

Xpert.press

Dirk Lewandowski

Suchmaschinen verstehen

2. Auflage

 Springer Vieweg

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



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 Nutzern 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 Inhalteanbieter, die ihre Inhalte „an den Mann“ bringen wollen. Hier 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. Dadurch, dass sie zentrale Knoten im Netz sind, spielen sie auch wirtschaftlich eine bedeutende Rolle.

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 potentiell 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 Nutzer nach ihren Aktivitäten im Internet fragt. Dabei sind Suchmaschinen sogar noch beliebter als das Schreiben und Lesen von E-Mails (Koch und Frees 2016). 83 Prozent aller deutschen Internetnutzer nutzen mindestens einmal pro Woche eine Suchmaschine, aber „nur“ 81 Prozent der Internetnutzer lesen bzw. schreiben in dieser Zeit mindestens eine E-Mail. Diese Daten stammen aus der ARD/ZDF-Onlinestudie (Koch und Frees 2016), welche jährlich die Nutzung des Internet in der deutschen Bevölkerung abfragt. Ähnliche Ergebnisse finden sich aber auch in vergleichbaren Studien (u. a. AGOF 2015); die Ergebnisse gleichen denen aus anderen Ländern (zum Beispiel Purcell et al. 2012, für die USA).

Sieht man sich in der ARD/ZDF-Onlinestudie an, welche Angebote im Internet sonst noch besonders häufig genutzt werden, so gibt es neben E-Mail- und Suchmaschinen-nutzung nur eine einzige Kategorie, welche auch auf mehr als 70 Prozent der Nutzer kommt, nämlich „Informationen suchen“ (77 Prozent der Nutzer tun dies mindestens einmal wöchentlich). Dabei fällt auf, dass es sich wieder um einen informationsorientierten Zweck handelt. Alle anderen Anwendungen im Netz kommen auf deutlich weniger als 50 Prozent bei der wöchentlichen Nutzung. Auch die Nutzung von Apps, Onlinecommunities, Videoportalen und Chats liegt in diesem Bereich.

Eine zweite Betrachtungsweise ist der Blick auf die beliebtesten Websites (Alexa.com 2017). Hier steht Google an erster Stelle, gefolgt von YouTube, Facebook, Amazon und Ebay.¹ 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

¹ Alexa.com weist die Zahlen für Google.de und Google.com gesondert aus, weshalb Google die Positionen 1 und 3 belegt.

(Internet Live Stats, & Statistic Brain Research Institute o.D.) – das sind mehr als eine Million Suchanfragen pro Sekunde!

Noch eine weitere Betrachtungsebene ergibt sich, wenn wir uns ansehen, wie Nutzer 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. Mittlerweile werden für Werbung im Internet allein in Deutschland jährlich etwa 6 Milliarden Euro ausgegeben. Davon entfallen mehr als 3 Milliarden Euro, also über die Hälfte, auf Textanzeigen in Suchmaschinen (AdEx Benchmark 2016 2017).

Diese Werbeform ist vor allem so attraktiv, weil Nutzer 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 Nutzer, 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 Nutzer 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 Werbeplattform zu nutzen, sondern auch durch die Möglichkeit, von Nutzern 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. Akteuren (vgl. Röhle 2010, S. 14) zu tun haben:

1. Suchmaschinenbetreiber: Das Interesse der Suchmaschinenbetreiber liegt auf der einen Seite darin, ihre Nutzer 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. Nutzer: Das Interesse der Nutzer 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 (potentiellen) Nutzern 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 Nutzer 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 Nutzern 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 des 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 (ComScore 2013; StatCounter o.D.).

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 Nutzer gewinnen, wenn sich die an Google gewöhnten Nutzer sofort zurecht finden; 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 zugrundeliegenden 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 auf hat 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 Medienwissenschaft (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 in Kap. 18 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.

