

Suchmaschinen- Optimierung

Für Webentwickler

Inhalt

Vorwort	11
1 Suchen im Web	13
1.1 Webkataloge	15
1.1.1 Auswahl der Rubrik	17
1.1.2 Die Titelwahl	17
1.1.3 Vorsicht beim Beschreibungstext	19
1.1.4 Stichwörter mit Sorgfalt wählen	20
1.1.5 Häufige Fehler	20
1.1.6 Submit-Tools	20
1.2 Suchmaschinen	21
1.2.1 User-Interface	22
1.2.2 Hürden	23
1.2.3 Funktionen und Komponenten	24
1.3 Metasuchmaschinen	26
1.3.1 Formale Kriterien	27
1.3.2 Einsatzgebiet	28
1.3.3 Operatoren	29
1.3.4 Präsentation der Suchergebnisse	29
2 Die Anatomie des World Wide Web	33
2.1 Exkurs in HTML	34
2.1.1 HTML-Dokumentstruktur	35
2.1.2 Tags	36
2.1.3 Meta-Tags	38
2.1.4 Sonstige Meta-Tags	45
2.1.5 Cascading Style Sheets	47
2.2 Trägermedium Internet	49
2.2.1 Das Client-Server-Prinzip	50
2.2.2 TCP/IP	52
2.2.3 Adressierung der Hosts	53
2.2.4 Funktion und Aufbau eines URL	53
2.3 HTTP	55
2.3.1 Request	58
2.3.2 Response	61
2.3.3 HTTP live erleben	63

3 Architektur von Suchmaschinen 65

3.1 Dokumentgewinnung mit dem Webcrawler-System.....	66
3.1.1 Dokumentenindex.....	67
3.1.2 Scheduler.....	68
3.1.3 Crawler.....	70
3.1.4 Storeserver.....	72
3.1.5 Repository.....	77
3.2 Datenaufbereitung und Dokumentanalyse.....	78
3.2.1 Datenaufbereitung durch den Parser.....	82
3.2.2 Daten normal isierung.....	84
3.2.3 Wortidentifikation durch den Tokenizer.....	85
3.2.4 Identifikation der natürlichen Sprache.....	87
3.2.5 Grundformreduzierung durch Word Stemming.....	90
3.2.6 Mehrwortgruppenidentifikation.....	94
3.2.7 Stoppwörter.....	95
3.2.8 Keyword-Extrahierung.....	97
3.2.9 URL-Verarbeitung.....	101
3.3 Datenstruktur.....	101
3.3.1 Hitlist.....	102
3.3.2 Direkter Index.....	105
3.3.3 Invertierter Index.....	107
3.3.4 Verteilte Datenstruktur.....	108

4 Gewichtung und Relevanz 111

4.1 Statistische Modelle.....	113
4.1.1 Boolesches Retrieval.....	113
4.1.2 Fuzzy-Logik.....	114
4.1.3 Vektorraummodell.....	115
4.1.4 Relative Worthäufigkeit (TF).....	118
4.1.5 Inverse Dokumenthäufigkeit (IDF).....	119
4.1.6 Bedeutung der Lage und Auszeichnung eines Terms.....	120
4.1.7 Betrachtung des URL.....	121
4.2 Page-Rank.....	121
4.2.1 Link-Popularity.....	122
4.2.2 Das Page-Rank-Konzept und der Random Surfer.....	123
4.2.3 Page-Rank-Formel.....	124
4.2.4 Ein Beispiel zur Page-Rank-Berechnung.....	125
4.2.5 Effekte des Page-Rank.....	127
4.2.6 Der intelligente Surfer und weitere Einflussfaktoren.....	129
4.2.7 Bad-Rank.....	131
4.3 Click-Popularity.....	133
4.4 Cluster-Verfahren.....	136
4.4.1 Cluster-Verfahren im Einsatz.....	137
4.4.2 Vivisimo - ein Pionier.....	138
4.4.3 Single-Pass-Methode.....	139

