

Webinhalte optimieren

*Kurzanleitung für Publisher für
das Internet*

Google™

Inhalt

Einleitung	2
Kurzübersicht über die Websuche	3
Was ist neu in Google Websuche?	4
Kann Google Ihre Website finden?	5
Kann Google Ihre Website indizieren?	6
<i>Kontrollieren, was Google indiziert</i>	7
<i>Robots.txt im Vergleich zu Meta-Tags</i>	9
<i>Kontrollieren von Caching und Snippets</i>	10
Enthält Ihre Website eindeutigen und nützlichen Content?	11
Steigern der Sichtbarkeit: bewährte Verfahren	12
Webmaster Central	13
<i>Sitemaps</i>	14
Häufig gestellte Fragen	15
Glossar	19

Einleitung

Wenn Sie Sichtbarkeit möchten, sind Sie im Internet genau richtig. Fragen Sie einen Inserenten, der seine Umsätze mit Online-Anzeigen steigern konnte. Oder einen Blogger, dessen Bekanntheit zu einem Buchauftrag geführt hat. Auch ein Zeitungsverlag, der seine Auflage dank der erhöhten Zugriffszahlen auf eine internationale Leserschaft ausdehnen konnte, wird Ihnen dies bestätigen.

Bei Google werden wir häufig gefragt, wie die Websuche funktioniert und wie Web-Publisher ihre Sichtbarkeit im Internet maximieren können.

Mit diesem kurzen Leitfaden möchten wir Ihnen näher bringen, wie eine Suchmaschine Ihren Content "sieht" und wie Sie Ihre Internetpräsenz optimal konzipieren, damit das, was gefunden werden soll, auch gefunden wird - und das, was verborgen bleiben soll, verborgen bleibt.

Dieser Leitfaden ist mit Tipps und Online-Tools für Webmaster, einer schrittweisen Anleitung und mit häufig gestellten Fragen für kleine Web-Publisher sowie für Eigentümer großer Websites konzipiert.

Parallel zur dramatischen Entwicklung des Internets in den letzten zehn Jahren hat Google seinen Ansatz für die Websuche und seine Beziehung zu Eigentümern von Websites ausgebaut. Wir haben zahlreiche Tools entwickelt, mit denen Webmaster die Sichtbarkeit ihres Contents maximieren sowie die Indizierung ihrer Webseiten steuern können. Aber es gibt immer noch etwas, was wir tun können. Wir hoffen, dass Sie dieser Leitfaden zu Feedback anregt und Sie uns wissen lassen, was wir tun können, um das Internet sowohl für Suchende als auch für Publisher noch attraktiver zu machen.

– Ihr Google Webmaster Team

Kurzübersicht über die Websuche:

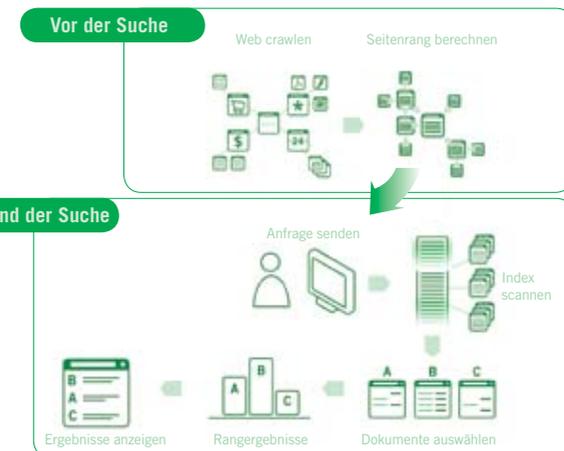
Funktionsweise

Stellen Sie sich das Internet einfach als sehr großes Buch mit einem eindrucksvollen Index vor, der genau Aufschluss darüber gibt, wo Sie das Gesuchte finden.

Google verfügt über eine Gruppe von Computern – "Googlebot" –, die fortwährend Milliarden von Seiten im Internet "crawlen" (durchsuchen). Dieser Crawl-Prozess ist algorithmisch: Computerprogramme bestimmen, welche Websites wie oft gecrawlt werden und wie viele Seiten von jeder Website abgerufen werden. Wir akzeptieren keine Bezahlung, damit eine Website häufiger gecrawlt wird. Außerdem trennen wir das Suchsegment unserer Geschäftstätigkeit klar von unserem gewinnorientierten AdWords-Service.

Der Crawl-Prozess von Google beginnt mit einer Liste von Webseiten-URLs. Beim Durchsuchen dieser Websites erkennt Googlebot Links auf jeder Seite und fügt sie zu seiner Liste der zu crawlenden Seiten hinzu. Googlebot kopiert jede gecrawlte Seite, um einen umfassenden Index aller Wörter, die das Programm sieht, zusammenzustellen. Diese Liste gibt außerdem an, wo jedes Wort auf der jeweiligen Seite steht.

Wenn ein Nutzer eine Suchabfrage eingibt, durchsuchen unsere Suchmaschinen den Index nach übereinstimmenden Seiten und zeigen dem Nutzer die relevantesten Ergebnisse an. Die Relevanz wird durch über 200 Faktoren bestimmt. Einer davon ist der Seitenrang (PageRank) für eine bestimmte Seite. Der Seitenrang ist die Maßeinheit der "Wichtigkeit" einer Seite basierend auf den eingehenden Links von anderen Seiten. Einfach ausgedrückt, jede Webseite, die mit der Webseite XYZ verknüpft ist, erhöht den Seitenrang der Webseite XYZ.



Was ist neu in Google Websuche?

Während die Grundlagen der Websuche im Großen und Ganzen gleich geblieben sind, arbeitet Google kontinuierlich an der Verbesserung der Suchergebnisse.

Welcher Unterschied besteht zwischen der heutigen Websuche und der Websuche beispielsweise vor fünf Jahren? Zunächst einmal ist der Suchvorgang sehr viel schneller geworden.

