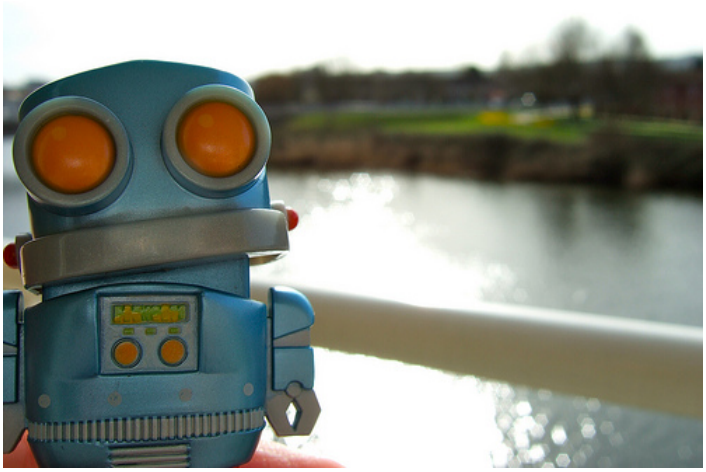
GOOGLEBOT VERSTEHEN: SO FUNKTIONIERT CRAWLING UND INDEXIERUNG BEI GOOGLE



Veröffentlicht am 25. April 2025 von Maya

"Es kann ein bisschen dauern, bis Google die Änderungen darstellt." So oder so ähnlich lauten die Antworten von Suchmaschinen-Profis [...]

Der Ausdruck Googlebot wird Dir zwangsläufig begegnen, wenn Du Dich mit der SEO (Search Engine Optimization) für Deine Website beschäftigst. Google ist dabei mit einem Marktanteil von über 90% die wichtigste Adresse. Entscheidend für die Indexierung Deiner Seite bei Google ist dabei der Vorgang des sogenannten **Crawlings** durch die Googlebots.

WAS IST DER GOOGLEBOT?

Die Websites werden von Google mithilfe des Googlebots gelesen. Er wird auch als Crawler oder Spider bezeichnet und identifiziert sich z.B. durch die User-Agent-Kennung Googlebot/2.1.

Der Name Googlebot setzt sich aus dem Firmennamen und der technischen Bezeichnung "Bot", kurz für Robot, zusammen. Eine spezielle Sorte Bots sind Webcrawler oder Searchbots, die so programmiert werden, dass sie das Internet durchsuchen und Webseiten analysieren. Google nutzt seinen Bot dazu, um Inhalte zu analysieren und auffindbar machen zu können.

Es existieren verschiedene Varianten, z.B.:

- **Googlebot-Video** für Videoinhalte
- **Googlebot-Image** für die Bildersuche

- **Googlebot-News** für News-Inhalte
- **Googlebot-Mobile** für mobile Websites

WAS MACHT DER GOOGLEBOT?

Der Googlebot besucht und crawlt Webseiten, indem er sich **von Link zu Link bewegt**. Dabei erkennt der Bot neue Inhalte, prüft Aktualisierungen bestehender Seiten und bewertet technische Aspekte wie Ladegeschwindigkeit oder Mobile-Freundlichkeit. Alle vom Robot gefundenen Inhalte werden heruntergeladen und je nach Relevanz im Googleindex abgelegt.

Das Crawling ist die Voraussetzung dafür, dass Deine Seite überhaupt **in Betracht gezogen** wird, bei Google gefunden zu werden. Erst nach dem Crawling entscheidet Google, ob eine Seite indexiert – also in den Suchindex aufgenommen – wird. Wurde die Seite kürzlich von einem der Bots gecrawlt, legt er die Informationen für die anderen Crawler in einem Cache zurück.

WIE OFT BESUCHT DER GOOGLEBOT EINE SEITE?

Wann der Googlebot wieder kommt, hängt von verschiedenen Faktoren ab. Der Bot bewegt sich anhand von Links. Deswegen sind der PageRank und die Anzahl und die Qualität der vorhandenen Backlinks entscheidend, bis der Googlebot ein erneutes Crawling der Seite unternimmt. Auch die Ladezeiten und die Struktur einer Webseite sowie die Aktualisierungsfrequenz der Inhalte spielen eine Rolle. **Ein Normwert lässt sich nicht ermitteln**.

