



Xpert.press



Dirk Lewandowski

Suchmaschinen verstehen

 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 dieser Reihe
<http://www.springer.com/series/4393>

Dirk Lewandowski

Suchmaschinen verstehen

Dirk Lewandowski
Hochschule für Angewandte Wissenschaften
Hamburg
Deutschland
E-Mail: dirk.lewandowski@haw-hamburg.de

ISSN 1439-5428

Xpert.press

ISBN 978-3-662-44013-1

ISBN 978-3-662-44014-8 (eBook)

DOI 10.1007/978-3-662-44014-8

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 2015

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.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Berlin Heidelberg ist Teil der Fachverlagsgruppe Springer Science+Business Media
(www.springer.com)

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 Danksagung	11
1.8 Zusammenfassung	11
Literatur	12
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 Spezielsuchmaschinen	19
2.4.3 Metasuchmaschinen	21
2.4.4 Web-Verzeichnisse	23
2.4.5 Social-Bookmarking-Dienste	24
2.4.6 Frage-Antwort-Dienste	25
2.5 Zusammenfassung	25
Literatur	26
3 Wie Suchmaschinen funktionieren	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	41
3.3.2	Willentlicher Ausschluss von Inhalten durch die Suchmaschinenbetreiber	43
3.3.3	Die Datenbasis der Suchmaschinen	44
3.3.4	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	61
	Literatur	62
4	Wie Suchmaschinen genutzt werden	65
4.1	Der Suchprozess	65
4.2	Erfassung von Nutzungsdaten	67
4.3	Anfragetypen	69
4.4	Sessions	72
4.5	Suchanfragen	74
4.5.1	Formulierung der Suchanfragen	74
4.5.2	Länge der Suchanfragen	76
4.5.3	Verteilung der Suchanfragen nach Häufigkeiten	78
4.5.4	Suchanfragen-Trends	80
4.5.5	Verwendung von Operatoren und Befehlen zur gezielten Suche	82
4.6	Themen	83
4.7	Zusammenfassung	84
	Literatur	86
5	Das Ranking der Suchergebnisse	89
5.1	Gruppen von Rankingfaktoren	91
5.2	Textstatistik	93
5.2.1	Ermittlung potenziell relevanter Dokumente	93
5.2.2	Berechnung von Häufigkeiten	95
5.2.3	Berücksichtigung von Strukturinformationen in Dokumenten	97
5.3	Popularität	99
5.3.1	Linktopologische Verfahren	100
5.3.2	Nutzungsstatistische Verfahren	105
5.4	Aktualität	113
5.5	Lokalität	115
5.6	Personalisierung der Suchergebnisse	118
5.7	Technische Rankingfaktoren	120
5.8	Ranking und Spam	121
5.9	Zusammenfassung	122
	Literatur	123

6 Die Präsentation der Suchergebnisse	125
6.1 Der Aufbau der Suchergebnisseiten	126
6.2 Elemente auf den Suchergebnisseiten	132
6.2.1 Organische Ergebnisse	133
6.2.2 Werbung	134
6.2.3 Universal-Search-Ergebnisse	136
6.2.4 Knowledge-Graph-Ergebnisse	138
6.2.5 Navigationselemente	139
6.2.6 Optionen auf der Suchergebnisseite	140
6.3 Der Aufbau der Trefferbeschreibungen	140
6.4 Optionen zu den einzelnen Ergebnissen	144
6.5 Auswahl von geeigneten Treffern auf den Suchergebnisseiten	145
6.6 Zusammenfassung	146
Literatur	147
7 Der Suchmaschinenmarkt	149
7.1 Das Geschäftsmodell der Suchmaschinen	149
7.2 Die Bedeutung der Suchmaschinen für die Online-Werbung	151
7.3 Marktanteile der Suchmaschinen	152
7.4 Bedeutende Suchmaschinen	155
7.5 Index vs. Suchmaschine	156
7.6 Zusammenfassung	159
Literatur	160
8 Suchmaschinenoptimierung	161
8.1 Die Bedeutung der Suchmaschinenoptimierung	162
8.2 Grundlegende Verfahren der Suchmaschinenoptimierung	165
8.2.1 On-The-Page-Optimierung	167
8.2.2 Off-The-Page-Optimierung	170
8.3 Suchmaschinenoptimierung und Spam	172
8.4 Die Position der Suchmaschinenbetreiber	172
8.5 Zusammenfassung	174
Literatur	175
9 Alternativen zu Google	177
9.1 Überschneidungen zwischen den Suchergebnissen verschiedener Suchmaschinen	178
9.2 Warum sollte man eine andere Suchmaschine als Google nutzen?	179
9.2.1 Einholen einer „zweiten Meinung“	180
9.2.2 Mehr bzw. zusätzliche Ergebnisse	181
9.2.3 Andere Ergebnisse	181
9.2.4 Bessere Ergebnisse	182

