



X . media . press

**Christian Wenz
Tobias Hauser (Hrsg.)**

X.media.press ist eine praxisorientierte Reihe zur Gestaltung und Produktion von Multimedia-Projekten sowie von Digital- und Printmedien.

2. Auflage

Websites optimieren – Das Handbuch



Springer Vieweg

X . media . press



X.media.press ist eine praxisorientierte Reihe zur Gestaltung und Produktion von Multimedia-Projekten sowie von Digital- und Printmedien.

Christian Wenz · Tobias Hauser
Herausgeber

Websites optimieren – Das Handbuch

2. Auflage

 Springer Vieweg

Herausgeber

Christian Wenz
Arrabiata Solutions GmbH
München, Deutschland

Tobias Hauser
Arrabiata Solutions GmbH
München, Deutschland

ISSN 1439-3107

ISBN 978-3-658-07261-2

DOI 10.1007/978-3-658-07262-9

ISBN 978-3-658-07262-9 (eBook)

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

Die erste Auflage des Werkes erschien 2013 unter dem Titel „Website-Optimierung“ im Addison-Wesley Verlag.

© Springer Fachmedien Wiesbaden 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 Vieweg ist eine Marke von Springer DE. Springer DE ist Teil der Fachverlagsgruppe Springer Science+Business Media www.springer-vieweg.de
(www.springer.com)

Geleitwort



Liebe Leserinnen, liebe Leser,

die ersten Jahre der Informationstechnologie wurden durch Meilensteine wie die Erfindung des Computers (1941), das erste Modem (1958), die erste Maus (1963) und schließlich IBMs ersten Personalcomputer (1981) sowie Microsofts erstes Betriebssystem (1981) geprägt. Mit dem World Wide Web (1990) begann die Ära der Vernetzung von Menschen mittels Informationstechnologie, die massive Verbreitung digitaler Güter und das Entstehen elektronischer Märkte.

1996 breiteten sich erste Websites in Deutschland aus, 1998 änderte sich das Suchverhalten dank einer Revolution namens Google, 2004 revolutionierte Facebook das Kommunikationsverhalten. Heute gehören Websites nicht nur zum guten Ton erfolgreicher Unternehmen und Privatanwender. Es wird von einer Website mehr erwartet als einfach nur da zu sein oder gut auszusehen. Sie muss funktional, leicht bedienbar, performant, sicher und gut zu finden sein.

Zu all diesen Themen gibt es individuelle Forschungsbereiche. Meine Arbeiten befassen sich beispielsweise mit der Funktionalität und Bedienbarkeit von Webseiten durch Auswertung des Nutzerverhaltens beim Surfen. Dies erfolgt unter anderem durch Blickbeobachtung mittels Eye-Tracking, kamerabasierter Emotionserkennung, Messung der Hautleitfähigkeit und Aufzeichnung der Gehirnaktivität mittels EEG. Ein Labor, das solche Messungen ermöglicht, ist heute nicht mehr nur wissenschaftlichen Zwecken vorbe-

halten. Aber natürlich ist der Aufwand dahinter beträchtlich. Und das betrifft nur einen kleinen Teil des zu beachtenden Spektrums an Anforderungen guter Websites.

Umso schwieriger ist die Aufgabe, dieses ganze Spektrum in einem Buch zu vereinen. Genau das gelingt den Autoren in dem vorliegenden Werk mit hohem praktischem Nutzen. Die verschiedenen Teildisziplinen werden kompakt und übersichtlich erklärt. Die Themen mischen dabei technisch anspruchsvolle Bereiche wie die Absicherung der eigenen Website und Website-Performance mit Marketingthemen rund um Newsletter und Webanalyse. So findet sich der Einsteiger in die Weboptimierung gut zurecht, aber auch Webprofessionals erhalten viele nützliche Tipps.

Ich freue mich auf viele bessere Websites!

Professor Dr. Ricardo Büttner

München, Juli 2014

Vorwort

Die Anforderungen an Websites steigen unaufhörlich an. In der ersten Hälfte der 1990er-Jahre war es zunächst wichtig, überhaupt eine Webpräsenz zu haben. Ob diese auf einer eigenen, (damals) teuren *.de*-Domain lag oder bei CompuServe oder Geocities gehostet wurde, erschien nebensächlich. Mit der zunehmenden Versorgung der Bevölkerung mit Internet- und dann Breitbandanschlüssen wurde die Latte höher gelegt: Es reichte nicht mehr, einfach eine in Photoshop erstellte Printanzeige in HTML zu exportieren und online zu stellen – eine echte Webanwendung musste her, mit richtiger Funktionalität und einer Bedienung wie im Web üblich. Der Fokus lag also und liegt immer noch darauf, was grafisch und technisch möglich ist und wie dies umgesetzt werden kann.

Heutzutage, wo jede Firma mit einer Website versorgt ist, werden zusätzliche Anforderungen auf die Agenda gesetzt: Wie wird die Website verbessert – möglichst ohne wieder ganz bei null anzufangen? Dass es Optimierungspotenzial gibt, zeigt sich an vielerlei Stellen.

Wer bei den relevanten Suchbegriffen nicht auf der ersten Seite bei Google & Co. auftaucht, existiert möglicherweise gar nicht und sollte deswegen die Site für Suchmaschinen optimieren und Suchmaschinenmarketing betreiben (Kap. 1 – Stefan Fischerländer, C. W., und Kap. 2 – Katja Heinemann). Damit Kunden nicht nur einmal, sondern öfters bei einer Website vorbeikommen, sind Newsletter ein wichtiges Mittel, wobei es zahlreiche Fallstricke gibt (Kap. 3 – Katja Heinemann). Ohne soziale Medien wie Facebook, Twitter und andere ist heutzutage kaum mehr eine Marketingkampagne denkbar (Kap. 4 – C. W.).

Eine Website kann erst dann richtig erfolgreich sein, wenn sie gut zu bedienen ist (Kap. 5 – T. H.), auch für Personen mit Einschränkungen (Kap. 6 – C. W.). Während wir uns normalerweise über jeden Benutzer freuen, sorgen Angreifer, die – teilweise automatisiert – nach Sicherheitslücken suchen, gerne für Ärger, weswegen wir vorsorglich Maßnahmen ergreifen sollten (Kap. 7 – C. W.). Letztendlich kommt es auch auf die Geschwindigkeit einer Website an, weil das unter anderem Auswirkungen auf die Verkaufszahlen und die Position bei Suchmaschinen hat (Kap. 8 – C. W.).

Die Optik darf bei alledem nicht vernachlässigt werden, wenn beispielsweise das Design optimiert werden soll oder gar ein Relaunch ansteht (Kap. 9 – T. H.). Ob alle genannten Maßnahmen dann auch tatsächlich Wirkung gezeigt haben, lässt sich nur durch entspre-

chendes Controlling feststellen (Kap. 10 – Thomas Kraehe). Dabei sind aus der Praxis adaptierte Beispiele hilfreich.

Wir sind mit diesem Buch am Puls der Zeit – klar also, dass sich das Aussehen oder sogar die Verfügbarkeit der erwähnten und gezeigten Websites und Tools seit der Manuskriptfertigstellung verändert haben. In so einem Fall freuen wir uns besonders über Ihre Rückmeldung, damit wir das in den Folgeauflage berücksichtigen können.

Wir decken also mit dem Buch eine große Bandbreite an Optimierungsaufgaben ab. Aus unserer jahrzehntelangen Erfahrung in Webprojekten kommen wir mit Themen der Website-Optimierung regelmäßig in Kontakt. Wir freuen uns jedoch besonders, dass wir bei einigen Themen Unterstützung von namhaften Branchenexperten erhalten haben, mit denen wir schon seit Jahren zusammenarbeiten: vielen Dank an Katja Heinemann, Thomas Kraehe und Stefan Fischerländer!

An dieser Stelle noch einen herzlichen Dank an das Team, das dieses Buch bereits in der Voraufgabe unterstützt hat: Franz Neumeier und Marcus Schoeberichts haben als Fachlektoren die Kapitel überprüft und dabei wertvolles Feedback gegeben. Marita Böhm hatte ein Auge auf Rechtschreibung, Grammatik und Formulierungen – Sie werden es daran merken, wie gut der Text zu lesen ist. Birgit Ellissen hat als zuständige Erstlektorin das Projekt von seiner Erstkonzeption bis hin zur Fertigstellung begleitet. Bei unserem neuen Zuhause, dem Springer-Verlag, haben Bernd Hansemann und Sybille Thelen die Betreuung nahtlos fortgeführt Herzlichen Dank natürlich auch an alle Kunden der Arrabiata Solutions GmbH (<http://www.arrabiata.de/>), die es uns in spannenden Projekten erst ermöglicht haben, Websites zu optimieren.