Darüber hinaus sind unsere Crawl- und Indexsysteme verglichen mit den vor fünf Jahren eingesetzten Systemen viel intelligenter. Wir durchsuchen Webseiten heute beispielsweise kontinuierlich und planen die Besuche der einzelnen Webseiten geschickter, sodass die Aktualität der Informationen maximiert wird. Dieser effizientere Ansatz berücksichtigt die Tatsache, dass die Online-Präsenz einer Zeitung beispielsweise häufiger gecrawlt werden muss als eine statische Website, die nur einmal im Monat aktualisiert wird. Mit unseren Webmaster-Tools können Webmaster darüber hinaus selbst steuern, wie häufig wir ihre Websites crawlen. Dies führt insgesamt zu einem aktuelleren und umfassenderen Index.

Obwohl die Websuche heute schneller und effizienter denn je ist, haben die wichtigsten Faktoren, die die Sichtbarkeit einer Website auf der Google Seite der Suchergebnisse bestimmen, seit Einführung unserer Suchmaschine nach wie vor Priorität.

[Kann Google Ihre Website finden?](#) (Seite 5)

[Kann Google Ihre Website indizieren?](#) (Seite 6)

[Enthält die Website eindeutigen und nützlichen Content?](#) (Seite 11)

Kann Google Ihre Website finden?

Die Einbeziehung in die Google Seite der Suchergebnisse ist kostenlos und kinderleicht. Sie brauchen Ihre Website nicht einmal bei Google einreichen. Die überwiegende Mehrheit der auf unserer Seite der Suchergebnisse aufgeführten Websites wird nicht manuell zur Einbeziehung eingereicht, sondern automatisch gefunden und hinzugefügt, wenn Googlebot das Internet crawlt.

Auch wenn Google Milliarden Seiten crawlt, lässt sich nicht vermeiden, dass einige Websites übersehen werden. Wenn Googlebot eine Website übersieht, geschieht dies häufig aus einem der folgenden Gründe:

- Die Website ist nicht ausreichend über mehrere Links mit anderen Websites im Internet verknüpft.
- Die Website wurde erst freigeschaltet, nachdem der letzte Crawl-Vorgang von Google abgeschlossen war.
- Die Website war vorübergehend nicht verfügbar oder wir erhielten eine Fehlermeldung, als wir versucht haben, sie zu crawlen.

Mit Google Webmaster Tools wie Sitemaps können Sie bestimmen, ob Ihre Website in den Index von Google einbezogen werden soll und ob wir bei dem Versuch, sie zu crawlen, eine Fehlermeldung erhalten (siehe Seite 14). Mit diesen Tools können Sie Ihre URL auch manuell zum Index von Google hinzufügen oder Google eine Sitemap-Datei zur Verfügung stellen, mit der wir einen besseren Einblick in Ihren Content erhalten. So unterstützen Sie uns dabei, neue Bereiche und neuen Content auf Ihrer Website zu finden.

Kann Google Ihre Website indizieren?

Mitunter stellen Webmaster fest, dass ihre Websites nicht auf der Seite der Suchergebnisse aufgeführt werden. Das Problem kann in diesem Fall die "Indizierbarkeit" sein, d. h., ob Googlebot eine Webseite zur Einbeziehung in unsere Seite der Suchergebnisse kopieren kann.

Struktur und Inhalt

Häufig liegt die Nichteinbeziehung in die Seite der Suchergebnisse in der Struktur und im Inhalt der Webseite begründet. Eine Seite, auf der ein Nutzer ein Formular ausfüllen muss, kann beispielsweise von Google möglicherweise nicht indiziert werden. Auch kann eine Seite, die "dynamischen Content" (Flash, JavaScript, Frames oder dynamisch erstellte URLs) enthält, nicht immer problemlos von Suchmaschinen indiziert werden. Wenn Sie wissen möchten, ob dies für Ihre Website zutrifft, zeigen Sie die Website in einem Textbrowser wie Lynx oder in einem Browser, in dem Bilder, Javascript und Flash deaktiviert sind, an. So können Sie feststellen, ob auf Ihren gesamten Content zugegriffen werden kann.

Wenn Ihre Website viele Bilder enthält, stellen Sie sicher, dass Sie den wichtigsten Inhalt jedes einzelnen Bildes im Text beschreiben. Dies ermöglicht Suchmaschinen nicht nur die korrekte Indizierung des Bildes, sondern ermöglicht sehbehinderten Nutzern den Zugriff auf das Bild. Für das Bild können Sie auch ALT Text und beschreibende Dateinamen nutzen, wie in folgendem Beispiel dargestellt (hierbei handelt es sich um das Logo einer Firma namens "Buffy's House of Pies")

```

```

URLs

Eine zusätzliche Hürde könnte die URL selbst sein. Wenn Sitzungs-IDs oder mehrere Parameter in der URL vorhanden sind oder die URL mehrfach umgeleitet wird, kann Google die Seite unter Umständen nicht indizieren.

Server und Network

Probleme mit dem Server oder Netzwerk können den Zugriff auf bestimmte Seiten Ihrer Website verhindern. Mit den unter Google Webmaster Central verfügbaren Tools können Publisher eine Liste ihrer Seiten anzeigen, auf die Googlebot nicht zugreifen kann. Weitere Informationen zu Webmaster Central erhalten Sie auf Seite 13.

Robots Exclusion Protocol

Mitunter werden Seiten durch das Robots Exclusion Protocol blockiert, einem technischen Standard, mit dem Web-Publisher Suchmaschinen "mitteilen" können, den Content ihrer Website nicht zu indizieren (siehe Seite 7). Wenn Ihre Website nicht auf der Google Seite der Suchergebnisse angezeigt wird, überprüfen Sie, ob der Zugriff für unsere Crawler durch die Datei robots.txt oder einen Meta-Tag blockiert wird.

