
Xpert.press

Die Reihe **Xpert.press** vermittelt Professionals in den Bereichen Softwareentwicklung, Internettechnologie und IT-Management aktuell und kompetent relevantes Fachwissen über Technologien und Produkte zur Entwicklung und Anwendung moderner Informationstechnologien.

Weitere Bände in der Reihe <http://www.springer.com/series/4393>

Dirk Lewandowski

Suchmaschinen verstehen

2., überarbeitete und erweiterte Auflage

Dirk Lewandowski
Hochschule für Angewandte Wissenschaften Hamburg
Hamburg
Deutschland

ISSN 1439-5428 ISSN 2522-0667 (electronic)
Xpert.press
ISBN 978-3-662-56410-3 ISBN 978-3-662-56411-0 (eBook)
<https://doi.org/10.1007/978-3-662-56411-0>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer Vieweg

© Springer-Verlag Berlin Heidelberg 2018

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften. Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Vieweg ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

Vorwort zur zweiten Auflage

Wenn ein Buch eine zweite Auflage erlebt, so ist das ein Grund zur Freude. Zum einen natürlich für den Autor, denn offensichtlich war sein Buch so erfolgreich, dass eine Neubearbeitung gewünscht wurde. Zum anderen ist eine überarbeitete Neuauflage im positiven Fall auch eine Freude für die Leserinnen und Leser: Alles wurde auf den aktuellen Stand gebracht, wo nötig, wurde erweitert, Fehler wurden ausgemerzt und Verbesserungen vorgenommen.

Das Ziel dieser zweiten Auflage von *Suchmaschinen verstehen* ist, den Leserinnen und Lesern in dieser Hinsicht ein erfreuliches Erlebnis zu bieten:

- Das Buch wurde erheblich erweitert. Neben zahlreichen neu hinzugenommenen Abschnitten findet sich ein komplett neues Kapitel zur Suchmaschinenwerbung. Damit wird das Themenspektrum um ein weiteres bedeutendes (und abseits des Marketings viel zu wenig behandelten) Thema erweitert.
- An vielen Stellen habe ich den Text auf der Basis von Anregungen von Leserinnen und Lesern oder aus eigener Einsicht verbessert. Ich habe mich, wie schon bei der ersten Auflage, bemüht, den Text möglichst verständlich zu halten. In der neuen Auflage finden sich neben zahlreichen verbesserten Erklärungen auch neue Abbildungen, um die Themen noch besser verständlich zu machen.
- Alle im Text genannten Zahlen, beispielsweise zur Internet- und Suchmaschinennutzung, wurden anhand der neuesten Studien und Erhebungen auf den aktuellen Stand gebracht. Auch die Screenshots der Suchergebnisseiten wurden durchgehend aktualisiert. Teilweise haben sich in der Ergebnisdarstellung nur geringfügige Änderungen ergeben; der Anspruch war allerdings, dass das Buch durchgehend den aktuellen Stand zum Zeitpunkt der Manuskriptabgabe zeigt. Ebenso wurden alle im Text genannten URLs überprüft und auf den aktuellen Stand gebracht.
- Nicht zuletzt dient eine Neuauflage auch dazu, Fehler zu beseitigen. Natürlich wird auch diese zweite Auflage nicht frei von Fehlern sein, ich habe mich aber redlich (und mit hohem Zeitaufwand) bemüht, sie auf ein Minimum zu reduzieren.

Trotz aller Erweiterungen habe ich darauf geachtet, dass es sich weiterhin um ein einführendes Buch handeln soll. Ich war mir der Gefahr bewusst, das Manuskript immer weiter auszubauen, was zwar die Detailtiefe stärken, die Lesbarkeit und die kompakte Darstellung aber erheblich einschränken würde. So habe ich mich bemüht, der Versuchung zu widerstehen, vor allem zu meinen „Lieblingsthemen“ weitere umfangreiche Ausführungen zu ergänzen. Wer sich tiefer für das ein oder andere Thema interessiert, sei auf die Empfehlungen für weiterführende Literatur am Ende eines jeden Kapitels verwiesen.

Mein besonderer Dank gilt allen, die die Überarbeitung des Manuskripts durch ihre Anmerkungen zur ersten Auflage und durch zahlreiche Diskussionen unterstützt haben. Der erste, der sich mit einer umfangreichen Liste von Verbesserungsvorschlägen (und leider auch von Fehlern) gemeldet hatte, war Markus Günther. Viele weitere, die wertvolle Anregungen geliefert haben, folgten. Ich möchte mich bei ihnen allen herzlich bedanken.

Für die kontinuierliche Diskussion über die letzten Jahre hinweg und die Freude, die sie mir bei meiner Tätigkeit an der HAW Hamburg geben, bedanke ich mich bei dem kleinen Search-Studies-Team: Friederike Hanisch, Sebastian Sünkler und Christiane Behnert.

Nicht vergessen werden dürfen auch diejenigen, die bei der Erstellung der ersten Auflage maßgeblich geholfen haben und deren Arbeit sich natürlich auch in der aktuellen Version wiederfindet: Dorothee Wagner, Alexandra Linhart und Jenny Krieger. Vielen Dank!

