

Crawl Budget

Der Begriff „Crawl Budget“ beschreibt den zeitlichen Rahmen, in welchem der Webcrawler von [Google](#) das Crawlen, Indizieren und Positionieren einer Website ansetzt.

Crawl Budget – Schlüsselbegriffe

Googlebot

Der Googlebot (Webcrawler) ist für die Analyse aller Websites im [WorldWideWeb](#) verantwortlich. Täglich führt der Googlebot einen [FreshCrawl](#) und im Abstand von circa 30 Tagen einen [Deep Crawl](#) durch.

1. Scannen

Zunächst wird die Zielseite Seite vom Googlebot (Webcrawler) „gescannt“. Hierbei kann es sich sowohl um eine neue Zielseite (Landingpage/Unterseite) oder eine bereits [indexierte Seite](#) handeln.

2. Indizierung

Simultan zum Scan-Vorgang wird die Seite zudem „indiziert“, was bedeutet, dass die [URL](#) als sichtbares Ergebnis in den Datenbanken der [SERPs](#) berücksichtigt wird.

3. Positionierung

Abschließend zum Crawl-Vorgang fällt die Suchmaschine anhand verschiedener Faktoren wie der [Ladezeit](#), [Term-Ausprägungen](#), [Verlinkungen](#) usw. die Entscheidung wo die Seite in der Google-Suche positioniert wird.

Crawl Budget – Definition

Das Crawl-Budget entscheidet darüber, wie oft die relevantesten Seiten einer Website analysiert werden. Je nach Qualität der Inhalte und anhand der [Usability](#) werden hochwertige Seiten öfter gecrawlt. Gemäß einer Aussage von Matt Cutts (Google) werden Websites mit gutem [PageRank](#) mit mehr Crawl-Budget versehen [{1}](#).

Unterscheidung – Crawl Budget und Index Budget

Das Crawl-Budget ist vom Index Budget zu unterscheiden. Das Index-Budget beschreibt das Volumen an URLs, welche in den Index aufgenommen werden können. Eine mangelnde Ausschöpfung des Index-Budgets kommt unter anderem dann zum tragen, wenn eine Website, Seiten beinhaltet, welche einen [404-Fehlercode](#) ausweisen. Jede Seite, welche analysiert wird, belastet das mögliche Crawl-Budget.

Crawl Budget – Mögliche Komplikationen

Durch das beinhalten vieler Seiten, welche einen 404-Fehler aufweisen, wird das Index Budget nicht vollumfänglich ausgeschöpft. So kann es zu Komplikationen beim crawlen der gesamten Website kommen und es besteht die Möglichkeit, dass lediglich ein Bruchteil aller Seiten gescannt wird. Dieses Problem tritt meist bei großen Websites mit zahllosen Unterseiten auf.

Zusammenhang: Crawl Budget und SEO

Durch den beschriebenen Umstand, hat das Crawl-Budget einen direkten Einfluss auf die Sichtbarkeit einer Website. So bildete sich in der [Suchmaschinenoptimierung](#) der Teilbereich „Crawl Budget Optimization“. Der Begriff beschreibt die Möglichkeit den Googlebot dahingehend zu lenken, dass dieser nur ausgewählte und inhaltlich elitäre Zielseiten besucht. Durch das Ausschließen von Seiten, welche einen Fehlercode beinhalten oder mangelhafte Inhalte aufweisen, kann das Crawl-Budget effizienter verwendet werden. Mithilfe der Selektierung von Prioritätsseiten können hochwertige Seiten so gestaltet werden, dass sie vom Googlebot als wichtig eingestuft werden.

Crawl Budget – Handhabung

Um das Crawl-Budget einer Website zu ermitteln, kann die [Google Search Console](#) verwendet werden. Durch das – von Google bereitgestellte – Tool kann das aktuelle Crawl-Budget und die vergangene Zeit des letzten Besuchs des Crawlers ermittelt werden.

Anhand der Diagramme lassen sich folgende Werte ablesen:

1. Die täglich gescannten Zielseiten.
2. Die heruntergeladenen Bytes.
3. Der zeitliche Rahmen, welcher für das Herunterladen genutzt wurde.