Kontrollieren, was Google indiziert

Jeder Web-Publisher verfolgt ein anderes Ziel im Hinblick darauf, was er im Internet erreichen möchte. Einige Zeitungsverlage haben sich beispielsweise entschlossen, kostenlosen Zugriff auf ihre aktuellen Artikel zu gewähren und den Zugriff auf die Archive über einen gebührenpflichtigen Premium-Service anzubieten. Einige Publisher erstreben Sichtbarkeit auf allen Google-eigenen Produkten der Suchmaschine (z. B. Google Mobile, Google Images usw.), während andere lediglich auf der Seite der Suchergebnisse erscheinen möchten.

Suchmaschinen sind bemüht, die Wünsche der Publisher zu erfüllen. Schließlich handelt es sich um ihren eigenen Content. Wir können jedoch keine Gedanken lesen, daher ist es wichtig, dass Webmaster uns mitteilen, wie ihr Content indiziert werden soll. Dies kann über das Robots Exclusion Protocol geschehen, eine etablierte technische Spezifikation, mit der Suchmaschinen mitgeteilt wird, welche Website oder welche Teile einer Website nicht durchsuchbar sein sollen und welche Teile auf der Seite der Suchergebnisse sichtbar bleiben sollen.

Robots.txt: website-weite Kontrolle

Der Kern des Robots Exclusion Protocol ist eine einfache Textdatei mit dem Namen robots.txt, die seit vielen Jahren zum Industriestandard gehört. Mit robots.txt können Sie den Zugriff auf mehreren Ebenen kontrollieren, angefangen von der gesamten Website über einzelne Verzeichnisse und Seiten eines bestimmten Typs bis hin zu einzelnen Seiten.

Einige Seiten meiner Website sollen nicht im Google Index erscheinen. Wie verhindere ich, dass sie auf Googles Seite der Suchergebnisse aufgeführt werden?

Im Allgemeinen möchten die meisten Eigentümer von Websites, dass Googlebot auf ihre Website zugreift, sodass ihre Webseiten von Nutzern, die die Google Suchmaschine verwenden, gefunden werden. Es kann jedoch vorkommen, dass einige Seiten nicht indiziert werden sollen. Dies gilt beispielsweise für interne Protokolle oder Nachrichtenartikel, deren Zugriff kostenpflichtig ist.

Sie können Seiten aus dem Google Index ausschließen, indem Sie eine robots.txt-Datei erstellen und sie im Stammverzeichnis Ihres Webservers ablegen. In der robots.txt-Datei sind die Seiten aufgeführt, die von Suchmaschinen nicht indiziert werden sollen. Das Erstellen einer robots.txt-Datei ist unkompliziert und gibt Publishern hinreichend Kontrolle darüber, wie Suchmaschinen auf ihre Websites zugreifen.

Wenn ein Webmaster beispielsweise die Indizierung seines internen Protokolls verhindern möchte, sollte die robots.txt-Datei Folgendes enthalten:

User-Agent: Googlebot – Die User-Agent-Zeile gibt an, dass es sich bei dem nächsten Abschnitt um eine Reihe von Anleitungen ausschließlich für Googlebot handelt.

Disallow: /logs/ – Die "Disallow"-Zeile teilt Googlebot mit, dass kein Zugriff auf die Dateien im Unterverzeichnis "logs" Ihrer Website erfolgen soll.

Der Eigentümer der Website hat festgelegt, dass keine der Seiten im Verzeichnis "logs" auf der Google Seite der Suchergebnisse enthalten sein soll.

Alle großen Suchmaschinen lesen und befolgen die Anleitungen in der robots.txt-Datei. Bei Bedarf können Sie verschiedene Regeln für verschiedene Suchmaschinen festlegen.

Meta-Tags: feinkörnige Kontrolle

Zusätzlich zur robots.txt-Datei, mit der Sie Anleitungen für viele Dateien auf Ihrer Website kurz und prägnant festlegen können, können Sie den Robots Meta-Tag verwenden. Mit diesem Tag können Sie eine feinkörnige Kontrolle einzelner Seiten Ihrer Website realisieren. Zum Implementieren dieses Tags fügen Sie einfach bestimmte Meta-Tags zu einer HTML-Seite hinzu, um die Indizierung einer Seite zu steuern. Gemeinsam geben Ihnen die robots.txt-Datei und die Meta-Tags die Flexibilität, komplexe Zugriffsrichtlinien relativ einfach auszudrücken.

Auf meiner Website befindet sich ein bestimmter Nachrichtenartikel, auf den nur registrierte Nutzer zugreifen können. Wie verhindere ich, dass dieser Artikel auf der Google Seite der Suchergebnisse erscheint?

Fügen Sie dazu einfach den Meta-Tag NOINDEX zum ersten <head>-Bereich des Artikels hinzu. Beispiel:

```
<html>
<head>
<meta name="googlebot" content="noindex">
[...]
```

Dadurch wird eine Indizierung dieser Datei durch Google unterbunden.

Dabei sollten Sie jedoch bedenken, dass Google diese Art von Seiten in einigen Fällen dennoch indizieren soll, beispielsweise einen im Archiv gespeicherten Nachrichtenartikel, den Nutzer gegen Gebühr online lesen können. Während diese Art von "Premium-Content" nicht auf der Google Seite der Suchergebnisse erscheint, beziehen bestimmte Google Services wie News Archive Search den Artikel mit einem deutlichen Hinweis auf die Kostenpflichtigkeit in ihre Indizes ein.

Robots.txt im Vergleich zu Meta-Tags

Im Allgemeinen ist die robots.txt-Datei eine gute Möglichkeit, website-weite Kontrolle auszuüben, während Meta-Tags eine feinkörnige Kontrolle einzelner Dateien ermöglichen. Meta-Tags sind besonders nützlich, wenn Sie berechtigt sind, einzelne Dateien, jedoch nicht die gesamte Website zu bearbeiten. Mit Meta-Tags können Sie außerdem komplexe Zugriffsrichtlinien auf Seitenbasis festlegen.

