



Irina Hey, Mario Fischer

Serienspecial: SEO fängt mit OnPage an Teil 5: Technische Aspekte der OnPage-Optimierung

Der letzte Beitrag unseres OnPage-Specials beschäftigte sich vorwiegend mit der Vermeidung doppelter Inhalte im eigenen Webauftritt. In diesem Beitrag sollen nun die technischen Aspekte der OnPage-Optimierung näher erläutert werden. Dabei handelt es sich um Optimierungsfaktoren technischer Natur wie Servereinstellungen, Programmieretechniken und die Performance von Webseiten, die bei der Optimierung eine große Rolle spielen und für den Gesamterfolg in Suchmaschinen mit verantwortlich sind.

- Teil 1: Meta-Tags und Snippetoptimierung
- Teil 2: Einzelseitenoptimierung im sichtbaren Contentbereich
- Teil 3: Interne Verlinkungsstruktur und -strategie
- Teil 4: Vermeidung von Duplicate Content
- ▶ Teil 5: Technische Aspekte der OnPage-Optimierung
- Teil 6: Fehlersuche und -behebung

Wer die eigene Webseite auf Vordermann bringen und fit für Google & Co. machen möchte, kommt an technischen Optimierungsfaktoren in der Regel nicht vorbei. Suchmaschinen basieren auf technisch komplexen Berechnungen und berücksichtigen nicht nur inhaltliche Kriterien, sondern auch das gesamte technische Set-up einer Webseite. Daher erfordert die Suchmaschinenoptimierung nicht nur das Grundverständnis des technischen Aufbaus einer Webseite, sondern setzt Kenntnisse in der Serverthematik und anderen Programmieretechniken voraus.

Für Suchmaschinen sind auch die technischen Faktoren, die über den HTML-Quellcode hinausgehen, durchaus interessant und ggf. für eine RankingEinstufung zu berücksichtigen. Sind Teile einer Webseite für Suchmaschinen gesperrt? Sind einzelne HTML-Dokumente für den Nutzer verfügbar? Und wie schnell reagieren der

Webserver und ggf. angeschlossene Datenbanken auf Benutzeranfragen? Dieser Teil des Serienspecials bringt die technischen Komponenten ans Licht, die den Webmastern helfen sollen, eine Webseite mit dem entsprechenden Know-how so zu optimieren, dass der Webauftritt ein besseres Ranking bei Suchmaschinen erzielt.

Am Anfang war der Server

Bevor man eine Webseite „ins Netz“ stellen kann, benötigt diese einen sogenannten Web-space und einen Webserver, der die einzelnen Webdokumente verwaltet und bei Anfragen dem Nutzer zurückgibt. Die Dienstleistung des Servers nennt man Dienst und der Datenaustausch zwischen Client und Server ist in der Regel durch ein Protokoll festgelegt. Wird also eine Anfrage (Request) an den Server gestellt, gibt er ein bestimmtes Dokument (z. B. HTML- oder PDF-Do-

DIE AUTORIN



Irina Hey ist Head of Marketing und Communications von OnPage.org und eine passionierte Suchmaschinenoptimiererin.

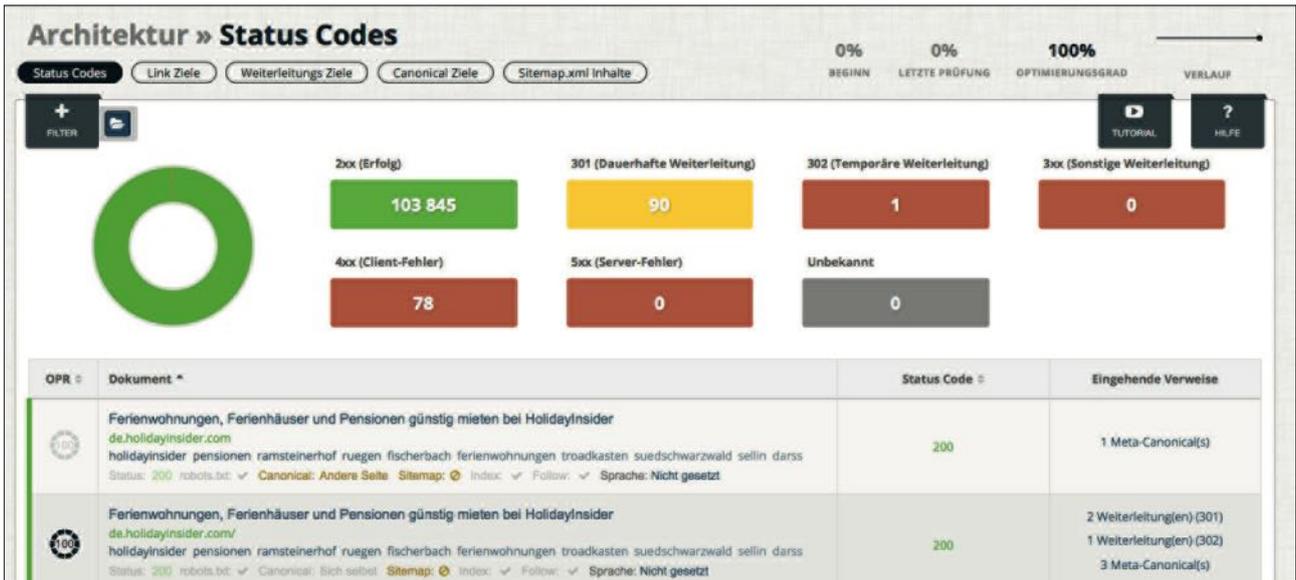


Abb. 1: Beispiel Server-Statuscodes eines Projekts (Quelle: OnPage.org Zoom!)

Code	Beschreibung
200 (Erfolgreich)	Die Anforderung wurde vom Server verarbeitet. Im Allgemeinen bedeutet dies, dass die angeforderte Seite vom Server zurückgegeben wurde. Falls dieser Status für Ihre Datei "robots.txt" angezeigt wird, wurde die Datei erfolgreich von Googlebot abgerufen.
201 (Erstellt)	Die Anforderung wurde erfolgreich ausgeführt. Der Server hat eine neue Ressource erstellt.
202 (Akzeptiert)	Die Anforderung wurde vom Server angenommen, aber noch nicht verarbeitet.
203 (Nicht verbindliche Informationen)	Die Anforderung wurde vom Server verarbeitet, die zurückgegebenen Informationen stammen jedoch möglicherweise von einer anderen Quelle.
204 (Kein Content)	Die Anforderung wurde vom Server verarbeitet, es wird jedoch kein Content zurückgegeben.
205 (Content zurücksetzen)	Die Anforderung wurde vom Server verarbeitet, es wird jedoch kein Content zurückgegeben. Anders als beim Statuscode 204 muss der Anfragende bei dieser Antwort die Dokumentenansicht zurücksetzen, z. B. getätigte Eingaben in einem Formular löschen.
206 (Teil-Content)	Der Server hat eine partielle GET-Anforderung verarbeitet.

