

Dirk Lewandowski

Suchmaschinen verstehen

3. Auflage

 Springer Vieweg



Suchmaschinen verstehen

Dirk Lewandowski

Suchmaschinen verstehen

3., vollständig überarbeitete und erweiterte Auflage

Dirk Lewandowski
Hochschule für Angewandte Wissenschaften Hamburg
Hamburg, Deutschland

ISBN 978-3-662-63190-4 ISBN 978-3-662-63191-1 (eBook)
<https://doi.org/10.1007/978-3-662-63191-1>

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 GmbH Deutschland, ein Teil von Springer Nature 2015, 2018, 2021

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 allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

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.

Planung: Petra Steinmüller

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 dritten Auflage

Es ist mir eine Freude, dass dieses Buch nun bereits in der dritten Auflage vorliegt. Die zahlreichen Rückmeldungen von Leserinnen und Lesern der vergangenen Auflagen haben mir nicht nur gezeigt, dass ein Buch, welches einen Überblick über alle wesentlichen Facetten des Themas Suchmaschinen und Web-Suche gibt, nötig ist. Vielmehr habe ich durch die vielen Rückmeldungen auch zahlreiche Hinweise bekommen, welche Teile noch verständlicher gemacht werden sollten, wo ergänzt, und wo gestrichen werden sollte.

Diese neue Auflage wurde aber nicht nur in den Details überarbeitet, sondern inhaltlich etwas umstrukturiert, um den aktuellen Gegebenheiten gerecht zu werden: So wurde das Kapitel zu den Spezialsuchmaschinen bzw. zu den speziellen Kollektionen, die in die Web-Suche eingehen, gemäß der immer weiter zunehmenden Bedeutung dieser Kollektionen als nun sechstes Kapitel verschoben und vollständig neu strukturiert und überarbeitet. Das Kapitel knüpft nun direkt an das Ranking der allgemeinen Web-Inhalte an und führt auf das folgende Kapitel zur Präsentation der Suchergebnisse hin. Das Kapitel zu Social Media wurde gekürzt und in das Kapitel zum Deep Web integriert. Zwar ist die Bedeutung von Social Media bei weitem nicht zurückgegangen, allerdings hat die noch vor einigen Jahren von vieler Seite angenommene, starke Einbindung der bekannten Social-Media-Dienste in die Suchmaschinen nicht stattgefunden. Daher handelt es sich bei den Inhalten dieser Dienste nun (weiterhin) um Inhalte des Deep Web, die allerdings aufgrund ihrer Bedeutung weiterhin in einem eigenen Abschnitt behandelt werden.

Die Literatur zum Thema Suchmaschinen ist mittlerweile fast unüberschaubar; vieles ist allerdings zu spezialisiert, um in ein Überblickswerk aufgenommen zu werden. Die Verweise auf die Literatur wurden vielfach erheblich ergänzt; ich habe aber darauf geachtet, dass diesem Buch nicht das Schicksal vieler einführender Werke widerfährt, nämlich, dass es immer weiter anwächst, letztlich aber keine gut lesbare Einführung mehr darstellt.

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 geringfügigen 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.

Und nicht zuletzt wurde der Text hin zu einer gendergerechten Sprache angepasst. Leider gibt es noch keine Regeln für solche Anpassungen, die auch zu einem sprachlich befriedigenden Text führen. Daher wurde in manchen Fällen das generische Maskulinum belassen.

Mein besonderer Dank gilt allen, die die Überarbeitung des Manuskripts durch ihre Anmerkungen zu den vergangenen Auflagen und durch zahlreiche Diskussionen unterstützt haben. 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, vor allem bei Sebastian Sünkler und Sebastian Schultheiß. Bei der Überarbeitung des Manuskripts haben auch wieder studentische Hilfskräfte mitgeholfen. Bei Nina Niesche bedanke ich mich für die Überarbeitung der Screenshots und zahlreicher Schaubilder; bei Ngan Pham für die Überarbeitung der Literaturverzeichnisse. Nicht vergessen werden dürfen natürlich auch diejenigen, die bei der Erstellung der ersten Auflagen maßgeblich geholfen haben und deren Arbeit sich natürlich auch in der aktuellen Version wiederfindet: Dorothee Wagner, Alexandra Linhart und Jenny Peters. Vielen Dank!

Hamburg, Deutschland
Februar 2021

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	11
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	20
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
2.4.7 Soziale Netzwerke	26
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	33
3.2 Content Acquisition	36
3.3 Web Crawling: Dokumente im Web aufspüren	38
3.3.1 Suchmaschinen steuern und ausschließen	42
3.3.2 Ausschluss von Inhalten durch die Suchmaschinenbetreiber	45

3.3.3	Aufbau der Datenbasis und Crawling für spezielle Kollektionen	47
3.4	Der Indexer: Dokumente für die Suche aufbereiten	48
3.4.1	Indexierung von Bildern, Audiodateien und Videos	54
3.4.2	Repräsentation von Web-Dokumenten in Suchmaschinen	55
3.5	Der Searcher: Suchanfragen verstehen	58
3.6	Zusammenfassung	62
	Literatur	63
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	78
4.5.3	Formulierung der Suchanfragen	79
4.5.4	Länge der Suchanfragen	81
4.5.5	Verteilung der Suchanfragen nach Häufigkeiten	83
4.5.6	Suchanfragen-Trends	86
4.5.7	Verwendung von Operatoren und Befehlen zur gezielten Suche. . .	87
4.6	Themen	88
4.7	Zusammenfassung	89
	Literatur	90
5	Das Ranking der Suchergebnisse	93
5.1	Gruppen von Rankingfaktoren	95
5.2	Textstatistik	96
5.2.1	Ermittlung potenziell relevanter Dokumente	97
5.2.2	Berechnung von Häufigkeiten	99
5.2.3	Berücksichtigung von Strukturinformationen in Dokumenten . . .	101
5.3	Popularität	103
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	124
5.7	Technische Rankingfaktoren	127
5.8	Ranking und Spam	128
5.9	Zusammenfassung	130
	Literatur	131

6 Vertikale Suche: Spezielle Kollektionen und ihre Einbindung in die allgemeine Suche	135
6.1 Spezialsuchmaschinen als Basis der Universal Search	137
6.2 Typen von Spezialsuchmaschinen	140
6.3 Beispiele für Kollektionen	142
6.3.1 Nachrichten	143
6.3.2 Wissenschaftliche Inhalte	147
6.3.3 Bilder	150
6.3.4 Videos	151
6.4 Einbindung von Spezialsuchmaschinen in die Universal Search	152
6.5 Zusammenfassung	153
Literatur.	154
7 Die Präsentation der Suchergebnisse	157
7.1 Der Einfluss von Endgerät und Bildschirmauflösung	158
7.2 Der Aufbau der Suchergebnisseiten.	159
7.3 Elemente auf den Suchergebnisseiten	168
7.3.1 Organische Ergebnisse.	168
7.3.2 Werbung	169
7.3.3 Universal-Search-Ergebnisse	170
7.3.4 Knowledge-Graph-Ergebnisse	172
7.3.5 Direkte Antworten	172
7.3.6 Einbindung von Transaktionen auf der Suchergebnisseite	175
7.3.7 Navigationselemente	175
7.3.8 Hilfestellungen zur Veränderung der Suchanfrage	175
7.3.9 Suchoptionen auf der Ergebnisseite.	178
7.4 Der Aufbau der Trefferbeschreibungen	178
7.5 Optionen zu den einzelnen Ergebnissen	182
7.6 Auswahl von geeigneten Treffern auf den Suchergebnisseiten	184
7.7 Zusammenfassung	185
Literatur.	186
8 Der Suchmaschinenmarkt	189
8.1 Das Geschäftsmodell der Suchmaschinen	189
8.2 Die Bedeutung der Suchmaschinen für die Online-Werbung	191
8.3 Marktanteile der Suchmaschinen.	192
8.4 Bedeutende Suchmaschinen.	194
8.5 Partnerschaften auf dem Suchmaschinenmarkt	195
8.6 Zusammenfassung	197
Literatur.	198
9 Suchmaschinenoptimierung	201
9.1 Die Bedeutung der Suchmaschinenoptimierung	202