Das dritte Kapitel erklärt dann den grundlegenden technischen Aufbau von algorithmischen Suchmaschinen. 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 Nutzer automatisch interpretiert und verarbeitet werden können.

Nach diesen beiden technischen Kapiteln betrachten wir dann die Nutzerseite: Was wird eigentlich in Suchmaschinen gesucht, wie werden Suchanfragen formuliert und wie wählen die Nutzer 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](#) widmet sich dann 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 Nutzer.

Dies bringt uns zu den wirtschaftlichen Gegebenheiten, die mit den Suchmaschinen verbunden sind. Im siebten Kapitel 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.

Das achte Kapitel widmet sich der Seite derjenigen, die ihre Inhalte über Suchmaschinen optimal bereitstellen möchten bzw. deren Helfern, den Suchmaschinenoptimierern. Diese verwenden ihre Kenntnisse über die Rankingverfahren der Suchmaschinen, um Inhalte besser auffindbar zu machen und ihren Kunden Traffic 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. 9](#) 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 Nutzern von den eigentlichen Suchergebnissen unterschieden werden können.

Mit den tatsächlich bestehenden Alternativen zu Google beschäftigt sich dann das zehnte Kapitel. 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.

Im elften Kapitel 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 Nutzer 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. 12](#) 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. 13](#) 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.

Weitere Instrumente der spezialisierten Recherche sind sogenannte Spezialsuchmaschinen ([Kap. 14](#)), die sich auf einen bestimmten Bereich spezialisieren. Alle bekannten Suchmaschinenbetreiber haben neben der allgemeinen Suche auch Spezialsuchmaschinen im Angebot, beispielsweise für Nachrichten, Videos oder Bilder.

Viel diskutiert wurde in den letzten Jahren der Bereich Social Media, auch im Kontext der Vermittlung von Informationen. Die Inhalte der sozialen Netzwerke sind umfangreich und spielen für viele Nutzer eine große Rolle. Allerdings gehören auch sie größtenteils zum Deep Web. Daher haben sich die großen Suchmaschinenbetreiber mehr oder weniger erfolgreich bemüht, entweder diese Inhalte im Rahmen von Kooperationen in ihre Suchmaschinen einzubauen. In [Kap. 15](#) werden Lösungsansätze in diesem Bereich vorgestellt.

Während sich die vorangegangenen Themen mit Aspekten aus den Bereichen Technik, Nutzung und Recherche beschäftigten, geht es in [Kap. 16](#) um die gesellschaftliche Rolle von Suchmaschinen. Welche Rolle spielen Suchmaschinen heute für den Wissenserwerb und welche Rolle sollten sie spielen?

Das abschließende Kapitel beschäftigt sich dann 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, 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, Nutzer, Inhalteanbieter 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 Nutzer selbst Inhalte erstellen können, welche dann durchsuchbar gemacht werden. Dazu kommen Inhalte aus der physischen Welt, die in die digitale Welt übertragen und in die Suche eingebunden werden.

Literatur

- AdEx Benchmark 2016. (2017). http://iabspain.es/wp-content/uploads/iabeurope_adexbenchmark_2016_full-report.pdf. Zugegriffen: 17. Okt. 2017.
- AGOF. (2015). AGOF Internet Facts 2015-05. https://www.agof.de/download/Downloads_Internet_Facts/Downloads_Internet_Facts_2015/Downloads_Internet_Facts_2015-05/05-2015_AGOF%20internet%20facts%202015-05.pdf?x74949. Zugegriffen: 17. Okt. 2017.
- Alexa.com. (2017). Top sites in Germany. <https://www.alexa.com/topsites/countries/DE>. Zugegriffen: 17. Okt. 2017.
- Bundesverband Digitale Wirtschaft. (Hrsg.). (2013). *OVK Online-Report 2013/02: Zahlen und Trends im Überblick*. Düsseldorf.
- 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: 16. Okt. 2017.
- Internet Live Stats, & Statistic Brain Research Institute. (o.D.). Anzahl der Suchanfragen bei Google weltweit in den Jahren 2000 bis 2016 (in Milliarden). In *Statista – Das Statistik-Portal*. <https://>