München/Starnberg, im August 2014

Christian Wenz, Tobias Hauser

PS: Wir freuen uns über Rückmeldungen zum Buch unter <http://www.arrabiata.de/optimieren>. Sofern (oder – eher – sobald) es Errata zum Buch gibt, veröffentlichen wir sie dort.

Inhaltsverzeichnis

| | | |
|----------|---|----------|
| 1 | Besser auffindbar – Suchmaschinenoptimierung | 1 |
| | Stefan Fischerländer und Christian Wenz | |
| 1.1 | Hartnäckige Mythen | 2 |
| 1.2 | Kennzahlen | 3 |
| 1.3 | Suchdienste im Web | 5 |
| 1.4 | So listen Suchmaschinen die Ergebnisse | 10 |
| 1.4.1 | Relevanz | 10 |
| 1.4.2 | Der HTML-Code wird zerlegt | 11 |
| 1.4.3 | Link Popularity | 13 |
| 1.4.4 | Google PageRank | 13 |
| 1.4.5 | Linktexte | 15 |
| 1.4.6 | Vertrauen zählt | 15 |
| 1.4.7 | Aktuelle Ranking-Kriterien | 17 |
| 1.5 | Die Optimierung durchführen | 18 |
| 1.5.1 | Auf die Ziele kommt es an | 18 |
| 1.5.2 | Keywords auswählen | 19 |
| 1.5.3 | Website-Struktur | 21 |
| 1.5.4 | Eine HTML-Seite optimieren | 22 |
| 1.5.5 | Die Verlinkung verbessern | 23 |
| 1.5.6 | Suchmaschineneintrag | 24 |
| 1.5.7 | Der richtige Einsatz von Domainnamen | 27 |
| 1.5.8 | Optimierung ohne SEO | 28 |
| 1.6 | Die Tücken der Technik | 28 |
| 1.6.1 | JavaScript, Flash & Co. | 28 |
| 1.6.2 | Frames und Iframes | 29 |
| 1.6.3 | Umbau der Website – Weiterleitungen | 31 |
| 1.6.4 | Dynamische Seiten – URL-Parameter | 32 |
| 1.6.5 | Die Geheimnisse der robots.txt-Datei | 33 |
| 1.6.6 | Unfreiwilliger Spam | 35 |

| | | |
|----------|--|-----------|
| 1.7 | Pay-per-Click-Werbung | 38 |
| 1.7.1 | Funktionsweise | 39 |
| 1.8 | 10 Top-Tipps zur Suchmaschinenoptimierung | 39 |
| 2 | Gut vermarktet – Online- und Suchmaschinenmarketing | 41 |
| | Katja Heinemann | |
| 2.1 | Suchmaschinenwerbung – Google AdWords & Co. | 41 |
| 2.1.1 | Google AdWords einrichten | 43 |
| 2.1.2 | Eine AdWords-Kampagne erstellen | 44 |
| 2.1.3 | Kostenkontrolle und Bietermodell | 49 |
| 2.2 | Themenrelevantes Marketing | 51 |
| 2.2.1 | Klassische Bannerwerbung | 51 |
| 2.2.2 | Affiliate-Marketing | 57 |
| 2.2.3 | Produktsuchmaschinen | 63 |
| 2.2.4 | Verzeichnisdienste & lokale Suche | 66 |
| 2.2.5 | Eintrag in Google Maps via Google My Business | 68 |
| 2.3 | Kosten- und Nutzenrechnung | 69 |
| 2.4 | 10 Top-Tipps zum besseren Marketing | 72 |
| 3 | Gut gesendet – E-Mail- und Permission-Marketing | 75 |
| | Katja Heinemann | |
| 3.1 | Erfolgsfaktoren | 76 |
| 3.1.1 | Adressliste | 76 |
| 3.1.2 | An- und Abmeldemechanismus | 79 |
| 3.1.3 | Rechtssicherheit | 81 |
| 3.1.4 | Der richtige Zeitpunkt | 83 |
| 3.1.5 | Format | 83 |
| 3.1.6 | Optimierung für mobile Endgeräte | 84 |
| 3.2 | Die richtigen Inhalte | 87 |
| 3.2.1 | Betreffzeile und Vorschaufenster | 87 |
| 3.2.2 | Transparente Herkunft | 87 |
| 3.2.3 | Themenschwerpunkte | 88 |
| 3.3 | Tools für E-Mail-Marketing | 91 |
| 3.3.1 | Gängige Mailprogramme wie Outlook | 91 |
| 3.3.2 | Beauftragung eines Dienstleisters | 93 |
| 3.3.3 | Bezug über SaaS | 93 |
| 3.3.4 | Plug-ins und Lizenzprodukte | 94 |
| 3.3.5 | Den passenden Anbieter finden | 95 |
| 3.4 | Erfolgskontrolle und Kennzahlen | 96 |
| 3.4.1 | Bounce-Rate | 97 |
| 3.4.2 | Zustellrate | 97 |
| 3.4.3 | Öffnungsrate | 97 |

| | | |
|----------|---|------------|
| 3.4.4 | Klickrate | 98 |
| 3.4.5 | Conversion-Rate | 98 |
| 3.4.6 | Return on Marketing Investment (ROMI) | 98 |
| 3.5 | 10 häufige Fehler beim E-Mail-Marketing | 99 |
| 4 | Erfolgreich im Netzwerk – Social Marketing | 101 |
| | Christian Wenz | |
| 4.1 | Facebook | 102 |
| 4.1.1 | Facebook-Grundlagen | 102 |
| 4.1.2 | Gefällt mir | 104 |
| 4.1.3 | Facebook-Seiten | 110 |
| 4.1.4 | Facebook-Anwendungen | 115 |
| 4.1.5 | Werben auf Facebook | 121 |
| 4.2 | Google+ | 127 |
| 4.2.1 | +1 | 128 |
| 4.2.2 | API-Programmierung | 130 |
| 4.2.3 | Seiten erstellen | 134 |
| 4.3 | Twitter | 134 |
| 4.3.1 | Grundlagen | 134 |
| 4.3.2 | Schaltflächen | 136 |
| 4.4 | Weitere Dienste | 138 |
| 4.4.1 | XING | 138 |
| 4.4.2 | LinkedIn | 138 |
| 4.4.3 | MySpace | 138 |
| 4.4.4 | VZ | 139 |
| 4.4.5 | Loklisten | 139 |
| 4.4.6 | Pinterest | 140 |
| 4.4.7 | So.cl | 141 |
| 4.4.8 | YouTube | 141 |
| 4.5 | Die 10 Top-Social-Media-Dienste | 143 |
| 5 | Usability – vom Nutzererlebnis zur Conversion-Rate | 145 |
| | Tobias Hauser | |
| 5.1 | Ziele finden | 145 |
| 5.1.1 | Nutzererlebnis | 146 |
| 5.1.2 | Zielgruppe | 146 |
| 5.1.3 | Transaktionen | 147 |
| 5.1.4 | Die Ziele-Checkliste | 154 |
| 5.2 | Nutzer in Schwierigkeiten – kritische Punkte | 158 |
| 5.2.1 | Auflösungen, Bildschirmgrößen und mobile Endgeräte | 158 |
| 5.2.2 | Navigation | 166 |
| 5.2.3 | Inhalte und Inhaltsdarstellung | 182 |