Abb. 2: Bedeutung der Statuscodes, die 2xx zurückgeben

kument) an den Client zurück. Über eine spezielle URL ruft man das angefragte Dokument auf dem Server ab und es erscheint z. B. im Browser. Diese Vorgehensweise passiert in Millisekunden und bildet die Grundlage des Webs.

Server können je nach Anforderung in unterschiedlichen Konfigurationen bei verschiedenen Anbietern gebucht werden (z. B. Strato, 1&1, Host-Europe, Goneo u. a.). Außerdem gibt es einige Serverarten, die sich v. a. im Leistungsvolumen und den Sicherheitseinstellungen unterscheiden. Es ist zu empfehlen, sich hierfür genauer über

alle Möglichkeiten des Servereinsatzes und die Kosten zu informieren, bevor man mit einem Webprojekt an den Start geht. Danach macht ein Umzug in der Regel deutlich mehr Aufwand. Abhängig von der Projektgröße gibt es vielfältige Angebote und viele Vergleichsmöglichkeiten.

Server-Statuscodes

Ein Server-Statuscode (auch HTTP-Statuscode) wird bei jeder Serveranfrage zurückgegeben. Statuscodes teilen mit, ob die Anfrage vom Server erfolgreich bearbeitet wurde. Im Fehlerfall

wird ein Code aus- bzw. zurückgegeben, der auf den Fehler hinweist. So können Webmaster einfach nachvollziehen, ob das Webdokument an den Anfragenden erfolgreich ausgeliefert oder weitergeleitet wurde oder ob der Request grundsätzlich fehlgeschlagen ist. Hierfür gibt es viele verschiedene Server-Statuscodes. Für die Suchmaschinenoptimierung sind einige wenige Statuscodes von besonderer Bedeutung. Statuscodes können mithilfe verschiedener freier und kostenpflichtiger Tools ermittelt werden.

Generell teilen sich die Statuscodes wie folgt auf:

1xx (Vorläufige Antwort)

Diese Statuscodes bedeuten, dass nur eine vorläufige Antwort zurückgegeben wurde und der Anfragende weitere Maßnahmen ergreifen muss.

2xx (Erfolgreich)

Diese Statuscodes geben an, dass die Anforderung vom Server erfolgreich verarbeitet wurde. Insbesondere der Statuscode 200 steht für die erfolgreiche Ausführung der Anforderung. Eine hohe Anzahl von 200er-Statuscodes bedeutet also, dass die angeforderten Webdokumente problemlos aufgerufen werden konnten und die Nutzer keine Schwierigkeiten damit hatten.

TIPP

Viele Contentmanagementsysteme leiten ohne entsprechende Modifikation interne Links leider häufig noch immer fehlerhaft mit dem Statuscode 302 um. Unter Umständen verliert man damit eine Menge interne Linkpower und nicht selten steigt das Ranking merklich an, wenn man diesen Fehler beseitigt. Eine einmalige Prüfung des eigenen Systems lohnt sich also auf alle Fälle.

3xx (Weitergeleitet)

Zur Ausführung dieser Art der Anforderung an den Server sind weitere Maßnahmen erforderlich. Diese Statuscodes werden häufig für Weiterleitungen verwendet. Weiterleitungen von URLs von einer Seite zu anderen bedeuten in der Regel nichts anderes, als dass die URLs einer Seite sich geändert haben. Je nach Art der Weiterleitung gibt ein Statuscode Auskunft darüber, wo sich der neue Speicherort des Webdokuments auf dem Server befindet.

Suchmaschinenbetreiber empfehlen, generell die serverseitigen 301-Weiterleitungen zu nutzen. Der 301-Statuscode (permanent Redirect) bedeutet, dass die angeforderte URL dauerhaft zu einem neuen Speicherort umgezogen ist. Somit können Nutzer und Suchmaschinen die richtige Adresse gleich finden. Aber nicht nur das, auch die Linkkraft wird bei einer Umleitung mit 301 vollständig vererbt. Wenn Links auf einer umgezogenen Seite vorhanden sind, empfiehlt es sich auf jeden Fall, mit einem permanenten Redirect umzuleiten.

301-Weiterleitungen sind in folgenden Fällen besonders empfehlenswert:

- » a) Wenn die Startseite eines Web-auftritts unter vielen Adressen aufrufbar ist:
http://beispiel.de/home, http://home.beispiel.de oder http://www.beispiel.com
In der Regel wird nur eine Variante bevorzugt (kanonische Variante).