9.2.5	Andere Trefferpräsentation	182
9.2.6	Andere Benutzerführung	183
9.2.7	Andere Suchmöglichkeiten	183
9.3	Wann sollte man eine andere Suchmaschine als Google nutzen?	184
9.4	Besonderheiten bei Google aufgrund seiner Marktdominanz	186
9.5	Zusammenfassung	188
	Literatur	189
10	Besser suchen	191
10.1	Quellenauswahl	193
10.2	Auswahl passender Suchbegriffe	195
10.3	Boolesche Operatoren	195
10.4	Verbindung von Suchanfragen mit Booleschen Operatoren	199
10.5	Formulare für die erweiterte Suche	200
10.6	Befehle	202
10.7	Komplexe Suchen	206
10.8	Zusammenfassung	207
	Literatur	207
11	Die Qualität der Suchergebnisse	209
11.1	Kriterien für die Bewertung von Texten im Web	209
11.2	Menschliche vs. maschinelle Qualitätsüberprüfung	211
11.3	Wissenschaftliche Evaluierung der Suchergebnisqualität	215
11.3.1	Standardaufbau von Tests zur Retrievaleffektivität	217
11.3.2	Messung der Retrievaleffektivität mittels Klickdaten	221
11.3.3	Evaluierung im Interaktiven Information Retrieval	222
11.4	Zusammenfassung	224
	Literatur	225
12	Das Deep Web	227
12.1	Die Inhalte des Deep Web	229
12.2	Quellen vs. Inhalte von Quellen, Zugänglichkeit von Inhalten <i>via</i> Web	231
12.3	Die Größe des Deep Web	235
12.4	Bereiche des Deep Web	236
12.5	Welche Rolle spielt das Deep Web heute?	237
12.6	Zusammenfassung	237
	Literatur	238
13	Spezialsuchmaschinen und ihre Recherchemöglichkeiten	239
13.1	Probleme der allgemeinen Suchmaschinen	240
13.2	Typen von Spezialsuchmaschinen	242
13.3	Nachrichtensuchmaschinen	244

13.4 Suchmaschinen für wissenschaftliche Inhalte	248
13.5 Faktensuchmaschinen	252
13.6 Einbindung von Spezialsuchmaschinen in der Universal Search	252
13.7 Zusammenfassung	255
Literatur	256
14 Suchmaschinen und Social Media	259
14.1 Typen von Social-Media-Angeboten	260
14.2 Social-Media-Angebote als Rechercheinstrumente	263
14.3 Integration von Daten aus Social-Media-Angeboten in Suchmaschinen	263
14.4 Zusammenfassung	267
Literatur	267
15 Suchmaschinen und ihre Rolle als Vermittler von Informationen	269
15.1 Die Interessen der Suchmaschinenbetreiber	270
15.2 Suchmaschinen-Bias	272
15.3 Interessengeleitete Präsentation von Suchergebnissen	273
15.4 Was würde „Suchneutralität“ bedeuten?	277
15.5 Zusammenfassung	278
Literatur	278
16 Die Zukunft der Suche	281
16.1 Suche als Basistechnologie	281
16.2 Der Wandel der Suchanfragen und der Dokumente	282
16.3 Besseres Verstehen von Dokumenten und Suchanfragen	285
16.4 Das Ende der „mobilen Suche“	285
16.5 Antworten anstelle von Dokumenten	286
16.6 Die wirtschaftliche Zukunft der Suchmaschinen	287
16.7 Die gesellschaftliche Zukunft der Suchmaschinen	288
16.8 Zusammenfassung	290
Literatur	290
Glossar	291
Sachverzeichnis	303

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, sowohl hinsichtlich des Suchprozesses als auch des Nutzerverhaltens. 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 und 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 meisten 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. 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. Da 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 auch 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 erklären aus der massenhaften Nutzung der Suchmaschinen einerseits und aus der Reihung und Darstellung der Suchergebnisse andererseits.