9.2	Grundlegende Verfahren der Suchmaschinenoptimierung.	204
9.2.1	Inhalte	207
9.2.2	Website-Architektur.	208
9.2.3	HTML	209
9.2.4	Vertrauen	209
9.2.5	Links	210
9.2.6	Personenbezogene Faktoren.	210
9.2.7	Schädliche Faktoren.	211
9.2.8	Spezielsuchen.	212
9.3	Suchmaschinenoptimierung und Spam	212
9.4	Die Bedeutung von Rankingupdates	213
9.5	Suchmaschinenoptimierung für besondere Kollektionen	214
9.6	Die Position der Suchmaschinenbetreiber	215
9.7	Zusammenfassung	216
	Literatur.	217
10	Suchmaschinenwerbung.	219
10.1	Besonderheiten der Suchmaschinenwerbung.	222
10.2	Funktionsweise und Ranking	224
10.3	Unterscheidbarkeit von Werbung und organischen Suchergebnissen	226
10.4	Werbung innerhalb der Universal-Search-Ergebnisse	227
10.5	Zusammenfassung	228
	Literatur.	229
11	Alternativen zu Google.	231
11.1	Überschneidungen zwischen den Suchergebnissen verschiedener Suchmaschinen	232
11.2	<i>Warum</i> sollte man eine andere Suchmaschine als Google nutzen?	233
11.2.1	Einholen einer „zweiten Meinung“	233
11.2.2	Mehr bzw. zusätzliche Ergebnisse	234
11.2.3	Andere Ergebnisse	235
11.2.4	Bessere Ergebnisse.	236
11.2.5	Andere Trefferpräsentation	236
11.2.6	Andere Benutzerführung	236
11.2.7	Verhinderung der Bildung von Nutzerprofilen	237
11.2.8	Andere Suchmöglichkeiten	237
11.3	<i>Wann</i> sollte man eine andere Suchmaschine als Google nutzen?	237
11.4	Besonderheiten bei Google aufgrund seiner Marktdominanz	239
11.5	Zusammenfassung	242
	Literatur.	243

12	Besser suchen	245
12.1	Quellenauswahl	247
12.2	Auswahl passender Suchbegriffe	249
12.3	Boolesche Operatoren	249
12.4	Verbindung von Suchanfragen mit Booleschen Operatoren	253
12.5	Formulare für die erweiterte Suche	254
12.6	Befehle	256
12.7	Komplexe Suchen	259
12.8	Zusammenfassung	260
	Literatur	261
13	Die Qualität der Suchergebnisse	263
13.1	Kriterien für die Bewertung von Texten im Web	263
13.2	Menschliche vs. maschinelle Qualitätsüberprüfung	266
13.3	Wissenschaftliche Evaluierung der Suchergebnisqualität	268
13.3.1	Standardaufbau von Tests zur Retrievaleffektivität	271
13.3.2	Messung der Retrievaleffektivität mittels Klickdaten	274
13.3.3	Evaluierung im Interaktiven Information Retrieval	276
13.4	Zusammenfassung	277
	Literatur	278
14	Das Deep Web	281
14.1	Die Inhalte des Deep Web	284
14.2	Quellen vs. Inhalte von Quellen, Zugänglichkeit von Inhalten <i>via</i> Web	285
14.3	Die Größe des Deep Web	289
14.4	Bereiche des Deep Web	290
14.5	Social Media als Inhalt des Deep Web	291
14.6	Welche Rolle spielt das Deep Web heute?	293
14.7	Zusammenfassung	294
	Literatur	294
15	Suchmaschinen und ihre Rolle als Vermittler von Informationen	297
15.1	Die Interessen der Suchmaschinenbetreiber	298
15.2	Suchmaschinen-Bias	300
15.3	Auswirkungen des Search Engine Bias auf die Suchergebnisse	301
15.4	Interessengeleitete Präsentation von Suchergebnissen	304
15.5	Was würde „Suchneutralität“ bedeuten?	308
15.6	Zusammenfassung	309
	Literatur	310
16	Die Zukunft der Suche	313
16.1	Suche als Basistechnologie	314
16.2	Der Wandel der Suchanfragen und der Dokumente	315
16.3	Besseres Verstehen von Dokumenten und Suchanfragen	316

16.4	Die wirtschaftliche Zukunft der Suchmaschinen	317
16.5	Die gesellschaftliche Zukunft der Suchmaschinen	318
16.6	Zusammenfassung	320
	Literatur	321
Glossar	323
Stichwortverzeichnis	337

Abbildungsverzeichnis

Abb. 2.1	Startseite des Firefox-Browsers mit Adresszeile und Google-Suche als voreingestellter Startseite (25.11.2020)	14
Abb. 2.2	Inhalte von Suchmaschinen	17
Abb. 2.3	Startseite der Suchmaschine AltaVista (1996); https://web.archive.org/web/1996_1023234631/http://altavista.digital.com/	19
Abb. 2.4	Inhalte von Spezialsuchmaschinen	21
Abb. 2.5	Inhalte von Hybridsuchmaschinen.	22
Abb. 2.6	Inhalte von Metasuchmaschinen	23
Abb. 2.7	Inhalte von Web-Verzeichnissen	24
Abb. 3.1	Grundlegender Aufbau einer Suchmaschine (aus Risvik und Michelsen 2002, S. 290)	31
Abb. 3.2	Die Struktur des Web („Bow-tie model“; Broder et al. 2000, S. 318).	35
Abb. 3.3	Strukturierte Darstellung von Produktinformationen vs. unstrukturierte Informationen in (HTML-)Dokumenten	37
Abb. 3.4	Hinweis auf die Cache-Kopie auf der Suchergebnisseite (Beispiel Google; 25.11.2020)	42
Abb. 3.5	Anzeige einer Cache-Kopie (Beispiel Google; 25.11.2020)	43
Abb. 3.6	Ausschnitte aus einer robots.txt-Datei (Beispiel Google.de (Ausschnitt); 26. November 2020)	44
Abb. 3.7	Gezielter Ausschluss von bestimmten Crawlern (Beispiel faz.net (Ausschnitt); 27. Januar 2021).	45
Abb. 3.8	Hinweis auf einen aufgrund des „Recht auf Vergessenwerden“ ausgeschlossenen Inhalt (Beispiel von Google; 30. November 2020).	46
Abb. 3.9	Buchregister als ein Beispiel für einen invertierten Index (Ausschnitt aus Lewandowski 2009).	49
Abb. 3.10	Einfacher invertierter Index	51
Abb. 3.11	Invertierter Index mit Angaben zu Worthäufigkeiten.	52
Abb. 3.12	Invertierter Index mit Positionsangaben	53
Abb. 3.13	Bild mit Umgebungstexten	54

Abb. 3.14	Repräsentation von Web-Dokumenten nach Bereichen und Feldern (beispielhaft)	56
Abb. 3.15	Zeitliche Stellung einer Suchanfrage (Lewandowski 2011, S. 63)	59
Abb. 3.16	Implizite Anfrageinterpretation (Beispiel Google (Ausschnitt); 26.11.2020)	61
Abb. 3.17	Explizite Anfrageinterpretation (Beispiel Wolfram Alpha; 26.11.2020)	61
Abb. 4.1	Der Suchprozess in der Websuche (Lewandowski 2012, S. 104)	68
Abb. 4.2	Verarbeitung von unterschiedlichen Arten von Suchanfragen und Ausgabe unterschiedlicher Arten von Dokumenten (beispielhaft)	76
Abb. 4.3	Suchvorschläge während der Eingabe (Beispiel Google; 26. November 2020)	78
Abb. 4.4	Korrekturvorschlag (oben) und durchgeführte Korrektur (unten) auf der Suchergebnisseite (Beispiel Google)	80
Abb. 4.5	Typische Verteilung der Suchanfragen nach ihrer Länge in Wörtern (nach Spink et al. 2001, S. 230)	82
Abb. 4.6	Verteilung von Suchanfragehäufigkeiten (beispielhaft)	84
Abb. 4.7	Wiederkehrendes Ereignis (Suchbegriffe „weihnachten“ und „ostern“; Grafik aus Google Trends; 3. Dezember 2020)	85
Abb. 4.8	Impuls (Suchbegriff „fukushima“; Grafik aus Google Trends; 3. Dezember 2020)	85
Abb. 4.9	Evergreen (Suchbegriff „sex“; Grafik aus Google Trends)	85
Abb. 4.10	Suchbegriff mit zwei Spitzen im Suchverlauf pro Jahr (Suchbegriff „basteln“; Grafik aus Google Trends; 29. Januar 2021)	86
Abb. 4.11	Themenverteilung bei Suchanfragen (Daten aus Lewandowski 2006)	89
Abb. 5.1	Bedeutung der Textstatistik für das Ranking	98
Abb. 5.2.	(Quasi-)Synonyme in der Trefferliste (Beispiel Google; Suchanfrage Samstag hannover; 6. Januar 2021)	98
Abb. 5.3	Wortverteilungen in Dokumenten (Luhn 1958, S. 161)	100
Abb. 5.4	Erster Treffer für die Suchanfrage „Paris Hilton“ (Beispiel Google; 3. Dezember 2020)	101
Abb. 5.5	Erster Treffer für die Suchanfrage „Hilton Paris“ (Beispiel Google; 3. Dezember 2020)	102
Abb. 5.6	In-Links/Out-Links (Lewandowski 2005, S. 118)	104
Abb. 5.7	Darstellung von Dokumenten mit Gewichtungen innerhalb des Linkgraphen (aus Page et al. 1999, S. 4)	107
Abb. 5.8	Auswertung der Klicks auf der Suchergebnisseite (fiktives Beispiel)	111
Abb. 5.9	Ausschnitt aus einem automatisch erstellten, nutzerspezifischen Themenprofil bei Google (6. Dezember 2020)	113
Abb. 5.10	Ebenen der Suchanfragenanalyse (aus Lewandowski 2011, S. 62)	114
Abb. 5.11	Für die Suchmaschine erfassbare Daten im Suchverlauf.	115
Abb. 5.12	Ausschnitt aus dem Webprotokoll von Google (30. November 2020)	116

