

Googlebot

Stand: 21.07.2022

Bei einer Suchanfrage müssen viele Informationen zusammenlaufen, damit der Nutzer ein Ergebnis erhält. Das geschieht bereits im Vorfeld mit dem Googlebot. Dieser trägt alle Informationen im Index zusammen und stellt so sicher, dass Nutzer Suchergebnisse erhalten, die möglichst exakt zu ihrer Anfrage passen.

Definition

Der Googlebot ist ein [Webcrawler](#) der Suchmaschine Google. Gelegentlich wird er auch als Searchbot oder Spider bezeichnet. Denn wie eine Spinne ihr Netz bis in den kleinsten Winkel überblickt, so kennt der Googlebot das World Wide Web.

So erstellt [Google](#) einen Katalog aus Daten, die u. a. für die Googlesuche verwendet werden: den Index. Google-Nutzer müssen dann nur ein Wort oder eine Kombination von Wörtern in die Suche eingeben. Die Suchmaschine gleicht diese mit ihrem Katalog ab und gibt eine Liste von Suchergebnissen aus. Diese basieren dann auf den gesuchten Wörtern und der Einschätzung des Googlebots darüber, für wie hilfreich er die Webseiten hält.

Je weiter vorn eine Webseite in den Suchergebnissen, den [SERPs \(Search Engine Result Pages\)](#), auftaucht, desto besser hat der Googlebot sie bewertet. Diese Bewertung nennt man Ranking. Ein besonders gutes Ranking sorgt dafür, dass viele Nutzer die Webseite finden, weil sie innerhalb der SERPs eine gute Platzierung innehat.

Aufgabe der Googlebots

Ihre primäre Aufgabe sieht die Suchmaschine darin, Nutzer zu unterstützen, ihre Suchanfrage so optimal wie möglich zu beantworten. Sind diese mit den Suchergebnissen zufrieden, verwenden sie Google weiterhin, erzeugen also neuen [Traffic](#). Dank dieses Traffics bleibt Google relevant und kann Werbeplätze verkaufen, wodurch sich der Konzern wiederum finanziert.

Vonseiten der Websitebetreiber besteht ebenfalls das Interesse an Traffic, sei es beispielsweise, um Informationen zu verbreiten oder Waren anzubieten. Deshalb möchten Websitebetreiber ebenfalls von Web-Nutzern gefunden werden und möglichst weit vorn in den Google-SERPs erscheinen.

Der Googlebot fungiert als Vermittler zwischen den Usern und der [Suchmaschine](#). Er prüft, ob die einzelnen Webseiten die Bedingungen der Suchmaschine erfüllen und wie sehr sie den Nutzern gefallen.

Funktionsweise von Googlebots

Beim Durchsuchen des Internets bewegt sich der Googlebot über Hyperlinks von einer Webseite zur nächsten. Dabei durchsucht er die Webseite gleich nach weiteren Hyperlinks. Die Webseite selbst speichert er mit allen vorhandenen Informationen in seinem Index ab.

Diese [indexierten Inhalte](#) stellen die Grundlage für die Suchanfragen von Usern dar. Dabei gleicht die Suchmaschine die Suchanfrage mit den indexierten Inhalten ab und generiert dann eine möglichst relevante Ergebnisseite.

Um den Index aktuell zu halten, ist der Googlebot ständig damit beschäftigt, neue Webseiten zu finden und aufzunehmen. Außerdem überprüft er bereits besuchte Websites auf neue Inhalte, Änderungen und veraltete Links. Im Zuge dieses Suchvorgangs werden alle gefundenen Informationen dem bestehenden Datensatz hinzugefügt.

Als Grundlage für dieses Vorgehen verwendet Google hochkomplexe Algorithmen. Diese legen nicht nur fest, wie, wann, was und wonach gesucht wird, sondern auch, welche Wertigkeit gefundene Informationen haben und wie diese sich auf das Ranking der Suchtreffer auswirken.

Bei Keywords beispielsweise sieht der Googlebot-Algorithmus ganz genau hin. Er prüft Menge, Art und Position der Keywords, ihre Anzahl und ob die Wahl von Synonymen, Satzstruktur und weiteren Aspekten sinnvoll erscheint. Auch die Aktualität der Informationen wird berücksichtigt: Ist eine Erläuterung veraltet, eine verlinkte Seite nicht mehr erreichbar oder wird ebenfalls als veraltet erkannt, fließt das in die Bewertung ein.

Beurteilt der Googlebot die Verwendung als unlogisch oder unnötig gehäuft, wird er Seiten als weniger relevant einstufen, seltener besuchen und nur weit hinten in den SERPs anzeigen.

Die verwendeten Algorithmen unterscheiden auch, um welche Art von Inhalt es sich auf der Website handelt und welcher Googlebot zuständig ist.

Arten von Googlebots

Neben dem Googlebot, der klassische Desktop-Webseiten durchsucht, gibt es weitere, spezialisiertere Webcrawler bei Google: Beispielsweise Bots, die sich den Smartphone-Webseiten widmen. Dadurch ist es möglich, dass responsive Webangebote öfter berücksichtigt werden.

Weitere Spezialisten aus der Googlebot-Familie nehmen Videos, Bilder und Nachrichten genauer unter die Lupe. Ein weiterer Crawler lädt Webseiten herunter, um passende Werbung zu ermitteln. Die verschiedenen Googlebots tauschen ihre Informationen dann miteinander aus: Durchsucht ein Bot eine Seite, legt er diese im Cache – einem speziellen Speicher – ab, sodass sie dort für andere Bots verfügbar ist.