| | | |
|----------|--|------------|
| 5.2.4 | Formulare | 190 |
| 5.2.5 | Suche | 193 |
| 5.2.6 | Prozesse | 195 |
| 5.2.7 | Hilfe und Fehler | 199 |
| 5.3 | Messen und Testen | 202 |
| 5.3.1 | Informationen aus Statistiken | 202 |
| 5.3.2 | Heatmaps, A/B-Tests und multivariate Tests | 206 |
| 5.3.3 | Nutzertests | 208 |
| 5.4 | 10 Top-Tipps für bessere Usability | 214 |
| 6 | Websites für alle – Barrierefreiheit | 217 |
| | Christian Wenz | |
| 6.1 | Standards | 219 |
| 6.2 | Web Content Accessibility Guidelines (WCAG) 2.0 | 220 |
| 6.2.1 | Wahrnehmbar | 221 |
| 6.2.2 | Bedienbar | 229 |
| 6.2.3 | Verständlich | 236 |
| 6.2.4 | Robust | 240 |
| 6.3 | Tools | 242 |
| 6.3.1 | Screenreader | 242 |
| 6.3.2 | Validierer | 242 |
| 6.3.3 | Browsererweiterungen | 243 |
| 6.4 | 10 Top-Checkpunkte für Barrierefreiheit | 245 |
| 7 | Das Scheunentor – Sicherheitslücken vermeiden | 247 |
| | Christian Wenz | |
| 7.1 | Hintergründe | 247 |
| 7.2 | Minimalitätsprinzip für Informationen | 248 |
| 7.3 | Cross-Site Scripting (XSS) | 252 |
| 7.3.1 | Die Same-Origin-Policy von JavaScript | 252 |
| 7.3.2 | XSS in Aktion | 253 |
| 7.3.3 | Beispiele für XSS-Angriffe | 256 |
| 7.3.4 | Gegenmaßnahmen | 261 |
| 7.4 | SQL Injection | 270 |
| 7.4.1 | SQL Injection in Aktion | 270 |
| 7.4.2 | Beispiele für SQL Injection | 273 |
| 7.4.3 | Gegenmaßnahmen | 275 |
| 7.5 | Session-Angriffe | 283 |
| 7.5.1 | Angriffe gegen Cookies | 284 |
| 7.5.2 | Session-Hijacking | 285 |
| 7.5.3 | Gegenmaßnahmen | 289 |
| 7.6 | Cross-Site Request Forgery (CSRF) | 293 |

| | | |
|----------|---|------------|
| 7.6.1 | CSRF in Aktion | 293 |
| 7.6.2 | Gegenmaßnahmen | 296 |
| 7.6.3 | CSRF-Schutz mit PHP | 296 |
| 7.6.4 | CSRF-Schutz mit ASP.NET | 299 |
| 7.7 | Clickjacking | 300 |
| 7.7.1 | Clickjacking in Aktion | 300 |
| 7.7.2 | Gegenmaßnahmen | 301 |
| 7.8 | RIA-Angriffe | 302 |
| 7.8.1 | Flash | 303 |
| 7.8.2 | Silverlight | 304 |
| 7.9 | Automatisierung und CAPTCHAs | 305 |
| 7.9.1 | CAPTCHAs in Aktion | 305 |
| 7.9.2 | reCAPTCHA mit PHP | 307 |
| 7.9.3 | reCAPTCHA mit ASP.NET | 310 |
| 7.10 | 10 Top-Sicherheitslücken | 311 |
| 8 | Wider die Langsamkeit – Performance-Optimierung | 313 |
| | Christian Wenz | |
| 8.1 | Weniger Daten | 314 |
| 8.1.1 | JavaScript komprimieren | 315 |
| 8.1.2 | CSS komprimieren | 322 |
| 8.1.3 | HTML komprimieren | 324 |
| 8.1.4 | Bilder komprimieren | 326 |
| 8.2 | Weniger HTTP-Verbindungen | 331 |
| 8.2.1 | Mehrere Domains | 333 |
| 8.2.2 | Dateien zusammenführen | 334 |
| 8.2.3 | CSS-Sprites | 335 |
| 8.2.4 | JavaScript asynchron laden | 340 |
| 8.3 | Caching | 341 |
| 8.3.1 | HTTP-Header | 342 |
| 8.3.2 | CDN | 345 |
| 8.4 | 10 Top-Tipps zur Performance-Optimierung | 346 |
| 9 | Alles neu? – Designoptimierung und Relaunch-Konzeption | 349 |
| | Tobias Hauser | |
| 9.1 | Relaunch – ja oder nein? | 349 |
| 9.2 | Konzeption | 351 |
| 9.2.1 | Konzeptionelle Struktur | 351 |
| 9.2.2 | Frühe Konzeption: Moods, Scribbles und Co. | 354 |
| 9.2.3 | Visuelle Bestandteile | 357 |
| 9.2.4 | Funktionalitäten und Anforderungen | 360 |
| 9.2.5 | Prozesse und Schnittstellen | 365 |

| | | |
|-----------|---|------------|
| 9.2.6 | Wireframes und Klickdummy | 370 |
| 9.2.7 | Migration von Inhalten | 374 |
| 9.3 | Optische Struktur | 375 |
| 9.3.1 | UI Design Patterns | 376 |
| 9.3.2 | Raster | 377 |
| 9.3.3 | Responsive Design | 380 |
| 9.4 | 10 Top-Anzeichen für einen Relaunch | 383 |
| 10 | E-Controlling – Erfolgskontrolle im Netz | 385 |
| | Thomas Kraehe | |
| 10.1 | Zum Begriff | 385 |
| 10.2 | Betriebswirtschaft | 386 |
| 10.3 | Technologie | 386 |
| 10.3.1 | Server-Logs | 387 |
| 10.3.2 | Tracking-Image | 388 |
| 10.3.3 | Tracking mit JavaScript | 389 |
| 10.3.4 | Cookies | 390 |
| 10.3.5 | API-Call | 391 |
| 10.4 | Tools | 391 |
| 10.4.1 | Google Analytics | 392 |
| 10.4.2 | Piwik | 394 |
| 10.4.3 | CMS-Integration | 396 |
| 10.5 | Standardfunktionsumfang | 396 |
| 10.5.1 | Visits | 396 |
| 10.5.2 | Page Impressions | 397 |
| 10.5.3 | Seiten | 398 |
| 10.5.4 | Zeitraum | 399 |
| 10.5.5 | Besucher im Detail | 399 |
| 10.5.6 | Herkunft der Besucher | 401 |
| 10.5.7 | Keywords | 402 |
| 10.5.8 | Technische Ausstattung der Besucher | 405 |
| 10.5.9 | Datenexport | 407 |
| 10.6 | Fortgeschrittenes Tracking | 407 |
| 10.6.1 | Besucherinteresse/Engagement | 408 |
| 10.6.2 | Demografie | 408 |
| 10.6.3 | Ziele | 409 |
| 10.6.4 | E-Commerce-Integration | 409 |
| 10.6.5 | Kampagnen | 411 |
| 10.6.6 | Landing Pages | 413 |
| 10.6.7 | Segmentierung | 413 |
| 10.6.8 | Link-Tracking | 414 |
| 10.6.9 | Heatmaps | 416 |

| | | |
|------------------------|--|------------|
| 10.6.10 | Klickpfade | 418 |
| 10.6.11 | Live-Tracking | 419 |
| 10.6.12 | Soziale Medien | 419 |
| 10.6.13 | Mobile-Tracking | 421 |
| 10.6.14 | Performance | 421 |
| 10.6.15 | Customizing | 422 |
| 10.7 | Apps | 423 |
| 10.8 | Facebook-Tracking | 424 |
| 10.9 | Zurück zur Betriebswirtschaft | 426 |
| 10.10 | Probleme bei der Erhebung | 428 |
| 10.11 | Datenschutz | 429 |
| 10.11.1 | Speicherung der IP-Adresse | 429 |
| 10.11.2 | Einsatz von Cookies | 432 |
| 10.11.3 | Do not track | 434 |
| 10.11.4 | Browser Stealth Mode | 435 |
| 10.11.5 | Canvas Fingerprinting | 438 |
| 10.12 | Fazit | 438 |
| 10.13 | 10 Top-Checkpunkte für das E-Controlling | 439 |
| Sachverzeichnis | | 441 |

Stefan Fischerländer und Christian Wenz

Vorsicht, Floskel: „Die beste Website nutzt nichts, wenn niemand sie besucht.“ So abgedroschen diese Aussage auch klingt, sie ist trotzdem wahr. Nun gibt es natürlich vollkommen unterschiedliche Motivationen, eine Website zu betreiben, und davon ist es abhängig, welche Marketingmaßnahmen sinnvoll sind. Es mag eine ganze Reihe von Szenarien geben, in denen es nicht nötig ist, für eine Website extra Werbung zu machen. Doch im Normalfall gilt heute im Web die einfache Regel: Je mehr Besucher, desto besser.