Abb. 5.13 Die Google.com-Startseite am 11. September 2001 (http://blogoscoped.com/files/google-on-911-large.png ; 6. Januar 2021)	118
Abb. 5.14 Erste Ergebnisse zur Suchanfrage bundesrat auf Google Deutschland (Google.de; 31. Januar 202).	120
Abb. 5.15 Erste Ergebnisse zur Suchanfrage bundesrat auf Google Schweiz (Google.ch; 31. Januar 2021).	121
Abb. 5.16 Suchergebnisseite (Ausschnitt) in der Desktopsuche von Google (Suchanfrage restaurant; 3. Dezember 2020)	122
Abb. 5.17 Suchergebnisseite (Ausschnitt) in der mobilen Suche von Google (Suchanfrage restaurant; 3. Dezember 2020)	123
Abb. 5.18 Original (links) und mit Werbung angereicherte Kopie eines Wikipedia-Artikels (7. Januar 2021)	129
Abb. 6.1 Aufbau einer Suchmaschine: Verschiedene Kollektionen („vertikale Suchmaschinen“) – beispielhaft ohne Anspruch auf Vollständigkeit	136
Abb. 6.2 Übersicht über archivierte Versionen einer URL in der Wayback Machine (4. Dezember 2020)	141
Abb. 6.3 Zusammenstellung von Nachrichten zu einem Thema mittels Topic Detection and Tracking (Beispiel Google News, 22. Januar 2021).	145
Abb. 6.4 Startseite von Google News (4. Dezember 2020)	146
Abb. 6.5 Beispiel einer Fehlzuordnung bei Google Scholar: Autorenangabe (6. Februar 2021)	148
Abb. 6.6 Beispiel einer Fehlzuordnung bei Google Scholar: Titelangabe (6. Februar 2021)	149
Abb. 6.7 Darstellung von Ergebnissen in Google Scholar	149
Abb. 6.8 Wege von der Suchanfrage zur Nachricht	152
Abb. 7.1 Suchergebnisseite von Google (sichtbarer Bereich) in verschiedenen Versionen (beispielhaft; v.l.n.r. Desktop, iPad quer, iPad hoch, Android-Smartphone hoch; 06.12.2020)	159
Abb. 7.2 Suchergebnisseite (Beispiel Google; 6. Dezember 2020)	160
Abb. 7.3 Suchergebnisseite mit Werbung (Beispiel Google; 1. Februar 2021)	161
Abb. 7.4 Beantwortung einer Faktenfrage auf der Suchergebnisseite (Beispiel Google; 6. Dezember 2020).	162
Abb. 7.5 Einbindung eines Faktencontainers auf der Suchergebnisseite (Beispiel Google; 3. Dezember 2020).	163
Abb. 7.6 Anzeige von verwandten Fragen auf der Suchergebnisseite (Beispiel Google; 8. Januar 2021).	164
Abb. 7.7 Schematische Darstellungen der wichtigsten Layouts von Suchergebnisseiten.	164
Abb. 7.8 Typische Blickhäufungen bei einer listenbasierten Trefferdarstellung	166
Abb. 7.9 Blickhäufungen auf einer Suchergebnisseite mit Universal-Search-Ergebnissen	167

Abb. 7.10 Platzierung des News-Containers nach Nachrichtenaufkommen (Ausschnitte von Google-Suchergebnisseiten vom 09.01–11.01.2021) . . .	171
Abb. 7.11 Knowledge-Graph-Ergebnisse (Beispiel Google; 3. Dezember 2020)	173
Abb. 7.12 Direkte Antwort auf eine Suchanfrage (Beispiel Google; 3. Dezember 2020)	174
Abb. 7.13 In die Suchergebnisseite integrierte Hotelbuchung (Beispiel Google; 1. Februar 2021).	176
Abb. 7.14 Vorschläge zur Verbesserung der Suchanfrage (Beispiel Google; 3. Dezember 2020).	177
Abb. 7.15 Vorschläge zur Veränderung der Suchanfrage im Knowledge Graph (Beispiel Google; 3. Dezember 2020)	177
Abb. 7.16 Vorschläge für weiterführende Fragen (Beispiel Google; 1. Februar 2021).	178
Abb. 7.17 Suchoptionen auf der Suchergebnisseite bei Google (6. Januar 2021)	179
Abb. 7.18 Trefferbeschreibung eines organischen Treffers (Google; 6. Dezember 2020)	180
Abb. 7.19 Erweiterte Trefferbeschreibung mit Sitelinks (Google; 6. Dezember 2020).	181
Abb. 7.20 Aktualitätshinweis in der Trefferbeschreibung (6. Dezember 2020).	181
Abb. 7.21 Autorenangabe, Zitationshinweis und Link auf eine maschinelle Übersetzung in der Trefferbeschreibung (6. Dezember 2020).	181
Abb. 7.22 Navigationspfad und Nutzerbewertung in der Trefferbeschreibung (6. Dezember 2020)	182
Abb. 7.23 Trefferbeschreibung mit aufgeklapptem Kontextmenü (16. Oktober 2017).	183
Abb. 7.24 Anzeige ähnlicher Websites auf Basis einer bereits bekannten Site (8. Januar 2021)	183
Abb. 8.1 Unterschiedliche Ergebnisdarstellung und -anreicherung auf Basis der gleichen organischen Ergebnisse (Yahoo, Bing; 15. Januar 2021)	196
Abb. 8.2 Das Beziehungsgeflecht der Suchmaschinen (Deutschland).	197
Abb. 9.1 The Periodic Table of SEO Success Factors (http://searchengineland.com/seotable ; 12. Januar 2021)	206
Abb. 10.1 Suchergebnisseite mit organischen Ergebnissen und Anzeigen (Desk- top; Ausschnitt, 4. Dezember 2020).	221
Abb. 10.2 Trefferbeschreibung eines Werbetreffers (oben) und eines organischen Treffers (unten) für die gleiche Website zur gleichen Suchanfrage (Beispiel Google; 6. Januar 2021).	222
Abb. 10.3 Suchergebnisseite mit organischen Ergebnissen und Anzeigen (Mobil- telefon; 4. Dezember 2020)	223
Abb. 11.1 Mögliche Strategien bei Unzufriedenheit mit Suchergebnissen	238