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 Reihung angezeigt werden. Wir können zwar theoretisch aus allen Ergebnissen auswählen, doch verlassen wir uns 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) oder

der „Society of the Query“ (so der Titel einer Konferenzreihe und eines Buchs; König und Rasch 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 Internets sind. Das Internet betrachten wir hier als eine Ansammlung von Protokollen und Diensten, zu denen beispielsweise auch E-Mail, Chat und 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 (van Eimeren und Frees 2014). 82% aller deutschen Internetnutzer nutzen mindestens einmal pro Woche eine Suchmaschine, aber „nur“ 79% der Internetnutzer lesen bzw. schreiben in dieser Zeit mindestens eine E-Mail. Diese Daten stammen aus der ARD/ZDF-Onlinestudie (van Eimeren und Frees 2014), welche jährlich die Nutzung des Internet in der deutschen Bevölkerung abfragt. Ähnliche Ergebnisse finden sich aber auch in vergleichbaren Studien (u. a. AGOF\ „Siehe Arbeitsgemeinschaft Online Forschung“ Internet Facts 2014); 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 Suchmaschinennutzung nur eine einzige Kategorie, welche auch auf mehr als 70% der Nutzer kommt, nämlich „Informationen suchen“ (82% der Nutzer tun dies mindestens einmal wöchentlich). Dabei fällt auf, dass es sich wieder um einen informationsorientierten Zweck handelt. Alle anderen Anwendungen im Netz (mit Ausnahme von Wetterinformationen, die von der Hälfte der Nutzer abgerufen werden) kommen auf deutlich weniger als 50% 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 (ComScore 2014). Hier steht Google an erster Stelle, gefolgt von Facebook und Ebay. Zwar lässt sich aufgrund dieser Erhebung nicht unterscheiden, inwieweit beispielsweise auf den Websites von Google tatsächlich die Suche oder ein anderer Dienst genutzt wurde, allerdings 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. Nach Erhebungen des Marktforschungsunternehmens ComScore werden weltweit jeden Monat mehr als 175 Mrd. Suchanfragen in die allgemeinen Suchmaschinen eingegeben (Sullivan 2013) – das sind mehr als 5,8 Mrd. Anfragen pro Tag bzw. mehr als 2 Mio. 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 von diesen. 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 Web-Verzeichnisse und Social-Media-Angebote, die uns zu Informationen führen. Doch keiner dieser Dienste hat eine den Suchmaschinen vergleichbare Bedeutung für den Informationszugang im Web erreicht.

Und nicht zuletzt haben Suchmaschinen eine große Bedeutung aufgrund des Online-Werbemarkts. Mittlerweile werden mit Werbung im Internet allein in Deutschland jährlich mehr als 7 Mrd. € umgesetzt. Davon entfallen rund 35 %, also mehr als 2,5 Mrd. €, auf Textanzeigen in Suchmaschinen (OVK Online-Report 2013/02 2013). 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, die zwar eine Werbung sehen, jedoch in dem Moment überhaupt kein Interesse daran haben, 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 aber auch, dass Suchmaschinenanbieter ihre Suchmaschinen nicht allein, wie oft behauptet wird, an den Ansprüchen und Bedürfnissen der Nutzer ausrichten, sondern natürlich auch an ihren eigenen 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 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. Inhalteanbieter: Wer Inhalte im Web anbietet, möchte auch von (potenziellen) Nutzern gefunden werden. Allerdings besteht ein weiteres Interesse vieler Inhalteanbieter 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 Inhalteanbietern 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 (Beispiele: Blogger, Youtube).
3. *Scan and serve*: Google bringt Dinge aus der echten Welt in die digitale (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 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. Der Anspruch der Suchmaschinen geht aber deutlich weiter, wie sich nicht zuletzt auch

an Googles Firmenmotto deutlich wird: „Das Ziel von Google ist es, die Informationen der Welt zu organisieren und für alle zu jeder Zeit zugänglich und nutzbar zu machen“ (Google Inc. 2014).