Nicht zuletzt der große Wirbel, der in den 1990er-Jahren zum Thema Internet inszeniert wurde und Anfang 2000 in einem regelrechten Hype endete, ließ offenbar viele Menschen glauben, es würde reichen, eine Homepage ins Web zu stellen, und schon kämen alle interessierten Surfer dort vorbei. Das große Missverständnis dabei war, dass das Web nicht einfach nur ein neuer Vertriebsweg ist, sondern ein Angebot an den Kunden darstellt. Und für dieses Angebot muss geworben werden, genauso wie niemand auf die Idee käme, einen Laden zu eröffnen, ohne Werbung dafür zu betreiben.

Neben allen aus der klassischen Werbung ins Netz übertragenen Methoden (vor allem Bannerwerbung und Werbemails) gibt es im Internet eine Möglichkeit der Promotion, für die es im „wirklichen Leben“ kein Pendant gibt: die Positionierung in Suchmaschinen. Während gewöhnliche Werbemaßnahmen, egal ob Plakate am Straßenrand oder Werbebanner im Web, das Interesse des Betrachters erst wecken müssen, hat der Nutzer einer Suchmaschine sein Interesse bereits mit der Eingabe eines einschlägigen Begriffs bekundet. Nun gilt es nur noch, ihn zur eigenen Seite zu lotsen – und dort die richtigen Infos zu bieten.

Stefan Fischerländer
Osterhofen, Deutschland
e-mail: stefan@gipfelstolz.de

Christian Wenz 
Arrabiata Solutions GmbH, München, Deutschland
e-mail: christian.wenz@arrabiata.de

Die Zahlen sprechen eine eindeutige Sprache. Die regelmäßigen Erhebungen der Firma Pew Internet Data führten zu interessanten Ergebnissen (jeweils bezogen auf den US-Markt): Anfang 2002 lag der Prozentsatz der Amerikaner, die eine Suchmaschine benutzt haben, bei etwas über 50 % – wohlgemerkt, aller Amerikaner. Zehn Jahre später sind daraus schon 73 % geworden. Jeden einzelnen Tag suchen knapp 60 % aller erwachsenen Internetnutzer Google & Co. auf. Einer anderen (und älteren) US-amerikanischen Studie zufolge finden bis zu 87 % der Erstbesucher einer Homepage über Suchmaschinen dorthin. Erfolgreiches Webpublishing bedeutet heute nicht nur, gute Inhalte mithilfe einer optimalen Technik bereitzustellen, sondern auch, interessierte Nutzer für die gebotenen Informationen oder Dienstleistungen zu finden.

Trotz dieser überzeugenden Argumente dauerte es einige Jahre, bis sich Suchmaschinen als Marketinginstrumente etablierten. Dies liegt wohl nicht nur an manchen Vorurteilen (siehe folgenden Abschnitt), die sich nach wie vor hartnäckig halten, sondern auch an den Zuständigkeiten in den Unternehmen. Die Zusammenarbeit von Marketing- und IT-Abteilung ist in vielen Betrieben noch verbesserungsfähig – doch die Suchmaschinenoptimierung muss genau an dieser Nahtstelle durchgeführt werden.

An dieser Stelle ist ein kurzer Hinweis zum Sprachgebrauch angebracht: Auch wenn es die Logik gebieten würde, Suchmaschinenmarketing als Überbegriff für alle Marketingmaßnahmen, die mithilfe von Suchmaschinen durchgeführt werden, zu benutzen, hatte sich in der Branche eingebürgert, als Suchmaschinenmarketing Werbung mit Pay-per-Click-Anzeigen zu verstehen. Abgekürzt wurde die englische Version SEM („Search Engine Marketing“) benutzt. Inzwischen kann sich für bezahlte Werbung in Google der korrekte Ausdruck *Search Engine Advertising (SEA)* zunehmend durchsetzen.

Suchmaschinenoptimierung hingegen bezeichnet Maßnahmen, um in den normalen – also nicht bezahlten – Suchergebnissen möglichst weit vorn aufzutauchen (genau die Maßnahmen, die im Folgenden ausführlicher besprochen werden). Suchmaschinenoptimierung wird mit SEO für „Search Engine Optimization“ abgekürzt; damit die Sprachverwirrung aber noch weiter steigt, steht das Kürzel SEO auch für Suchmaschinenoptimierer. Wenn Ihnen die Aussage „Ein SEO macht immer SEO, oft aber auch SEA“ sinnvoll erscheint, haben Sie die wichtigsten Abkürzungen verinnerlicht.

1.1 Hartnäckige Mythen

In den Weiten des Internets – und auch in Branchenforen wie etwa der bekannten, vom Hauptautor dieses Kapitels betriebenen Website „SuchmaschinenTricks“ (<http://www.suchmaschinentricks.de/>) – finden sich immer wieder Vorurteile zum Thema Suchmaschinenoptimierung. Auf die hartnäckigsten dieser Mythen wird im Folgenden kurz eingegangen, bevor wir uns mit der eigentlichen Thematik intensiver beschäftigen.

Mythos 1: Topplatzierungen in den Suchmaschinen sind nur durch Zufall zu erreichen.

Probieren Sie es aus. Sie werden sehen, es hat nicht nur mit Zufall zu tun. Die folgenden Abschnitte sollten genügend Informationen liefern, um eine Website auch für beliebte Suchbegriffe weit nach oben zu bringen.

Mythos 2: Topplatzierungen in den Suchmaschinen werden heimlich an zahlende Kunden verkauft.

Auch hier lautet der Ratschlag wieder: Ausprobieren. Sie werden sehen, dass auch für wichtige Begriffe Platz eins ganz ohne Bezahlung möglich ist. Verwechseln Sie das aber nicht mit Pay-per-Click-Werbung (siehe Abschn. 1.7 „*Pay-per-Click-Werbung*“).

Mythos 3: Wirklich gute Seiten setzen sich auch ohne Optimierungen durch.

Natürlich gibt es viele Möglichkeiten, eine Website bekannt zu machen. Suchmaschinen aber bringen qualifizierte Nutzer, die sich durch die Eingabe eines einschlägigen Suchbegriffs als interessiert ausgewiesen haben, auf die Site.

Mythos 4: Suchmaschinenoptimierung ist unmoralisch und wird nur von dubiosen Websites benutzt.

In der Tat haben sich in der Vergangenheit insbesondere Anbieter aus den Schmutzecken des Web mit wenig ehrenhaften Trickereien hervorgetan. Doch ist deshalb der Umkehrschluss noch lange nicht statthaft. Optimierte Seiten helfen den Internetnutzern sogar, gewünschte Informationen leichter zu finden. Das haben inzwischen fast alle renommierten Unternehmen erkannt und lassen sich zielgerichtet über Suchmaschinen Besucher zuführen.

Mythos 5: Die besten Platzierungen erreicht man nur, wenn man für jede Suchmaschine eine eigens optimierte Seite erstellt.

Dies mag früher vielleicht so gewesen sein. Heute haben sich die Methoden der Suchmaschinen so weit angeglichen, dass das nicht mehr nötig ist. Im Gegenteil ist es sogar problematisch, mehrere inhaltlich identische Versionen der gleichen Seite zu veröffentlichen (siehe Abschn. 1.6.6 „*Unfreiwilliger Spam*“).

Mythos 6: Die Angaben im Keywords-Meta-Tag sind wichtig.

Spätestens seit Googles Erfolg ab etwa 2000 spielen die Angaben im Keywords-Meta-Tag keine Rolle mehr bei der Sortierung der Suchergebnisse. Google beachtet den Inhalt dieses Tags nicht. Wenn Sie also von Tipps bezüglich Meta-Keywords lesen, sind diese möglicherweise schon seit Langem veraltet.

1.2 Kennzahlen

Die Zugriffszahlen für Websites sind die Einschaltquoten des Web. Für Angebote, deren Finanzierung schwerpunktmäßig über Werbung erreicht werden soll, sind diese Zahlen existenziell. Doch auch die Internetabteilungen in Unternehmen stehen oft genug vor der Aufgabe, den Erfolg ihrer Website mit Zahlen zu belegen. Die technischen Möglichkeiten des Mediums verführen zu der Annahme, im Web seien alle Zugriffe genauestens messbar.

Dies stimmt im Prinzip auch, doch die tatsächlichen Gegebenheiten lassen eine wirklich korrekte Bestimmung nicht zu.