Hamburg, im November 2017

Dirk Lewandowski

Inhaltsverzeichnis

1	Einführung	1
1.1	Die Bedeutung der Suchmaschinen	2
1.2	Ein Buch über Google?	6
1.3	Ziel dieses Buchs	7
1.4	Über Suchmaschinen sprechen	8
1.5	Aufbau des Buchs	8
1.6	Aufbau der Kapitel und Markierungen innerhalb des Texts	10
1.7	Zusammenfassung	11
	Literatur	11
2	Formen der Suche im Web	13
2.1	Suche nach einer Website vs. Suche nach Informationen zu einem Thema	13
2.2	Was ist ein Dokument?	15
2.3	Wo wird gesucht?	16
2.4	Zugänge zu den Informationen im World Wide Web	16
2.4.1	Suchmaschinen	17
2.4.2	Spezialsuchmaschinen	19
2.4.3	Metasuchmaschinen	22
2.4.4	Web-Verzeichnisse	23
2.4.5	Social-Bookmarking-Dienste	25
2.4.6	Frage-Antwort-Dienste	25
2.5	Zusammenfassung	26
	Literatur	27
3	Wie Suchmaschinen die Inhalte des Web erfassen und aufbereiten	29
3.1	Das World Wide Web und seine Erfassung durch Suchmaschinen	32
3.2	Content Acquisition	36
3.3	Web Crawling: Dokumente im Web aufspüren	37
3.3.1	Suchmaschinen steuern und ausschließen	42
3.3.2	Willentlicher Ausschluss von Inhalten durch die Suchmaschinenbetreiber	44

3.3.3	Aufbau der Datenbasis aus unterschiedlichen Kollektionen des Web . . .	46
3.3.4	Crawling für spezielle Kollektionen	48
3.4	Der Indexer: Dokumente für die Suche aufbereiten	49
3.4.1	Indexierung von Bildern, Audiodateien und Videos	55
3.4.2	Repräsentation von Web-Dokumenten in Suchmaschinen	56
3.5	Der Searcher: Suchanfragen verstehen	59
3.6	Zusammenfassung	62
	Literatur	64
4	Wie Suchmaschinen genutzt werden.	67
4.1	Der Suchprozess	67
4.2	Erfassung von Nutzungsdaten	69
4.3	Anfragetypen	70
4.4	Sessions	73
4.5	Suchanfragen	75
4.5.1	Eingabe der Suchanfragen	75
4.5.2	Suchvorschläge während der Eingabe	77
4.5.3	Formulierung der Suchanfragen	78
4.5.4	Länge der Suchanfragen	80
4.5.5	Verteilung der Suchanfragen nach Häufigkeiten	82
4.5.6	Suchanfragen-Trends	84
4.5.7	Verwendung von Operatoren und Befehlen zur gezielten Suche.	86
4.6	Themen	87
4.7	Zusammenfassung	88
	Literatur	90
5	Das Ranking der Suchergebnisse	93
5.1	Gruppen von Rankingfaktoren	95
5.2	Textstatistik	96
5.2.1	Ermittlung potentiell relevanter Dokumente	97
5.2.2	Berechnung von Häufigkeiten	98
5.2.3	Berücksichtigung von Strukturinformationen in Dokumenten	100
5.3	Popularität	102
5.3.1	Linktopologische Verfahren	104
5.3.2	Nutzungsstatistische Verfahren	109
5.4	Aktualität	117
5.5	Lokalität	119
5.6	Personalisierung der Suchergebnisse	123
5.7	Technische Rankingfaktoren	125
5.8	Ranking und Spam	126
5.9	Zusammenfassung	127
	Literatur	128

6 Die Präsentation der Suchergebnisse	131
6.1 Der Einfluss von Endgerät und Bildschirmauflösung	132
6.2 Der Aufbau der Suchergebnisseiten	133
6.3 Elemente auf den Suchergebnisseiten	140
6.3.1 Organische Ergebnisse	140
6.3.2 Werbung	141
6.3.3 Universal-Search-Ergebnisse	142
6.3.4 Knowledge-Graph-Ergebnisse	144
6.3.5 Direkte Antworten	146
6.3.6 Navigationselemente	147
6.3.7 Hilfestellungen zur Veränderung der Suchanfrage	147
6.3.8 Suchoptionen auf der Ergebnisseite	148
6.4 Der Aufbau der Trefferbeschreibungen	149
6.5 Optionen zu den einzelnen Ergebnissen	152
6.6 Auswahl von geeigneten Treffern auf den Suchergebnisseiten	154
6.7 Zusammenfassung	155
Literatur	156
7 Der Suchmaschinenmarkt	159
7.1 Das Geschäftsmodell der Suchmaschinen	159
7.2 Die Bedeutung der Suchmaschinen für die Online-Werbung	161
7.3 Marktanteile der Suchmaschinen	161
7.4 Bedeutende Suchmaschinen	163
7.5 Partnerschaften auf dem Suchmaschinenmarkt	164
7.6 Zusammenfassung	167
Literatur	168
8 Suchmaschinenoptimierung	171
8.1 Die Bedeutung der Suchmaschinenoptimierung	172
8.2 Grundlegende Verfahren der Suchmaschinenoptimierung	174
8.2.1 On-The-Page-Optimierung	176
8.2.2 Off-The-Page-Optimierung	180
8.3 Suchmaschinenoptimierung und Spam	182
8.4 Die Bedeutung von Rankingupdates	182
8.5 Suchmaschinenoptimierung für besondere Kollektionen	183
8.6 Die Position der Suchmaschinenbetreiber	184
8.7 Zusammenfassung	185
Literatur	186
9 Suchmaschinenwerbung	189
9.1 Besonderheiten der Suchmaschinenwerbung	193
9.2 Funktionsweise und Ranking	193
9.3 Unterscheidbarkeit von Werbung und organischen Suchergebnissen	195

