

## Deep Crawl

### Was ist ein Deep Crawl?

Per Definition bezeichnet **Deep Crawl** die Fähigkeit eines [Suchmaschinen Crawlers](#), selbstständig Unterseiten einer Website nach Inhalten zu durchsuchen und diese zu indexieren. Bei einigen Suchmaschinen ist die Indexierung von Inhalten auf eine bestimmte Verzeichnistiefe der Internetseite begrenzt. Der Deep Crawl erreicht, dass auch Unterseiten einer Website als eigenständige Seiten bewertet und indexiert werden. Ein Deep Crawl kann bei umfangreichen Internetseiten mehrere Stunden in Anspruch nehmen.

### Ziel des Deep Crawls

Ein Deep Crawl verfolgt die Absicht, Tiefen-Verlinkungen (sogenannte [Deep Links](#)) aufzuspüren, um immer tiefer in die Website-Strukturen einzudringen. Dem Crawler gelingt es auf diese Weise sämtliche Unterseiten einer Domain zu durchsuchen, sie als eigenständige Seiten zu werten und in der Folge zu indexieren. Unterseiten können durch Ihre [Indexierung](#) selbstständig ranken und werden bei einer Suchanfrage als eigenständige Bestandteile einer Website ausgegeben.

### Wie funktioniert ein Deep Crawl?

Im Grunde sorgt der Deep-Crawl für eine Tiefen-Indexierung von Websites. Der Crawler sucht auf den unterschiedlichsten Website-Ebenen nach Deep Links, denen er folgen kann. Deep Links bezeichnen hierbei Verweise, die tieferliegende Unterseiten miteinander verknüpfen und nicht auf die Startseite einer Webpräsenz verweisen.

Zum Beispiel: Der Lexikonbeitrag [seo-kueche.de/lexikon/crawler/](https://seo-kueche.de/lexikon/crawler/), der eine Unterseite dieser Website darstellt, verweist auf die Unterseite [seo-kueche.de/lexikon/algorithmus/](https://seo-kueche.de/lexikon/algorithmus/). Dieser Verweis stellt einen solchen „Deep Link“ dar, dem Google folgt. Im Lexikonbeitrag von „Algorithmus“ befinden sich nun weitere Verlinkungen, die vom Crawler untersucht werden. Auf diese Weise handelt sich Google von Seite zu Seite und verschafft sich einen Überblick über die Website.

Die Indexierung der Unterseiten erfolgt nach altbekannter Art. Der Quellcode der Website wird von Google ausgelesen und auf Basis der Ranking-Faktoren ausgewertet. An dieser Stelle greifen die

[Ranking-Kriterien](#) der jeweiligen Suchmaschine. Die Webseite wird bewertet und im Idealfall mit einem guten Ranking bei relevanten Suchanfragen belohnt.

## Rolle des Crawl Budget beim Deep Crawl

Durch den Deep Crawl haben auch tieferliegende Unterseiten komplexer und umfangreicher Websites die Chance, eine gute Platzierung in den Ergebnislisten einer Suchmaschine zu erreichen. **Ob Google eine einzelne Unterseite überhaupt crawlt, kann von mehreren Faktoren abhängen.** Einer davon ist das Crawl Budget. Dieser Wert gibt die Anzahl der Seiten an, die Google auf einer Domain maximal durchsucht. Websites, die wenige Tausend URLs aufweisen, haben in der Regel kein Problem mit einem unzureichenden Crawl Budget.

Anders sieht es bei umfangreichen Websites aus. Normalerweise crawlt Google hier nur einen Teil der Unterseiten. Der Seitenbetreiber muss dann befürchten, dass ihm wichtiger Traffic durch eine fehlende Indexierung der Unterseiten entgeht. Mit gezielten Maßnahmen kann man dem Problem entgegenwirken und so verhindern, dass Google bei einem Deep Crawl wichtige Inhalte übersieht. Dazu gehören beispielsweise die Entwicklung einer flachen Seitenstruktur oder das Ausschließen unwichtiger Unterseiten vom Crawling.

## Worin unterscheiden sich Deep Crawl und Fresh Crawl?

Mit dem Fresh Crawl und Deep Crawl setzt Google zwei verschiedene Aktualisierungszyklen um. Der Google Deep Crawl untersucht Websites intensiv und folgt jedem Link, um möglichst den gesamten Inhalt einer Website zu erfassen. Das Crawling nimmt bei umfangreichen Websites mehrere Stunden in Anspruch und wird nicht selten in Teilen über mehrere Tage hinweg durchgeführt. Es ist davon auszugehen, dass der Deep Crawl neue Webseiten nicht sofort untersucht.

Dem gegenüber steht der [Fresh Crawl](#), der Webseiten permanent crawlt. Das Ziel von Google ist es, das Internet ständig nach neuen Inhalten zu durchsuchen, und relevante, aktuelle Inhalte schnellstmöglich in den eigenen Index aufzunehmen. Um in kurzer Zeit möglichst viele Websites zu durchforsten, werden diese nur oberflächlich erfasst und auf Veränderungen hin überprüft. Wie oft eine Webpräsenz durch den Web Crawler besucht wird, hängt von Faktoren wie beispielsweise der Stärke des Brands oder ihrer Aktualisierungsrate ab.

## Neue Indexierungs-Struktur durch Google Caffeine

Um den Index noch aktueller zu halten, aktivierte Google im Jahr 2010 eine überarbeitete Datenbank- und Indexstruktur. Anstatt Webseiten – wie bis dahin üblich – in Ebenen zu unterteilen, verwendet Google seither eine netzartige Struktur. Doch welchen Vorteil bietet diese neue Infrastruktur?

Google kann seitdem viel schneller und vor allem flexibler Websites und Informationen crawlen und indexieren. Da täglich mehrere Hundert Gigabyte an Daten hinzukommen, ergibt dieses Update großen Sinn. Darüber hinaus kann erst durch die neue Caffeine-Struktur in Echtzeit auf Inhalte aus Social Media und Nachrichten-Plattformen zugegriffen werden. Google-Nutzer profitieren von dauerhaft hochaktuellen Ergebnislisten.

Weiterführende Informationen:

[What Crawl Budget Means for Googlebot](#)