Grundsätzlich gelten die folgenden Aussagen:

- Je höher die Anzahl der gescannten Zielseiten, desto besser.
- Für die Werte in Diagramm 2) und 3) ist das Gegenteil der Fall.
- Ein schnelles Herunterladen bedeutet, dass mit dem angesetzten Crawl-Budget mehr Zielseiten geprüft werden können.

Crawl Budget Optimization

Zur Erweiterung und Optimierung des angesetzten Crawl-Budgets für eine Website, haben SEOs und Webmaster folgende Möglichkeiten um Besuche des Googlebots zu fördern:

- **Geringe Klicktiefe:** Eine [flache Architektur](#) der Seiten hilft dabei, Zielseiten schneller zu erreichen und wird auch vom Googlebot beim scannen und indizieren der Seite wertgeschätzt.
- **Interne Verlinkungen:** Sinnvolle [interne Verlinkungen](#) erhöhen nicht nur die Usability, sondern geben auch dem Googlebot Aufschluss darüber, welche Seiten besonders wichtig sind.
- **Zweitrangige Seiten ausschließen:** Zweitrangige oder unwichtige Seiten können mit Hilfe der [robots.txt](#) von der Indexierung ausgeschlossen werden (Das empfiehlt sich unter anderem für: Kontaktformular- oder Login-Seiten)
- **Bereitstellung einer Sitemap:** Mit der Bereitstellung einer XML-Sitemap, welche die relevantesten und wichtigsten Seiten beinhaltet, wird sowohl dem Webmaster als auch dem Googlebot ein großer Gefallen getan.
- **Vermeidung von [Duplicate Content](#):** Zugunsten der Suchmaschinenoptimierung und des Crawl-Budgets-, sollten Seiten mit identischem Inhalt aus der Seiten-Kartei entfernt werden und durch eine 301-Weiterleitung umgeleitet werden. Doppelte Inhalte sind nicht nur schlecht für die SEO, sondern verschwenden auch das angesetzte Crawl-Budget.
- **Regelmäßig neue Inhalte:** Neue Inhalte oder auch „Fresh Content“ sorgen in der Regel dafür, dass sich der Googlebot mehr Zeit für eine Seite nimmt und können genutzt werden, um das Crawl-Budget weiter auszubauen. (Aktualisierte oder ergänzte Inhalte auf einer Seite einzufügen, sorgt aus Sicht von Google auch für mehr Relevanz in Bezug auf die Aktualität der Website).
- **Hilfreiche Informationstexte:** Hochwertige Inhalte werden von Google als elitär angesehen und sorgen für gute Ranking-Signale. Guter Content = Guter PageRank = Mehr Crawl-Budget für eine Website.
- **Neue Strukturen:** Neue oder verbesserte Strukturen einer Seite wirken unterstützend für eine optimale Ausschöpfung des Crawl-Budgets. Damit sind zum einen die Klicktiefe einer Zielseite (Jede Seite sollte unabhängig von welchem Ausgangspunkt mit lediglich 3 Klicks erreichbar sein)

und die Vermeidung von Sackgassen auf einer Webseite gemeint.

- **Schneller Server:** Der zeitliche Rahmen, den der Googlebot auf einer Webseite verbringt ist begrenzt. Daher ist es umso wichtiger auch einen schnellen Server in Anspruch zu nehmen, welcher die Ladezeiten der Seiten als solche und auch die Downloadgeschwindigkeit des Googlebots so niedrig wie möglich hält.
- Frequentiertes publizieren von qualitativ hochwertigen Inhalten.
- Überarbeitung von veraltetem [Content](#) oder Veredelung von SEO-Texten.
- Implementierung von WhitePaper-Dateien/Info-Sheets (PDF-Downloads), Videos oder Bildern auf einer Seite (Auch diese Elemente werden von Google indirekt als Content wahrgenommen).
- Synchronisierung von Sitemaps in der Google Search Console.
- Interne Verlinkungen auf die Zielseite
- Zuwachs von [Do-Follow Backlinks](#) (Je mehr Links zu einer Seite führen, desto relevanter wird Google diese einschätzen).