- de.statista.com/statistik/daten/studie/71769/umfrage/anzahl-der-google-suchanfragen-pro-jahr/. Zugegriffen: 17. Okt. 2017.
- 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.
- König, R., & Rasch, M. (Hrsg.). (2014). *Society of the query reader*. Amsterdam: Institute of Network Cultures.
- Lehmann, K., & Schetsche, M. (Hrsg.). (2005). *Die Google-Gesellschaft: Vom digitalen Wandel des Wissens*. Bielefeld: Transcript.
- Purcell, K., Brenner, J., & Raine, L. (2012). *Search engine use 2012*. Washington, DC: Pew Research Center. http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_Search_Engine_Use_2012.pdf. Zugegriffen: 16. Okt. 2017.
- Röhle T. (2010). *Der Google-Komplex: Über Macht im Zeitalter des Internets*. Bielefeld: Transcript.
- Stark, B., Dörr, D., & Aufenanger, S. (Hrsg.). (2014). Die Googleisierung der Informationssuche – Suchmaschinen im Spannungsfeld zwischen Nutzung und Regulierung. Berlin (u.a.): De Gruyter.
- StatCounter. (o.D.). Marktanteile von Suchmaschinen bei der Desktop-Suche und bei der mobilen Suche in Deutschland im September 2017. *Statista – Das Statistik-Portal*. <https://de.statista.com/statistik/daten/studie/301012/umfrage/marktanteile-der-suchmaschinen-und-marktanteile-mobile-suche/>. Zugegriffen: 17. Okt. 2017.
- Vaidhyanathan, S. (2011). *The googlization of everything (and why we should worry)*. Berkeley: University of California Press.



Vordergründig mag uns die Suche im Web erst einmal trivial erscheinen: Wir geben eine Suchanfrage ein und erhalten daraufhin eine Suchergebnisseite, auf der wir einen Treffer auswählen. Doch bei diesem Vorgehen handelt es sich nur um eine der vielen Möglichkeiten, an Informationen im Web zu gelangen. In diesem Kapitel werden die verschiedenen Zugänge zu den Informationen im Web beschrieben und es wird erklärt, warum sich der Zugang mittels Suchmaschinen fast exklusiv durchgesetzt hat.

2.1 Suche nach einer Website vs. Suche nach Informationen zu einem Thema

Zunächst einmal ist zu fragen, was man mit einer Suche eigentlich erreichen will bzw. kann. Fürs erste genügt es hier, drei Fälle zu unterscheiden. Diese Fälle werden anhand der in [Abb. 2.1](#) gezeigten Browser-Startseite erläutert:

1. Ein Nutzer möchte eine konkrete, ihm bereits bekannte Website ansteuern und gibt die URL direkt in die Adresszeile seines Browsers ein (s. [Abb. 2.1](#), obere Zeile links). Auf der Website liest er dann entweder direkt etwas, sucht dort oder klickt auf weitere Dokumente. Ein solcher Vorgang hat erst einmal wenig mit unserem intuitiven Verständnis von Suche zu tun, ist aber ein Mittel, zu (gesuchten) Informationen zu gelangen. Beispielsweise kann ein Nutzer, der sich für Nachrichten zu einem aktuellen Thema interessiert, direkt eine Nachrichtenwebsite anwählen und dort relevante Artikel entweder direkt auf der Startseite lesen, dort anklicken oder mittels der internen Suchfunktion der Nachrichtenwebsite suchen.
2. Ein Nutzer möchte eine konkrete, ihm bereits bekannte oder noch nicht bekannte Website ansteuern und sucht dazu über die URL- und Suchzeile im Browser (in [Abb. 2.1](#) obere Zeile) oder über eine zuvor ausgewählte Suchmaschine (das kann auch