Dies liegt insbesondere an den verschiedenen Caching-Methoden, die im Internet im Einsatz sind. So speichert jeder moderne Browser einmal angeforderte Seiten und Grafiken zwischen, um bei einer erneuten Anforderung derselben Seite durch den Benutzer die Seite nicht nochmals über das Internet laden zu müssen. Ähnliche Schwierigkeiten ergeben sich durch den Einsatz von Proxyservern in Firmen und bei Providern. Zum einen sind auch Proxyserver meist mit einem Cache ausgestattet, zum anderen erscheinen alle Benutzer hinter dem Proxy mit derselben IP-Adresse. Verschiedene Nutzer sind so nicht mehr zu unterscheiden.

Es kann aber umgekehrt auch vorkommen, dass ein Besucher während des Besuchs einer Website mit unterschiedlichen IP-Nummern im Logfile erscheint: Große Provider haben mehrere Proxys (mit jeweils eigener IP) im Einsatz, die zur Lastverteilung dienen. So kann ein Besucher die erste Seite über den einen und die zweite Seite über einen anderen Proxy angefordert haben. Eine Unterscheidung von einzelnen Visits bzw. PageViews wird aufgrund dieser Probleme erschwert oder gar unmöglich gemacht.

Die wichtigsten Begriffe zur Messung von Kennzahlen werden im Folgenden erklärt.

- *Hits*: Die Anzahl der von einem Webserver abgerufenen Dateien. Jede HTML-, aber auch Grafik-, CSS- oder JavaScript-Datei erzeugt einen Hit. Somit kann eine einzelne Seite mit vielen eingebetteten Grafiken eine ganze Reihe von Hits erzeugen, obwohl nur eine einzige (HTML-)Seite angezeigt wurde. Mit anderen Worten: Die Anzahl der Hits sagt nichts aus über die tatsächlichen Besucherzahlen auf einer Website.
- *PageViews/Page Impressions*: Gibt die Zahl der tatsächlich besuchten Seiten an. Dabei werden nur Seiten mit Inhalt gezählt; reine Weiterleitungsseiten etwa werden nicht gewertet. Bei Websites ohne Frames entspricht das einfach der Anzahl der vom Webserver abgerufenen HTML-Seiten. Sites mit Frames müssen anders gezählt werden: Jedes aufgerufene Frameset zählt als ein PageView, jede weitere vom Benutzer angeforderte Seite ergibt einen weiteren PageView.
- *Ad-Views/AdImpressions*: Die Anzahl der den Besuchern gezeigten Werbebanner. Durch Speicherung von IP-Nummern oder Session-Cookies versucht man zu verhindern, einem einzelnen Besucher dasselbe Banner mehrmals zu präsentieren. Diese Kennzahl hat allerdings in den vergangenen Jahren stark an Bedeutung verloren.
- *Visits*: Ein zusammenhängender Besuchsvorgang. Jedes Mal wenn ein Besucher mit einer neuen IP-Adresse die Site besucht, wird ein zusätzlicher Visit verzeichnet. Dabei ist es unerheblich, ob der Besucher nur eine oder mehrere Seiten der Website ansieht. Da viele IP-Adressen dynamisch vergeben werden, wird nach einer gewissen Zeitspanne (meist 30 Minuten) eine IP wieder als neu angesehen.
- *Rechner/Unique IP*: Die Anzahl der in einem Zeitraum festgestellten unterschiedlichen IP-Nummern, von denen aus eine Website besucht wurde.

Für gewöhnlich lässt sich folgende Relation dieser Zahlen feststellen:

Rechner < Visits < PageViews < Hits

Meist werden Visits und PageViews als Kennzahlen für die Beliebtheit einer Website veröffentlicht, da diese am ehesten Aufschluss über die tatsächlichen Besucherzahlen geben. Außer der absoluten Größe der Zahlen ist auch das Verhältnis von PageViews zu Visits interessant: Je größer diese Zahl ist, umso intensiver wird das Angebot von den Besuchern genutzt.

Alle aufgeführten Zahlen können aus den Logfiles der Webserver gewonnen werden. Dazu nutzt man spezialisierte Programme, die diese Logfiles analysieren und entsprechende Statistiken daraus ermitteln. Einige bekannte Anwendungen sind Webtrekk (<http://www.webtrekk.com/de/home.html>) oder AnalogX (<http://analog.gsp.com/>).

In der Nutzung meist einfacher sind webbasierte Dienste zum Tracken (Nachverfolgen) des Nutzerverhaltens auf einer Website. Bekannte Anbieter solcher Dienste sind z. B. etracker (<http://www.etracker.de/>) oder Google Analytics (http://www.google.com/intl/de_ALL/analytics/) oder selbst zu hostende Alternativen wie Piwik (<http://piwik.org/>).

- **Tip** Einen ausführlicheren Einblick in die Erfolgskontrolle erhalten Sie in Kap. 10, „E-Controlling – Erfolgskontrolle im Netz“.

1.3 Suchdienste im Web

Wer heute von Suchmaschinen spricht, meint in der Regel Volltextsuchmaschinen, die den gesamten Inhalt von Millionen oder gar Milliarden Webseiten in ihre Datenbank aufgenommen haben.

Als Urtyp dieser Suchmaschinen wie Google oder Yahoo! kann Archie angesehen werden. Archie begann seine Tätigkeit 1990 und durchsuchte alle bekannten FTP-Server. Das Programm legte die gefundenen Dateien in einer Datenbank ab, die von den Nutzern abgefragt werden konnte. Einen ähnlichen Weg ging das Gespann Gopher und Veronica. Die Betreiber von FTP-Servern konnten eine speziell formatierte Textdatei ablegen, in der sie Informationen über die gespeicherten Dokumente auf dem Server zugänglich machten.

Die erste automatisch erstellte Datenbank, die Adressen (URLs) von Webseiten enthielt, war *The World Wide Web Wanderer*. Ursprünglich sollte dieses Programm lediglich das Wachstum des Web ermitteln, und so zählte dieser erste Robot nur die Webserver. Kurz nach seinem Start wurde ihm dann beigebracht, die gefundenen Adressen abzuspeichern, und damit war die erste Suchmaschine geboren.

Robots

Der Begriff „Robot“ wird in Zusammenhang mit Suchmaschinen des Öfteren auftauchen. Robots im Internet haben wenig mit ihren Namenskollegen in der Automobilfabrik zu tun. Es handelt sich vielmehr um Programme, die auf einem Server laufen und selbstständig Daten mit anderen Rechnern



Abb. 1.1 Der einstige Pionier WebCrawler ist heute lediglich Oberfläche für die Metasuchmaschine Infospace

im Netz austauschen. Neben den Robots von Suchmaschinen, die oft auch als Spider oder Crawler bezeichnet werden, weil sie sich im Netz wie eine Spinne von Rechner zu Rechner bewegen, gibt es noch andere Arten. Mailbots etwa sammeln Mailadressen von Webseiten, um so an eine Vielzahl von Opfern für Werbemails zu gelangen. Andere Programme wiederum testen regelmäßig, ob ein Webserver zu erreichen ist, und informieren bei einem Ausfall den Administrator.

Allerdings waren die Suchmöglichkeiten sehr begrenzt, da nur nach Adressbestandteilen gesucht werden konnte. Diese Schwäche überwand als Erstes WebCrawler (siehe Abb. 1.1). Diese Suchmaschine wurde 1994 von einem Studenten an der Universität Washington entwickelt. WebCrawler bot als erste Suchmaschine die Möglichkeit, den kompletten Text der gespeicherten Websites zu durchsuchen. Alle danach initiierten Suchmaschinen wie etwa Lycos oder Infoseek hatten WebCrawler als Vorbild und boten keine konzeptionellen Verbesserungen. Auch die heute führenden Suchmaschinen wie Google oder Bing sind im Wesentlichen von WebCrawler beeinflusst.