Mitunter können beide Tools zur Lösung des gleichen Problems dienen.

Wie kann ich sicherstellen, dass der Text einer Seite zwar indiziert wird, jedoch nicht die Bilder?

Eine Möglichkeit wäre, den Zugriff auf Bilder mit robots.txt auf Ihrer Website nach Dateierweiterung zu blockieren. Die folgenden Zeilen einer robots.txt-Datei weisen Google an, keine Dateien mit der Dateierweiterung *.jpg oder *.jpeg zu indizieren:

```
User-agent: Googlebot
Disallow: /*.jpg#
Disallow: /*.jpeg#
```

Eine andere Möglichkeit besteht darin, das entsprechende Verzeichnis auszuschließen, sofern Ihr Content Management System (CMS) Bilder in einem separaten Verzeichnis ablegt. Wenn Ihre Bilder in einem Verzeichnis namens "/images" gespeichert sind, können Sie dieses Verzeichnis mit der folgenden Zeile aus allen Suchmaschinen ausschließen:

```
User-agent: *
Disallow: /images/
```

Möglich ist auch, einen NOINDEX-Tag zu jeder Datei hinzuzufügen, die ein Bild enthält.

Alle diese Ansätze gewährleisten, dass Ihre Bilder nicht indiziert werden. Die einzige Frage ist, wie weitreichend dieses Ausschließen von Bildern sein soll.

Kontrollieren von Caching und Snippets

In den Suchergebnissen wird normalerweise ein "Im Cache"-Link auf die entsprechende Seite und ein Textauszug aus den Ergebnisseiten (Snippet) angezeigt. Das folgende Beispiel zeigt eines der ersten Suchergebnisse für das Suchwort "Stockente":

Stockente im Natur Lexikon

Eine bebilderte Beschreibung der **Stockente** im Natur-Lexikon. Dieser Text enthält alle wichtigen Informationen zu der **Stockente**. Die **Stockente** ist in diesem ...
www.natur-lexikon.com/Texte/HWG/001/00075/HWG00075.html - 20k -
[Im Cache](#) - [Ähnliche Seiten](#)

Snippet – ein Textauszug aus der Webseite

Im-Cache-Link – Über diesen Link gelangt der Nutzer zu einer Kopie der Seite, die auf Google Servern gespeichert ist.

Wozu dient ein Snippet? Nutzer besuchen eine Website eher, wenn die Suchergebnisse einen Textauszug aus der Website enthalten. Der Grund hierfür besteht darin, dass Nutzer anhand der Snippets die Relevanz des Ergebnisses ihrer Suchabfrage leichter beurteilen können. Wenn ein Nutzer nicht in der Lage ist, dies schnell zu entscheiden, fährt er meist mit dem nächsten Suchergebnis fort.

Wozu dient ein "Im Cache"-Link? Der "Im Cache"-Link ist in vielen Fällen nützlich, z. B. wenn Websites vorübergehend nicht verfügbar sind, wenn Nachrichtenwebsites nach einem wichtigen Ereignis überlastet sind oder wenn Websites versehentlich gelöscht werden. Ein weiterer Vorteil besteht darin, dass in der Google Kopie im Cache die Suchbegriffe hervorgehoben werden, sodass der Nutzer die Relevanz der Seite schnell beurteilen kann.

Die meisten Web-Publisher möchten, dass Google sowohl den Snippet als auch den "Im Cache"-Link anzeigt. Es gibt mitunter jedoch Fälle, in denen der Eigentümer einer Website eine dieser Anzeigen oder beide deaktivieren möchte.

Mein Nachrichten-Content ändert sich mehrmals am Tag. Anscheinend indiziert Googlebot den Content nicht so schnell, wie wir ihn aktualisieren. Der "Im Cache"-Link verweist daher auf eine Seite, die nicht mehr aktuell ist. Wie kann ich verhindern, dass Google einen "Im Cache"-Link erstellt?

Um zu verhindern, dass dieser "Im Cache"-Link auf der Seite der Suchergebnisse angezeigt wird, kann der Eigentümer der Nachrichten-Website den NOARCHIVE-Tag zu dieser Seite hinzufügen:

```
<META NAME="GOOGLEBOT" CONTENT="NOARCHIVE">
```

Analog können Sie Google mit dem NOSNIPPET-Tag anweisen, kein Snippet für eine Seite anzuzeigen:

```
<META NAME="GOOGLEBOT" CONTENT="NOSNIPPET">
```

Hinweis: Durch Hinzufügen des NOSNIPPET-Tags wird gleichzeitig auch verhindert, dass ein "Im Cache"-Link angezeigt wird. Wenn Sie NOSNIPPET festlegen, erhalten Sie daher automatisch auch NOARCHIVE.

Enthält Ihre Website eindeutigen und nützlichen Content?

Sobald die Website auffindbar und indizierbar ist, stellt sich die Schlussfrage, ob der Content der Webseiten eindeutig und nützlich ist.

Schauen Sie sich zunächst den Text als Ganzes an. Dienen Titel und Textlinks zur Beschreibung? Ist der Textfluss natürlich, klar und intuitiv?

Ebenso wie das Kapitel eines Buches nach bestimmten Gebieten und Themen gegliedert ist, so sollte im Mittelpunkt jeder Webseite ein bestimmtes Gebiet oder Thema stehen. Aus einem solchen Aufbau ergeben sich Keywords und Wortgruppen auf natürliche Weise. Nutzer bleiben weitaus eher auf einer Webseite, die relevanten Content und relevante Links enthält.

Stellen Sie jedoch sicher, dass die von Ihnen verfassten Wortgruppen die Wörter enthalten, nach denen Besucher wahrscheinlich suchen. Wenn Ihre Website die Internetpräsenz eines Lieberhaberclubs der Automarke MG ist, müssen die Wörter "MG" und "Autos" auf jeden Fall im Text enthalten sein. Nur Begriffe wie "britische Autos" reichen nicht aus.