Code	Beschreibung
300 (Mehrfachauswahl)	Der Server hat bei dieser Anforderung die Wahl zwischen mehreren Aktionen. Entweder wählt der Server eine vom Anfragenden (User-Agent) angeforderte Aktion aus oder stellt dem Anfragenden eine Liste mit Aktionen zur Wahl.
301 (Dauerhaft verschoben)	Die gewünschte Seite wurde dauerhaft an einen neuen Speicherort verschoben. Gibt der Server diese Meldung als Antwort auf eine GET- oder HEAD-Anforderung zurück, wird der Anfragende automatisch an den neuen Speicherort weitergeleitet. Verwende diesen Code, um Googlebot zu informieren, dass eine Seite oder Website dauerhaft an einen neuen Speicherort verschoben wurde.
302 (Zeitweilig verschoben)	Der Server gibt zurzeit auf die Anforderung eine Seite von einem anderen Speicherort zurück. Der Anfragende soll jedoch auch bei zukünftigen Anforderungen den ursprünglichen Speicherort verwenden. Dieser Code ist dem Statuscode 301 ähnlich, da der Anfragende auch hier bei einer GET- oder HEAD-Anforderung zu einem anderen Speicherort geleitet wird. Verwende diesen Code jedoch nicht, um Googlebot über den neuen Speicherort einer Seite oder Website zu informieren, da Googlebot bei diesem Code auch weiterhin den ursprünglichen Speicherort durchsucht und indiziert.
303 (Anderen Speicherort aufrufen)	Dieser Code wird vom Server zurückgegeben, wenn der Anfragende zum Abrufen der Antwort eine separate GET-Anforderung an einen anderen Speicherort senden muss. Alle Anforderungen mit Ausnahme von HEAD-Anforderungen werden automatisch zu diesem anderen Speicherort weitergeleitet.
304 (Nicht geändert)	Die gewünschte Seite wurde seit der letzten Anforderung nicht geändert. Bei dieser Antwort des Servers wird kein Seiten-Content zurückgegeben. Konfiguriere Ihren Server so, dass diese Antwort ("If-Modified-Since"-HTTP-Header genannt) zurückgegeben wird, wenn eine Seite seit der letzten Anforderung durch den Anfragenden nicht geändert wurde. Dies spart Bandbreite und Overhead, da Ihr Server Googlebot so mitteilen kann, dass eine Seite seit dem letzten Crawling nicht geändert wurde.
305 (Proxy verwenden)	Der Anfragende kann nur über einen Proxy-Server auf die gewünschte Seite zugreifen. Bei dieser Antwort des Servers wird auch der zu verwendende Proxy angegeben.
307 (Vorübergehende Weiterleitung)	Der Server gibt zurzeit auf die Anforderung eine Seite von einem anderen Speicherort zurück. Der Anfragende soll jedoch auch bei zukünftigen Anforderungen den ursprünglichen Speicherort verwenden. Dieser Code ist dem Statuscode 301 ähnlich, da der Anfragende auch hier bei einer GET- oder HEAD-Anforderung zu einem anderen Speicherort geleitet wird. Verwende diesen Code jedoch nicht, um Googlebot über den neuen Speicherort einer Seite oder Website zu informieren, da Googlebot bei diesem Code auch weiterhin den ursprünglichen Speicherort durchsucht und indiziert.

Abb. 3: Bedeutung der Statuscodes, die 3xx zurückgeben

301 Weiterleitung einer Domain-URL von ohne auf mit www:

```
RewriteEngine On
RewriteCond %{HTTP_HOST} ^domain.de [NC]
RewriteRule ^(.*)$ http://www.domain.de/$1 [L,R=301]

Alternative Schreibweisen:

RewriteCond %{HTTP_HOST} !^www\.domain\.de$ [NC]
RewriteRule ^(.*)$ http://www.domain.de/$1 [L,R=301]

Oder

RewriteCond %{HTTP_HOST} !^www.domain.de$ [NC]
RewriteRule ^(.*)$ http://www.domain.de/$1 [L,R=301]
```

Abb. 4: 301-Weiterleitungsbeispiel von ohne www auf mit www.

- Alle anderen sollten permanent (mit 301) weitergeleitet werden. Auch in den Webmastertools kann die kanonische URL festgelegt werden.
- » b) Beim Umzug von einer Domain zu einer anderen. Auch hier empfiehlt sich der Einsatz von 301-Redirects, um sicherzustellen, dass auch die Linkkraft der alten Seite auf die neue Seite übertragen wird.

Wie leite ich richtig weiter?

Je nachdem, welchen Server man im Einsatz hat, gibt es verschiedene Vorgehensweisen. Wenn ein Apache-Server vorliegt, gibt es dort eine Möglichkeit eine htaccess-Datei zu hinterlegen und dort die Weiterleitungen in Textform einzutragen. Für Webtechniker sollte dies in der Regel ohne Probleme zu bewerkstelligen sein.



301 REDIRECT

Suchmaschinen wissen, dass die URL dauerhaft mit dem 301-Redirect umgezogen ist. Die Kraft der eingehenden Links wird mit übertragen.

302 REDIRECT

Beim 302-Redirect wird die Seite nur vorübergehend weitergeleitet. Suchmaschinen vermuten, dass die alte Adresse noch besteht. Linkkraft wird nicht übertragen.

Abb. 5: Der wesentliche Unterschied zwischen 301- und 302-Redirect

Code	Beschreibung
400 (Ungültige Anfrage)	Der Server konnte die Syntax der Anforderung nicht interpretieren.
401 (Nicht autorisiert)	Die Anfrage erfordert eine Authentifizierung. Der Server gibt diese Antwort unter Umständen für eine Seite nach einer Anmeldung zurück.
403 (Verboten)	Die Anforderung wurde vom Server zurückgewiesen. Falls dieser Statuscode beim Crawlen gültiger Seiten auf Ihrer Website an Googlebot zurückgegeben wird (dies wird in den Google Webmaster-Tools auf der Seite Crawling-Fehler unter Status angezeigt), wird der Zugriff durch den Googlebot möglicherweise von Ihrem Server oder Host blockiert.
404 (Nicht gefunden)	Die gewünschte Seite wurde vom Server nicht gefunden. Dieser Code wird häufig zurückgegeben, wenn eine Seite angefordert wird, die nicht auf dem Server vorhanden ist. Falls Ihre Website über keine robots.txt-Datei verfügt und dieser Statuscode in den Google Webmaster-Tools auf der Seite Blockierte URLs angezeigt wird, ist der Status korrekt. Falls dieser Statuscode zurückgegeben wird, obwohl eine robots.txt-Datei vorhanden ist, weist Ihre robots.txt-Datei möglicherweise einen falschen Namen oder Speicherort auf. Die Datei muss unter dem Namen robots.txt auf der obersten Ebene der Domain gespeichert sein. Wenn Du den Status für durch den Googlebot gecrawte URLs siehst, ist der Googlebot sehr wahrscheinlich einem ungültigen, veralteten oder falsch geschriebenen Link auf einer anderen Seite gefolgt.

Abb. 6: Bedeutung der Statuscodes, die 4xx zurückgeben

Hinweis: Vorsicht bei 302-Weiterleitungen!

Allgemein sind die 302-Weiterleitungen nicht empfehlenswert. Sie bedeuten, dass der Speicherort des Webdokuments nur vorübergehend unter einer anderen Adresse erreichbar ist und irgendwann wieder „zurückziehen“ wird. Deshalb betrachtet Google diese URLs als „vorübergehend“ und auch der Linkjuice, also die Kraft und/oder Bedeutung, die ein Link einer empfangenden Seite übertragen kann, wird hier eben nicht weitergegeben. Achten Sie bewusst drauf, dass Ihre Webseite so wenig wie möglich oder gar keine 302-Redirects beinhaltet. Dann kann nichts passieren.