Im Prinzip besteht jede Suchmaschine aus drei unabhängigen Bestandteilen. Der *Robot* besucht der Reihe nach die Webseiten, die auf einer Liste verzeichnet sind, und speichert die gefundenen Daten ab. Diese werden dann von einem *Indizierer* bearbeitet und so in einer Datenbank (dem *Index*) abgelegt, dass diese möglichst schnell und effektiv durchsucht werden kann. (Dabei werden gefundene Links, die auf bisher unbekannte Seiten verweisen, wiederum dem Robot mitgeteilt.) Das Durchsuchen dieser Datenbank übernimmt das *Frontend* der Suchmaschine, das die Suchanfrage eines Nutzers entgegennimmt und die

Tab. 1.1 Die wichtigsten Suchmaschinen für den deutschsprachigen Raum mit eigener Datenbank

| Suchmaschine | URL |
|----------------|---|
| Yandex | http://www.yandex.com/ |
| Google | http://www.google.de/ |
| Microsoft Bing | http://www.bing.de/ |

Tab. 1.2 Die interessantesten Metasuchmaschinen

| Metasuchmaschine | URL |
|------------------|---|
| Metager | http://www.metager.de/ |
| IxQuick | https://www.ixquick.de/ |

Treffer in einer möglichst sinnvollen Art und Weise listet. Tabelle 1.1 zeigt die wichtigsten Suchmaschinen für den deutschsprachigen Raum mit eigener Datenbank.

Doch es gibt noch weitere Arten von Suchdiensten, die vor einigen Jahren noch recht bedeutend waren, inzwischen aber viel von ihrer einstigen Popularität einbüßen mussten.

Metasuchmaschinen

Anders als Volltextsuchmaschinen pflegen Metasuchmaschinen keine eigene Datenbank. Sie beantworten Suchanfragen, indem sie die Anfragen an andere Suchdienste weiterreichen und die von dort erhaltenen Trefferlisten entgegennehmen. Die so erhaltenen Ergebnisse werden aufbereitet – mehrfache Treffer etwa aussortiert – und als eigenes Ergebnis dargestellt.

Früher waren Metasuchmaschinen recht beliebt, da sie durch die Kombination mehrerer Suchmaschinen einen größeren Datenbestand für Abfragen zur Verfügung hatten. Heute spielt allerdings die Frage der Indexgrößen kaum mehr eine Rolle; damit haben Metasuchmaschinen ihren größten Vorteil eingebüßt. Tabelle 1.2 zeigt eine Übersicht.

Seit den Veröffentlichungen zur weltweiten Überwachung des Internets durch die NSA und andere Geheimdienste haben Metasuchmaschinen wie IxQuick, die sich besonders dem Datenschutz verschrieben haben, einen deutlichen Zulauf erfahren.

Webkataloge

Es gibt im Web konstruktionsbedingt keine zentrale Stelle, die einen Überblick über alle vorhandenen Seiten behält. So stellte sich schon sehr früh für viele Nutzer die Frage, wie sie im Web die gesuchten Informationen finden sollten.

Zunächst versuchten einzelne Personen, Listen interessanter Websites zu führen und damit dem Informationsbedürfnis der anderen Nutzer zu entsprechen. Doch diese Listen wurden schnell sehr lang und damit unhandlich. Also begann man, die Listen inhaltlich zu sortieren und auf einzelne Seiten zu verteilen. Der hierarchisch gegliederte Webkatalog war geboren. Der wichtigste derartige Webkatalog war lange Zeit Yahoo!, der bereits 1994 gestartet wurde, wenngleich Yahoo! heute den Internetnutzern vor allem als Betreiber einer Volltextsuchmaschine und eines sehr populären Portals ein Begriff sein dürfte.

Tab. 1.3 Die wichtigsten deutschsprachigen Webkataloge

| Webkatalog | URL |
|------------|---|
| Allesklar | http://www.allesklar.de/ |
| Dmoz (ODP) | http://dmoz.org/World/Deutsch/ |
| Web.de | http://dir.web.de/ |

Mittlerweile hat Yahoo! allerdings einen Strategiewechsel vollzogen: der Webkatalog ist mittlerweile abgeschaltet, die Suche setzt mittlerweile auf Bing, auch wenn das nur durch einen kleinen Hinweis auf der Seite erkennbar ist.

Bestanden die ersten Webkataloge aus wenigen hundert oder tausend verzeichneten Seiten, so haben die großen internationalen Kataloge inzwischen Millionen von Websites in ihre Kategorien eingeordnet und beschrieben. Die größten deutschsprachigen Webkataloge liegen derzeit bei über eine halbe Million Websites.

Webkataloge werden von Surfern gern genutzt, um sich einen Überblick über die zu einem Thema vorhandenen Websites zu verschaffen. Im Gegensatz zu den (Volltext-) Suchmaschinen nehmen Kataloge keine einzelnen (HTML-)Seiten auf, sondern beschreiben pro Eintrag lediglich eine thematisch abgeschlossene Website.

Zu ausgefallenen Suchanfragen können Webkataloge wegen des relativ geringen Datenbestands oftmals keine Ergebnisse liefern. Deshalb liefern Kataloge heute meist zusätzlich zu den eigenen Treffern die Ergebnisse einer Volltextsuchmaschine. Tabelle 1.3 zeigt eine Übersicht.

Spezielsuchmaschinen

Haben Sie schon mal mit einer normalen Suchmaschine wie etwa Google nach Informationen zu einem thematisch sehr eng umgrenzten Gebiet gesucht? Wenn ja, dann wissen Sie, wie viele völlig unpassende Ergebnisse Sie erhalten, da die menschliche Sprache einfach zu vieldeutig ist. Meinen Sie mit *Kohl* das Gemüse oder den Exkanzler? Interessieren Sie sich bei der Frage nach *Bank* für Geldinstitute oder Sitzgelegenheiten?

Um diesen Problemen aus dem Weg zu gehen, gab es lange Zeit Spezielsuchmaschinen wie Klug Suchen und Sucharchiv, die nur Seiten aus dem gewählten Spezialgebiet beinhalten. Solche Spezielsucher existierten für nahezu alle Themen. Gut sortierte Listen, auf denen diese Experten verzeichnet sind, helfen bei der Suche ungemein. Der Kreis der Anbieter wird allerdings immer kleiner, weil Google und Bing immer besser werden und so in der Tat häufig den Unterschied zwischen Kohl und Kohl erkennen können.

Webportale

Die großen Webportale wie etwa AOL versuchen, möglichst alles anzubieten, was ein Surfer im Onlineleben so brauchen könnte – und dazu zählen auch Suchdienste. Da die Portale zu den meistbenutzten Seiten im Web gehören, sind die dadurch möglichen Besucherzahlen auch enorm. Deshalb ist es wichtig zu wissen, welches Portal die Daten welcher Suchmaschine nutzt. Tabelle 1.4 zeigt eine Übersicht.

Tab. 1.4 Die meistbesuchten deutschen Portalseiten

| Webportal | URL | Suchdienst |
|-----------|---|------------|
| AOL | http://www.aol.de/ | Google |
| Lycos | http://www.lycos.de/ | Bing |
| MSN | http://www.msn.de/ | Bing |
| T-Online | http://www.t-online.de/ | Google |
| Yahoo! | http://www.yahoo.de/ | Bing |

Marktanteile

Suchmaschinenoptimierung heißt also, eine Website so zu gestalten, dass sie in Suchmaschinen möglichst gut gefunden wird. Dabei ist offensichtlich, dass die beste Platzierung sinnlos ist, wenn niemand die fragliche Suchmaschine benutzt. Also müssen genau die Suchmaschinen, die am häufigsten befragt werden, im Zentrum jeglicher Optimierungsmaßnahmen stehen. Doch welche Suchmaschinen sind das?

Um die Marktanteile der Suchmaschinen in Deutschland festzustellen, hilft ein Blick auf das Webbarometer (<http://www.webhits.de/deutsch/index.shtml?webstats.html>) des Webstatistik-Dienstleisters WebHits (siehe Abb. 1.2). Dort wird täglich dargestellt, über welche Suchmaschinen die Nutzer deutscher Websites die an WebHits angeschlossenen Seiten gefunden haben.