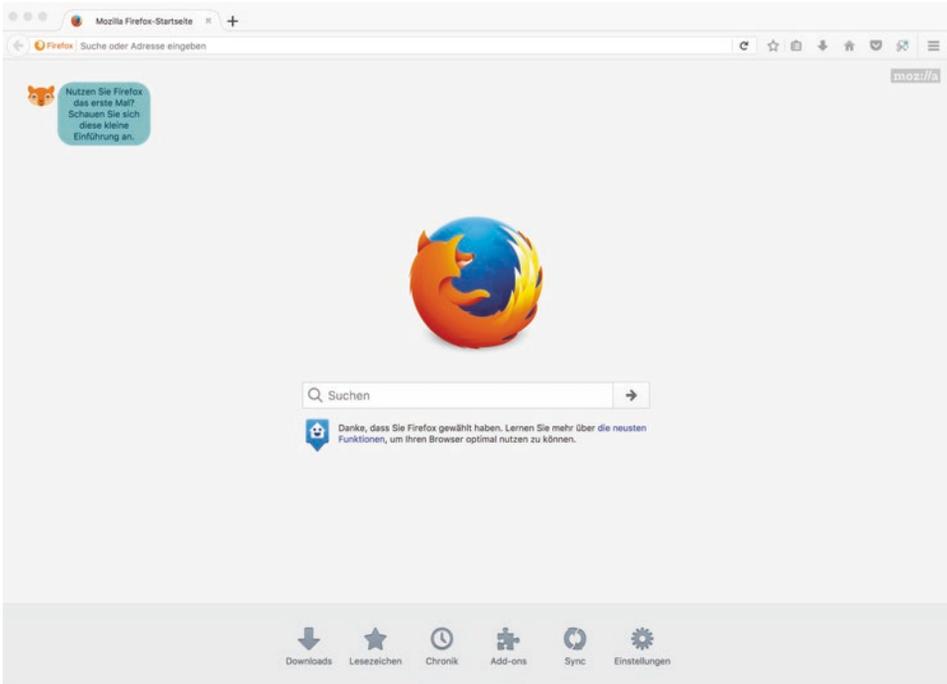


Abb. 2.1 Startseite des Firefox-Browsers mit Adresszeile und Google-Suche als voreingestellter Startseite

die Suchmaschine auf der Startseite des Browsers sein, sofern dort eine eingestellt ist) nach dieser Website. Dabei kann es sich um eine bekannte Website handeln – in diesem Fall ist die Suche nur eine „Abkürzung“ der direkten Eingabe in die Adresszeile (also beispielsweise die Eingabe von „spiegel“ im Suchfeld statt „www.spiegel.de“ in der Adresszeile) oder eine Hilfe, wenn man sich nicht mehr an die exakte Adresse der gesuchten Website erinnern kann (wenn man beispielsweise nicht mehr weiß, ob eine Website mit der Endung.de oder Endung.org zu finden ist). Bei der direkten Suche nach einer noch nicht bekannten Website nimmt der Nutzer zumindest an, dass eine solche Website existiert und sucht entsprechend.

3. Der letzte Fall ist der eines Nutzers, der auf der Suche nach ihm bislang noch nicht bekannten Informationen ist. Dieser Fall unterscheidet sich fundamental von den beiden vorangegangenen Fällen, da hier nicht nach einer konkreten Website gesucht wird, sondern nach Informationen zu einem Thema. Dabei kann nicht mit Sicherheit vorausgesehen werden, ob sich diese Informationen auf einer bestimmten Website befinden oder ob überhaupt die Informationen von einer einzigen Website ausreichend sind, um das Informationsbedürfnis zu befriedigen.