9.4	Werbung innerhalb der Universal-Search-Ergebnisse	197
9.5	Zusammenfassung	198
	Literatur	199
10	Alternativen zu Google	201
10.1	Überschnidungen zwischen den Suchergebnissen verschiedener Suchmaschinen	202
10.2	Warum sollte man eine andere Suchmaschine als Google nutzen?	203
10.2.1	Einholen einer „zweiten Meinung“	204
10.2.2	Mehr bzw. zusätzliche Ergebnisse	205
10.2.3	Andere Ergebnisse	205
10.2.4	Bessere Ergebnisse	206
10.2.5	Andere Trefferpräsentation	206
10.2.6	Andere Benutzerführung	207
10.2.7	Andere Suchmöglichkeiten	207
10.3	Wann sollte man eine andere Suchmaschine als Google nutzen?	208
10.4	Besonderheiten bei Google aufgrund seiner Marktdominanz	209
10.5	Zusammenfassung	212
	Literatur	213
11	Besser suchen	215
11.1	Quellenauswahl	217
11.2	Auswahl passender Suchbegriffe	219
11.3	Boolesche Operatoren	219
11.4	Verbindung von Suchanfragen mit Booleschen Operatoren	223
11.5	Formulare für die erweiterte Suche	224
11.6	Befehle	226
11.7	Komplexe Suchen	229
11.8	Zusammenfassung	230
	Literatur	231
12	Die Qualität der Suchergebnisse	233
12.1	Kriterien für die Bewertung von Texten im Web	233
12.2	Menschliche vs. maschinelle Qualitätsüberprüfung	235
12.3	Wissenschaftliche Evaluierung der Suchergebnisqualität	239
12.3.1	Standardaufbau von Tests zur Retrievaleffektivität	242
12.3.2	Messung der Retrievaleffektivität mittels Klickdaten	245
12.3.3	Evaluierung im Interaktiven Information Retrieval	246
12.4	Zusammenfassung	248
	Literatur	249
13	Das Deep Web	251
13.1	Die Inhalt des Deep Web	253
13.2	Quellen vs. Inhalte von Quellen, Zugänglichkeit von Inhalten <i>via</i> Web	255

13.3 Die Größe des Deep Web	259
13.4 Bereiche des Deep Web	260
13.5 Welche Rolle spielt das Deep Web heute?	261
13.6 Zusammenfassung	261
Literatur	262
14 Spezialsuchmaschinen und ihre Recherchemöglichkeiten	263
14.1 Probleme der allgemeinen Suchmaschinen	264
14.2 Typen von Spezialsuchmaschinen	266
14.3 Nachrichtensuchmaschinen	268
14.4 Suchmaschinen für wissenschaftliche Inhalte	272
14.5 Faktensuchmaschinen	275
14.6 Einbindung von Spezialsuchmaschinen in die Universal Search	277
14.7 Zusammenfassung	279
Literatur	280
15 Suchmaschinen und Social Media	281
15.1 Typen von Social-Media-Angeboten	282
15.2 Social-Media-Angebote als Rechercheinstrumente.	284
15.3 Integration von Daten aus sozialen Netzwerken in Suchmaschinen	285
15.4 Zusammenfassung	287
Literatur	288
16 Suchmaschinen und ihre Rolle als Vermittler von Informationen	289
16.1 Die Interessen der Suchmaschinenbetreiber	290
16.2 Suchmaschinen-Bias	292
16.3 Auswirkungen des Search Engine Bias auf die Suchergebnisse	293
16.4 Interessengeleitete Präsentation von Suchergebnissen	296
16.5 Was würde „Suchneutralität“ bedeuten?	299
16.6 Zusammenfassung	300
Literatur	301
17 Die Zukunft der Suche	303
17.1 Suche als Basistechnologie	304
17.2 Der Wandel der Suchanfragen und der Dokumente.	305
17.3 Besseres Verstehen von Dokumenten und Suchanfragen	306
17.4 Die wirtschaftliche Zukunft der Suchmaschinen.	307
17.5 Die gesellschaftliche Zukunft der Suchmaschinen	308
17.6 Zusammenfassung	310
Literatur	311
Glossar	313
Stichwortverzeichnis.	327