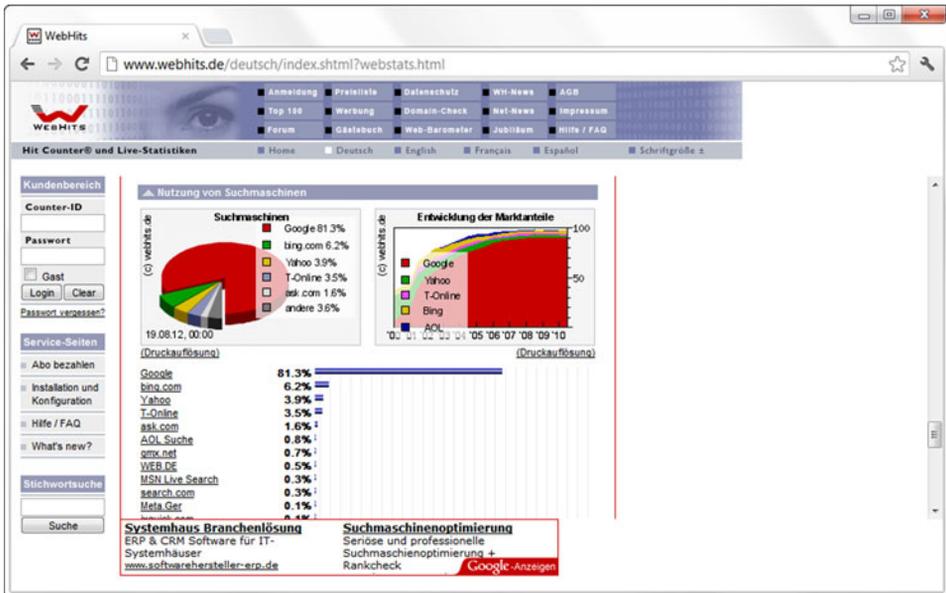


Abb. 1.2 Nutzungsintensitäten der Suchmaschinen in Deutschland (Quelle: webhits.de)

Damit ergibt sich ein recht gutes Bild der Marktanteile – und die Erkenntnis, dass Google weit mehr als 80 % aller Suchanfragen in Deutschland bedient. Die Frage nach der Wichtigkeit beantwortet sich somit von selbst: Aus Marketingsicht ist lediglich Google bedeutend, entsprechend sind auch die Betrachtungen in diesem Kapitel auf Google zugeschnitten.

1.4 So listen Suchmaschinen die Ergebnisse

Sicherlich haben Sie sich bei der Benutzung von Suchmaschinen auch schon gefragt, wie so denn nun ausgerechnet diese zehn Ergebnisse als Antwort auf Ihre Anfrage gezeigt werden. Die passende Reihung (engl.: Ranking) der Suchergebnisse gehört zu den zentralen Aufgaben des Frontends einer Suchmaschine. Wie wichtig und wie schwierig das ist, kann man schon beim Blick auf die Anzahl der Ergebnisse erahnen: Eine Suche nach *fußball* etwa fördert bei Google mehr als 50 Millionen Fundstellen zutage.

Diese Unmenge an Ergebnissen in eine sinnvolle Reihenfolge zu bringen – und das innerhalb von Sekundenbruchteilen – ist das Problem, das eine erfolgreiche Suchmaschine zu lösen hat.

1.4.1 Relevanz

Die größten Suchmaschinen haben eine unvorstellbar große Anzahl von Webseiten gespeichert. Google etwa spricht von „1 trillion“; das ist der amerikanische Begriff für eine Billion, also 1.000.000.000.000¹! Da ist es leicht nachvollziehbar, dass sie für die meisten Suchanfragen Tausende von Ergebnissen liefern. Ein Verfahren, um die vielen gefundenen Seiten in eine sinnvolle Reihenfolge zu bringen, nennt man *Ranking-Algorithmus*.

Dabei bedeutet „sinnvoll“, dass die Ergebnisse, die am besten zur Suchanfrage passen, möglichst weit vorne erscheinen sollen. Mit anderen Worten: Es soll nach Relevanz sortiert werden.

Diese Aufgabe bereitet mehrere Schwierigkeiten. Suchanfragen sind sehr oft nicht eindeutig. Wir hatten ja vorher schon einige Problemfälle vorgestellt: Welche Ergebnisse erwartet jemand, der einfach nach *fußball* sucht? Welche Bedeutung von *kohl* meint der Suchende? Und selbst der Politikernamen *schröder* ist nicht eindeutig – ehemaliger Bundeskanzler, Verteidigungsminister oder Familienministerin?

Abgesehen davon ist es problematisch, die Relevanz eines Textes (Bilder und andere Multimediaelemente werden von den heute üblichen Suchmaschinen komplett ignoriert) nur nach den darin vorkommenden Schlüsselbegriffen zu bewerten. So kann eine Goethe-Biografie durchaus das entscheidende Wort „Goethe“ nur ganz selten benutzen;

¹ Und diese Zahl ist aus dem Jahr 2008 (<http://googleblog.blogspot.de/2008/07/we-knew-web-was-big.html>)! Mittlerweile sind es also noch unvorstellbar viele mehr.

um Wiederholungen zu vermeiden, werden Synonyme eingesetzt: Dichter, Dichterst, Schriftsteller. Wie soll eine Suchmaschine erkennen, dass sich dieser Text vollständig um Goethe dreht, während auf einer anderen Seite das Wort Goethe vielfach eingesetzt wird, aber nur einen sehr geringen Bezug zum Dichterst hat?

Der wesentliche Trick – für die Suchmaschine wie für den Webmaster – besteht darin, nicht nur die Anzahl und Häufigkeit der einzelnen Wörter zu berücksichtigen, sondern auch die Position. HTML ist als Auszeichnungssprache ja gerade dazu erfunden worden, Text durch entsprechende Tags zu strukturieren. Dies kann für eine Relevanzbestimmung benutzt werden.

Mit anderen Worten: Suchmaschinen bewerten das Vorkommen von Begriffen innerhalb „wichtiger“ Tags (vor allem Titel, Überschriften) höher als im normalen Fließtext. Zusammen mit der Anzahl und relativen Häufigkeit des Begriffs kann so eine Bewertung vorgenommen werden, nach der bei einer Suchanfrage dann sortiert wird.

Spätestens seit sich im Web Geld verdienen lässt, zeigt diese Methode deutliche Schwächen. Böse Trickser nutzen die beschriebene und relativ einfache Methode der Relevanzbestimmung aus, um ihre eigenen Seiten bei wichtigen Suchbegriffen ganz nach oben zu bringen – unabhängig davon, ob der Seiteninhalt überhaupt zum gewünschten Begriff passt.

Um diesen Spammern das Handwerk zu legen, berücksichtigen Suchmaschinen heute in erster Linie andere Kriterien, die zur Sortierung der Ergebnisse benutzt werden. Die Folge ist, dass kaum mehr zu durchschauen ist, warum eine Website auf den vordersten Plätzen einer Suchmaschine liegt. Deshalb bedeutet Suchmaschinenoptimierung auch: vermuten, ausprobieren und korrigieren.

- ▶ **Tipp** Sicher kennen Sie den Begriff „Spamming“ von den unerwünschten Massensendungen von Werbemails. Im Zusammenhang mit der Suchmaschinenoptimierung wird dieser Begriff auch als Begriff für den Missbrauch der Suchmaschineneinträge benutzt. (Manchmal wird auch die Bezeichnung „Spamdexing“ verwendet.)

1.4.2 Der HTML-Code wird zerlegt

Welches sind denn nun die konkreten Stellen im HTML-Code, die wichtig sind für gute Positionen? Zwar hat hier jede Suchmaschine ihre eigenen Wertungen, trotzdem gibt es grundlegende Gemeinsamkeiten, die bei der Reihung der Ergebnisse berücksichtigt werden.

Titel

Die wichtigste Position innerhalb des HTML-Codes ist zweifelsohne der Titel. Nur wenn der Suchbegriff hier auftaucht, wird die Seite eine gute Position erreichen. Suchen Sie einfach bei einer Suchmaschine Ihrer Wahl nach einem beliebigen Begriff und kontrollie-

ren Sie dann, an welcher Position das erste Ergebnis auftaucht, das den gewählten Begriff nicht im Titel hat. Sie werden sehr schnell feststellen, welche zentrale Bedeutung der Titel für eine gute Position hat.

Anzahl und Häufigkeit

Ähnlich wichtig sind die (absolute) Anzahl und die (relative) Häufigkeit des gesuchten Begriffs auf den gefundenen Seiten. Je höher die jeweiligen Werte, umso weiter vorne wird die Seite zu finden sein. Allerdings haben die meisten Suchmaschinen hier Grenzwerte festgelegt, da vielfache Wiederholung desselben Begriffs zu den beliebtesten Tricks der Spammer gehört.