Wir werden auf die Unterteilung von Suchanfragen nach Intentionen bzw. Informationsbedürfnissen in [Abschn. 4.3](#) noch ausführlicher zu sprechen kommen. Vorläufig reicht erst

einmal die Unterscheidung zwischen der Suche nach bekannten *Websites* und der Suche nach *noch unbekanntem* Informationen. Um die Möglichkeiten verschiedener Zugänge zu den Informationen im Web einschätzen zu können, ist es wichtig, dass wir hier schon zwischen diesen Fällen unterscheiden können.

In unserem Beispiel haben wir schon gesehen, dass Suchanfragen an unterschiedlichen Stellen eingegeben werden können. Wir werden auf die Bedeutung der im Suchfeld bzw. in der Adresszeile des Browsers voreingestellten Suchmaschine und der voreingestellten Startseite im Kapitel zum Suchmaschinenmarkt ([Kap. 7](#)) zurückkommen.

Beispiel: Ist die Suche im Web die Suche nach einer Nadel im Heuhaufen?

Gerne wird die Suche im Web mit der Suche nach einer Nadel im Heuhaufen verglichen. Das Bild soll uns verdeutlichen, dass es aufgrund der riesigen Menge der vorhandenen Informationen (dem Heuhaufen) schwierig ist, die richtige (die Nadel) zu finden.

Doch dieses Bild ist schief: Zwar suchen wir in der kaum überschaubaren Menge der Informationen im Web, aber wir haben nur in bestimmten Fällen überhaupt eine genaue Vorstellung davon, was wir finden möchten. Nur in den Fällen, in denen wir nach einem konkreten Dokument suchen, passt der Vergleich: Wir wissen, wie die Nadel aussieht, und es gibt auch nur eine einzige Nadel, sodass wir erkennen können, wann unsere Suche beendet ist.

Im Fall der Suche nach bislang unbekanntem Informationen passt der Vergleich nicht mehr: Es könnte beispielsweise mehrere Nadeln geben, die auch noch unseren Zweck unterschiedlich gut erfüllen könnten. Und es könnte auch sein, dass wir erst zufrieden sein können, wenn wir mehrere Nadeln gefunden haben, die sich gegenseitig ergänzen oder bestätigen.

2.2 Was ist ein Dokument?

Das Web ist schon von seiner Anlage her multimedial und enthält viel mehr als nur Texte. Insofern sind auch Suchmaschinen nicht nur dazu da, Texte im Web aufzufinden, sondern auch andere Dokumententypen – auch wenn die Suche heute (noch) weitgehend textbasiert ist. Doch unabhängig davon, ob sich um einen Text, ein Bild oder ein Video handelt, wollen wir von einem Dokument bzw. alternativ von einem Informationsobjekt sprechen.

Was ist also ein Dokument? Bei Dokumenten denken wir vielleicht als erstes an offizielle Dokumente, wie sie in Behörden mit Stempel und Unterschrift ausgestellt werden. Im informationswissenschaftlichen Sinne sind Dokumente aber noch viel mehr: Es handelt sich schlicht um eine Aufzeichnung von Informationen, gleich ob dies in schriftlicher Form (Textdokument) oder beispielsweise in bildlicher Form (Bildokument) geschieht. Bezogen auf Suchmaschinen bedeutet dies, dass alle von ihnen angezeigten Inhalte (Texte, Bilder, Videos, usw.) Dokumente sind. Manchmal wird stattdessen auch von Informationsobjekten gesprochen, um zu verdeutlichen, dass es sich nicht nur um textuelle Dokumente handeln muss.

Wenn also in diesem Buch von Dokumenten gesprochen wird, kann es sich um jede dieser Formen handeln. Da der Hauptanteil der Dokumente, wie sie von Suchmaschinen erfasst werden, immer noch Texte sind, beschäftigen wir uns vor allem mit diesen Dokumenten, gehen aber, wo es sinnvoll ist, auch auf andere Dokumententypen ein.