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 zu antizipieren und auf Basis der Vorlieben des Nutzers und 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. 16 darauf zurückkommen; um die Bedeutung der Suchmaschinen und die Vielfalt ihrer Anwendungskontexte zu verstehen, sollten die Möglichkeiten aber bereits hier zumindest kurz angerissen werden.

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% aller Anfragen an allgemeine Suchmaschinen an Google gestellt, andere Suchmaschinen spielen nur eine untergeordnete Rolle (ComScore 2013).

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 dieses 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 zu-recht finden; andererseits müssen sie sich von Google unterscheiden, um überhaupt einen Mehrwert gegenüber dieser Suchmaschine darstellen zu können.

1.3 Ziel dieses Buchs

Dieses Buch 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 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 auch auf andere Sachverhalte übertragen werden kann.

Überhaupt geht es in diesem Buch um eine Übertragungsleistung: Das, was man anhand einer bzw. einiger Suchmaschinen lernen kann, 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 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 des Buchs 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 dem 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 3. 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 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. Und 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.

Kapitel 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 siebenten 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. Und natürlich klingt auch hier schon die Frage an, ob eine solche Situation wünschenswert ist und wie sie gegebenenfalls zu ändern wäre.

Mit den tatsächlich bestehenden Alternativen zu Google beschäftigt sich dann das neunte Kapitel. Zuerst einmal gilt, 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 10. 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, welcher sich Kap. 11 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.

Das 12. Kapitel beschäftigt sich mit den Inhalten des World Wide 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 finden lassen. Wir werden sehen, aus welchen Gründen diese Inhalte mit Suchmaschinen nicht auffindbar sind und mit welchen Methoden wir sie dennoch recherchieren können.

Weitere Instrumente der spezialisierten Recherche sind sogenannte Spezialsuchmaschinen, 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. Es gibt aber auch kleinere Anbieter, die sich exklusiv auf einen Bereich spezialisiert haben und für diesen oft bessere Recherchemöglichkeiten und ein individualisiertes Ranking anbieten.

Viel diskutiert wurde in den letzten Jahren der Bereich der 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 oder aber eigene Soziale Netzwerke aufzubauen. In Kap. 13 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. 15 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 es vielleicht auch nie 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 kommentierte Hinweise auf weiterführende Literatur für diejenigen, die sich tiefer in das Thema des jeweiligen Kapitels einarbeiten möchten.

Unter dem Stichwort „Exkurs“ finden sich Beispiele, die das im Haupttext Gesagte illustrieren und vertiefen, für das Verständnis des Haupttexts aber nicht unbedingt erforderlich sind.

1.7 Danksagung

Ich möchte mich vor allem bei meinen studentischen Hilfskräften, die mich auf ganz unterschiedliche Weise bei diesem Buch unterstützt haben, bedanken: Dorothee Wagner hat die meisten Schaubilder auf Basis meiner hingekritzelten Entwürfe gestaltet, Alexandra Linhart hat vor allem den ersten Entwurf des Glossars zusammengestellt. Jenny Krieger hat die Literaturlisten korrigiert und ergänzend an den Schaubildern gearbeitet. Alle drei haben mich auch sonst bei der Arbeit an diesem Buch unterstützt, indem sie all die kleinen Aufgaben, die einen Autor bei seiner Arbeit aufhalten, schnell und zuverlässig erledigt haben.

Dem kleinen Suche-Team an der HAW, bestehend aus Friederike Kerkmann, Sebastian Sünkler und Christiane Behnert, möchte ich für die zahlreichen Anregungen und Diskussionen der letzten Jahre danken. Sie werden vieles von dem, was wir im Lauf der Zeit besprochen haben, in diesem Buch wiederfinden.

Nicht zuletzt möchte ich auch den Autorinnen und Autoren des Handbuchs „Internet-Suchmaschinen“ danken, die viele Themen, die auch in diesem Buch besprochen werden, ausführlich in kompetenten Artikeln durchleuchtet haben. Von diesen habe ich beim Schreiben des vorliegenden Buchs erheblich profitiert.

1.8 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, Inhabeanbieter 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 eingescannt und in die Suche eingebunden werden.