5	Suchprozess	141
5.1	Arbeitsschritte des Query-Prozessors	142
5.1.1	Tokenizing.....	142
5.1.2	Parsing.....	142
5.1.3	Stoppwörter und Stemming.....	143
5.1.4	Erzeugung der Query.....	143
5.1.5	Verwendung eines Thesaurus.....	144
5.1.6	Matching und Gewichtung.....	144
5.1.7	Darstellung der Trefferliste.....	145
5.2	Suchoperatoren	146
5.2.1	Boolesche Ausdrücke.....	146
5.2.2	Phrasen.....	148
5.2.3	Wortabstand.....	148
5.2.4	Trunkierung.....	149
5.3	Erweiterte Suchmöglichkeiten	149
5.3.1	Sprachfilter.....	151
5.3.2	Positionierung.....	152
5.3.3	Aktualität.....	152
5.3.4	Domainfilter.....	152
5.3.5	Dateityp.....	153
5.3.6	Sonstige Suchmöglichkeiten.....	153
5.4	Nutzerverhalten im Web	154
5.4.1	Suchaktivitäten.....	155
5.4.2	Suchmodi.....	157
5.4.3	Welche Suchmaschine wird genutzt?.....	159
5.4.4	Was wird gesucht?.....	162
6	Onpage-Optimierung	167
6.1	Entwicklung eines Konzepts als erster Schritt	168
6.1.1	Zielgruppe und Zielsetzung.....	169
6.1.2	Durchführung.....	170
6.1.3	Spezielle Situation bei einem Relaunch.....	170
6.2	Strukturelle Vorbereitungen	172
6.2.1	Gültiges HTML.....	172
6.2.2	Einsatz von CSS.....	175
6.2.3	Seitenstruktur.....	176
6.2.4	Navigation.....	180
6.2.5	Frames.....	182
6.2.6	Die Startseite.....	189
6.2.7	Dateityp und dynamische Seiten.....	191
6.3	Schlüsselwort-Strategien	196
6.3.1	Erstes Brainstorming.....	199
6.3.2	Logbücher nutzen.....	200
6.3.3	Mitbewerber analysieren.....	200

6.3.4	Das Umfeld: Freunde, Kollegen und Bekannte.....	202
6.3.5	IDF überprüfen.....	203
6.3.6	Erste Bereinigung.....	204
6.3.7	Liste erweitern.....	205
6.3.8	Eigenschaften der Schlüsselwörter.....	208
6.3.9	Falsche orthografische Schreibweise.....	210
6.3.10	Getrennt oder zusammen?.....	211
6.3.11	Wortkombinationen und Wortnähe.....	212
6.3.12	Liste bereinigen.....	214
6.3.13	Finale Auswahl.....	215
6.4	Optimierung durch Tags.....	216
6.4.1	Title.....	216
6.4.2	Fließtext und die Keyword-Dichte.....	219
6.4.3	Aufzählungen.....	222
6.4.4	Texthervorhebungen.....	223
6.4.5	Überschriften.....	225
6.4.6	Links und Anchor-Text.....	226
6.4.7	Tabellen.....	229
6.4.8	Bilder und Image-Maps.....	231
6.4.9	Phantom-Pixel.....	232
6.4.10	Comment.....	233
6.4.11	Form und Input.....	234
6.4.12	Noscript.....	234
6.4.13	Iframe.....	235
6.5	PDF-Dokumente optimieren.....	237

7 Offpage-Optimierung 239

7.1	Webserver und Restriktionen.....	239
7.1.1	Webhosting.....	239
7.1.2	Restriktionen.....	241
7.2	Domainname und Verzeichnisse.....	242
7.2.1	Domainname.....	242
7.2.2	Verzeichnis- und Dateinamen.....	244
7.2.3	Verzeichnistiefe und Aktualität.....	246
7.3	Sitestructur.....	249
7.3.1	Redirects korrekt umsetzen.....	250
7.3.2	Deep Web.....	252
7.3.3	Seiten ausschließen (robots.txt).....	255
7.4	Link-Popularity erhöhen.....	257
7.4.1	Interne Verlinkung optimieren.....	258
7.4.2	Das KAKADU-Prinzip.....	258
7.4.3	Qualitätskriterien potenzieller Linkpartner.....	260
7.4.4	An andere Webautoren herantreten.....	261
7.4.5	Eingehende Links erzielen.....	262
7.4.6	Link-Farmen und Google-Bomben.....	265
7.5	Click-Popularity erhöhen.....	266

8	Spam	269
8.1	Keyword-Stuffing.....	270
8.2	Unsichtbare und kleine Texte.....	272
8.3	Hidden-Links.....	278
8.4	Meta-Spam.....	279
8.5	Doorway-Pages.....	280
8.6	Cloaking.....	283
8.7	Bait-And-Switch.....	285
8.8	Domain-Dubletten.....	286
8.9	Page-Jacking.....	288
8.10	Sonstige Spammethoden.....	289
9	Aufnahme in die Suchmaschine	291
9.1	Suchmaschinen-Kooperationen.....	291
9.2	Die Anmeldung.....	293
9.2.1	Manuelle Anmeldung.....	295
9.2.2	Automatische Anmeldung.....	298
9.2.3	Aufnahmedauer.....	299
9.3	Kostenpflichtige Leistungen.....	301
9.3.1	Payed-Inclusion-Programme.....	302
9.3.2	Pay-Per-Click (PPC).....	304
10	Monitoring und Controlling	309
10.1	Server-Monitoring.....	310
10.2	Logfile-Analyse.....	313
10.2.1	Anfragen pro Tag und Monat.....	315
10.2.2	Herkunftsland der Besucher.....	317
10.2.3	Seitenbesuche.....	317
10.2.4	Herkunft der Besucher.....	318
10.2.5	Besuche über Suchmaschinen.....	319
10.2.6	Suchbegriffe.....	320
10.2.7	Sonstige Informationen.....	321
10.3	Rank-Monitoring.....	322