2.3 Wo wird gesucht?

Die Zeiten, in denen Suchmaschinen weitgehend im gleichen Kontext, nämlich auf stationären Rechnern oder Laptops, verwendet wurden, sind vorbei. Mittlerweile findet Suche auf einer Vielzahl von Geräten statt: Wir verwenden heute beispielsweise selbstverständlich Smartphones und Tablets, auf denen wir dann natürlich auch suchen. Auch hier zeigt sich aber wieder die Unterscheidung zwischen der Websuche und der Suche innerhalb einer speziellen Anwendung, im Fall des Smartphones bzw. des Tablets meist innerhalb von Apps: Während in Apps nur ein beschränkter Datenbestand (innerhalb der App) durchsucht werden muss, geht es bei der Websuche wiederum „ums Ganze“, also eine möglichst vollständige Abbildung des Web. Egal, welches Gerät wir zur Suche verwenden: Die (Web-)Suche ist ein zentraler Bestandteil unserer Internetnutzung. Wie wir aber sehen werden, unterscheidet sich das Nutzerverhalten je nach Kontext (beispielsweise unterwegs vs. zu Hause) und Gerät (beispielsweise großer Bildschirm auf dem Laptop vs. kleiner Smartphone-Bildschirm). Suchmaschinen sind entsprechend angepasst, sodass sie auf unterschiedlichen Geräten und in unterschiedlichen Kontexten jeweils angepasste Ergebnisse und Ergebnisdarstellungen liefern (ausführlich dazu in [Abschn. 6.1](#)).

2.4 Zugänge zu den Informationen im World Wide Web

Suchmaschinen sind bei weitem nicht der einzige Zugang zu den Informationen im Web. Im Folgenden werden die verschiedenen Zugangsarten und –systeme vorgestellt¹ und in Relation zu den Suchmaschinen, um die es dann in den folgenden Kapiteln wieder exklusiv gehen soll, gesetzt. Dabei beginnen wir mit den Web-Suchmaschinen selbst, da sie unseren Ausgangspunkt bilden und wir dann die anderen Systeme in ihren Vor- und Nachteilen mit diesen vergleichen wollen.

Grundlegend lässt sich zwischen verschiedenen Typen von Suchmaschinen und anderen Systemen unterscheiden:

- Zu den Suchmaschinen zählen die allgemeinen Suchmaschinen, Spezialsuchmaschinen, Hybridsuchmaschinen und Metasuchmaschinen.
- Zu den anderen Systemen gehören Web-Verzeichnisse, Social-Bookmarking-Dienste und Frage-Antwort-Dienste.

¹ Die Gruppierung der Dienste folgt (Griesbaum et al. 2009).

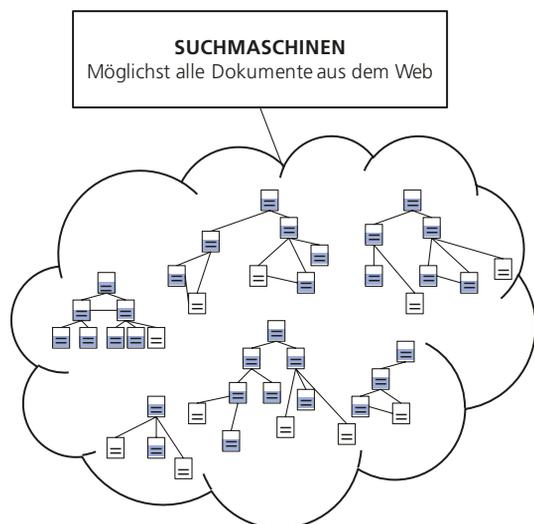
Wichtig für das Verständnis des Konzepts der Suchmaschinen ist, dass die unterschiedlichen Zugänge zu den Inhalten des Web auch unterschiedliche Zielsetzungen verfolgen. So wäre es etwa unfair, den Umfang der Datenbestände von Suchmaschinen und Web-Verzeichnissen zu vergleichen, da sie einen ganz unterschiedlichen Anspruch an die Vollständigkeit ihrer Datenbestände haben.