4xx (Anfragefehler)

Diese Statuscodes deuten darauf hin, dass die Anforderung wahrscheinlich einen Fehler enthält, der die Verarbeitung durch den Server verhindert.

404-Fehler-Statuscodes: der Schrecken der Nutzer?

Wenn eine Serveranfrage aus bestimmten Gründen scheitert, bekommt der Besucher eine Fehlermeldung „Objekt nicht gefunden“ (404) im Browser angezeigt. Das ist für den Nutzer besonders ärgerlich. Die mögliche Folge: Der User klickt zurück oder ist gar frustriert und schließt die Seite.

Die 404-Seiten können aber auch als ein ganz normaler Teil der Navigation einer Webseite betrachtet werden.

Es gibt keinen Grund, sich über sie zu ärgern, sondern man muss sie finden und den Nutzer damit im Idealfall begeistern. Dies könnte folgendermaßen aussehen:

1. Automatische Weiterleitung auf Startseite/Kategorie
2. Physikalisches Löschen des Webdokuments vom Server
3. Einrichtung einer aussagekräftigen oder (warum nicht?) auch etwas lustigen 404-Seite und einer Angabe der verschiedenen Möglichkeiten zur weiteren Navigation

Wann können 404-Statuscodes wirklich schaden?

Für Suchmaschinen sind viele 404-Fehler in der Regel ein Indiz dafür, dass der Webauftritt nicht ausreichend gepflegt wird. Eine schlecht gepflegte Webseite liefert nachvollziehbar auch eine schlechte User Experience. Es ist durchaus wahrscheinlich und viele Experten gehen auch davon aus, dass sich dies insgesamt negativ auf die Gesamtperformance in Suchmaschinen auswirken kann. Die einzelnen Fehlerseiten werden aus dem Ranking bewusst entfernt, da die Suchmaschine ihren Nutzern voll funktionierende und schnelle Informationen in den Suchergebnissen präsentieren und nicht auf Fehlerseiten verweisen möchte.

Vor allem bei einem ungewöhnlich hohen Vorkommen von 404-Seiten kann es kritisch werden. Wenn eine Webseite www.meinedomain.de beispielsweise 45.000 einzelne Seiten besitzt, aber man bei 10.000 davon einen 404-Fehler erhält – dann scheint da etwas nicht in Ordnung zu sein. Das könnte ein Hinweis auf ein größeres Problem mit dem technischen SEO sein. Das Nutzerengagement und der fallende Traffic können eine Senkung der Sichtbarkeit des gesamten Webauftritts hervorrufen.

Eine weitere mögliche negative Auswirkung der 404-Seiten kann eintreten,

TIPP

Bitte beachten Sie, dass das Löschen einer Seite in einem CMS oder Shop-system die Seite nicht immer auch tatsächlich vom Server löscht, sondern nur ihre Erreichbarkeit über die Benutzernavigation unterbindet. Bleibt die Seite aber weiterhin durch den Direktaufruf der Adresse (URL) erreichbar, wird sie auch zukünftig von den Bots der Suchmaschinen auf- bzw. abgerufen. Das allein stellt noch kein Problem dar. Dieses entsteht vielmehr erst im Lauf der Zeit, und zwar dann, wenn die von dieser vermeintlich gelöschten Seite angelinkten Inhalte wie z. B. Bilder irgendwann nicht mehr verfügbar sind. Die Wahrscheinlichkeit steigt von Jahr zu Jahr, dass immer mehr abgehende Links von dieser Seite nicht mehr da sind und damit 404-Fehler erzeugen. Bei vereinzeltem Auftreten stellt auch dies kein gravierendes Problem dar. „Pflegt“ man aber mit einem solchen – nicht gelöschten – Websystem über Jahre hinweg seinen Webauftritt, können leicht Tausende verwaiste Seiten auf dem Webserver liegen bleiben – auch und gerade nach einem Systemwechsel. Suchmaschinen können in der Regel nicht erkennen, dass derart entstandene Informationsruinen nicht Bestandteil des Webauftritts sind, und errechnen über Kennzahlen, dass der Sitebetreiber offenbar nicht die nötige Sorgfalt walten lässt. Wenn rein rechnerisch jeder zehnte Klick einen 404-Fehler erzeugt, macht es wohl weniger Sinn, Suchende auf eine solche Domain zu schicken.



Abb. 7: Lustiges Beispiel einer 404-Seite von Southpark.com – ganz im Stile der Zeichentrickserie

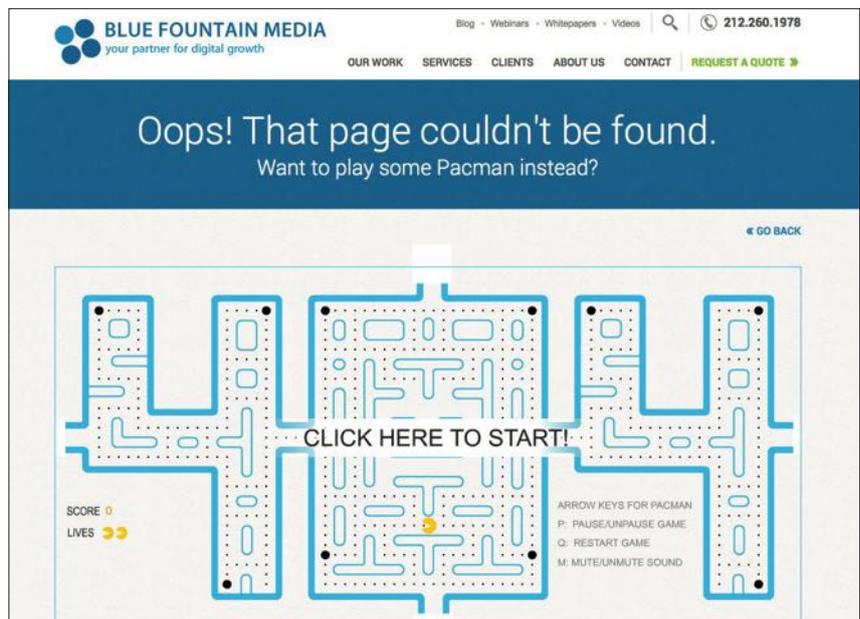


Abb. 8: Ein Retro-Spiel als Zeitvertreib auf der 404-Seite – nicht zielführend, jedoch ist die vollständige Navigation zum Weitermachen vorhanden (Quelle: <http://www.bluefountainmedia.com/irgendwas>)

wenn eine URL wertvolle interne Links enthält. Jede URL, die Links enthält, sollte im Idealfall auf funktionierende Seiten umgeleitet werden. Gibt es keinen einzigen Link auf der betroffenen Unterseite, dann kann diese vom Server direkt gelöscht werden.

