Indexierung

Stand: 13.10.2022

Websites liefern viele wichtige Informationen zu interessanten Themen, seien es Neuigkeiten, angebotene Produkte oder Dienstleistungen. Das Ganze trägt aber nur Früchte, wenn Nutzer die Seiten kennen und sie beispielsweise über bekannte Suchmaschinen finden können. Andernfalls stehen die Chancen vergleichbar schlecht, dass sie Traffic generieren. Um Sichtbarkeit in den Ergebnissen der Suchmaschinen generieren zu können, müssen sie erst einmal in deren Verzeichnis stehen. Das geschieht mithilfe der Indexierung.

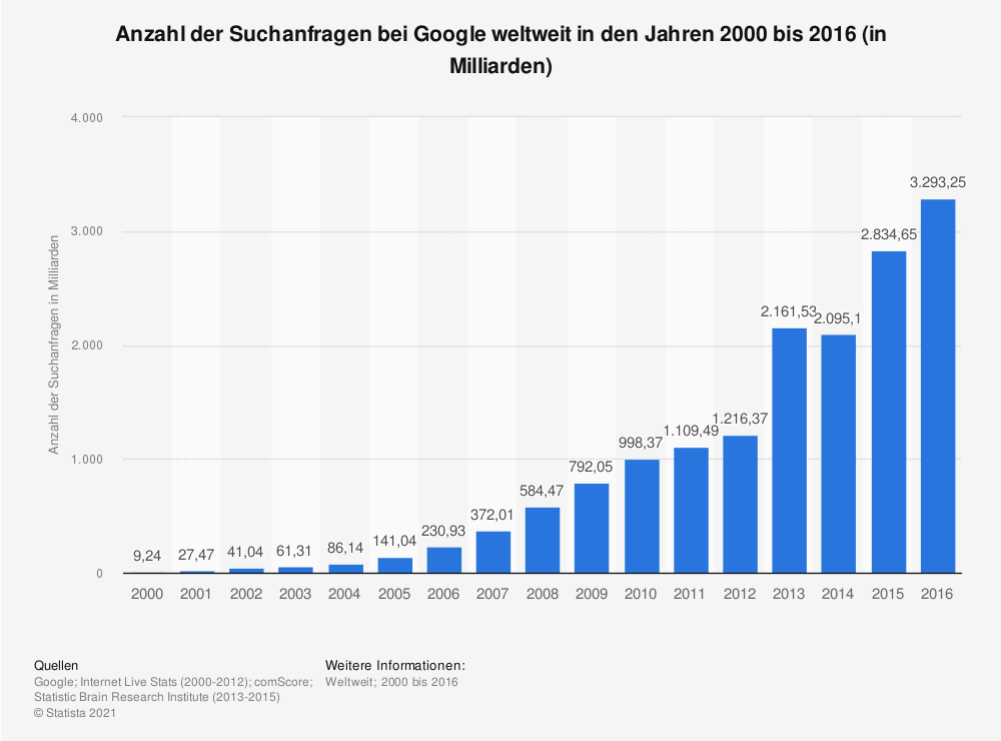
Definition

Die Indexierung ist ein mögliches Verfahren der **Informationsaufnahme**. Im Onlinezusammenhang geht es konkret darum, wie **Suchmaschinen** Daten gewinnen. Dafür nutzen sie einen Crawler, um Websites in den eigenen Index aufzunehmen. Dieser durchsucht die Informationen der Haupt- und Unterseiten.

Der **Index** ist dabei ein großes **strukturiertes Verzeichnis**, aus dem der Google-Algorithmus seine Informationen bezieht. Er zeigt also nur die Websites in den SERPs an, die darin gespeichert sind. Die Sammlung wird **ständig überarbeitet**, da das Programm neue Websites hinzufügt und bei Verstößen ebenfalls bestehende aus dem Index streicht.

Sobald eine Website im Index gelistet ist, kann sie oder eine passende Unterseite überhaupt erst in den Suchergebnissen auftauchen. So werden Nutzer auf sie aufmerksam und die Chancen einer Interaktion steigen, wenn die Suchmaschine sie auf den vorderen Plätzen anzeigt.

Ablauf einer Indexierung



Quelle: Statista

Doch wie schafft es jede der vielen Seiten im Internet auf diese Liste? *Das Durchsuchen der Seiten wird auch crawlen genannt*. Der eingesetzte Crawler **prüft**, ob die Inhalte der Website den **Richtlinien** entsprechen, und **nimmt** sie bei **Übereinstimmung auf**, indem er die Daten als Dokument mit spezifischen Stichwörtern in seinem Verzeichnis speichert.