Position

Die Position innerhalb des Textes ist aus zweierlei Gründen wichtig: Zum einen bewerten viele Suchmaschinen das Vorkommen eines Begriffs im Text möglichst weit oben als positiv, zum anderen ist es bei Suchanfragen, die aus mehreren Wörtern bestehen, sehr wichtig, dass diese Wörter auch innerhalb der Seite möglichst nahe beieinanderstehen; idealerweise kommen sie sogar in derselben Reihenfolge vor.

Überschriften

Überschriften spielen in HTML, das als Auszeichnungssprache konzipiert ist, eine wesentliche Rolle zur Gliederung von Texten. Deshalb werden Begriffe in großen Überschriften (`<h1>` und `<h2>`) oft als besonders relevant bewertet.

URL

Früher spielte die URL eine recht große Rolle, heute wird sie aber gern überschätzt. Eine Suche nach *suchmaschinen* bringt in den ersten zehn Treffern von Google zwar sieben Seiten, in denen der Suchbegriff im Titel vorkommt, aber lediglich in drei Treffern kommt das Suchwort in der URL vor.

Sonstiges

Andere Auszeichnungsmöglichkeiten von HTML, wie etwa ``, `<u>` oder `<i>`, können von Suchmaschinen ebenso zur Bewertung benutzt werden wie etwa das `alt`-Attribut des ``-Tags. Doch spielen diese Faktoren, wenn überhaupt, nur eine untergeordnete Rolle.

Semantische Elemente in HTML5

HTML5 ist die künftige Version des HTML-Standards. Die Verabschiedung ist zwar erst für 2014 angesetzt, aber bereits seit einigen Jahren unterstützen Browser die dazugehörigen Neuerungen – manche mehr, manche weniger. Eines der Features von HTML5 sind sogenannte semantische Elemente, die insbesondere der Überhandnahme von `<div>`-Elementen im HTML-Markup entgegenzutreten wollen. So gibt es unter anderem die folgenden neuen Tags:

- `<section>`,
- `<article>`,

- `<nav>`,
- `<aside>`,
- `<header>`,
- `<hgroup>`,
- `<footer>`.

Somit ist es denkbar, dass Suchmaschinen zukünftig die semantischen Informationen der Tags auslesen und beispielsweise den Inhalt eines `<aside>`-Elements (Hinweiskasten) oder von `<nav>` (Navigationsbereich) nicht mehr so viel Bedeutung beimessen.

1.4.3 Link Popularity

Die im vorhergehenden Abschnitt beschriebenen Faktoren zur Bestimmung der Relevanz haben den Nachteil, dass sie vom Autor einer Website relativ leicht beeinflusst werden können. Nehmen Sie ein beliebiges Stichwort, stellen Sie es an möglichst alle Positionen (Titel, Überschriften etc.) innerhalb der Seite, und Ihre Seite wird dafür vermutlich – je nachdem, wie umkämpft die Suchbegriffe sind – eine gute Position in den Suchmaschinen erreichen.

Wann begann die Suchmaschinenoptimierung?

So, wie eben beschrieben, begannen schlaue Leute ihre Seiten zu optimieren, um in den Suchmaschinen bei oft vorkommenden Anfragen ganz oben zu erscheinen. Wann die Zunft der Suchmaschinenoptimierer loslegte, lässt sich nicht genau feststellen. Sicher ist nur, dass bereits 1995 erste Beschwerden im Usenet auftauchten, dass Sexshops auch bei komplett unerotischen Suchanfragen weit vorne gelistet wurden. 1996 begannen Suchmaschinen, sich gegen allzu trickreiche Websitebetreiber zu wehren, und installierten Filter, die zumindest die einfachsten Tricks erkennen sollten.

Eine Anleitung zur Suchmaschinenoptimierung aus dem Jahre 1996 ist übrigens nach wie vor im Web zugänglich: <http://www.epage.com/faq/webannounce.html> – eine faszinierende Reise in die Vergangenheit.

Mit dem beginnenden Internetboom wurden gute Platzierungen in den Suchmaschinen plötzlich kommerziell wichtig. Die Folge war, dass die Ergebnisseiten von Suchmaschinen oft nicht die besten, also relevantesten Websites listeten, sondern die mit den besten Tricks im HTML-Code.

1.4.4 Google PageRank

Die Entwickler der Suchmaschine Google fanden ein zusätzliches Kriterium, das sie zur Relevanzbestimmung heranzogen: Die Anzahl der Links, die auf die zu bewertende Seite verweisen, wird als Qualitätsmerkmal gewertet. Dahinter steckt der Gedanke, dass nur qualitativ wertvolle Seiten viele Links von anderen Websites erhalten. Auf diese Weise

wird das Urteil vieler Menschen im Internet mit ins Ranking der einzelnen Suchmaschinen integriert und Spamming-Tricks die Grundlage entzogen. Der Erfolg dieser Idee lässt sich an der ungeheuren Popularität erkennen, die Google heute im Web genießt – und daran, dass quasi alle anderen Suchmaschinen diesen Ansatz ebenfalls aufgegriffen haben.

Der von Google benutzte Algorithmus berücksichtigt aber nicht nur die Anzahl der Links, die auf eine Seite zeigen, sondern auch, welche Qualität diese verweisenden Seiten selbst haben. Dabei drückt Google die Qualität einer Seite mit einem Zahlenwert zwischen null und zehn aus. Zehn steht dabei für die höchste Qualität und wird nur von wenigen Seiten erreicht. Google nennt diesen Zahlenwert *PageRank* und macht ihn über die Toolbar (<http://toolbar.google.de/>) zugänglich.

Seiten mit maximalem PageRank

Die Seite <http://www.searchenginegenie.com/pagerank-10-sites.htm> veröffentlichte regelmäßig eine Liste der Sites, die einen maximalen PageRank von zehn erreicht haben. Dazu gehören (Stand Mai 2012 – danach gab es leider kein Update mehr) neben den Downloadseiten des Adobe Flash Player und Adobe Reader auch die Website der US-Regierung, von Twitter, der Vereinten Nationen sowie einige möglicherweise weniger bekannte Sites wie die „European University Association“. Von PageRank 10 auf 9 abgerutscht sind Facebook – und Google selbst!

Die Berechnung des PageRank ist recht aufwendig, da zur Berechnung des Werts für eine Seite die PageRank-Werte aller Seiten, die dorthin verlinken, berücksichtigt werden müssen. Das heißt, dass diese Berechnung rekursiv durchzuführen ist, denn ein veränderter Wert einer Seite hat durch die Verlinkung im Web Auswirkungen auf alle anderen Seiten.

- ▶ **Tipp** Verfallen Sie nicht dem PageRank-Wahn! Die Jagd nach einem möglichst hohen PageRank kann von den eigentlichen Optimierungszielen ablenken – zudem gibt es wichtigere Faktoren als den PageRank. Außerdem ist die PageRank-Anzeige der Google Toolbar oftmals inkorrekt, da Google den dargestellten Wert nur noch recht selten aktualisiert.

Sichtbarkeitsindex

Die Bonner SEO-Agentur Sistrix (<http://www.sistrix.de/>) bietet als Teil ihrer Produkte einen selbst geschaffenen Indexwert, den „Sichtbarkeitsindex“, an. Dieser gibt an, wie „sichtbar“ eine Domain im Index von Google ist. Dazu werden automatisiert die Suchergebnisse für eine Viertelmillion Begriffe überprüft und – in Abhängigkeit von der Popularität dieser Suchbegriffe (siehe auch Abschn. 1.5.2 „Keywords auswählen“) – die Resultate gewichtet. Zahlreiche Agenturen beschwören die Aussagekraft dieses Index auf geradezu religiöse Art und Weise.

Ganz so einfach ist es aber nicht: Der Sichtbarkeitsindex hängt auch stark von der Themenvielfalt einer Webseite ab. Bei Nischen-Websites ist der Indexwert in der Regel gering, da weniger Suchbegriffe in der Auswahl vorhanden und auch die Suchvolumen entsprechend niedrig sind. Die Suchvolumen wiederum sagen nichts über den realen Traffic auf einer Website aus; ein niedriger Sichtbarkeitsindex bedeutet also nicht automatisch, dass eine Website geringe Abrufzahlen hat.

Insgesamt ist der Sichtbarkeitsindex ein nützlicher Indikator, aber keineswegs ein Allheilmittel.