

Indexierte Webseiten

Was sind indexierte Webseiten?

Indexiert sind laut Definition jene Seiten einer Website, die in den Datenbestand und somit in den [Index](#) einer Suchmaschine aufgenommen worden sind. Folglich sind indexierte Webseiten die Voraussetzung, um online über Ergebnislisten der Suchmaschinen gefunden zu werden.

Damit Webseiten in den Index einer Suchmaschine aufgenommen werden können, müssen sie zuvor von der Suchmaschine gecrawlt werden. Die Bots und [Crawler](#), die diese Aufgabe übernehmen, handeln sich unter anderem von Link zu Link und rufen dabei eine Vielzahl von Webseiten auf. Diese werden hinsichtlich ihrer Relevanz bewertet und in den Index aufgenommen. Dabei gilt: Nur indexierte Webseiten können auf den Suchergebnis-Seiten erscheinen und ranken. Und nur jene Webseiten, die ranken, bringen der Domain den erwünschten SEO-Traffic.

Tipps für die Indexierung von Webseiten

Möchte man als Website-Betreiber in den Google-Index aufgenommen werden, ist es notwendig die eigene Domain so übersichtlich und klar zu strukturieren, dass sich der Google-Crawler schnell zurechtfindet. Im Idealfall werden sämtliche relevanten Unterseiten vom Crawler zuverlässig erfasst und indexiert, also über die Suchmaschine auffindbar gemacht. Kann im Gegenzug ein Google-Bot eine Webseite nicht indexieren, werden keine Links zu der entsprechenden Seite in den Google [SERPs](#) ausgegeben.

Folgende Hinweise helfen dabei, dass sich der Google-Bot besser und schneller auf Ihrer Website zurechtfindet:

1. Verwenden Sie eine durchdachte Website-Architektur und achten Sie auf eine gute interne Verlinkung.
2. Erstellen Sie eine stets aktuell gehaltene [XML](#)-Sitemap und reichen Sie diese bei den [Google Webmaster Tools](#) ein.
3. „Räumen“ Sie regelmäßig auf, indem Sie in den Webmaster Tools gemeldete Fehlerseiten prüfen und gegebenenfalls die Ursache beheben.
4. Löschen Sie verwaiste Seiten oder verlinken Sie diese neu.
5. Führen Sie ein kontinuierliches Monitoring durch, um Probleme bei der Indexierung frühzeitig

festzustellen und zu beheben.

Darüber hinaus spielt die Benennung Ihrer [URLs](#) eine wichtige Rolle. Im besten Fall ändern sich die Namen niemals, sondern bleiben dauerhaft bestehen. Machen aber beispielsweise Veränderungen in der Menüstruktur eine Anpassung unumgänglich, müssen technische Vorkehrungen getroffen werden, um Duplikate oder Fehlerseiten zu vermeiden.

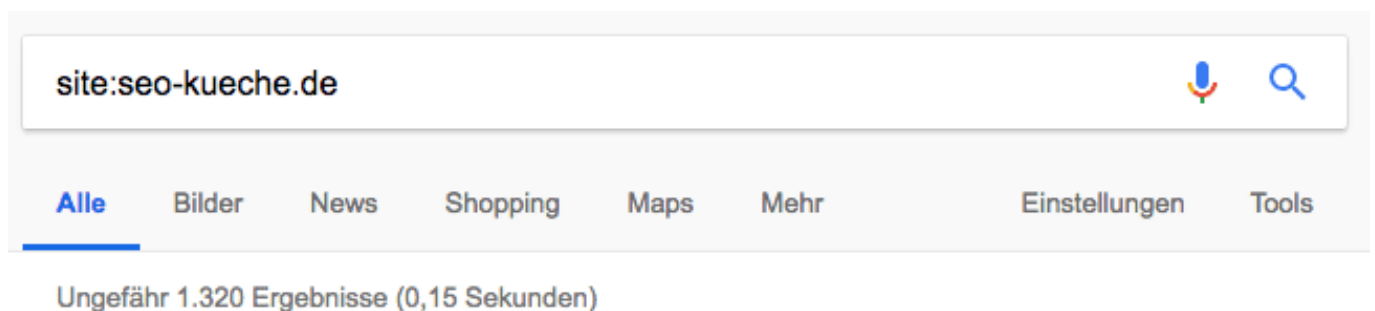
Fallstricke, die es zu beachten gibt, da sonst das Risiko eines kompletten Ranking-Verlusts besteht, treten besonders häufig im Rahmen eines Relaunchs auf. Wägen Sie in einem solchen Fall zunächst sorgfältig ab, welchen Nutzen und Aufwand eine nachträgliche Umstellung der URLs einer ganzen Website besitzt.