2.4.1 Suchmaschinen

Wenn wir von Suchmaschinen sprechen, meinen wir meist Web-Suchmaschinen (auch: allgemeine Suchmaschinen, Universalsuchmaschinen, algorithmische Suchmaschinen). Diese haben den Anspruch, die Inhalte des Web möglichst vollständig abzudecken und ggf. mit weiteren Inhalten anzureichern (s. [Abschn. 3.2](#)). In [Abb. 2.2](#) ist schematisch dargestellt, welche Inhalte des Web Suchmaschinen erfassen. Die Wolke stellt das Universum des Web dar. Darin befindet sich eine Vielzahl von Dokumenten, welche innerhalb von Websites (verdeutlicht durch die hierarchische Struktur aus Dokumenten) abgelegt sind. Hervorgehoben sind nun die Inhalte, die von der Suchmaschine überhaupt erfasst werden. Zwar ist es das Ziel von Suchmaschinen, die Inhalte des Web vollständig zu erfassen; dieses Ziel wird jedoch nicht erreicht bzw. kann auch gar nicht erreicht werden. Wir werden uns in [Kap. 3](#) genauer mit den Gründen dafür beschäftigen.

Dennoch erreichen Suchmaschinen eine größere Abdeckung des Web als andere Suchsysteme. Dies liegt zum einen an ihrem universalen Anspruch, zum anderen daran, dass sie die Inhalte automatisiert erfassen. Auch dieser Prozess wird in [Kap. 3](#) genau beschrieben; an dieser Stelle soll erst einmal genügen, dass Suchmaschinen in der Lage sind, eine riesige Zahl von Dokumenten im Web zu erfassen und durchsuchbar zu machen.

Abb. 2.2 Inhalte von Suchmaschinen



Das Konzept der algorithmischen Suchmaschine in den 90er Jahren

Die Idee der Suchmaschine, wie wir sie kennen, hat sich schon in der Frühzeit des Web entwickelt. Frühe Suchmaschinen wie Lycos und Webcrawler arbeiteten schon nach dem gleichen Prinzip wie heute Google: Die im Web vorhandenen Seiten werden durch das Verfolgen von Links erfasst und auf Suchanfragen hin werden gerankte Ergebnislisten zurückgegeben. Dieser Prozess läuft vollautomatisch.

Am besten lässt sich die Ähnlichkeit früher und heutiger Suchmaschinen vielleicht am Beispiel der Startseite von AltaVista, der damals führenden Suchmaschine, aus dem Jahr 1996 verdeutlichen (Abb. 2.3).

Zuerst einmal fällt die Ähnlichkeit mit heutigen Suchmaschinen wie Google auf: Es gibt ein zentral platziertes Suchfeld, daneben einen Button, mit dem sich die Suche abschicken lässt. Nutzer können dabei im Prinzip eingeben, was sie wollen, ohne dass sie eine besondere Abfragesprache lernen müssen. Ob einzelne Wörter, ganze Sätze oder Fragen: Es liegt in der automatischen Verarbeitung durch die Suchmaschine, zu den Suchanfragen passende Ergebnisse zu liefern.

Zum zweiten finden sich auf der AltaVista-Startseite Informationen über die Größe des Datenbestands. Es werden 30 Millionen Dokumente angegeben, eine zum damaligen Zeitpunkt große Zahl, wenn man bedenkt, dass das Web noch in den Kinderschuhen steckte. Mittlerweile ist das Web um ein Vielfaches größer, geblieben ist jedoch die Herausforderung, seine Inhalte vollständig und aktuell zu erfassen und in der Suche verfügbar zu machen (s. Kap. 3).

Zum dritten soll noch darauf hingewiesen werden, dass AltaVista bereits 1996 seine Suchergebnisse auch über andere Portale verfügbar machte, darunter u. a.



Abb. 2.3 Startseite der Suchmaschine AltaVista (1996); <https://web.archive.org/web/19961023234631/http://altavista.digital.com/>