Insgesamt lässt sich sagen, dass der Server-Statuscode 404 für SEO eine besondere Rolle einnimmt. Hier ist es entscheidend, dem Nutzer die beste Navigationserfahrung zu bieten. Sollte der Webseitenbesucher auf eine Fehlerseite stoßen, darf er nicht verzweifeln, sondern soll abgeholt und entsprechend im Webauftritt weitergeleitet werden. Schöne und ganz abgedrehte Beispiele

von 404-Seiten findet man bei den Kollegen aus den USA vermehrt.

5xx (Serverfehler)

Diese Statuscodes bedeuten, dass beim Verarbeiten der Anforderung ein interner Fehler auf dem Server aufgetreten ist. In der Regel werden diese Fehler vom Server selbst und nicht von der Anforderung verursacht. Hier empfiehlt es sich, den Hoster des Servers direkt zu kontaktieren. Für die Seitenoptimierung sind solche Fehler unvorteilhaft. Auch hier bleiben die Besucher aus, wenn sie solche Seiten ansurfen.

Die Kenntnisse über die Statuscodes der eigenen Webseiten helfen enorm

bei der Optimierung und Identifizierung technischer Fehler im Zusammenhang mit Serveranfragen.

Accessibility

Bevor eine Webseite in den Suchergebnissen ranken kann, muss diese von den Suchmaschinen zunächst gefunden werden. Ein typisches Business-Szenario hierzu kann folgendermaßen ablaufen: Ein Webmaster möchte einen Online-Shop eröffnen. Die ersten Schritte sind getan: Der Server ist gebucht, die Webseite steht. Der frischgebackene Shopbetreiber ist gut gestimmt und erwartet einen großen Andrang auf seinen Online-Shop. Doch

schnell kommt die Ernüchterung – die Besucher bleiben aus! Auch in den Suchergebnissen findet man die Seite einfach nicht. Einsteiger im Web kämpfen anfangs meist mit einer der Grundvoraussetzungen für gutes Ranking bei Suchmaschinen: Gefunden werden. Doch wie erreicht man die sogenannte Sichtbarkeit (Visibility) bei diesen Suchmaschinen im Web? Zunächst sollte geprüft werden, ob die Webseite überhaupt gut zugänglich ist und die einzelnen Dokumente durch die Suchmaschine problemlos gecrawlt werden können.

Um das festzustellen, gibt es zahlreiche Tools. In den Webmastertools von Google wird gemeldet, wann und ob der Suchmaschinenbot vorbeikommt. Nachfolgend werden einige Techniken und Vorgehensweisen vorgestellt, deren Einsatz einen Einfluss auf die Zugänglichkeit einer Webseite haben kann.

**Komm' rein!
Die robots.txt-Datei**

Die robots.txt ist eine gewöhnliche Textdatei, die in jedem beliebigen Texteditor erstellt werden kann. Diese Textdatei kann aber auch eine große Rolle für die Zugänglichkeit einer Webseite spielen. In einer robots.txt kann der Webseitenbetreiber festlegen, welche Verzeichnisse durch Suchmaschinen aufgenommen oder für diese gesperrt werden. Bevor der Suchmaschinenbot die Webseite crawlt, schaut er meistens in erster Linie nach, ob eine robots.txt vorhanden ist und welche Informationen für den Bot* dort hinterlegt sind.

Die Gefahr, dass man Webseitenverzeichnisse zum Teil nicht findet, besteht dann, wenn man versehentlich wichtige Verzeichnisse aussperrt. Logischerweise bleiben für die gesperrten Verzeichnisse die Sichtbarkeit und die Rankings aus. Die Datei enthält also jegliche Anweisungen für Webrobots. Suchmaschinen halten sich auch meistens an die Anweisungen.

Code	Beschreibung
500 (Interner Serverfehler)	Der Server kann die Anforderung aufgrund eines Fehlers nicht ausführen.
501 (Nicht implementiert)	Der Server kann die Anforderung aufgrund fehlender Funktionen nicht ausführen. Dieser Code kann beispielsweise zurückgegeben werden, wenn der Server die Anfragemethode nicht erkennt.
502 (Falsches Gateway)	Der Server hat als Gateway oder Proxy agiert und von einem vorgeschalteten Server eine ungültige Antwort erhalten.
503 (Service nicht verfügbar)	Der Server ist derzeit aufgrund von Überlastung oder Wartungsarbeiten nicht verfügbar. Im Allgemeinen ist dies ein vorübergehender Zustand.
504 (Gateway-Timeout)	Der Server hat als Gateway oder Proxy agiert und nicht rechtzeitig eine Anforderung vom vorgeschalteten Server erhalten.
505 (HTTP-Version nicht unterstützt)	Die in der Anforderung verwendete Version des HTTP-Protokolls wird vom Server nicht unterstützt.

Abb. 9: Bedeutung der Statuscodes, die 5xx zurückgeben

```

Beispiel: robots.txt

# robots.txt zu http://www.beispiel.de/

User-agent: UniversalRobot/1.0
User-agent: mein-Robot
Disallow: /quellen/dtd/

User-agent: *
Disallow: /unsinn/
Disallow: /temp/
Disallow: /newsticker.shtml
    
```

Abb. 10: Beispiel einer robots.txt-Datei

Erstellen einer robots.txt

Die robots.txt muss auf dem sog. Root-Verzeichnis einer Domain abgelegt werden. Der Dateiname lautet robots.txt und muss zwingend kleingeschrieben werden. Die Datei kann über den Browser jederzeit und von jedermann aufgerufen werden, z. B. www.meinedomain.de/robots.txt. Mithilfe dieser Datei können die Suchmaschinenbots Verzeichnisse aussparen, die für das Ranking uninteressant oder unwichtig sind (z. B. Check-out-Prozess, Druckseiten, [Duplicate Content*](#) etc.).