Steigern der Sichtbarkeit: bewährte Verfahren

Häufig werden wir von Eigentümern von Websites nach den besten Verfahren zur Steigerung der Sichtbarkeit und des Rankings ihrer Websites auf unserer Seite für Suchergebnisse gefragt. Unsere Antwort lautet einfach: "Versetzen Sie sich in den Nutzer. Wir tun nichts anderes."

Was bedeutet dies in der Praxis? Stellen Sie zunächst sicher, dass Nutzer die Informationen erhalten, nach denen sie suchen, denn die Relevanz zieht Besucher auf Ihre Website und hilft Ihnen, die Besucherzahlen zu halten.

Viele Eigentümer von Websites fixieren sich darauf, wie gut ihre entsprechenden Webseiten im Ranking abschneiden. Das Ranking wird jedoch neben dem PageRank durch über 200 Kriterien bestimmt. Es ist wesentlich besser, Ihre Zeit darauf zu verwenden, die Qualität Ihres Contents und dessen Zugänglichkeit zu optimieren, anstatt Möglichkeiten zu suchen, den Algorithmus einer Suchmaschine auszutricksen. Wenn eine Website unseren Qualitätsrichtlinien nicht entspricht, kann sie für den Index gesperrt werden.

Vorgehensweise:

1. Erstellen Sie relevanten, ins Auge fallenden Content: Besucher gelangen über verschiedene Links auf Ihre Seiten. Stellen Sie daher sicher, dass jede Seite die Aufmerksamkeit auf sich zieht.
2. Beziehen Sie Nutzer ein: Können Sie einen Bereich für Kommentare oder einen Blog zu Ihrer Website hinzufügen? Durch Aufbau einer Community können Sie die regelmäßige Nutzung Ihrer Website fördern. Durch Einbeziehung Ihrer Nutzer wird die Sichtbarkeit und die Nutzertreue gesteigert.
3. Überwachen Sie Ihre Website: Nutzen Sie Webmaster Central (siehe Seite 13), um festzustellen, welche Suchabfragen Besucher auf Ihre Website bringen, oder um Änderungen im Ranking auf der Seite der Suchergebnisse in Relation zu größeren Änderungen der Website zu verfolgen.
4. Setzen Sie sich qualitativ hochwertige Inbound-Links zum Ziel.
5. Geben Sie Klartext-Links an: Platzieren Sie Textlinks sinnvoll auf Ihre Website und stellen Sie sicher, dass sie Begriffe enthalten, die das Thema beschreiben.

Vermeiden Sie Folgendes:

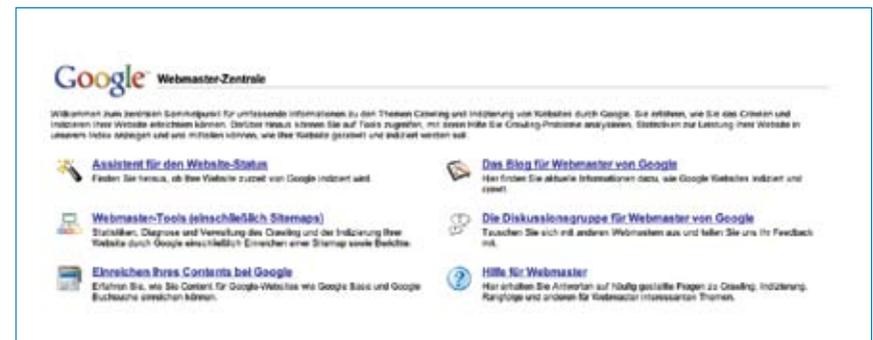
1. Füllen Sie Ihre Seiten nicht mit endlosen Listen von Keywords.
2. Versuchen Sie nicht, Ihre Seiten zu "verhüllen" (Cloaking), indem Sie Text schreiben, der zwar von Suchmaschinen, nicht aber von Nutzern gesehen werden kann.
3. Richten Sie keine Seiten "nur für Crawler" ein, indem Sie Seiten oder Links integrieren, deren einziger Zweck darin besteht, Suchmaschinen in die Irre zu führen.
4. Verwenden Sie keine Bilder, um wichtige Namen, Links oder wichtigen Content anzuzeigen. Bilder können von Suchmaschinen nicht "gelesen" werden.
5. Erstellen Sie nicht mehrere Kopien einer Seite unter verschiedenen URLs mit dem Ziel, Suchmaschinen in die Irre zu führen.

Wenn Sie sich unsicher sind, ziehen Sie unsere Richtlinien für Webmaster zu Rate. Weitere Informationen zu diesen Richtlinien erhalten Sie unter: google.de/webmasters/guidelines.html

Webmaster Central

Als Unternehmen, das sich die Bereitstellung von möglichst relevanten und nützlichen Suchergebnissen im Internet zum Ziel gesetzt hat, sind wir bestrebt, skalierbaren und gleichberechtigten Support für alle Webmaster und alle Websites zur Verfügung zu stellen, unabhängig von der jeweiligen Größe. Aus diesem Grund haben wir den Webmaster Central Blog entwickelt. Sie finden ihn unter: google.de/webmasters.

Webmaster Central stellt eine großartige Ressource für alle Web-Publisher dar. Fragen zum Crawling, Indizieren und Ranking werden umfassend beantwortet. Es schafft einen Zugang für Feedback und Probleme und bietet Diagnose-Tools, mit denen Webmaster potenzielle Crawling-Probleme debuggen können.



Nachfolgend eine kleine Kostprobe, welche Funktionen Sie im Webmaster Central Blog finden können.