Eine Seite, die viele hochwertige Backlinks besitzt, kann durchaus Im Minutentakt vom Googlebot ausgelesen werden. Bei kleineren Seiten mit wenig Backlinks kann es auch mal einen Monat oder länger dauern, bis der Crawler zurückkehrt und die Seite wieder von Google indiziert wird.

Grundsätzlich solltest Du für einen **nachhaltigen Linkaufbau** sorgen und Deine Inhalte regelmäßig aktualisieren. Halte Deine Themen relevant und hochwertig, um regelmäßig vom Crawler besucht zu werden. Sorge für eine **suchmaschinenfreundliche Struktur** der Webseiten-Navigation und halte die Ladezeiten durch professionelles Webdesign gering. Kurz zusammengefasst heißt das:

- Inhalte kontinuierlich **aktualisieren**
- **Hochwertigen, relevanten** Content bieten

- Eine **klare Seitenstruktur** schaffen
- **Schnelle Ladezeiten** sicherstellen
- **Interne Verlinkungen** strategisch einsetzen

Vermeide veraltete Black-Hat-SEO-Methoden wie Keyword-Stuffing oder künstlich erzeugte Backlinks. Google erkennt Manipulationsversuche mithilfe von Programmen wie **Google Penguin, Google Panda** oder **Helpful Content Update**. Google bietet jedoch weitere Möglichkeiten, um die Häufigkeit der Abfrage durch die Googlebots zu verbessern. Außerdem setzt seit März 2021 Google für fast alle Websites standardmäßig auf Mobile-First-Indexierung – das heißt: Die mobile Version Deiner Website ist für die Indexierung entscheidend. Achte also auf Responsive Design und Mobile Usability.

WEITERE TIPPS FÜR BESSERES CRAWLING

1. GOOGLE SEARCH CONSOLE NUTZEN

Wenn Du neue Inhalte veröffentlicht hast, kannst Du diese direkt über die Google Search Console zur Indexierung einreichen. Auf diese Weise brauchst Du nicht bis zum nächsten Crawlen durch den Googlebot warten, sondern **forderst seinen Besuch quasi selbst an**. Das **beschleunigt** die Sichtbarkeit Deiner Seite – besonders bei aktuellen Inhalten oder Produktneuheiten.

2. SITEMAP GEZIELT EINSETZEN

Pflege eine XML-Sitemap und hinterlege dort Prioritäten (z.B. Startseite mit hoher Priorität). Damit gibst Du an, welche Seiten besonders häufig aktualisiert werden und deshalb auch **öfter** vom Googlebot **analysiert werden sollten**. Ist die Startseite also beispielsweise an erster Stelle, ist sie mit **besonders hoher Priorität** versehen und würde daher auch tendenziell öfter vom Googlebot gecrawlt werden.

3. ROBOTS.TXT, NOINDEX- UND CANONICAL-TAGS

KONFIGURIEREN

Mit einer robots.txt-Datei steuert der Webmaster, welche Seiten **gecrawlt werden dürfen**. Nutze das gezielt, um unwichtige oder unfertige Inhalte (z.B. Admin-Bereiche, Filter-URLs) auszuschließen. Aber Achtung: Wird eine Seite in der robots.txt blockiert, kann sie zwar nicht gecrawlt werden – aber unter Umständen trotzdem indexiert werden, wenn z.B. externe Links auf sie verweisen.

Um sicherzugehen, dass eine Seite **nicht indexiert** wird, sollte **zusätzlich ein noindex-Tag** gesetzt werden. Dieses Meta-Tag wird direkt im HTML-Code der Seite eingebunden und signalisiert Google eindeutig: Diese Seite **soll zwar gecrawlt, aber nicht indexiert werden**. Das ist besonders hilfreich bei rechtlich sensiblen Seiten, internen Suchergebnissen oder Duplicate Content.

Der Einsatz von **Canonical-Tags** hilft im Gegensatz dazu, wenn es mehrere Versionen einer Seite mit sehr ähnlichem Inhalt gibt – zum Beispiel bei Produktvarianten oder Tracking-URLs. Mit dem Canonical-Tag wird Google mitgeteilt, welche Version als **originale und bevorzugte** Seite indexiert werden soll. So lässt sich **Duplicate Content vermeiden** und die **Autorität der „Hauptseite“ stärken**.