Was ist mit Flash-Webseiten, JavaScript, CSS & AJAX?

Noch immer bevorzugen nicht wenige Webmaster den Einsatz von Flash auf ihren Webseiten – vor allem im Kreativbereich ist diese Technologie häufiger anzutreffen. Während es in

der Vergangenheit der Suchmaschine nicht möglich war, Flash zu lesen und zu indexieren, ist das heute nicht mehr der Fall. Der Googlebot kann die Flash-SWF-Dateien indexieren, die Inhalte aus diesen Dateien auslesen, die URLs in SWF-Dateien erkennen und den darin enthaltenen Links folgen. Eine Garantie besteht natürlich nicht, dass die Suchmaschine alles vollständig erfasst. Google schränkt daher ein: „Wir können jedoch keine Garantie geben, dass sämtliche Inhalte gecrawlt und indexiert werden, ganz gleich, ob es sich um Flash handelt oder nicht“ (<http://einfach.st/gflash>). Das Gleiche gilt für JavaScript/CSS- und AJAX-Technologien. Während Google früher noch Schwierigkeiten hatte, solche Dateien zu interpretieren, ist es heutzutage kein nennenswertes Problem mehr. Viele andere Suchmaschinen können solche Inhalte allerdings

*siehe Online-Glossar unter www.websiteboosting.com

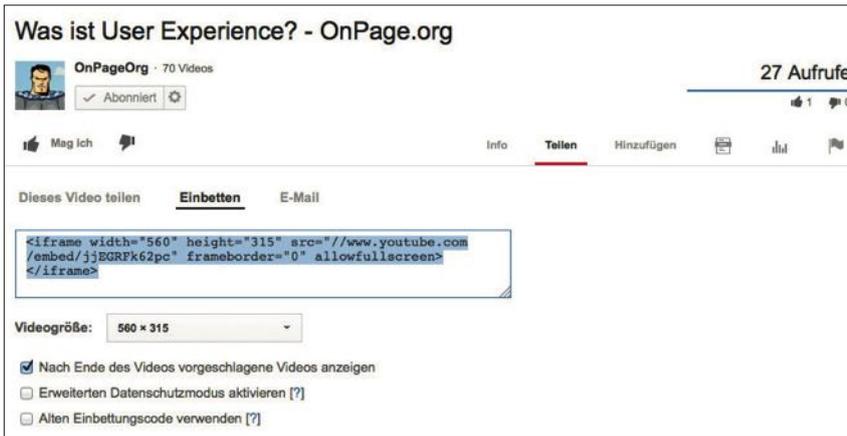


Abb. 11: Generierter iFrame-Code bei YouTube

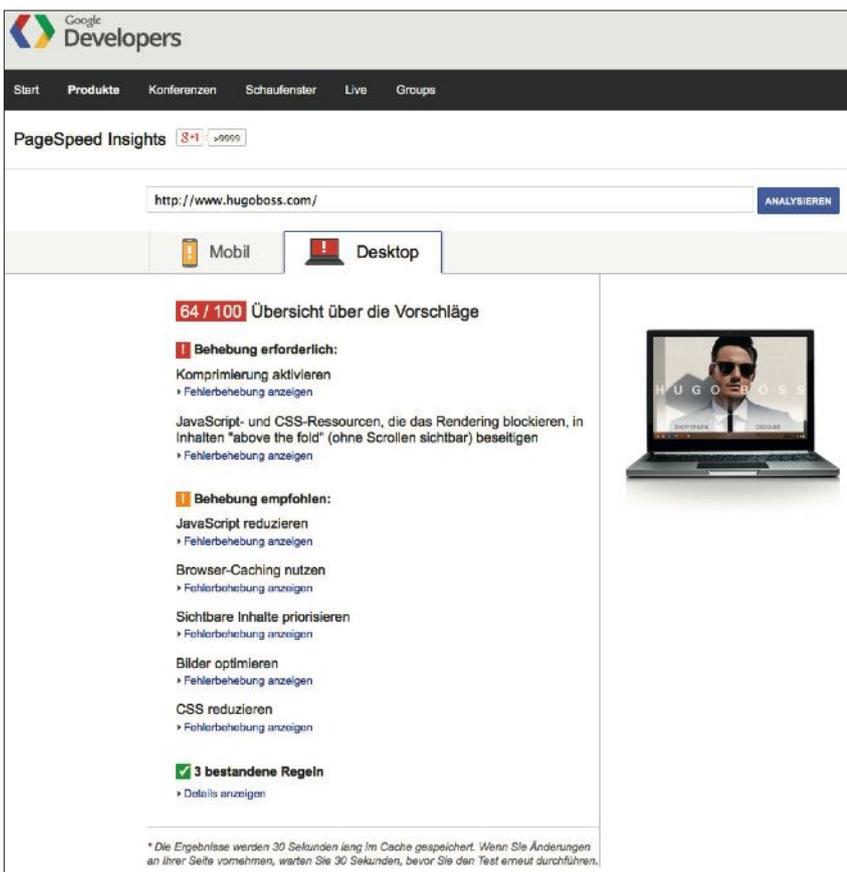


Abb. 12: Beispiel einer PageSpeed-Analyse bei Google PageSpeed Insights

nicht so frei lesen und interpretieren.

Technologien wie Flash, JavaScript und CSS sind daher eher für Grafiken und visuelle Darstellungen zu verwenden. Für die Inhalte hingegen, die bevorzugt in Suchmaschinen gefunden werden sollen, ist HTML die bessere Lösung.

Verwendung von iFrames

iFrames (Abk. für Inlineframes) sind Elemente zur Darstellung von Inhalten innerhalb eines HTML-Dokuments. An

sich sind sie als Rahmen zu verstehen, in denen sich Text-, Bild- oder Videoinhalte befinden. Das HTML-Element ist sehr gängig bei der Einbindung von YouTube-Videos. An der Stelle sei jedoch gesagt, dass die Inhalte in iFrames von Suchmaschinen weder erkannt noch indexiert werden. Der Einsatz von iFrames ist also nicht empfehlenswert, da die Suchmaschine damit nichts anfangen kann. Wenn es nicht möglich ist, auf iFrames gänzlich zu verzichten, dann besteht die

Möglichkeit, zusätzliche textbasierte Links zu erstellen, die weitere Informationen zu den angezeigten Inhalten im iFrame anbieten. So kann die Suchmaschine diese problemlos erkennen, crawlen und indexieren.