Abb. 12.1 Verbesserung der Ergebnisqualität durch elaborierte Suche (Stock 2007, S. 64)	247
Abb. 12.2 Mengendiagramme zur Verdeutlichung der Wirkung der Booleschen Operatoren	250
Abb. 12.3 Formular für die erweiterte Suche (Beispiel Google; https://www.google.de/advanced_search ; 6. Januar 2021).	255
Abb. 12.4 Erweiterte Bildersuche bei Google (https://www.google.de/advanced_image_search ; Ausschnitt, 4. Dezember 2020)	256
Abb. 13.1 Wikipedia-Artikel zu „Männerrock“ mit Warnhinweis (Ausschnitt; 4. Februar 2021).	268
Abb. 13.2 Precisiongraph aus einer Suchmaschinen-Evaluierung (Lewandowski 2015, S. 1771)	269
Abb. 13.3 Vergleich zweier Suchmaschinen für navigationsorientierte Suchanfragen (Lewandowski 2015, S. 1769).	270
Abb. 14.1 Schematische Darstellung des Oberflächenweb (Bergman 2001).	282
Abb. 14.2 Schematische Darstellung des Deep Web (Bergman 2001).	282
Abb. 14.3 Unterscheidung zwischen einer Website und einer Invisible-Web-Quelle	286
Abb. 14.4 Taxonomie der Online-Informationen (Stock 2003, S. 27)	286
Abb. 14.5 Taxonomie der Online-Informationen (Stock 2003, S. 27), ergänzt um Beispieldienste	288
Abb. 14.6 Trefferbeschreibung einer Quelle des Proprietary Web (Beispiel Google; 11. November 2017)	291
Abb. 15.1 Suchergebnisseite von Google für die Suchanfrage „Sportlerinnen“ (Ausschnitt; 6. Januar 2021)	302
Abb. 15.2 Klickverteilung auf einer Universal-Search-Suchergebnisseite (Beispiel 1) ohne Unterlegung; (Lewandowski und Sünkler 2013).	306
Abb. 15.3 Klickverteilung auf einer Universal-Search-Suchergebnisseite (Beispiel 2) mit grauer Unterlegung des mittleren Containers; (Lewandowski und Sünkler 2013)	307

Tabellenverzeichnis

Tab. 4.1	Unterscheidung zwischen konkretem und problemorientiertem Informationsbedarf (angepasst nach Frants et al. 1997, S. 38, deutsch aus Stock 2007, S. 51 f.)	71
Tab. 4.2	Anfrageintentionen in der Websuche ^a	72
Tab. 4.3	Anteil der Suchanfragen nach Anfragetyp (aus Lewandowski 2014, S. 48)	73
Tab. 4.4	Suchanfragen nach Segmenten von jeweils 10 Prozent (modifiziert aus Lewandowski 2015).	85
Tab. 6.1	Kollektionen am Beispiel von Google (Auswahl; angepasst aus Lewandowski 2013, S. 499).	143
Tab. 12.1	Bewertbarkeit des Sucherfolgs nach Anfragetyp (aus Lewandowski 2014, S. 46 f.)	246
Tab. 12.2	Wichtige Suchbefehle bei Google und Bing*	257
Tab. 13.1	CARS-Checkliste (übersetzt aus Harris 2020)	265
Tab. 13.2	Vergleich der Testverfahren (Lewandowski 2011, S. 224)	277



In diesem Buch geht es darum, die Suchwerkzeuge, die wir selbst jeden Tag benutzen, besser zu verstehen. Erst wenn wir grundlegend verstehen, wie Suchmaschinen aufgebaut sind und wie sie funktionieren, können wir sie effektiv in unseren Recherchen einsetzen.

Aber nicht nur der Einsatz bestehender Suchmaschinen ist hier von Relevanz, sondern es soll auch darum gehen, was wir von den bekannten Suchmaschinen wie Google lernen können, wenn wir eigene Suchsysteme aufbauen möchten. Der Ausgangspunkt ist dabei, dass die Suchmaschinen des World Wide Web zurzeit die technisch führenden Systeme sind, die sowohl hinsichtlich des Suchprozesses als auch hinsichtlich des Nutzerverhaltens die Vorgaben machen. Das bedeutet, dass wir uns, wenn wir selbst Suchsysteme aufbauen, an den durch die Web-Suchmaschinen geprägten Gewohnheiten orientieren müssen, ob wir das wollen oder nicht.

Dieses Buch ist der Versuch, das Thema Suchmaschinen umfassend zu behandeln im Sinne einer Betrachtung aus unterschiedlichen Blickwinkeln:

1. Technik: Zuerst einmal sind Suchmaschinen technische Systeme. Hierbei geht es um die Erfassung der Inhalte des Web sowie um das Ranking und die Präsentation der Suchergebnisse.
2. Nutzung: Suchmaschinen werden nicht nur von ihren Entwicklern, sondern auch von ihren Nutzenden geprägt. Da die bei der Nutzung anfallenden Daten wiederum in das Ranking der Suchergebnisse und die Gestaltung der Benutzerführung eingehen, hat die Nutzung einen großen Einfluss darauf, wie die Suchmaschinen gestaltet sind.
3. Recherche: Zwar werden Suchmaschinen in den weit überwiegenden Fällen auf recht simple Weise genutzt – und viel mehr ist oft ja auch für eine erfolgreiche Recherche gar nicht nötig –, allerdings sind Suchmaschinen auch Werkzeuge zur professionellen Recherche nach Informationen. Dass Suchmaschinen leicht für jedermann zu bedienen sind, bedeutet nicht, dass man auch jede Rechercheaufgabe leicht mit ihnen lösen kann.

4. Wirtschaft: Suchmaschinen haben eine große Bedeutung für Inhaltenanbieter, die ihre Inhalte „an den Mann“ bringen wollen. Dadurch, dass sie zentrale Knoten im Netz sind, spielen sie auch wirtschaftlich eine bedeutende Rolle. In diesem Bereich geht es vor allem um die Sichtbarkeit in den Suchmaschinen, die durch verschiedene Maßnahmen des Online Marketings (wie Suchmaschinenoptimierung und das Schalten von Anzeigen) erreicht werden kann.
5. Gesellschaft: Da Suchmaschinen das bevorzugte Mittel zur Informationssuche sind und täglich massenhaft genutzt werden, haben sie auch eine enorme Bedeutung für den Wissenserwerb in der Gesellschaft. Hier stellt sich unter anderem die Frage nach der Glaubwürdigkeit der Suchergebnisse und der Rolle der Suchmaschinen bei der Verbreitung von Misinformationen und Desinformationen, was häufig unter dem Label „Fake News“ behandelt wird.

Die grundlegende These dabei ist, dass das eine ohne das andere nicht möglich ist: Wir können Suchmaschinen nicht als technische Systeme begreifen, wenn wir nicht um ihre gesellschaftliche Bedeutung wissen. Ebenso wenig können wir ihre gesellschaftlichen Auswirkungen verstehen, wenn wir nicht die zugrunde liegende Technik kennen. Natürlich muss man nicht in allen Bereichen das gleiche Detailwissen haben; eine solide Basis sollte man aber auf jeden Fall erreichen.

Natürlich können die genannten Themen in einem Einführungsbuch nicht allumfassend behandelt werden. Es geht mir vielmehr darum, die für die Diskussion über Suchmaschinen zentralen Konzepte vorzustellen und das grundlegende Wissen zu vermitteln, das eine fundierte Diskussion über Suchmaschinen überhaupt erst möglich macht.

1.1 Die Bedeutung der Suchmaschinen

Ich vertrete in diesem Buch die Auffassung, dass Suchmaschinen eine enorme gesellschaftliche Bedeutung haben. Diese lässt sich einerseits aus ihrer massenhaften Nutzung und andererseits aus der Reihung und Darstellung der Suchergebnisse erklären.

Suchmaschinen werden (wie andere Dienste des Internet) massenhaft genutzt. Ihre Bedeutung liegt dabei darin, dass wir sie nutzen, um *aktiv* nach Informationen zu suchen. Mit jeder Eingabe einer Suchanfrage offenbaren wir unsere Interessen, und mit jeder Suchergebnisseite, die uns eine Suchmaschine zurückgibt, findet eine (technisch vermittelte) Interpretation sowohl der Suchanfrage als auch der Menge der gefundenen und potenziell relevanten Ergebnisse statt. Dadurch, dass eine Suchmaschine diese Interpretationen auf eine bestimmte Weise durchführt, vermittelt sie ein bestimmtes Bild der Informationswelt des World Wide Web.

Zu jeder Suchanfrage wird eine Ergebnisseite angezeigt, auf der die Ergebnisse in einer bestimmten Anordnung angezeigt werden. Wir können zwar theoretisch aus allen Ergebnissen auswählen, doch verlassen wir uns doch stark auf die von der Suchmaschine

vorgegebene Reihung. De facto wählen wir also nicht aus den unter Umständen Millionen von gefundenen Treffern aus, sondern nur aus den wenigen zuerst angezeigten.

Zieht man dies in Betracht, so ergeben sich gesellschaftliche Fragen wie die nach der Vielfalt auf dem Suchmaschinenmarkt: Ist es in Ordnung, wenn eigentlich nur eine einzige Suchmaschine verwendet wird und diese uns zu jeder Suchanfrage nur eine einzige von vielen möglichen Sichten auf das Informationsuniversum bietet?

