.....	132
Abb. 6.4.	Lage von Text in einer HTML-Tabelle.....	147
Abb. 6.5.	Beispiel einer Robots.txt-Datei.....	167
Abb. 7.1.	Datumsangabe der Indexierung bei Altavista.....	172
Abb. 7.2.	Datumsangabe der Indexierung bei Fireball - verkürzte Darstellung	173
Abb. 7.3.	Google Toolbar mit PageRank Information - verkürzte Darstellung	179
Abb. 7.4.	Zirkelbezug bei Link Popularity.....	181
Abb. 9.1.	Darstellung der anfragenden Suchmaschinen mittels Server Logs. . .	209
Abb. 9.2.	Listing Enhancement bei Altavista.....	223

Abb. 10.1.	Listing bei Overture.....	243
Abb. 10.2.	Darstellung des Payed Placement von Overture bei Altavista.....	246
Abb. 10.3.	Espotting Listing bei Yahoo.....	250
Abb. 10.4.	C-P-C Preis Darstellung bei QualiGo.....	254
Abb. 11.1.	Monatliche Seitenabrufe - Balkendiagramm.....	262
Abb. 11.2.	Monatliche Seitenabrufe, Besucheranzahl und Anfragen.....	263
Abb. 11.3.	Anfrageverteilung über Wochentage-Balkendiagramm.....	263
Abb. 11.4.	Tägliche Seitenabrufe und Anfragen.....	264
Abb. 11.5.	Anfragen nach Webseiten.....	265
Abb. 11.6.	Bericht der Top 50 Suchbegriffe.....	266
Abb. 11.7.	Verweisende Websites.....	267
Abb. 11.8.	Anfragen nach Suchmaschinen.....	268

Tabellenverzeichnis

Tabelle 1.1.	Portal-Partner von Overture in Deutschland.....	10
Tabelle 1.2.	Portal-Partner von Overture in USA.....	10
Tabelle 1.3.	Portal-Partner von Espotting in Deutschland.....	10
Tabelle 3.1.	Darstellung einer Datenbanktabelle.....	40
Tabelle 6.1.	Suchmaschinen und Meta-Tags.....	138
Tabelle 6.2.	Zulässige Textdateiformate bei Suchmaschinen.....	164
Tabelle 9.1.	Nutzerzahlen von Suchportalen in Deutschland.....	208
Tabelle 9.2.	Nutzerzahlen von Suchportalen in der Schweiz.....	208
Tabelle 9.3.	Kooperationen von Suchmaschinen und Webkatalogen in Deutschland.....	213
Tabelle 9.4.	Rooperationen von Suchmaschinen und Webkatalogen in USA ..	213
Tabelle 9.5.	Restriktionen bei der URL-Anmeldung bei Suchmaschinen	216
Tabelle 9.6.	Restriktionen bei der URL-Anmeldung bei Webkatalogen.....	216
Tabelle 7.7.	Landesindexe bei Altavista.....	222
Tabelle 9.8.	PartnerSite - Partner von FAST.....	223
Tabelle 9.9.	Portal-Partner von FAST.....	225
Tabelle 9.10	Portal-Partner von Inktomi.....	226
Tabelle 10.1	Portal-Partner von Google für AdWords Werbung.....	238
Tabelle 10.2	Portal-Partner von Overture in Deutschland.....	245
Tabelle 10.3	Portal-Partner von Overture in USA.....	245
Tabelle 10.4	Rangposition bei Overture Portal-Partnern.....	246
Tabelle 10.5	Partner von Espotting in Deutschland.....	248
Tabelle 10.6	Mindestkosten für die Aufnahme bei QualiGo.....	252
Tabelle 10.7	QualiGo Partner in der Schweiz.....	253
Tabelle 10.8	QualiGo Partner in Österreich.....	253
Tabelle 10.9	QualiGo Partner in Deutschland.....	253