Ladezeit: PageSpeed is King

Eine der wohl wichtigsten Optimierungsfaktoren auf der technischen Seite ist die Anpassung der Seitengeschwindigkeit. Der PageSpeed ist ein von der Suchmaschine Google bestätigter Rankingfaktor und kann nachweislich mehr Erfolg im Ranking bei Suchergebnissen bringen. Aber wieso ist die Geschwindigkeit so wichtig? Die Ladezeit eines Webdokumentes von der Serverabfrage zur Browseranzeige soll in der Regel etwa ein bis zwei Sekunden dauern. Das ist ein Richtwert, den die Nutzer als schnell empfinden. Sobald die Geschwindigkeit nachlässt und eine Seite nach dem Eintippen einer URL etwa länger als fünf Sekunden zum Laden benötigt, wird der Nutzer ungeduldig und unzufrieden. Eine schon 2009 durchgeführte Studie für E-Commerce-Webseiten von Forrester Consulting (<http://einfach.st/forr2>), die von Akamai in Auftrag gegeben wurde, zeigte, wie ungeduldig die Nutzer eigentlich sind. Die Befragungen belegen, dass der durchschnittliche Online-Einkäufer bei einer aufgerufenen Webseite eine Ladezeit von maximal nur zwei Sekunden toleriert. Bei einer Wartezeit von drei Sekunden verlassen bereits 40 Prozent aller Befragten die Seite, was eine hohe [Bounce-Rate](#)* für die Seite verursachen kann, die sich wiederum negativ auf das Ranking auswirken kann.

Dieses Beispiel zeigt, dass die User Experience durch lange Ladezeiten verschlechtert wird, was wiederum für die Suchmaschine ein Indiz für unzureichende Seitenqualität ist. Es ist also empfehlenswert, die Ladezeit so gering wie möglich zu halten, um die User Experience auf der Webseite zu verbessern.

*siehe Online-Glossar unter www.websiteboosting.com

Tipp: Um die Geschwindigkeit zu überprüfen, benutzen Sie am besten das Tool PageSpeed Insights von Google Developers (kostenlos unter <http://einfach.st/gps>). Hier bekommen Sie Einblicke und Tipps, wie Sie Ihre Webseite oft mit den einfachsten Mitteln noch verbessern können. Die Analyse wird bei Mobil und Desktop durchgeführt und zeigt Fehler auf, die eine Webseite verlangsamen können. Zudem wird ein Optimierungsgrad ermittelt, z. B. 64/100 erreichte Punkte.

Es gibt einige Praktiken, die im Allgemeinen helfen können, Webseiten schneller zu machen. Diese Checkliste zeigt die wichtigsten Stellschrauben, die justiert werden können, um einen Webauftritt hinsichtlich der Geschwindigkeit zu optimieren.

1) Minimieren von HTTP-Requests

Die Seiten werden insgesamt schneller geladen, wenn sie weniger HTTP-Anfragen erhalten. Eine Webseite nimmt mit ihren Komponenten wie Bildern, Stylesheets, Scripts, Flash usw. ganze 80 % der Gesamtladezeit in Anspruch. Allein durch das Entfernen überflüssiger Designs und nicht benötigter Bilder und Grafiken kann eine enorme Geschwindigkeitsverbesserung erreicht werden.

Ein Beispiel zeigt, wie man mithilfe von CSS-Sprites die Ladezeit verbessert: CSS-Sprite ist eine Grafikdatei, in der z. B. mehrere Hintergrundbilder und Icons zusammengefügt sind. Mithilfe der CSS-Einstellungen kann die Position jedes Bildausschnitts definiert und später durch den Browser abgerufen werden.

Damit der Browser die genaue Position der einzelnen Bildfragmente auslesen und anzeigen kann, sollte in den CSS-Einstellungen die jeweilige horizontale und vertikale Ausrichtung hinzugefügt werden.

Der CSS-Code dazu kann beispielsweise wie folgt aussehen:

```
.introduction {
    background-image: url(bg.gif);
    background-position: [horizontal position] [vertical position]
}
```

Der CSS-Code mit Information zu horizontaler und vertikaler Position:

```
ul li {
    background-image: url(bg.gif);
    background-position: 19px 85px;
}
```

2) JavaScript/CSS auslagern

Der Quelltext einer Seite besteht zum größten Teil aus CSS und JavaScript. Um die Größe des Quellcodes zu reduzieren, können die CSS- und JavaScript-Dateien in einem einzigen Dokument zusammengeführt werden.

Wenn der Browser die externe Datei zusätzlich cacht (zwischenspeichert), das heißt, sie nicht mehr beim jedem Seitenaufruf, sondern nur beim Aufruf der ersten Seite einmal lädt, kann noch mehr Ladezeit eingespart werden.

Damit das Caching gut funktioniert, sollte ein Expires-Header oder ein Cache-Control: max-age eingefügt werden. Dieser enthält einen Zeitraum, bis zu welchem die Datei gecacht, also aus dem Browserspeicher geladen werden darf. Daher braucht der Browser diese Angabe zur Gültigkeitsüberprüfung. Google empfiehlt hierzu explizit, den Expires-Header zu nutzen und das Auslaufdatum auf mindestens eine Woche bis zu einem Jahr festzulegen.

Praxisbeispiel: Die JavaScript-Datei wird aus dem HTML-Code entfernt und in einem externen Texteditor (WordPad oder SimpleText) in ein neues Dokument eingefügt. Dieses Dokument soll z. B. mit dem Namen „main.js“ im Ordner abgelegt werden, in dem die übrigen HTML-Dokumente

gespeichert sind. Anschließend sollten im HTML-Code folgende Befehle entfernt werden:

```
<script language="JavaScript">
und
</script>
```

Stattdessen wird dann der folgende Befehl als Verweis auf die nun ausgelagerte Datei eingefügt:

```
<script language="JavaScript"
type = „text/javascript“
src="main.js"></script>
```

3) Bildoptimierung

Bilder zählen zu den Elementen, die den größten Teil einer Webseite ausmachen. Bilddateien können schon bei der Bearbeitung mit Photoshop für die Webseite optimiert werden, indem sie im Webseitenformat abgespeichert werden. Die Bilder haben dadurch eine reduzierte Datengröße und laden schneller. Zur Bildgrößenoptimierung können Tools wie jpegtran, jpegoptim (für JPEG-Dateien) oder OptiPNG, PNGOUT (für PNG-Dateien) hilfreich sein.

