

Deep Crawl

Stand: 29.11.2022

Ein wichtiges Hilfsmittel, um die Sichtbarkeit der eigenen Website im Internet zu erhöhen, ist, einen Platz möglichst weit vorn in den SERPs bei Google zu ergattern. Ein Großteil der Nutzer beschränkt sich bei der Auswahl des passenden Ergebnisses auf die ersten Einträge, weniger als 5 % klicken überhaupt auf die zweite Seite. Dementsprechend sind die oberen Listenplätze hart umkämpft.

Damit Ihre Website eine Chance hat, ebenfalls weiter oben aufzutauchen, muss der Google-Algorithmus als erste Hürde auf sie aufmerksam werden. Der dazugehörige **Crawler** prüft den Inhalt und nimmt die Seite in den **Index** auf, wenn diese die Kriterien von Google erfüllt.

Was ist ein Deep Crawl?

Ihre Website beinhaltet, je nach Ausrichtung, hilfreiche Informationen oder nützliche Produkte, die Sie der Welt präsentieren möchten. Dabei erstellen Sie keine ellenlange Landing Page, sondern teilen die Inhalte **sinnvoll auf Unterseiten** auf. Dadurch setzen Sie **thematische Schwerpunkte** und erleichtern den Nutzern die Navigation.

Durch die verzweigte Struktur ist es wichtig, dass der Crawler jede der Unterseiten erfasst, sodass diejenige in den SERPs erscheint, die für den jeweiligen Suchende **relevant** ist. Für eine vollständige Indizierung führt das Programm einen **Deep Crawl** durch. Der Crawler begreift dabei jede Unterseite **losgelöst** von den anderen. Der Vorgang dauert teilweise bis zu mehreren Stunden.

Funktionsweise des Deep Crawls

Grundsätzlich verfährt der Crawler bei einem Deep Crawl wie bei einem einfachen Crawling. Das Programm sieht den **Quellcode** ein und beurteilt den **Inhalt** anhand von Vorgaben dahingehend, ob es die Seite indexiert und wie hoch es sie rankt.

Bei dieser Variante fokussiert der Crawler sich allerdings auf Unterseiten, die miteinander durch **Deep Links** vernetzt sind. Das ist zum Beispiel bei diesem Wiki der Fall. So arbeitet er sich Schritt für Schritt vor.

Ziel des Deep Crawls

Gibt ein Nutzer einen Begriff in die Suchmaschine ein, erwartet er eine **präzise Auflistung** nützlicher **Geschäftsleitung**: Hartmut Deiwick • Gerichtsstand: AG Braunschweig • Registernummer: HRB 205088

• Ust-IdNr.: DE 250 332 694 • St.-Nr.: 14/201/16808

Bankverbindung: Volksbank Braunschweig • IBAN: DE61 2699 1066 185 2167 000 • BIC: GENODEF1WOB

Websites, die genau auf die Keywords Bezug nehmen. Oft behandeln die Autoren bestimmte Themen aber auf Unterseiten, nicht auf der Startseite. Daher müssen auch diese in den SERPs auftauchen. Durch den Deep Crawl **erschließt** das Programm die **Strukturen** einer Website immer weiter, sodass der Prozess genau das ermöglicht.

Zusammenhang zwischen Crawl Budget und Deep Crawl

Den Crawling-Vorgang führt die jeweilige Suchmaschine **selbstständig** durch, ohne dass der Seitenbetreiber etwas vorbereiten muss. Für jede Website gibt es jedoch eine **Höchstanzahl an Links**, die der Crawler berücksichtigen kann. Das ist das sogenannte Crawl Budget. Es setzt sich aus dem **Crawling-Bedarf** und der **-Frequenz** zusammen.

Das Programm kann jede Website, deren Linkstruktur sich auf wenige Tausend URLs beschränkt, problemlos erfassen. Bei komplexeren Gliederungen ist das Crawl Budget irgendwann aufgebraucht, weswegen der Crawler **nicht alle Unterseiten** in den Index aufnehmen kann. Folglich haben nicht alle einen Eintrag in den SERPs. Sie existieren trotzdem weiterhin, die Leser müssen sich dafür aber durch die Website klicken, um sie zu öffnen.

Als Seitenbetreiber können Sie das unvollständigen Erfassen Ihrer Website vermeiden, indem Sie die **Hierarchie weniger komplex** gestalten. Folgende Punkte erleichtern das Crawlen:

- Achten Sie darauf, dass Content **nur einmalig** auf Ihrer Seite steht und somit nur unter einem Link zu finden ist.
- In der **txt-Datei** können Sie bestimmte Seiten wie Warenkörbe vom Crawling ausklammern und so Platz für wichtigere Unterseiten in den SERPs schaffen.
- Prüfen Sie Ihre Unterseiten insgesamt auf die **Relevanz** für Nutzer und legen Sie gegebenenfalls Content (und damit Links) zusammen.

Deep Crawl vs. Fresh Crawl

Bei der Indexierung von Websites gibt es **nicht nur einen Crawling-Vorgang**. Je nach Funktion unterscheiden sich die Crawling-Prozesse. Der **Deep Crawl** stellt eine **bestimmte Website** in den Mittelpunkt und durchforstet diese nach tieferen Verlinkungen. Internetangebote, die gerade erst online gegangen sind, bezieht er vermutlich nicht direkt mit ein.

Im Gegensatz dazu zielt ein **Fresh Crawl** darauf ab, aktualisierte Inhalte auf einer Website zu erfassen. Dadurch kann die Suchmaschine gegenwärtige Entwicklungen oder Ereignisse Nutzern schnell zur Verfügung stellen. Dafür crawlt das Programm **Homepages dauerhaft** an der **Oberfläche**, ohne in die Hierarchie einzusteigen. Die Seiten, die User regelmäßig aufrufen, untersucht der Crawler bevorzugt.

Löwenstark Digital Group GmbH

Geschäftsführung: Hartmut Deiwick • Gerichtsstand: AG Braunschweig • Registernummer: HRB 205088

• Ust-IdNr.: DE 250 332 694 • St.-Nr.: 14/201/16808

Bankverbindung: Volksbank Braunschweig • IBAN: DE61 2699 1066 185 2167 000 • BIC:

GENODEF1WOB

Bedeutung des Deep Crawls für die SEO

Dank des Deep-Crawl-Prozesses kann der [Algorithmus](#) jede Unterseite einer Website in den SERPs darstellen. Interessierte finden mit einem schnellen Klick dadurch die Informationen, nach denen sie gesucht haben.

Das bedeutet, dass eine Homepage bei jeder indexierten Unterseite die Chance hat, bei einem bestimmten Keyword weit oben in den Listenplätzen zu erscheinen. Dadurch kann der Betreiber **aus mehreren Richtungen Traffic** auf der Seite **generieren**, indem die Suchmaschine unterschiedliche Unterseiten in Abhängigkeit des Begriffes anzeigt.

Um zu erreichen, dass die eigene Website durch die dazugehörigen Unterseiten oft in den Ergebnissen vertreten ist, müssen Sie allerdings **jede** der gewünschten Seiten dafür **optimieren**. Nur, wenn die Inhalte möglichst viele [Rankingfaktoren](#) erfüllen, belohnt der Algorithmus sie mit einer hohen Sichtbarkeit.

Betreiber neu erstellter Websites müssen sich etwas gedulden, bis das Programm all ihre Unterseiten crawlt und ihren [Pagerank](#), also die Position in der Ergebnisliste, durch den Vorgang bestimmt.