Die Bedeutung der Suchmaschinen wurde schon in schlagkräftige Titel wie „Google-Gesellschaft“ (so der Titel eines Buchs von 2005; Lehmann und Schetsche 2005), der „Society of the Query“ (so der Titel einer Konferenzreihe und eines Buchs; König und Rasch 2014) und „Die Googleisierung der Informationssuche“ (Stark et al. 2014) gebracht. Vielleicht muss man nicht gleich so weit gehen, Suchmaschinen (oder die Suchanfragen) als bestimmenden Faktor unserer Gesellschaft auszurufen; die enorme Bedeutung der Suchmaschinen für unseren Wissenserwerb lässt sich allerdings nicht mehr abstreiten.

Wenn wir die nüchternen Zahlen betrachten, sehen wir zunächst einmal, dass Suchmaschinen der beliebteste Dienst des Internet sind. Das Internet betrachten wir hier als eine Ansammlung von Protokollen und Diensten, zu denen beispielsweise auch E-Mail, Chat und das File Transfer Protocol (FTP) gehören. Erstaunlich ist nun, dass die Nutzung von Suchmaschinen an der Spitze steht, wenn man Nutzende nach ihren Aktivitäten im Internet fragt. Dabei sind Suchmaschinen sogar noch beliebter als das Schreiben und Lesen von E-Mails (Koch und Frees 2016). 76 Prozent aller Deutschen nutzen mindestens einmal pro Woche eine Suchmaschine, aber „nur“ 65 Prozent lesen bzw. schreiben in dieser Zeit mindestens eine E-Mail. Diese Daten stammen aus der ARD/ZDF-Onlinestudie (Beisch und Schäfer 2020), welche jährlich die Nutzung des Internet in der deutschen Bevölkerung abfragt. Vergleichbare Untersuchungen bestätigen die hohe Nutzungsfrequenz von Suchmaschinen: So zeigt die Eurobarometer Studie (European Commission 2016), dass in Deutschland 85 Prozent aller Internetnutzer mindestens einmal in der Woche eine Suchmaschine nutzen; bei der täglichen Nutzung sind es immerhin noch 48 Prozent. Deutschland liegt damit unter dem Durchschnitt der EU-Länder (88 bzw. 57 Prozent).

Sieht man sich in der ARD/ZDF-Onlinestudie an, welche Angebote im Internet sonst noch besonders häufig genutzt werden, so sind dies neben E-Mail- und Suchmaschinen die Messenger (wohl vor allem WhatsApp). Social-Media-Dienste erreichen dagegen nur 36 Prozent.

Eine zweite Betrachtungsweise ist der Blick auf die beliebtesten Websites (Alexa.com 2021). Hier steht Google an erster und dritter Stelle (google.com und google.de), gefolgt von YouTube, Amazon und Ebay. Es ist auffällig, dass nicht nur Google auf Platz 1 steht, sondern mit Ebay und Amazon auch zwei E-Commerce-Unternehmen vertreten sind, die zwar auch zahlreiche Möglichkeiten zum Stöbern bieten, aber natürlich auch bei der (Produkt-)Suche eine große Rolle spielen.

Dass Suchmaschinen ein Massenphänomen sind, lässt sich auch anhand der Zahl der täglich an sie gestellten Suchanfragen zeigen. Marktforschungsunternehmen schätzen die Zahl der allein an Google gestellten Suchanfragen auf etwa 3,3 Billionen im Jahr 2016

(Internet Live Stats und Statistic Brain Research Institute 2017) – das sind mehr als eine Million Suchanfragen pro Sekunde!

Noch eine weitere Betrachtungsebene ergibt sich, wenn wir uns ansehen, wie Nutzende im World Wide Web an Informationen gelangen. Zwar gibt es theoretisch viele Zugänge zu den Informationen im Netz, doch Suchmaschinen sind der weit bedeutendste. Auf der einen Seite lassen sich Webseiten natürlich direkt aufrufen, indem man die Adresse (*Uniform Resource Locator; URL*) in die Browserzeile eingibt. Und dann gibt es noch andere Dienste wie zum Beispiel Social-Media-Angebote, die uns zu Webseiten führen. Doch keiner dieser Dienste hat eine den Suchmaschinen vergleichbare Bedeutung für den Informationszugang im Web erreicht, und es ist auch nicht abzusehen, dass sich diese Situation in absehbarer Zeit ändern wird.

Nicht zuletzt haben Suchmaschinen eine große Bedeutung auch aufgrund des Online-Werbemarkts. Die Vermarktung von Anzeigen in Suchmaschinen (Anzeigen als Antwort auf eine Suchanfrage) hat hier einen Anteil von 40 Prozent des Markts (Zenith 2021); in Deutschland wurden mit Anzeigen in Suchmaschinen im Jahr 2019 4,1 Milliarden Euro umgesetzt (Statista 2021).

Diese Werbeform ist vor allem so attraktiv, weil Nutzende mit jeder ihrer Suchanfragen verraten, was sie finden möchten – und damit auch, ob und was sie eventuell kaufen möchten. So können Werbetreibende leicht entscheiden, wann sie einem Nutzer tatsächlich ihr Produkt anbieten möchten. Streuverluste, also der Anteil der Nutzenden, der zwar eine Werbung sieht, jedoch in dem Moment überhaupt kein Interesse daran hat, lassen sich auf diese Weise erheblich reduzieren bzw. gar vermeiden.

Suchmaschinenbetreiber müssen, wie andere Unternehmen auch, Geld verdienen. Das einzige Modell, mit dem sich Suchmaschinen bislang tragen, ist die Einblendung von Werbung, die in Form von Textanzeigen rund um die Suchergebnisse platziert ist. Andere Erlösmodelle haben sich nicht durchgesetzt. Das bedeutet auch, dass Suchmaschinenanbieter ihre Suchmaschinen nicht, wie oft behauptet wird, allein an den Ansprüchen und Bedürfnissen der Nutzenden ausrichten, sondern natürlich auch an ihren eigenen Gewinnabsichten und denen ihrer Werbekunden.

Für Unternehmen ergibt sich die Bedeutung der Suchmaschinen aber nicht nur aufgrund der Möglichkeit, Suchmaschinen als Werbepattform zu nutzen, sondern auch durch die Möglichkeit, von Nutzenden in den regulären Suchergebnissen gefunden zu werden. Die Verfahren, die dazu dienen, die Wahrscheinlichkeit dieses Auffindens zu erhöhen, werden unter dem Titel Suchmaschinenoptimierung zusammengefasst.

Wir sehen also schon an dieser Stelle, dass wir es, wenn wir Suchmaschinen nicht nur als technische Systeme, sondern als gesellschaftlich relevant betrachten, mit mindestens vier Interessengruppen bzw. Akteursverbänden (vgl. Röhle 2010, S. 14) zu tun haben:

1. Suchmaschinenbetreiber: Das Interesse der Suchmaschinenbetreiber liegt auf der einen Seite darin, ihre Nutzenden zufriedenzustellen. Dabei geht es einerseits um die Qualität der Suchergebnisse, andererseits um das Nutzererleben. Das zweite große (oder gar

- größere?) Interesse der Suchmaschinenbetreiber liegt darin, ihren Werbekunden ein attraktives Umfeld zu bieten und mit der Werbung möglichst viel Geld zu verdienen.
2. Nutzende: Das Interesse der Nutzenden liegt darin, mit geringem Aufwand zufriedenstellende Suchergebnisse zu erhalten und in ihrem Suchprozess nicht zu sehr gestört zu werden, beispielsweise durch aufdringliche Werbung.
 3. Inhaltenanbieter: Wer Inhalte im Web anbietet, möchte auch von (potenziellen) Nutzenden gefunden werden. Allerdings besteht ein weiteres Interesse vieler Inhaltenanbieter auch darin, mit ihren Inhalten Geld zu verdienen. Das wiederum bedeutet, dass es nicht unbedingt in ihrem Interesse liegt, ihre Inhalte den Suchmaschinen *vollständig* zur Verfügung zu stellen.
 4. Suchmaschinenoptimierer: Suchmaschinenoptimierer sorgen im Auftrag von Inhaltenanbietern dafür, dass Angebote im Web auffindbar werden; dies in erster Linie in Suchmaschinen. Durch ihr Wissen über die Rankingverfahren der Suchmaschinen und dessen Ausnutzung zur Platzierung „ihrer“ Websites beeinflussen sie wiederum die Suchmaschinenbetreiber, die sich vor Manipulationen schützen möchten.