Praxistipp: Damit der Browser die Maße eines Bildes nicht selbst berechnen muss, sollte die Bildgröße im Quellcode angegeben werden.

```

```

Dabei soll die Originalgröße die gleiche wie die im Quellcode angegebene Größe sein, da Browser häufig die Bilder als Original herunterladen und dann herunterskalieren. Auch das spart Zeit. Muss man solche Änderungen von Hand durchführen, sollte man diese Ergänzung zumindest bei Bildern, die von vielen Seiten verwendet werden, und auf Seiten, die oft aufgerufen werden, vornehmen. Die Informationen, welche Objekte das sind, bekommt

man aus seinem Webanalytics-Programm wie z. B. Google Analytics.

4) Serverleistung optimieren

Die Antwortzeit Ihres Servers gibt an, wie lange es dauert, den für den Beginn des Renderns (also des technischen Aufbereitens) der Seite erforderlichen HTML-Code von Ihrem Server zu laden. Dabei wird die Netzwerklatenz (Verzögerung bis zur Reaktion einzelner Systeme) zwischen Google und Ihrem Server berücksichtigt. Die einzelnen Messergebnisse können variieren, aber die Unterschiede sollten nicht sehr groß sein. Große Abweichungen der Server-Antwortzeit deuten unter Umständen auf ein Leistungsproblem hin.

Praxistipp: Es gibt viele Gegebenheiten, die die Antwort eines Servers verringern können, wie, z. B. langsame Datenbank-Requests, Speicherplatzmangel oder Bibliotheken. Das Monitoring des Servers ist daher sehr empfehlenswert, um kritische Faktoren zu ermitteln und rechtzeitig eingreifen zu können. Schafft man es, die Server-Antwortzeit auf 200 Millisekunden zu minimieren, dann kann es bereits eine spürbare Verbesserung in der Gesamtperformance geben.

5) Einsatz von CDN

Der Einsatz eines Content Delivery Networks kann ebenfalls helfen, die Performance zu steigern.

Ein Content Delivery Network ist ein Verbund von Servern, um von Nutzern angefragte Daten abhängig vom geografischen Standort schnell verfügbar zu machen. Die Antwortzeit des Servers wird verkürzt, indem nicht stur der zentrale eigene Webserver, sondern eine Kopie auf einem dem Nutzer näher gelegenen Server angesteuert wird. Andere Server, die auf der ganzen Welt verteilt sind, beziehen von diesem Server die jeweiligen Daten. Sie spie-

TIPP

Schlechte Nachrichten für alle, die Fehler aus der Vergangenheit mit dem Mantel der stillen 301-Weiterleitung bedecken: Zu viele Weiterleitungen können der Gesamtperformance einer Site schaden. Insofern bekommt die alte SEO-Regel, eine einmal publizierte URL im Idealfall nie mehr zu ändern, wieder sichtbar Gewicht. Agenturen und deren Kunden mit einem Hang zu ständiger „Redesigneritis“ sahen oft in der einfach realisierbaren Weiterleitung das Heilmittel, beliebig und drauflos immer und immer wieder neue URLs zu produzieren. Damit hilft man zwar den Robots der Suchmaschinen, aber wenn Umleitung auf Umleitung auf Umleitung erfolgt, kann das am Ende eine indirekte negative Wirkung haben. Man sollte also darauf achten, keine Weiterleitungskaskaden zu erzeugen, sondern immer maximal eine Weiterleitung verwenden. Und natürlich ist es immer besser, die ursprünglichen URL beizubehalten. Falls das nicht möglich ist, sollte man wie immer prüfen, ob man dies nicht wenigstens für die wirklich sehr häufig aufgerufenen Seiten tun kann. Die Startseite aufgrund eines Systemwechsels dauerhaft umzuleiten, erscheint unter dieser Perspektive ebenfalls als keine gute Idee.

geln diese und fragen in regelmäßigen Abständen nach, ob es neue oder aktualisierte Dateien gibt. Ein CDN kann mit Diensten wie GDN von Google oder Amazon Cloudfront umgesetzt werden. Gerade für Websites, die auch international häufiger abgerufen werden, ist der Einsatz eines CDN durchaus überlegenwert.

6) 301-Redirects

Auch Weiterleitungen können der Grund für eine langsame Seite sein. Durch eine Minimierung der Weiterleitungsketten lässt sich die Ladezeit enorm verringern. Die interne Verlinkung und die Weiterleitungsregeln einer Webseite können für Klarheit

sorgen. Es gibt eine Reihe von Tools, die bei der Optimierung der internen Verlinkung helfen. Weitere Infos hält u. a. auch Google zur Verfügung unter <http://einfach.st/avrdr>.

Fazit

Wer seinen Webauftritt optimal gestalten und für Suchmaschinen in eine gute Ausgangsposition bringen will, muss zum einen für den reibungslosen Zugang durch die Suchmaschine sorgen, damit das Crawling und die Indexierbarkeit gegeben sind, und zum anderen den eigenen Server permanent überwachen, um die tatsächliche Erreichbarkeit und Reaktionsgeschwindigkeit der Webseite zu gewährleisten.

Durch Optimierung verschiedener technischer Faktoren kann einer der wichtigsten Rankingfaktoren – die Ladezeit einer Seite – verbessert werden und somit eine performante User Experience gewährleistet werden. Die technische Webseitenoptimierung ist alles andere als einfach und für den Einsteiger besonders herausfordernd. Halten Sie also durch und happy optimizing!

In der nächsten Ausgabe finden Sie den sechsten und letzten Teil unserer OnPage-Optimierungsserie, bei dem es um die Fehlersuche und -behebung geht. ¶



TIPP

OnPage.org stellt unter <http://einfach.st/opobook> ein von Irina Hey verfasstes E-Book mit 120 Seiten, den Ratgeber „Bessere Webseiten“ mit einer Schritt-für-Schritt-Anleitung, kostenlos zum Download zur Verfügung.