Literatur

- AGOF e. V. (2014) Internet facts 2014–2007 http://www.agof.de/download/Downloads_Internet_Facts/Downloads_Internet_Facts_2014/Downloads_Internet_Facts_2014-07/07-2014_AGOF%20internet%20facts%202014-07.pdf?1f74f1. Zugegriffen: 24. Sept. 2014.
- 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: 30. Sept. 2014.
- ComScore Inc. (2014). Germany Top 20. March 2014. <http://www.comscore.com/Insights/Market-Rankings/Germany-Top-20-March-2014>. Zugegriffen: 23. Sept. 2014.
- Google Inc. (2014). Über Google. <https://www.google.de/intl/de/about/>. Zugegriffen: 23. Sept. 2014.
- 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: 23. Sept. 2014
- Röhle, T. (2010). *Der Google-Komplex: Über Macht im Zeitalter des Internets*. Bielefeld: Transcript.
- Sullivan, D. (2013). Google still world's most popular search engine by far, but share of unique searchers dips slightly. Search Engine Land. <http://searchengineland.com/google-worlds-most-popular-search-engine-148089>. Zugegriffen: 30. Sept. 2014
- Vaidhyanathan, S. (2011). *The googlization of everything (and why we should worry)*. Berkeley: University of California Press.
- Van Eimeren, B., & Frees, B. (2014). Ergebnisse der ARD/ZDF-Onlinestudie 2014: 79% der Deutschen online – Zuwachs bei mobiler Internetnutzung und Bewegtbild. *Media Perspektiven*, 45(7–8), 378–396.

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 das Suchfeld im Browser (in Abb. 2.1 obere Zeile

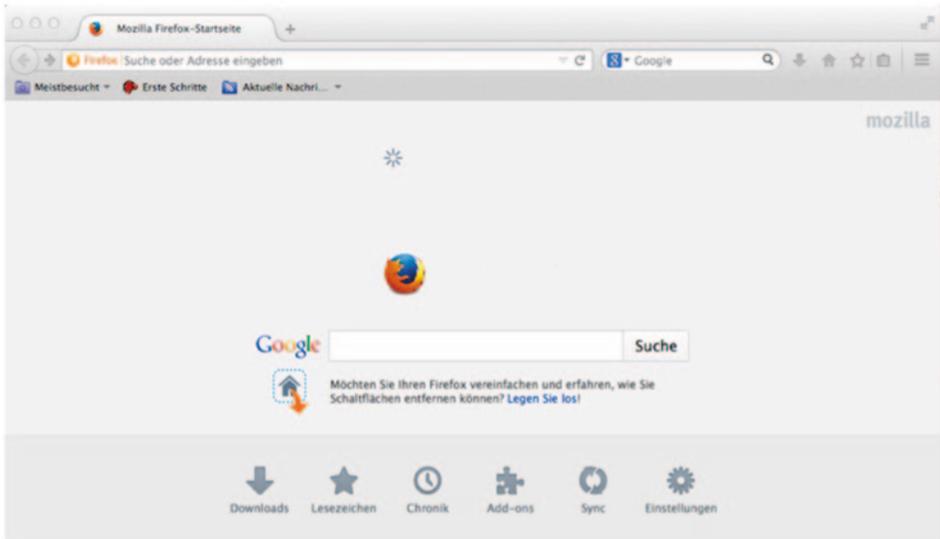


Abb. 2.1 Startseite des Firefox-Browsers mit Adresszeile, Suchfeld und Google als voreingestellter Startseite

rechts) oder über eine zuvor ausgewählte Suchmaschine (das kann auch 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 „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 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. Im Moment reicht erst einmal die Unterscheidung zwischen der Suche nach bekannten *Websites* und der Suche nach *noch unbekanntem Informationen*. Um die Möglichkeiten verschiedener Zu-

gänge zu den Informationen im Web einschätzen zu können, ist es aber 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, das 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 Dokumenttypen – 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 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 Inhalte, die von diesen angezeigt werden (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 eine 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 dies sinnvoll ist, auch auf andere Dokumenttypen ein.