Schon aus dieser knappen Erläuterung der Akteure wird deutlich, dass das Zusammenspiel zu Konflikten führen kann. Suchmaschinenbetreiber müssen abwägen zwischen den Interessen ihrer Nutzenden und ihrer Werbekunden; Suchmaschinenoptimierer müssen für eine bestmögliche Sichtbarkeit der Angebote ihrer Kunden sorgen, dürfen ihr Wissen über die Funktionsweise der Suchmaschinen aber nicht so weit ausnutzen, dass sie von den Suchmaschinenbetreibern wegen Manipulation abgestraft werden.

Wir sehen, dass wir es auf dem Suchmaschinenmarkt mit komplexen Interaktionen zu tun haben. Nur, wenn wir Suchmaschinen aus unterschiedlichen Blickwinkeln betrachten, können wir diese Interaktionen einordnen und verstehen, warum Suchmaschinen so gestaltet sind, wie sie es nun einmal sind. Suchmaschinen müssen den Ansprüchen der verschiedenen Nutzergruppen gerecht werden; es reicht nicht aus, wenn sie sich auf eine dieser Gruppen beschränken.

Wenn über Suchmaschinen gesprochen wird und ihre Bedeutung für den Zugang zu Informationen, dann werden meist nur die Inhalte, die originär für das Web produziert wurden, berücksichtigt. Doch schon seit Jahren bemühen sich die Suchmaschinen, auch Inhalte aus der „echten“, also der physischen Welt, zu erfassen und in ihre Suchsysteme zu integrieren. Vaidhyanathan (2011) unterscheidet drei Bereiche von Inhalten, die Suchmaschinen wie Google erfassen:

1. *Scan and link*: Fremde Inhalte werden erfasst, aggregiert und zur Suche zur Verfügung gestellt (Beispiel: Websuche).
2. *Host and serve*: Von den Nutzenden selbst erstellte Inhalte werden auf der eigenen Plattform gesammelt und gehostet (Beispiel: Youtube).
3. *Scan and serve*: Dinge aus der echten Welt werden vom Suchmaschinenbetreiber in die digitale Welt überführt (Beispiele: Google Books, Google Street View)

Vaidhyanathan (2011) fasst dies unter „The Googleization of Everything“ (so auch der Titel seines Buchs) zusammen und verdeutlicht damit nicht nur, dass die Inhalte der Suchmaschinen deutlich über die Inhalte des Web hinausgehen (auch wenn diese weiterhin die Basis bilden), sondern auch, dass wir in Bezug auf die Entwicklung der Suchmaschinen immer noch am Anfang stehen: Bislang ist nur ein kleiner Teil all der Informationen, die für Suchmaschinen von Interesse sind, überhaupt digitalisiert und damit für die Suche zugänglich.

Und es gibt noch eine zweite, weitgehend als selbstverständlich angenommene Annahme, nämlich, dass ein Suchvorgang notwendigerweise eine vom Nutzer eingegebene Suchanfrage enthalten müsse. Allerdings sehen wir, dass Suchmaschinen Suchanfragen zunehmend selbst generieren können, indem sie das Verhalten eines Nutzers beobachten und dann Informationen anbieten, die dieser Nutzer mit einer hohen Wahrscheinlichkeit gebrauchen kann. Wenn ein Nutzer beispielsweise mit seinem Smartphone in der Tasche durch eine Stadt spaziert, dann ist es ein leichtes, zur Mittagszeit seinen Wunsch nach einer Essensmöglichkeit vorherzusehen und auf Basis der aus der Vergangenheit bekannten Vorlieben dieses Nutzers sowie seines aktuellen Standorts ein Restaurant vorzuschlagen. Um dies zu tun, ist zwar eine Suchanfrage (die sich aus den genannten Informationen zusammensetzt) vonnöten, der Nutzer muss diese aber nicht selbst eingeben. Wir werden im Kap. 4 darauf zurückkommen.

1.2 Ein Buch über Google?

Denkt man an Suchmaschinen, so denkt man in erster Linie an Google. Wir alle benutzen diese Suchmaschine nahezu jeden Tag, und meist auch für alle möglichen Recherchezwecke. Auch hier sprechen die Zahlen eine deutliche Sprache: In Deutschland werden weit mehr als 90 Prozent aller Anfragen an allgemeine Suchmaschinen an Google gestellt, andere Suchmaschinen spielen nur eine untergeordnete Rolle (Statcounter 2021).

Daher setzt dieses Buch an der Alltagserfahrung mit Google an und versucht anhand dieses bekannten Beispiels den Aufbau und die Verwendung von Suchmaschinen zu erklären. Allerdings geht der Anspruch dieses Buchs weiter: Es geht auch darum zu zeigen, welche Alternativen zu Google es gibt und wann es sich lohnt, diese zu verwenden. Eine Beschreibung aller möglicher Suchmaschinen wird sich in diesem Buch aber nicht finden; vielmehr geht es darum, anhand von Beispielen andere Suchmaschinen vorzustellen und so den Leser erst einmal auf den Gedanken zu bringen, vor komplexeren Recherchen zu überlegen, ob Google überhaupt die beste Suchmaschine für genau diese Recherche ist.

Ein Stück weit kann man auch sagen, dass man, wenn man eine Suchmaschine kennt, mit allen anderen auch besser wird umgehen können. Wir werden den grundlegenden Aufbau von Suchmaschinen und ihre wichtigsten Funktionen anhand der Suchmaschine, die wir alle bereits zumindest von der Nutzerseite her kennen, kennenlernen. Das erworbene Wissen lässt sich dann leicht auf andere Suchmaschinen übertragen.

Auch die meisten Recherchebeispiele und Screenshots stammen von Google. Die Beispiele lassen sich allerdings in den allermeisten Fällen auf andere Suchmaschinen übertragen. Wo dies nicht der Fall ist, wird gesondert darauf hingewiesen.

Zu der Ähnlichkeit zwischen den verschiedenen Suchmaschinen kann man generell sagen, dass sich die Google-Konkurrenten in einem Dilemma befinden: Auch, wenn sie selbst innovative Funktionen anbieten und versuchen, Dinge anders als Google zu machen, so orientieren sie sich doch grundlegend an der von Google geprägten Vorstellung davon, wie eine Suchmaschine auszusehen und zu funktionieren hat. Diese Orientierung an Google ist den anderen Suchmaschinenanbietern nicht anzulasten, denn einerseits können sie nur Nutzende gewinnen, wenn sich die an Google gewöhnten Nutzenden sofort zurechtfinden; andererseits müssen sie sich von Google unterscheiden, um überhaupt einen Mehrwert gegenüber dieser Suchmaschine darstellen zu können.

1.3 Ziel dieses Buchs

Naturgemäß ist dieses Buch in seiner Funktion als Einführungsbuch beschränkt, und es versteht sich als Überblickswerk. Das bedeutet auch, dass viele Themen nicht im Detail behandelt werden können, sondern wir naturgemäß „an der Oberfläche“ bleiben müssen. Das heißt aber nicht, dass die Inhalte deshalb „flach“ sein müssen. Ich habe mich bemüht, die Inhalte so einfach wie möglich darzustellen, ohne es dabei aber an der erforderlichen Genauigkeit fehlen zu lassen. Manche Themen habe ich exemplarisch herausgegriffen, d. h. an einem Beispiel (wie etwa einer Spezialsuchmaschine) wird etwas ausführlicher erläutert, was dann auf andere Sachverhalte übertragen werden kann.

Überhaupt geht es in diesem Buch um eine Übertragungsleistung: Das, was man anhand einer bzw. einiger Suchmaschinen lernt, sollte man auf andere übertragen können. Daher macht es auch nichts, dass einige der in diesem Buch gezeigten Inhalte – vor allem, wenn es um Details bei einer bestimmten Suchmaschine geht – sich bereits bei Erscheinen dieses Buchs schon wieder verändert haben können. Gerade in sich schnell verändernden Bereichen ist dies unvermeidlich; das Ziel ist es aber, grundsätzliches Wissen über Suchmaschinen zu vermitteln, das dann auf alle Suchmaschinen übertragbar ist.

Dieses Buch ersetzt keine Einführungswerke beispielsweise in das Information Retrieval oder die Recherche in Suchmaschinen, auch wenn Themen aus diesen Bereichen behandelt werden. Die relevante Einführungsliteratur zu den entsprechenden Themen wird im jeweils passenden Kapitel genannt. Das Ziel dieses Buchs ist der Überblick und eine Betrachtung verschiedener Sichtweisen auf Suchmaschinen, nicht die allumfassende Darstellung der einzelnen Themenbereiche.