Im Index selbst bilden die Stichwörter die Basis für die Gruppierung thematisch gleicher Websites. Zusätzlich reichert ein Programm die Dokumente mit **Deskriptoren** an, in diesem Fall sind das unter anderem die Keywords oder Rankingfaktoren. Dadurch kann der Algorithmus die Seiten bei einer Suchanfrage auf ihre Relevanz bewerten und ausgeben, sobald die Suchbegriffe zu den jeweiligen Keywords passen.

Bedeutsamkeit der Website-Struktur für die Indexierung

Einer der wichtigsten Aspekte für eine Website, die ein hohes Trafficaufkommen und eine gesteigerte Conversion Rate hat, ist die **Übersichtlichkeit**. Die User sollen sich auf den ersten Blick dort zurechtfinden und alle für sie relevanten Informationen schnell erfassen können, ohne sich durch zahlreiche Reiter und Unterseiten zu klicken.

Das Gleiche gilt für die eingesetzten Bots. Auch sie müssen die präsentierten Inhalte lückenlos erfassen

können, um zeitnah zu entscheiden, ob sie die Seite im Verzeichnis aufnehmen, da das **Crawl Budget begrenzt** ist. Das betrifft unter anderem die nachvollziehbare Kategorisierung der einzelnen Domain-Unterseiten.

Um den Suchmaschinen-Bots das Crawlen zu erleichtern, ist es deshalb hilfreich, eine **XML-Sitemap** zur Verfügung zu stellen, auf der alle URLs der Website aufgeführt sind, die der Algorithmus berücksichtigen soll. Sie kann in der Google Search Console eingereicht werden. Gleichzeitig erleichtert Google Webmastern so die SEO-Indexierung.

Wann ist eine Website indexiert?

In der Regel indexiert Google **eigenständig**, indem die Crawler die Linkstrukturen verfolgen und dabei neue Websites beziehungsweise Inhalte aufnehmen. Darüber hinaus kann ein **Webseitenbesitzer** ebenfalls selbst **aktiv** werden und die Suchmaschine auf die eigene Website aufmerksam machen. Dafür bieten sich ihm drei Möglichkeiten:

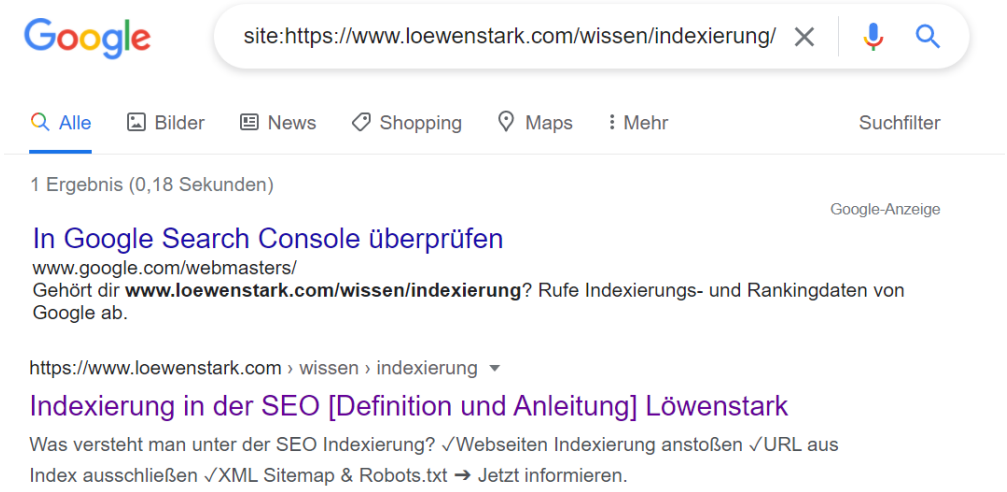
1. Die **Sitemap** in der Google Search Console hinterlegen: Dadurch crawlt der Dienst die eingetragenen URLs meist innerhalb der nächsten 24 Stunden. Dank einer Übersicht haben Webmaster den Überblick über den Crawlingverlauf.
2. Mithilfe des Google-Kontos können Webseitenbetreiber eine **Anfrage** stellen, damit ihre Seite indexiert wird.
3. Betrifft die Indexierung nur bestimmte **Unterseiten**, können sie in der Google Search Console an Google übermittelt werden.