- Diagnose potenzieller Probleme beim Zugriff auf Seiten und entsprechende Lösungsvorschläge
- Anforderung, bestimmte Seiten aus unserem Index zu entfernen
- Sicherstellen, dass Ihre robots.txt-Datei die Seiten wie erwartet zulässt bzw. blockiert
- Statistische Daten zu Abfragen und Seiten in Bezug auf Ihre Website anzeigen
- Statistische Abfragedaten: Ermitteln Sie, mit welchen Suchabfragen Ihre Website die höchsten Besucherzahlen verzeichnet und welche Themen Sie auf Ihrer Website ausbauen können, um die Besucherzahlen zu erhöhen.
- Seitenanalyse: Zeigen Sie Ihre Webseite so an, wie Google sie sieht. Zeigen Sie die

häufigsten Wörter Ihrer Website, Inbound-Links zur Website und eine Beschreibung Ihrer Website an, die von anderen Nutzern, die über einen Link auf Ihre Website gelangt sind, stammt.

- Crawl-Rate: Stellen Sie fest, wie häufig Ihre Website von Googlebot gecrawlt wird, und teilen Sie Google mit, ob das Crawling schneller oder langsamer ausgeführt werden soll.

Sitemaps

Webmaster Central bietet Publishern außerdem Sitemaps für die Ergebnisse der Websuche, von Mobile und News.

Sitemaps ist ein Protokoll, das wir mit anderen Suchmaschinen unterstützen, um Webmastern dabei zu helfen, uns mehr Informationen über ihre Webseiten zukommen zu lassen. Das Sitemaps-Protokoll ergänzt vorhandene Standardmechanismen zum Crawlen des Internets. Anhand dieses Protokolls können Nutzer Google Informationen über die Seiten ihrer Website zukommen lassen, um das Crawling und die Sichtbarkeit ihrer Seiten auf der Google Seite der Suchergebnisse zu verbessern.

Neben Sitemaps für die Websuche bieten wir außerdem Google Mobile Sitemaps, mit dem Publisher URLs einreichen können, die unseren mobilen Index mit Content für mobile Geräte versorgen.

Und für Publisher, deren Nachrichten-Website in Google News integriert ist, können News Sitemaps zur Erhebung statistischer Daten über die Artikel von Publishern beitragen, angefangen von Abfragen bis hin zur Häufigkeit des Erscheinens. In Verbindung mit den Diagnosetools von Webmaster Central kann News Sitemaps außerdem Fehlerberichte liefern, anhand derer Probleme geklärt werden können, auf die Google beim Crawlen oder Extrahieren von Nachrichtenartikeln auf der Website eines Publishers stößt. Darüber hinaus kann ein Publisher eine News Sitemap-Datei mit URLs einreichen, die für die Einbeziehung in Google News berücksichtigt werden sollen. Im Gegensatz zu Web Sitemaps und Mobile Sitemaps ist News Sitemaps zurzeit nur in englischer Sprache verfügbar. Wir hoffen aber, es in Kürze auch in anderen Sprachen zur Verfügung stellen zu können.

Häufig gestellte Fragen

Warum kann Google keinen individuellen Support für meine Website durchführen?

Es befinden sich schätzungsweise 100 Mio. Websites im Internet. Jede einzelne dieser Websites ist für uns wichtig, denn ohne sie – egal wie groß oder klein – wäre unser Index weniger umfangreich und letztendlich weniger nützlich für unsere Nutzer.

Webmaster Central stellt eine großartige Supportquelle für alle Arten von Websites dar. Wir posten und beantworten Fragen von Publishern, sodass jeder Nutzer von den Informationen profitieren kann. Im Webmaster Central finden Sie außerdem eine freundliche und hilfreiche Gemeinschaft von Webmastern, die Tipps geben und bei der Fehlerbehebung helfen.

Beeinflussen die von Ihnen geschalteten Anzeigen das Ranking? Ist Ihre Anzeigenschaltung völlig von den Suchergebnissen getrennt?

Anzeigen- und Suchrankings sind in keiner Weise miteinander verknüpft. Bei uns arbeiten sogar völlig verschiedene Teams an den Rankings, sodass es keine Beeinflussung gibt. Wir sind der Meinung, dass die Objektivität unserer Suchergebnisse wichtig dafür ist, den Nutzern den bestmöglichen Service zu bieten.

Wie füge ich eine Website zum Suchindex von Google hinzu?

Die Einbeziehung in die Suchergebnisse von Google ist kostenlos und kinderleicht. Es ist keine manuelle Einreichung der Website bei Google erforderlich. Google ist eine vollständig automatisierte Suchmaschine, die das Web regelmäßig crawlt und Websites findet, um sie unserem Index hinzuzufügen. Die überwiegende Mehrheit der auf unserer Seite der Suchergebnisse aufgeführten Websites wird nicht manuell zur Einbeziehung eingereicht, sondern automatisch gefunden und hinzugefügt, wenn unsere Spider das Internet crawlen.

Darüber hinaus bieten Google Webmaster Tools (im Webmaster Central) Webmastern eine einfache Möglichkeit, eine Sitemap oder ihre URLs für den Google Index einzureichen und ausführliche Berichte über die Sichtbarkeit ihrer Seiten bei Google zu erhalten. Mit Google Webmaster Tools können sich Eigentümer von Websites Google automatisch über alle aktuellen Seiten und Updates dieser Seiten auf dem Laufenden halten.

Wie lang dauert es im Durchschnitt, bis Google eine neu erstellte Website findet, und wie häufig crawlt Google das Internet im Allgemeinen?

Es gibt keine Zeitangabe, wie lange Google braucht, bis es eine neue Website findet.

Googlebot crawlt das Internet regelmäßig, um unseren Index neu aufzubauen. Mit Webmaster Central kann ein Webmaster sehen, wie häufig seine Website von Googlebot gecrawlt wird, und Google mitteilen, ob schneller oder langsamer gecrawlt werden soll.