Gerade bei Studierenden besteht oft die Angst, das Thema Suchmaschinen könne man nur verstehen, wenn man sich detailliert in Algorithmen und technische Details einarbeite. In diesem Buch sollen zwar die wesentlichen Verfahren knapp und verständlich beschrieben werden, das Ziel ist aber vor allem, die den technischen Verfahren zugrunde liegenden Ideen zu verstehen. Damit kann man einschätzen, warum Suchmaschinen so gut oder

schlecht funktionieren, wie sie das zurzeit tun und welche Perspektiven ihrer Weiterentwicklung es gibt.

Es ist nur natürlich, dass bei jedem Versuch, ein Thema aus unterschiedlichen Perspektiven zu betrachten, man doch die Brille des eigenen Fachs aufhat und bei den Interessen des eigenen Fachs Schwerpunkte setzt. So folgen mein Interesse und der Schwerpunkt meiner Betrachtung naturgemäß dem Themenfeld und den Methoden der Informationswissenschaft, die technische Informationssysteme immer (auch) aus der Perspektive des Menschen betrachtet. Ich habe mich allerdings bemüht, auch die Perspektive anderer Fächer wie der Informatik und der Medien- und Kommunikationswissenschaft (auch in ihrer Literatur) zu berücksichtigen.

1.4 Über Suchmaschinen sprechen

Um über einen Gegenstand zu sprechen, bedarf es eines einheitlichen Vokabulars. Man muss wissen, dass man, wenn man bestimmte Begriffe verwendet, über das gleiche spricht. Um nicht aneinander vorbeizureden, muss man sich daher auf eine Terminologie einigen. Da im Bereich der Suchmaschinen bislang eine solche anerkannte Terminologie noch nicht existiert, und etwa Suchmaschinenoptimierer, Informationswissenschaftler und Kommunikationswissenschaftler jeweils eine „eigene Sprache“ sprechen, soll mit diesem Buch auch ein Beitrag zur Verständigung geleistet werden. So findet sich am Ende auch ein Glossar, das alle wichtigen Begriffe noch einmal in alphabetischer Reihenfolge auflistet und erklärt. Ich habe mich bemüht, dort auch synonyme und ähnliche Bezeichnungen zu berücksichtigen, damit Leser, die bereits Vorkenntnisse aus der Literatur gewonnen haben, „ihre“ Begriffe auch wiederfinden und sich schnell an die von mir verwendete Terminologie gewöhnen können.

1.5 Aufbau des Buchs

Man kann dieses Buch natürlich von vorne bis hinten durchlesen, und das ist auch meine primäre Intention beim Schreiben gewesen. Wenn man sich allerdings nur zu einem bestimmten Themenbereich informieren möchte, so ist dies aufgrund der Kapitelstruktur auch möglich.

Nach diesem einführenden Kapitel beginnt das Buch mit einem Kapitel über die Formen der Suche im Web. Natürlich sind Suchmaschinen wie Google nicht die einzige Form des Zugangs zu den Informationen im Web, auch wenn sich die Form der „algorithmischen Universalsuchmaschine“ weitgehend durchgesetzt hat. Die verschiedenen Formen von Suchsystemen werden kurz vorgestellt und in ihrer Bedeutung in den Kontext der Recherche im Web eingeordnet.

In Kap. 3 wird dann der grundlegende technische Aufbau von algorithmischen Suchmaschinen erklärt. Dabei wird erläutert, wie die Suchmaschinen an die Inhalte aus dem

Web gelangen, wie diese so aufbereitet werden, dass sie effizient durchsuchbar sind, und wie die Suchanfragen der Nutzenden automatisch interpretiert und verarbeitet werden können.

Nach diesen beiden technischen Kapiteln betrachten wir dann die Nutzerseite (Kap. 4): Was wird eigentlich in Suchmaschinen gesucht, wie werden Suchanfragen formuliert und wie wählen die Nutzenden die für sie passenden Ergebnisse aus?

Eng damit verbunden sind die Verfahren des Rankings, also der Anordnung der Suchergebnisse. In Kap. 5 werden die grundlegenden Verfahren beschrieben und in ihrer Bedeutung eingeordnet. Zwar wird immer wieder behauptet, das Ranking der Suchergebnisse wäre das große Geheimnis jeder Suchmaschine, doch mit einer Kenntnis der wichtigsten Rankingfaktoren kann man die Zusammenstellung der Suchergebnisse zumindest grundlegend erklären, auch wenn die konkrete Reihung von einer Vielzahl von Gewichtungen abhängig ist, die im Detail nicht nachvollzogen werden können. Dieses Verständnis wiederum kann uns sowohl bei der Recherche als auch bei der Aufbereitung unserer eigenen Inhalte für Suchmaschinen oder auch der Erstellung von eigenen Informationssystemen helfen.

Kap. 6 zeigt dann, wie die Suchergebnisse aus dem allgemeinen Web-Index durch die Hinzunahme von speziellen Kollektionen wie Nachrichten, Bildern und Videos erweitert werden. Dazu haben die bekannten Suchmaschinen zahlreiche Spezialsuchmaschinen aufgebaut, deren Ergebnisse auf den Suchergebnisseiten der allgemeinen Web-Suche angezeigt werden.

Kap. 7 widmet sich der Präsentation der Suchergebnisse. Seit einigen Jahren sind die bekannten Suchmaschinen von der üblichen Listenform der Suchergebnisdarstellung abgewichen und haben mit Konzepten wie *Universal Search* und *Knowledge Graph* neue Formen der Zusammenstellung von Suchergebnissen etabliert. Dadurch wurden nicht nur die Ergebnisseiten attraktiver gemacht und die Auswahlmöglichkeiten auf diesen Seiten erhöht, sondern die Suchmaschinen steuern mit dieser Art der Ergebnispräsentation auch gezielt die Aufmerksamkeit der Nutzenden.

Dies bringt uns zu den wirtschaftlichen Gegebenheiten, die mit den Suchmaschinen verbunden sind. In Kap. 8 beschäftigen wir uns mit dem Suchmaschinenmarkt und damit u. a. mit der Frage, wie es Google gelungen ist, den Suchmaschinenmarkt (zumindest in Europa) fast vollständig zu beherrschen. Natürlich klingt auch hier schon die Frage an, ob eine solche Situation wünschenswert ist und wie sie gegebenenfalls zu ändern wäre.

Kap. 9 widmet sich der Seite derjenigen, die ihre Inhalte über Suchmaschinen optimal bereitstellen möchten bzw. deren Helfern, den Suchmaschinenoptimierern. Diese verwenden ihre Kenntnisse über die Indexierungs- und Rankingverfahren der Suchmaschinen, um Inhalte besser auffindbar zu machen und ihren Kunden Traffic, also Besucher, zuzuführen. Die dazu verwendeten Techniken reichen von der einfachen Veränderung von Texten bis hin zu komplexen Verfahren, die die Verlinkungsstruktur im Web berücksichtigen.

In Kap. 10 wird die in Suchmaschinen angezeigte Werbung ausführlich beschrieben. Dabei geht es zum einen um die Betrachtung der auf den Suchergebnisseiten gezeigten Anzeigen als einem Typ von Suchergebnissen, andererseits aber auch um die Frage,

inwieweit diese Anzeigen von den Nutzenden von den eigentlichen Suchergebnissen unterschieden werden können.

Mit den tatsächlich bestehenden Alternativen zu Google beschäftigt sich dann Kap. 11. Zuerst einmal gilt es, die Frage zu beantworten, was eine Suchmaschine überhaupt zu einer alternativen Suchmaschine macht. Reicht es aus, dass es sich einfach um eine andere Suchmaschine als Google handelt? Dann wird basierend auf grundsätzlichen Erwägungen und konkreten Situationen im Rechercheprozess dargestellt, in welchen Fällen sich der Wechsel zu einer anderen Suchmaschine lohnt.

In Kap. 12 wechseln wir wieder die Perspektive und betrachten Suchmaschinen diesmal als Instrumente für eine fortgeschrittene Internetrecherche. Im Kapitel über das Nutzerverhalten wurde klar, dass die meisten Nutzenden wenig Energie in die Formulierung ihrer Suchanfragen und die Sichtung der Treffer stecken. Nun wollen wir sehen, mit welchen Strategien und mit welchen Befehlen sich das Beste aus den Suchmaschinen herausholen lässt.