Es ist nicht möglich, einen der Googlebots (mobil oder Desktop-Computer) gezielt zum Durchsuchen der

Website auszuwählen. Allerdings greift der Grundsatz „Mobile First“ inzwischen auch bei dem Desktop-Webcrawler. Der Grund dafür ist das Verhalten der Nutzer: Da sie seit einigen Jahren deutlich öfter mit dem Smartphone als via Desktop-Computer auf Webseiten zugreifen, testet auch der Googlebot zunächst, ob Webseiten dieser Anforderung gerecht werden.

Ob Googlebots eine Seite besucht haben, zeigt sich im Protokoll an der User-Agent-Kennung. Das heißt: Beginnen Zugriffsanfragen mit dem Namen „Googlebot“, ist es wahrscheinlich, dass dieser auf die Website zugegriffen hat. Die Anfragen werden in der Logdatei bzw. der Protokolldatei gespeichert und können von dort aus nachvollzogen werden.

Vorgehen des Bots

Der Googlebot ist so entwickelt, dass er über mehrere tausend Computer ausgeführt werden kann. Das Ziel ist es, die Bandbreite nicht zu sehr zu beanspruchen. Dadurch kann es jedoch passieren, dass von mehreren Computern aus mehrere Webcrawler zugreifen. Die Anfragen der Crawler werden in den Logfiles des Servers protokolliert und erlauben es, nachzuvollziehen, wer Anfragen an den Server gestellt hat.

Um sich von Webseite zu Webseite zu bewegen, folgt der Googlebot Links. Dafür erkennt der Bot SRC- und Href-Links, also Ressourcen und Verbindungen, und analysiert diese.

Laut Google greift der Bot im Durchschnitt einmal alle paar Sekunden auf eine Webseite zu. Die Frequenz ist u. a. abhängig von der Anzahl der externen Links, die auf eine Seite verweisen, bzw. vom PageRank der Seite. Weniger stark verlinkte Webseiten werden u. U. nur alle paar Tage oder noch seltener vom Bot besucht.

Den Googlebot aussperren

Sobald ein Link zu einer Webseite irgendwo auftaucht, wird der Googlebot darauf aufmerksam. Dafür reicht es aus, wenn nur ein einziger Nutzer den Link auf einer öffentlich zugänglichen Plattform teilt. Ein Verbergen von Webseiten ist daher nahezu unmöglich.

Trotzdem kann es natürlich gute Gründe dafür geben, dem Googlebot den Zugriff zu verweigern, beispielsweise aus Datenschutzgründen. Hier muss der Bot explizit darum gebeten werden, Inhalte vom Crawling auszusparen.

Eine Möglichkeit besteht darin, dem Stammverzeichnis der eigenen Webseiten eine robots.txt-Datei hinzuzufügen. Diese Datei weist den Bot an, welche Bereiche der Website durchsucht werden dürfen und welche nicht.

Die Verwendung einer [robots.txt-Datei](#) verhindert jedoch nicht zuverlässig, dass eine Webseite nicht in der Google-Suche erscheint, da der Webcrawler diese manchmal ignoriert. Besser geeignet ist ein Robots-Meta-Tag im Head-Element einer Webseite.

Dieses weist alle Crawler an, die betreffende Seite nicht in den Suchergebnissen darzustellen. Soll nur der Googlebot ausgeschlossen werden, muss im name-Attribut „robots“ durch „googlebot“ ersetzt werden.

Außerdem kann das Meta-Tag „[nofollow](#)“ verwendet werden; damit unterbindet man die Verfolgung von jeglichen Links auf der eigenen Webseite durch den Bot. Soll der Bot dagegen nur bestimmten Links nicht folgen, fügt man das Attribut rel=“nofollow“ nur zum jeweiligen Link hinzu.

Verdacht auf Missbrauch

Manche Web-Nutzer geben sich, in der Hoffnung auf werbefreie Seiten oder Seiten mit besser zugänglichen Informationen, als Webcrawler aus. Dabei wird die User-Agent-Kennung so verfälscht, dass dort wie oben erwähnt „googlebot“ steht.

Wenn der Verdacht besteht, dass es sich um gefälschte Googlebots-Zugriffe handelt, ist es möglich, dies über eine umgekehrte DNS-Anfrage zu klären. DNS bedeutet Data Name Server und die Anfrage funktioniert ähnlich der umgekehrten Telefonabfrage. Nach der Dateneingabe werden die Daten des Absenders ausgegeben.

Handelt es sich wirklich um den Googlebot, sollte der Name auf „googlebot.com“ enden. Im zweiten Schritt wird dann eine reguläre DNS-Anfrage durchgeführt, um herauszufinden, ob sie wieder die ursprüngliche IP-Adresse enthält. Ist dies der Fall, kann man davon ausgehen, dass der Besucher wirklich der Googlebot ist.

Bedeutung von Googlebots für die Suchmaschinenoptimierung

Für die [Suchmaschinenoptimierung](#) (SEO) ist es wichtig, sich mit der Funktionsweise des Googlebots auszukennen, z. B., um neue Inhalte möglichst schnell vom Bot indexieren zu lassen. Eine andere Umschreibung für das schnelle Zugänglichmachen von neuen Informationen ist „seeden“.

Weitere wichtige Punkte für die Suchmaschinenoptimierung:

- Der Googlebot kann nicht zwangsläufig alle Inhalte auswerten. Einige neuere Bild- oder Videoformate kennt er beispielsweise nicht.
- Die Wahl und Position von Keywords können negative Auswirkung auf das Ranking in den SERPs haben.
- Es gibt eine Reihe von Möglichkeiten, den Crawling-Vorgang zu begünstigen, zum Beispiel mit einer Sitemap.

Als [SEO-Agentur](#) stehen wir Ihnen mit unserer Expertise zur Seite. Kontaktieren Sie uns gerne!