Was ist, wenn meine Website zwar in den Suchergebnissen der Websuche erscheinen soll, aber nicht in einzelnen Services wie Google News oder Google Bildsuche?

Google gibt Web-Publishern stets die Möglichkeit, sich bei den Services abzumelden. Ein Publisher kann sich dazu an das Supportteam eines bestimmten Produkts wenden.

Wie bereits in diesem Leitfaden erwähnt, kann die Indizierung sowohl von Bildern als auch von Webseiten mit dem Robots Exclusion Protocol blockiert werden. Die in Webmaster Central verfügbare Funktion zum Entfernen einer URL kann ebenfalls zu diesem Zweck verwendet werden und deckt die Websuche sowie die Bildsuche ab.

Da Googlebot auf verschiedene Bots zurückgreift, können Sie außerdem das Blockierungsziel festlegen:

- Googlebot: Crawl Seiten für unseren Webindex und unseren Nachrichtenindex.
- Googlebot-Mobile: Crawl Seiten für unseren Mobile-Index.
- Googlebot-Image: Crawl Seiten für unseren Bildindex.
- Mediapartners-Google: Crawl Seiten, um AdSense-Content festzulegen. Diesen Bot nutzen wir nur, um Ihre Website zu crawlen, wenn Sie AdSense-Anzeigen auf Ihrer Website schalten.
- Adsbot-Google: Crawl Seiten, um die Qualität der AdWords-Zielseite zu bewerten. Diesen Bot verwenden wir nur, wenn Sie Ihre Website mit Google AdWords bewerben.

Um Googlebot beispielsweise vollständig zu blockieren, können Sie die folgende Syntax verwenden:

```
User-agent: Googlebot
```

```
Disallow:
```

Kann ich den Text für den Snippet selbst auswählen?

Nein. Dies ist weder aus der Sicht des Nutzers noch aus der Sicht des Erstellers des Contents ratsam. Wir wählen einen Textauszug aus der Website aus, der die Abfrage des Suchenden im Kontext zeigt, was wiederum die Relevanz des Suchergebnisses veranschaulicht.

Studien haben ergeben, dass Nutzer eine Website eher besuchen, wenn die Suchergebnisse einen Snippet enthalten. Dies liegt daran, dass Snippets es Nutzern erleichtern zu erkennen, warum das Ergebnis für ihre Abfrage relevant ist. Wenn der Nutzer dies nicht auf Anhieb erkennen kann, geht er normalerweise zum nächsten Suchergebnis weiter.

Web-Publisher können einen Meta-Tag in ihre Seiten integrieren, um Google mit zusätzlichen Informationen zu versorgen für den Fall, dass wir mit unserem Algorithmus keinen sinnvollen Snippet aus dem Content der Seite generieren können. Fügen Sie dazu Folgendes zum <head>-Bereich der Seite hinzu:

```
<meta name="description" content="Warum mag Anja keine Hasen? Gleich wissen wir mehr.">
```

Web-Publisher, die nicht möchten, dass ein Snippet aus ihren Seiten generiert wird, können den NOSNIPPET-Tag wie folgt verwenden:

```
<meta name="robots" content="nosnippet">
```

Mitunter verwenden wir auch die Beschreibung einer Seite aus Open Directory Project für den Snippet auf der Seite der Suchergebnisse. Wenn diese Beschreibung nicht verwendet werden soll, fügen Sie einfach den folgenden Meta-Tag hinzu:

```
<meta name="robots" content="noodp">
```

Aktuelle Nachrichtenschlagzeilen erscheinen auf meiner Website nur einige Stunden. Dann werden sie aktualisiert und in einen Standardbereich verschoben. Ich möchte, dass zwar der gesamte Artikel im Google Index erscheint, nicht aber die aktuellen Nachrichtenschlagzeilen.

Verzeichnis abzulegen und anhand der robots.txt-Datei den Zugriff auf dieses Verzeichnis durch Googlebot zu sperren.

Eine andere Möglichkeit ist, den NOFOLLOW-Tag zum <HEAD>-Bereich der HTML-Datei Ihres Bereichs für aktuelle Nachrichtenschlagzeilen hinzuzufügen. Damit wird Googlebot angewiesen, keinen Links, die es auf dieser Seite findet, zu folgen: Beachten Sie, dass der NOFOLLOW-Tag Googlebot nur davon abhält, Links von einer Seite auf eine andere zu folgen: Wenn eine andere Website einen Link zu diesem Artikel enthält, kann Google ihn beim Indizieren trotzdem finden.

Wenn ich mehrere Domain-Namen habe und den gleichen Content von diesen verschiedenen Domains aus ausführe, werde ich dann aus Ihren Suchergebnissen ausgeschlossen?

Auch wenn einige Publisher versuchen, Suchmaschinen in die Irre zu führen, indem sie Content duplizieren und gespiegelte Websites unterhalten, gibt es auch rechtmäßigen Content, für dessen Duplizierung es gute Gründe gibt. Google möchte diese Websites nicht schlechter stellen. Wir behandeln beispielsweise Content in verschiedenen Sprachen (z. B. Englisch auf einer Website und Französisch auf einer anderen) nicht als duplizierten Content.

Wenn der gleiche Content auf mehreren Websites (z. B. bei überregionaler Verbreitung von Artikeln) vorhanden ist, bedeutet dies nicht unbedingt, dass eine oder einige dieser Websites vollständig von unserer Seite der Suchergebnisse entfernt werden. Beachten Sie jedoch, dass jede Instanz des Artikels wahrscheinlich weiter unten im Ranking angezeigt wird, da sie nur einen Bruchteil der eingehenden Links aufweist wie eine einzelne Kopie. Im Allgemeinen wird eine einzelne Kopie eines Artikels im Ranking höher eingestuft und daher von mehr Nutzern gesehen als mehrere Kopien des gleichen Contents.