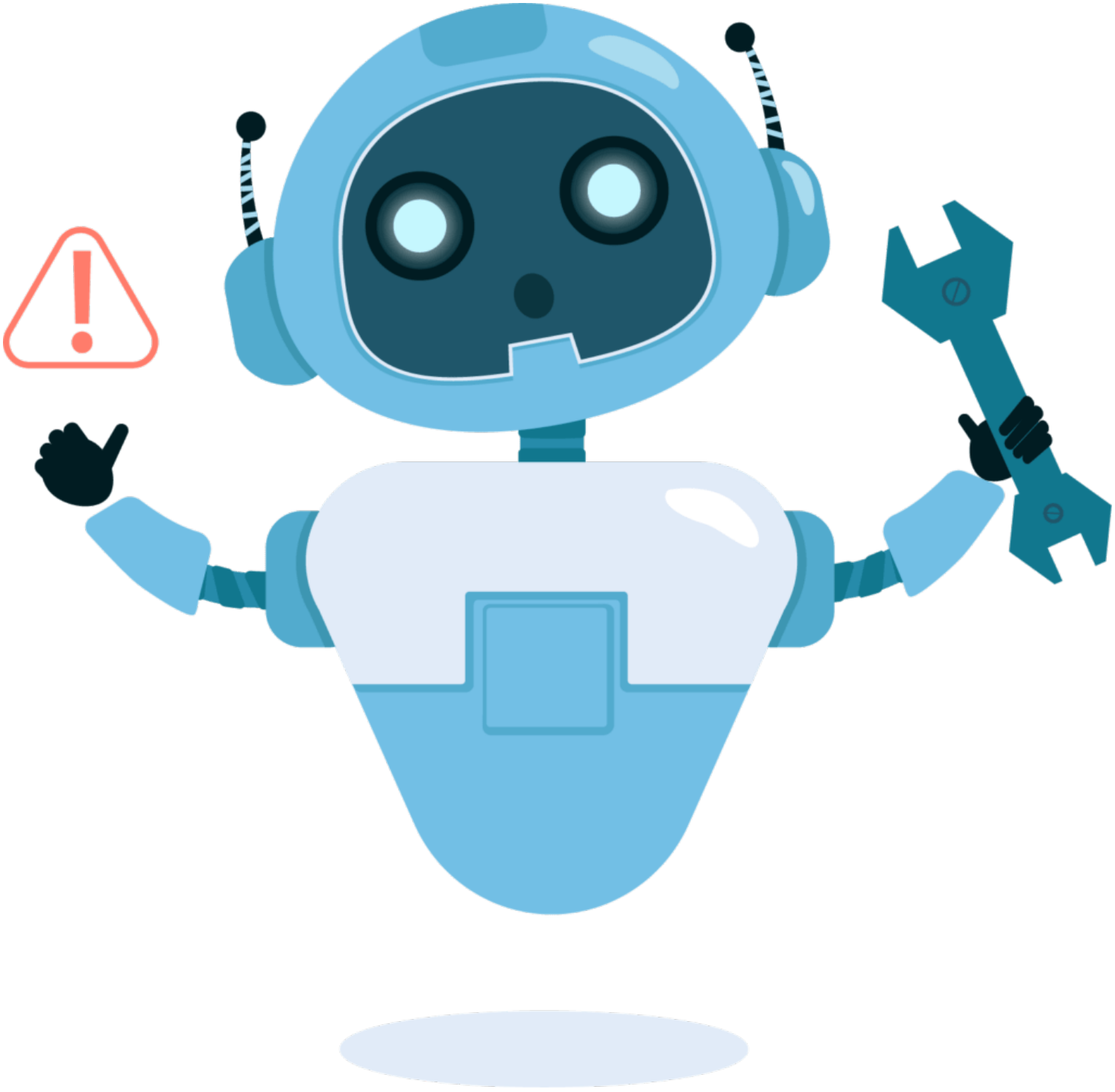
Das heißt: Verwende noindex, wenn eine Seite **gar nicht erscheinen soll**, und canonical, wenn Du **ähnliche Seiten zusammenfassen** willst, ohne Inhalte komplett auszuschließen.