Auch ein Thema der Recherche, aber ebenso der generellen Bewertung von Suchmaschinen, ist die Frage nach der Qualität der Suchergebnisse, der sich Kap. 13 widmet. Die Qualität der Suchergebnisse lässt sich dabei aus zwei unterschiedlichen Perspektiven betrachten: Einmal geht es um die individuelle Trefferbewertung durch den Nutzer im Lauf seiner Recherche; das andere Mal um den wissenschaftlichen Vergleich der Trefferqualität unterschiedlicher Suchmaschinen.

Kap. 14 beschäftigt sich mit den Inhalten des Web, die für die allgemeinen Suchmaschinen nicht zugänglich sind, dem sog. *Deep Web*. Dort findet sich ein wahrer Schatz an Informationen, die sich mit Google und ähnlichen Suchmaschinen nicht oder nur eingeschränkt auffinden lassen. Wir werden sehen, aus welchen Gründen diese Inhalte den Suchmaschinen verborgen bleiben und mit welchen Methoden wir dennoch an sie gelangen können.

Während sich die vorangegangenen Themen mit Aspekten aus den Bereichen Technik, Nutzung und Recherche beschäftigten, geht es in Kap. 15 um die gesellschaftliche Rolle von Suchmaschinen. Welche Rolle spielen Suchmaschinen heute für den Wissenserwerb und welche Rolle sollten sie spielen?

Das abschließende Kap. 16 beschäftigt sich mit der Zukunft der Suche. Ein Buch wie dieses kann immer nur eine Momentaufnahme bieten, und vor zehn Jahren hätte es ein anderes Bild geliefert als heute. Das „Problem“ der Suche ist allerdings keineswegs gelöst (und wird vielleicht auch nie gelöst werden), daher lohnt es sich, die heutigen Suchmaschinen nicht nur in ihrer Entwicklung hin zum gegenwärtigen Stand zu betrachten, sondern auch einen Blick in die (nähere) Zukunft zu wagen.

1.6 Aufbau der Kapitel und Markierungen innerhalb des Texts

Es liegt in der Natur der Sache, dass Kapitel zu unterschiedlichen Themen auch verschieden aufgebaut sein müssen. Gewisse Gemeinsamkeiten haben die Kapitel in diesem Buch dennoch: Am Anfang steht jeweils eine Einleitung, die das Themenfeld absteckt und kurz auf die Bedeutung des Themas innerhalb des Buchs eingeht. Darauf folgen dann die ausführlichen Erläuterungen; am Ende eines jeden Kapitels steht eine Zusammenfassung, die die wichtigsten Punkte wiederholt. Ebenfalls am Ende jedes Kapitels findet sich ein Literaturverzeichnis sowie in einem Kasten kommentierte Hinweise auf weiterführende Literatur für diejenigen, die sich tiefer in das jeweilige Thema einarbeiten möchten.

Ebenfalls in Kästen finden sich innerhalb des Texts Beispiele, die das im Haupttext Gesagte illustrieren und vertiefen, für das Verständnis des Haupttexts aber nicht unbedingt erforderlich sind.

1.7 Zusammenfassung

Suchmaschinen sind bedeutende Recherchewerkzeuge für den Zugang zu Informationen im World Wide Web. Sie werden in diesem Buch aus den Perspektiven der Technik, der Nutzung, ihrer wirtschaftlichen Aspekte, der Recherche und der Gesellschaft betrachtet.

Die Bedeutung der Suchmaschinen ergibt sich aus ihrer massenhaften Nutzung und dadurch, dass sie der bei weitem bevorzugte Zugang zu den Informationen im World Wide Web sind. Genutzt wird für die Recherche allerdings vor allem eine Suchmaschine, Google.

Suchmaschinen sind nicht nur als technische Systeme zu betrachten. Aufgrund des Zusammenspiels unterschiedlicher Akteure (Suchmaschinenbetreiber, Nutzende, Inhaltanbieter und Suchmaschinenoptimierer) ergeben sich umfangreiche Einflüsse auf die Suchergebnisse, die nicht allein von den Suchmaschinenbetreibern gesteuert werden.

Inhaltlich erfassen Suchmaschinen nicht mehr nur die Inhalte des Webs, sondern bieten auch Plattformen, auf denen Nutzende selbst Inhalte erstellen können, welche dann durchsuchbar gemacht werden. Weiterhin bieten die Suchmaschinenbetreiber neben der Websuche verschiedene Spezialsuchmaschinen an, deren Ergebnisse aber wiederum in die allgemeinen Suchergebnisseiten eingehen. Zuletzt kommen noch Inhalte aus der physischen Welt hinzu, die in die digitale Welt übertragen und in die Suche eingebunden werden.

Literatur

AdEx Benchmark 2016. (2017). <https://www.iab.it/wp-content/uploads/2017/08/IABEurope-AdEx-Benchmark-2016-full-report.pdf>. Zugegriffen am 17.01.2021.

Alexa.com. (2021). Top sites in Germany. <https://www.alexa.com/topsites/countries/DE>. Zugegriffen am 02.02.2021.

- Beisch, N., & Schäfer, C. (2020). Ergebnisse der ARD/ZDF-Onlinestudie 2020: Internetnutzung mit großer Dynamik: Medien, Kommunikation, Social Media. *Media Perspektiven*, 51(9), 462–481.
- Bundesverband Digitale Wirtschaft. (Hrsg.). (2013). *OVK Online-Report 2013/02: Zahlen und Trends im Überblick*. Düsseldorf. https://www.bvdw.org/presseserver/bvdw_ovk_online_report_2013_02/report_ovk_report_2013_02.pdf. Zugegriffen am 17.01.2021.
- ComScore. (2013). Future in focus: Digitales Deutschland 2013. <https://www.comscore.com/ger/Insights/Presentations-and-Whitepapers/2013/2013-Future-in-Focus-Digitales-Deutschland>. Zugegriffen am 17.01.2021.
- European Commission. (2016). *Special Eurobarometer 447 – Online platforms*. European Commission. <https://doi.org/10.2759/937517>.
- Internet Live Stats, & Statistic Brain Research Institute. (2017). Anzahl der Suchanfragen bei Google weltweit in den Jahren 2000 bis 2016 (in Milliarden). *Statista – Das Statistik-Portal*. <https://de.statista.com/statistik/daten/studie/71769/umfrage/anzahl-der-google-suchanfragen-pro-jahr/>. Zugegriffen am 17.01.2021.
- Koch, W., & Frees, B. (2016). Dynamische Entwicklung bei mobiler Internetnutzung sowie Audios und Videos – Ergebnisse der ARD/ZDF-Onlinestudie 2016. *Media Perspektiven*, 47(9), 418–437. https://www.ard-zdf-onlinestudie.de/files/2016/0916_Koch_Frees.pdf. Zugegriffen am 17.01.2021.
- König, R., & Rasch, M. (Hrsg.). (2014). *Society of the query reader: Reflecons on web search*. Amsterdam: Institute of Network Cultures.
- Lehmann, K., & Schetsche, M. (Hrsg.). (2005). *Die Google-Gesellschaft: Vom digitalen Wandel des Wissens*. Transcript.
- Röhle, T. (2010). *Der Google-Komplex: Über Macht im Zeitalter des Internets*. Transcript. <https://doi.org/10.14361/transcript.9783839414781>.
- Stark, B., Dörr, D., & Aufenanger, S. (Hrsg.). (2014). *Die Googleisierung der Informationssuche – Suchmaschinen zwischen Nutzung und Regulierung*. Berlin/Boston: de Gruyter. <https://doi.org/10.1515/9783110338218>.
- Statcounter. (2021). Search engine market share. <https://gs.statcounter.com/search-engine-market-share>. Zugegriffen am 27.01.2021.
- Statista. (2021). Prognose der Umsätze mit Suchmaschinenwerbung in Deutschland in den Jahren 2017 bis 2025 (in Millionen Euro). *Statista – Das Statistik-Portal*. <https://de.statista.com/prognosen/456188/umsaetze-mit-suchmaschinenwerbung-in-deutschland>. Zugegriffen am 02.02.2021.
- Vaidhyanathan, S. (2011). *The Googlization of Everything (and why we should worry)*. University of California Press. <https://doi.org/10.1525/9780520948693>.
- Zenith. (2021). Prognose zu den Investitionen in Internetwerbung weltweit in den Jahren 2018 bis 2022 nach Segmenten (in Milliarden US-Dollar). *Statista – Das Statistik-Portal*. <https://de.statista.com/statistik/daten/studie/209291/umfrage/investitionen-in-internetwerbung-weltweit-nach-segmenten/>. Zugegriffen am 26.01.2021.