2.3 Wo wird gesucht?

Wenn wir von Suchmaschinen sprechen, denken wir meist an die Suche auf einem PC, sei es am Schreibtisch oder auf einem Laptop. Mittlerweile findet Suche aber auch in anderen Kontexten und auf anderen Geräten statt: Wir verwenden heute 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 oder des Tablets meist innerhalb von Apps: Während im letztgenannten Fall 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 (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 Kap. 6).

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).

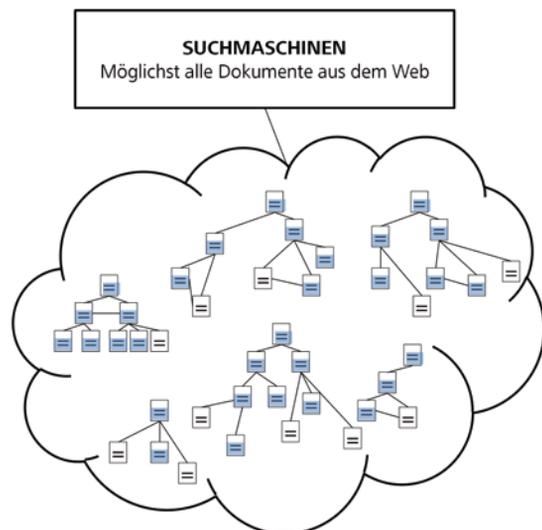
Wichtig für das Verständnis des Konzepts der Suchmaschinen ist, dass die unterschiedlichen Zugänge zu den Inhalten des Webs auch unterschiedliche Zielsetzungen verfolgen. So wäre es etwa unfair, den Umfang der Datenbestände von Suchmaschinen und Web-Verzeichnissen zu vergleichen, da sie einen ganz unterschiedlichen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

2.4.1 Suchmaschinen

Wenn wir von Suchmaschinen sprechen, meinen wir meist Web-Suchmaschinen (auch: allgemeine Suchmaschinen, Universalsuchmaschinen, algorithmische Suchmaschine). Diese haben den Anspruch, die Inhalte des Webs möglichst vollständig abzudecken. 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 Webs 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 einem 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



Exkurs: Das Konzept der algorithmischen Suchmaschine in den 1990er-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 früherer 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 müssten. 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 Mio. 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). <http://web.archive.org/web/19961023234631/http://altavista.digital.com/>

Yahoo. Schon damals bauten viele Anbieter keine eigenen Suchmaschinen auf, sondern bedienten sich im Rahmen von Kooperationen der Ergebnisse der großen Suchmaschinen. Wir werden in Kap. 7 auf solche Kooperation zurückkommen und sehen, welchen Einfluss sie auf den aktuellen Suchmaschinenmarkt haben.

Doch auch die Unterschiede zwischen damals und heute sollen nicht verschwiegen werden: Schon oberhalb des Suchfelds finden sich einige Dinge, die anders sind als bei heutigen Suchmaschinen: Bei AltaVista konnte man direkt auf der Startseite zwischen verschiedenen Suchmodi auswählen, in diesem Fall zwischen der voreingestellten einfachen Suche mit nur einem Suchfeld und einer erweiterten Suche. Schon in der einfachen Suche konnte man die zu durchsuchende Kollektion (voreingestellt: Web) auswählen und das Format, in dem die Suchergebnisse angezeigt werden. Wir werden die erweiterte Suche und unterschiedliche Ergebnisdarstellungen in späteren Kapiteln ausführlich kennenlernen.