Kriterium	**robots.txt**	**noindex-Tag**	**Canonical-Tag**
Zweck	Verhindert das **Crawling** bestimmter Seiten	Verhindert die **Indexierung**, erlaubt aber Crawling	Verweist auf die **bevorzugte Version** bei ähnlichem Inhalt
Einbindung	In die Datei /robots.txt im Root-Verzeichnis	Im <head> der HTML-Seite oder per HTTP-Header	Im <head> der HTML-Seite
Crawler-Zugriff	Bot darf Seite **nicht besuchen**	Bot darf Seite besuchen, aber **nicht indexieren**	Bot darf alle Seiten crawlen, folgt aber der Canonical-Angabe
Sichtbarkeit in Google	Kann **trotz Blockade** indexiert werden, z.B. über Links	Seite wird **nicht** in den Suchergebnissen angezeigt	Nur die **kanonische URL** erscheint in den Suchergebnissen
Einsatz bei Duplicate Content	Eingeschränkt sinnvoll (z.B. für Sortierfilter)	Ja, wenn Inhalte gar nicht gelistet werden sollen	Ja, wenn ähnliche Seiten **zusammengefasst** werden sollen
Typische Anwendung	Admin-Bereiche, Ressourcen, Filter-Parameter	Interne Suchergebnisse, AGB, Danke-Seiten	Produktvarianten, Sortieroptionen, Tracking-Parameter

Kriterium	robots.txt	noindex-Tag	Canonical-Tag
Risiken bei falscher Anwendung	Wichtige Seiten könnten **nicht gecrawlt** werden	Wichtige Seiten könnten **aus dem Index verschwinden**	Relevante Seiten könnten ihre **Rankingkraft verlieren**

Soll das Crawling **optimal gesteuert** werden, kann es sich anbieten alle drei Methoden gezielt und in Kombination einzusetzen.

Diese Angaben sind **technische Empfehlungen**, keine harten Verbote. Während seriöse Suchmaschinen wie Google oder Bing sich an robots.txt, noindex und Canonical-Tags halten, gibt es Bots, die sich **nicht darum scheren** und trotzdem crawlen oder indexieren. Diese lassen sich über solche Mittel leider **nicht zuverlässig aussperren**. Wer sensible Bereiche schützen will, sollte zusätzlich auf **Login-Schranken oder serverseitige Sperren** setzen.