Zur Gewährleistung der Suchqualität bezieht Google darüber hinaus nicht mehrere Kopien einer Seite in die Suchergebnisse ein. Stattdessen wählen wir meistens nur eine Version der Seite zum Anzeigen aus. Webmaster können Google jedoch ihre bevorzugte Version mitteilen, indem sie die Kopien, die nicht in unseren Suchergebnissen erscheinen sollen, mit der robots.txt-Datei oder einem Meta-Tag blockieren.

Warum wird meine Website aus dem Google Index ausgeschlossen?

Möglicherweise wird Ihre Website gar nicht ausgeschlossen. Es gibt viele Gründe, warum eine Website nicht auf unserer Seite der Suchergebnisse erscheint (siehe Seiten 5-11).

Wenn Ihre Website die Erkennung oder Indizierung nicht aus irgendeinem Grund selbst erschwert, kann es jedoch durchaus sein, dass sie gesperrt ist. Websites können für unseren Index gesperrt werden, falls sie nicht den in unseren Webmaster-Richtlinien dargelegten Qualitätsstandards (verfügbar im Webmaster Central) entsprechen. Dies ist meist der Fall, wenn eine Website unfaire Methoden einsetzt, um ein höheres Ranking in den Suchergebnissen zu erzielen. Zu den häufigen Verstößen gegen die Richtlinien gehört das "Cloaking" (Texte werden so verfasst, dass sie zwar von Suchmaschinen, nicht aber von Nutzern gesehen werden) oder das Einrichten von Seiten bzw. Links zu dem einzigen Zweck, Suchmaschinen in die Irre zu führen und die Suchergebnisse zu beeinflussen.

Wenn Webmaster den Verdacht haben, dass ihre Websites gegen unsere Qualitätsrichtlinien verstoßen, können sie ihre Website entsprechend ändern und uns über den Link zur Anforderung der erneuten Einbeziehung in Webmaster Tools zur erneuten Bewertung der Website auffordern.

Glossar

"Im Cache"-Link

Ein Schnappschuss der Seite, wie sie beim letzten Besuch von Google angezeigt wurde. Eine im Cache gespeicherte Kopie ermöglicht Nutzern die Anzeige einer Seite, auch wenn die Online-Version nicht verfügbar ist. Allerdings kann der Content etwas unterschiedlich sein. Klicken Sie zum Anzeigen der Kopie im Cache auf den "Im Cache"-Link unter einem Suchergebnis.

Cloaking

Suchmaschinen bekommen anderen Content angezeigt als Nutzer.

CMS (Content Management System)

Ein Softwaresystem, mit dem der Inhalt von Computer-, Bilder- und Audiodateien bis hin zu Webcontent verwaltet wird.

Crawler

Software, mit der URLs im Internet oder in einem Intranet erkannt und indiziert werden.

Crawling

Der Prozess, mit dem Suchmaschinen Seiten aus dem Internet sammeln.

Dateierweiterung

Der Name einer Computerdatei (.doc, .txt, .pdf usw.), der häufig auf die Art der in der Datei gespeicherten Daten verweist.

Dynamischer Content

Content wie Bilder, Animationen oder Videos, für die Flash, JavaScript, Frames oder dynamisch erstellte URLs erforderlich sind.

Gespiegelte Website

Eine duplizierte Webseite, die mitunter dazu benutzt wird, Suchmaschinen in die Irre zu führen und zu versuchen, die Indizierung und das Ranking einer Website im Internet zu optimieren.

HTML (Hypertext Markup Language)

Eine Auszeichnungssprache, die im Internet zum Strukturieren von Text dient.

Indizieren

Der Vorgang, mit dem Content Ihrer Website zu einer Suchmaschine hinzugefügt wird.

Keyword

Ein Wort, das in das Suchfeld einer Suchmaschine eingegeben wird. Die Suchmaschine sucht anschließend nach Seiten, die das Wort oder die Wortgruppe enthalten.

Meta-Tags

Ein Tag in HTML, der den Content einer Webseite beschreibt. Mit Meta-Tags kann die Indizierung einzelner Seiten einer Website gesteuert werden.

PageRank

Der PageRank oder Seitenrang ist eine Funktion von Google, mit der die Rangfolge einer Website auf unserer Seite der Suchergebnisse bestimmt wird. Der Seitenrang basiert auf der einzigartigen demokratischen Beschaffenheit des Internets, indem die immense Linkstruktur als Indikator des Werts einer einzelnen Seite dient. Wichtige, qualitativ hochwertige Websites erhalten einen höheren Seitenrangwert, den Google sich bei jedem Durchführen einer Suche merkt. Google kombiniert den Seitenrang mit anspruchsvollen Textzuordnungsverfahren, um Seiten zu finden, die sowohl wichtig als auch relevant für Ihre Suchabfragen sind.

Robots Exclusion Protocol

Eine technische Spezifikation, mit der Suchmaschinen mitgeteilt wird, welche Website oder Teile einer Website nicht durchsuchbar sein sollen und welche Teile auf der Seite der Suchergebnisse sichtbar bleiben sollen.

Robots.txt

Eine Textdatei, mit der Web-Publisher den Zugang auf ihre Website auf mehreren Ebenen kontrollieren können, angefangen von der gesamten Website über einzelne Verzeichnisse und Seiten eines bestimmten Typs bis hin zu einzelnen Seiten. Diese Datei teilt Crawlern mit, welche Verzeichnisse gecrawlt bzw. nicht gecrawlt werden sollen.

Stammverzeichnis

Das oberste Verzeichnis in der Struktur eines Computerdateisystems.

URL (Uniform Resource Locator)

Die Adresse einer Website im Internet, bestehend aus dem Zugangsprotokoll (http), dem Domain-Namen (www.google.de) und in einigen Fällen dem Speicherort einer anderen Datei (www.google.de/webmaster).

Weitere Informationen zu Webmaster Central erhalten Sie unter:

google.de/webmasters/

Google™