Ist die Website erfolgreich indexiert, zeigt Google in der Search Console an, wenn die Website verifiziert ist. Außerdem nutzen Webmaster die **site:-Abfrage** dafür. Statt in der Suchmaschine mit dem Titel der URL nach der Website zu suchen, setzen sie den Parameter „site:“ vor die genaue Adresse in der Suchanfrage, sodass wirklich nur die Website angezeigt wird statt anderer ähnlicher Suchtreffer, die die gleichen Keywords enthalten. Taucht die Website dann in den SERPs auf, ist sie indexiert.

Hier ein Beispiel:

site:https://www.loewenstark.com/wissen/indexierung/

Und so sieht die site-Abfrage einer indexierten Seiten in Google aus:



Wichtige Seiteneigenschaften für eine erfolgreiche Indexierung

Damit das eingesetzte Programm die Struktur und Inhalte einer Website schnell erfassen kann, gibt es für Webseitenbetreiber einige Aspekte, auf die sie achten sollten:

- Innerhalb der Navigation sollten sie möglichst **auf JavaScript verzichten**, da Google häufig Probleme mit der Skriptsprache hat.
- Der Crawler erfasst **statische URLs** beziehungsweise solche mit wenigen oder keinen Session-IDs sowie übersichtlichen Parametern leichter.
- Die URLs sollten einfach strukturiert aufgebaut sein und eine eindeutige Interpunktion haben. Auf Inlineframes sollten Webmaster verzichten, da der Bot sie nicht crawlen kann.
- **Sitemaps** sollten immer **aktuell** gehalten werden. Das bedeutet, jede einzelne URL, die neu hinzukommt, sollte der Webmaster in die Sitemap aufnehmen.

Indexierung beschleunigen

Da jeden Tag unzählige Websites auf der ganzen Welt online gehen und auch bereits indexierte Onlineauftritte Seitenstrukturen aktualisieren, kann es eine **gewisse Zeit dauern**, bis Google auf eine Website stößt und sie entsprechend crawlt. Selbst, wenn die Suchmaschinenroboter von Google unaufhörlich das Netz durchforsten. Nicht jede Website, die online geht, ist automatisch im Index.

Die Indexierung kann durch die Search Console beschleunigt werden. Legen Webmaster dort eine **HTML-Datei** im Stammverzeichnis der Domain ab, verifiziert Google die Seite und indexiert sie parallel dazu schneller.

Indexierung bestimmter URLs verhindern

Wie bereits angedeutet, hat jede Website ein Crawl Budget. Dies gibt unter anderem die Anzahl der Unterseiten vor, die der Crawlbot in den Index aufnimmt. Da die Indexierung im Normalfall automatisch abläuft, hat der Webseitenbesitzer nur wenig Einfluss darauf.

Das ändert sich, sobald er eine **robots.txt-Datei** im Stammverzeichnis hinterlegt. Dadurch können **Seiten** der URL **explizit** aus dem Index von Google **ausgeschlossen** werden, wenn der Webmaster die **„disallow„-Funktion** dafür nutzt. Der Crawler nimmt die Daten der URL oder der Unterseite in dem Fall meist nicht auf. Neben der Datei kann ein Webmaster den **Meta-Tag „noindex„** im HTML-Dokument setzen.

Ebenso nützlich für die Indexierung ist der **Canonical-Tag**. Hierüber schließt der Verantwortliche Inhalte, die als doppelt interpretiert werden könnten, für den Crawler aus, indem er im Quellcode der Website auf eine kanonische URL als Standardressource verweist. Darüber werden Suchmaschinen informiert, bei welcher URL sich um das Original handelt und welche Webseiten sie für die Suche ignorieren können.

Fazit

Eine Website ist das Aushängeschild und die Visitenkarte für jeden Unternehmer. Daher sollte jeder darauf achten, dass sie bei den Suchmaschinen, insbesondere beim Riesen Google, gelistet ist und Nutzer sie finden können.

Durch die Indexierung ist ein kleiner Teil der **Onpage-Optimierung** schon geschafft. Dadurch kennt der Google-Bot die Website bereits. Jetzt müssen die User ebenfalls auf die Seite aufmerksam werden, indem sie in den Top 10 der SERPs auftaucht. Dafür sind weitere SEO-Maßnahmen erforderlich. Webseitenbetreiber sollten dabei auf eine einfache Navigation innerhalb der Seiten und auf hochwertigen Content Wert legen. Zudem sollten sie ihre Sitemap im Auge behalten und regelmäßig überprüfen, denn das Monitoring ist ein grundlegendes Element der Suchmaschinenoptimierung. Dann steht einem Erfolg im Netz nichts mehr im Weg.