CRAWL BUDGET

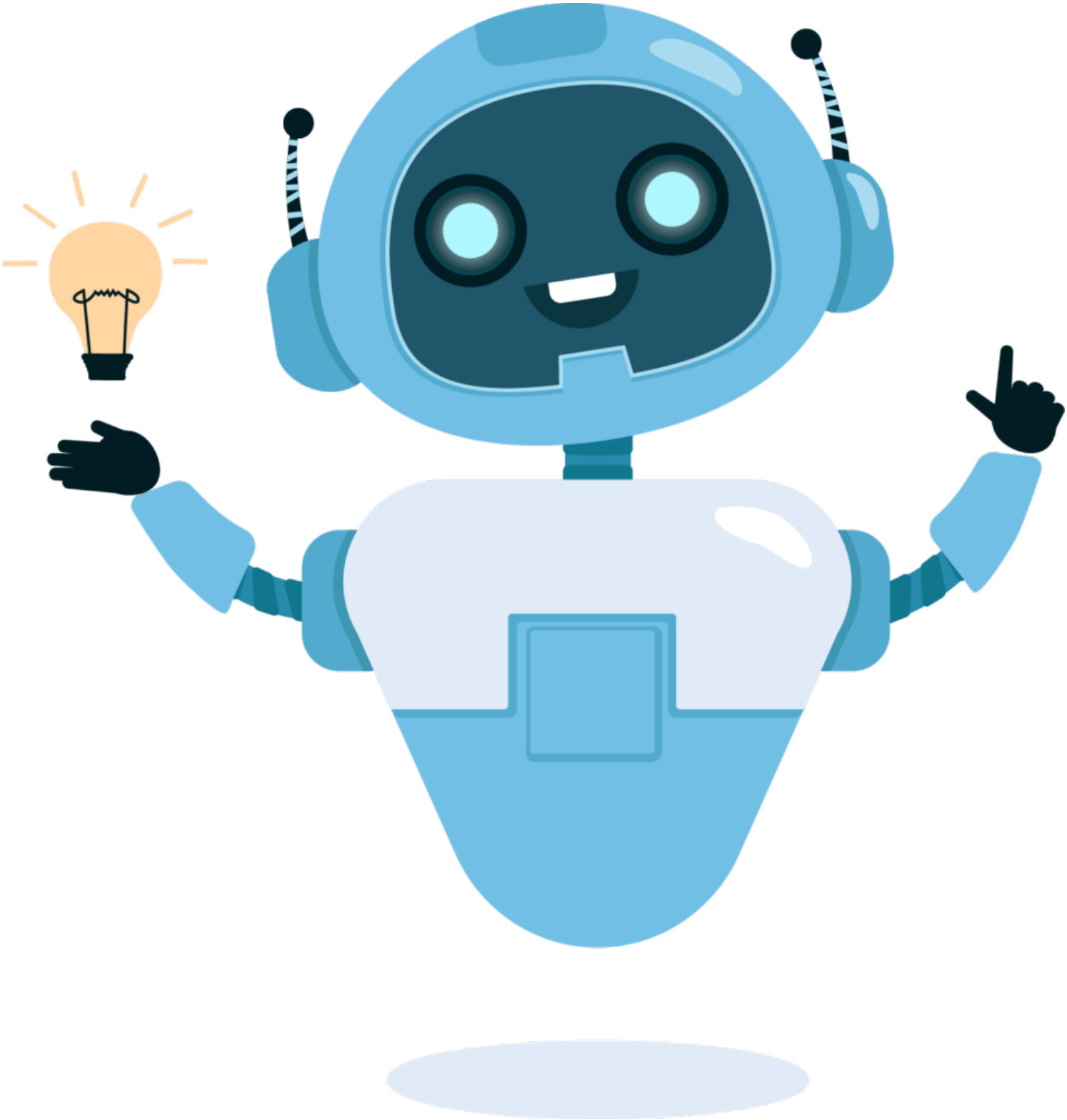
Unter dem Begriff „Crawl Budget“ versteht man die Anzahl an Seiten, die der Googlebot innerhalb eines bestimmten Zeitraums auf Deiner Website crawlen kann und will. Dieses Budget ergibt sich

aus zwei Faktoren: der **Crawl-Rate** (also wie viele Anfragen der Bot pro Sekunde stellen darf) und der **Crawl-Demand** (wie groß das Interesse an Deinen Inhalten ist).

Technische Probleme auf Deiner Website **können dieses Budget negativ beeinflussen**. Dazu zählen etwa lange Ladezeiten, zu viele Weiterleitungen oder fehlerhafte Seiten. Auch sogenannte HTTP-Statuscodes spielen hier eine Rolle:

- **404** bedeutet, dass eine Seite nicht gefunden wurde,
- **301** steht für eine permanente Weiterleitung,
- **500** signalisiert einen Serverfehler.

Wenn solche Meldungen zu häufig auftreten, kann das den Googlebot **ausbremsen** oder dafür sorgen, dass Deine Inhalte **seltener berücksichtigt** werden. Um das Crawling optimal zu unterstützen, sollten diese Fehler möglichst vermieden und regelmäßig mithilfe von Tools wie der Google Search Console überprüft werden.



301-Weiterleitungen (permanente Redirects) sind dabei **keine Fehler** – im Gegenteil: Sie sind ein wichtiges Mittel, um alte URLs sauber auf neue umzuleiten. **Problematisch wird es erst, wenn sogenannte Redirect-Ketten** entstehen, also z.B. eine Weiterleitung von A B C D. Solche Ketten kosten Zeit und Crawling-Ressourcen. Besser ist es, direkt von A D zu leiten.

FAZIT: GOOGLEBOT VERSTEHEN HEISST SEO VERBESSERN

Der Googlebot ist das **Tor zu Deinem Ranking bei Google**. Nur was gecrawlt und indexiert wurde, kann auch gefunden werden. Durch **technische Optimierung, hochwertigen Content und eine gezielte Strategie** in Google Search Console und Sitemap kann die Sichtbarkeit nachhaltig verbessert werden.

Du willst Deine Website gezielt für den Googlebot und die Suchmaschine optimieren, weißt aber nicht so recht, wo Du anfangen sollst? Unser Team unterstützt Dich gerne mit fundierter Erfahrung im Bereich SEO und Online-Marketing!