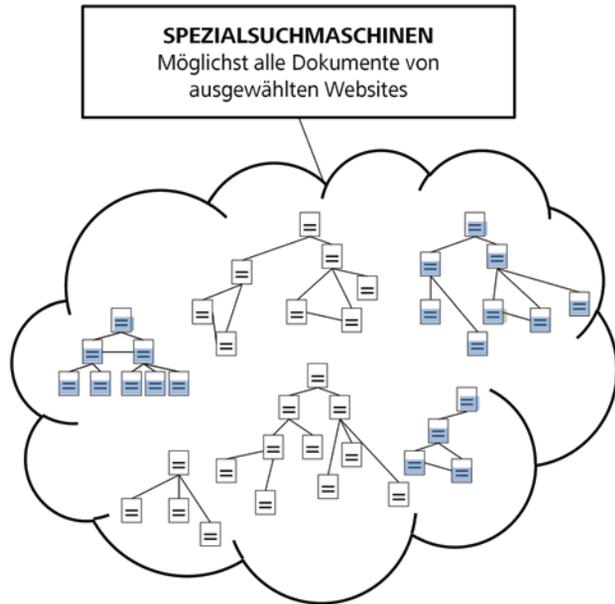
Erhellend sind auch die Texte, die sich auf der AltaVista-Startseite rund um das Suchfeld finden. Zum einen wird ein Link auf eine *Mirror Site* geboten; solche „Spiegel“ sind nichts anderes als Kopien von Web-Angeboten, die an einem anderen geografischen Ort vorliegen, in diesem Fall in Australien. Die Internetverbindungen im Jahr 1996 waren wesentlich weniger ausgebaut als heute, und man musste auf die Antworten entfernter Webserver oft recht lange warten. Spiegel wurden erstellt, um diese Wartezeiten zu verkürzen. Heute haben Suchmaschinen über die Welt verteilte *Data Centers*, die den Datenbestand der Suchmaschine und die Verarbeitung der Suchanfragen verteilen. Allerdings müssen Nutzer nicht mehr explizit eines dieser Data Centers auswählen, sondern sowohl der Index als auch die Verarbeitung der Suchanfragen werden automatisch verteilt.

2.4.2 Spezialsuchmaschinen

Von den allgemeinen Suchmaschinen sind die Spezialsuchmaschinen zu unterscheiden (Lewandowski 2009). Spezialsuchmaschinen verfolgen das Ziel, möglichst alle Dokumente von *ausgewählten Websites* zu erfassen.

Spezialsuchmaschinen sind thematisch beschränkt und machen damit eine zielgenaue Recherche möglich, wobei das Ranking speziell auf die von ihnen erschlossenen Dokumente angepasst werden kann, ebenso wie die sachliche Erschließung der Dokumente. Schließlich ergeben sich auch Vorteile bei der Darstellung der Ergebnisse, welche auf den individuellen Zweck der Spezialsuchmaschinen sowie auf das Niveau der Zielgruppe angepasst werden kann (Lewandowski 2009, S. 56). Dass Spezialsuchmaschinen nicht durch Universalsuchmaschinen ersetzt werden können, ergibt sich aus den Problemen der letzteren (eine ausführliche Erläuterung dazu findet sich in Kap. 13):

Abb. 2.4 Inhalte von
Spezielsuchmaschinen



1. Universalsuchmaschinen haben technische Beschränkungen und können (trotz des Etiketts „Universal“) nicht das gesamte Web abdecken,
2. es bestehen finanzielle Hürden, die die Erfassung der Inhalte und ihre Erschließung beschränken,
3. Universalsuchmaschinen richten sich an den Durchschnittsnutzern aus,
4. sie müssen eine einheitliche Erschließung aller indextierten Inhalte betreiben, damit alles gemeinsam durchsuchbar ist.

Spezielsuchmaschinen beschränken sich also willentlich auf einen bestimmten Bereich des Web (s. Abb. 2.4). In der Regel erfolgt die Beschränkung auf bestimmte Quellen, also Websites. Diese Websites werden in der Regel von Hand ausgewählt. Wenn man beispielsweise eine Spezielsuchmaschine für Nachrichten aufbauen möchte, ist es sinnvoll, zuerst die relevanten Nachrichten-Websites zusammenzustellen, die dann von der Suchmaschine kontinuierlich nach neuen Inhalten (Seiten) abgefragt werden.

Eine *Website* ist ein in sich geschlossenes Angebot im Web, das mehrere Webseiten beinhalten kann. Die Abgrenzung erfolgt über die Domain (zum Beispiel spiegel.de) oder über Subdomains bzw. Verzeichnisse (news.spiegel.de bzw. spiegel.de/news).

Bei einer *Webseite* (auch: *Seite*) dagegen handelt es sich um ein einzelnes Dokument, das in der Regel aus Text und dazu gehörenden medialen Elementen (Bilder, Videos, usw.) zusammengesetzt ist.

Die Unterscheidung Website – Webseite ist im Deutschen leider nicht besonders griffig. Im Englischen wird zwischen *website* und *web page* unterschieden; dort sind also auch die Begriffe weniger leicht zu verwechseln.