Indexierungsstatus einer Website einsehen

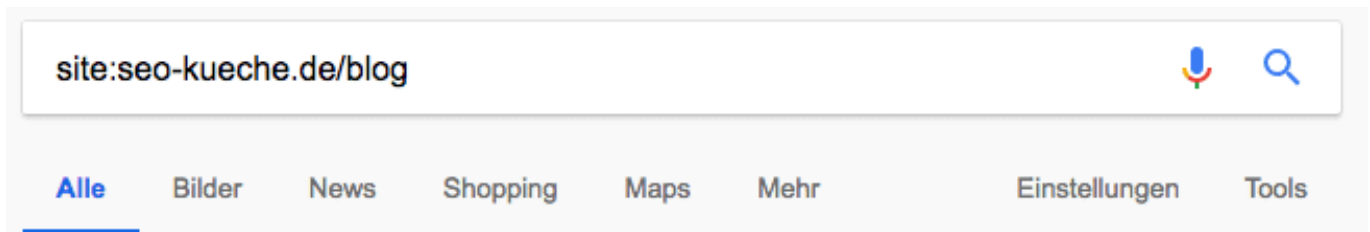
Google bietet Website-Betreibern die Möglichkeit einzusehen, wie viele Unterseiten einer Domain von der Suchmaschine indiziert wurden. Zur Auswahl stehen die `site:-`Abfrage in der Suchmaschine selbst und das Prüfen des Indexierungsstatus in den Google Webmaster Tools als die zuverlässigere Variante.

Indexierungsstatus mit der `site:-`Abfrage beobachten

Geben Sie im Suchfeld von Google den Befehl `site:name-der-website.tld` ein. Also beispielsweise `site:seo-kueche.de`



Sie bekommen nun die Zahl der indizierten Webseiten für die gesamte Domain zurück. Sie können die Suche eingrenzen, indem Sie den Website-Namen zusammen mit einer Oberkategorie eingeben. Also beispielsweise `site:seo-kueche.de/blog`



Ungefähr 716 Ergebnisse (0,17 Sekunden)

Indexierungsstatus in den Google Webmaster Tools einsehen

Eine site:-Abfrage dient in erster Linie der groben Orientierung. Insbesondere bei größeren Domains bietet sich eine Abfrage des Indexierungsstatus über die Google Webmaster Tools an. Zwar können Sie keine Webseiten einsehen, dafür allerdings die Gesamtanzahl der indexierten Seiten. Und nicht nur das: Sie können mithilfe der Webmaster Tools nachvollziehen, wie sich die indexierten Seiten Ihrer Domain in den vergangenen zwölf Monaten entwickelt haben.

Überprüfen Sie dabei, ob die von Google ausgegebene Anzahl an indexierten URLs dem Wert entspricht, den Sie erwartet haben. Weicht Ihr erwarteter Wert erheblich vom tatsächlichen ab, kann ein strukturelles Problem die Ursache sein. Auch bei einem plötzlich stattfindenden erheblichen Anstieg oder Rückgang der indexierten Webseiten, kann das auf Probleme hindeuten, die gelöst werden müssen.

Erwartete und tatsächliche Anzahl indizierter Webseiten

Der Indexierungsstatus ist erst dann von Wert, wenn Website-Betreiber wissen, wie viele URLs tatsächlich indexiert sein sollten. Die von Ihnen erwartete Gesamtanzahl an indexierten Webseiten können Sie anschließend mit dem Indexierungsstatus abgleichen.

Weicht der erwartete Wert deutlich vom tatsächlichen ab, kann das unterschiedliche Gründe zur Ursache haben, die hier beispielhaft dargestellt werden.

Mehr URLs indexiert als erwartet:

- Duplikate durch unterschiedliche Schreibweisen in URL (z.B. Nicht-Beachtung der Klein- und

Großschreibung)

- nicht mehr existente URLs werden als valides Ergebnis (200-Statuscode) ausgegeben, statt mit 404- oder 410-Statuscode gekennzeichnet
- Verwendung der Session-ID in URLs
- URLs mit unbrauchbaren Parametern, die nicht von der Indexierung ausgeschlossen wurden
- veraltete, fehlerhafte Sitemap

Weniger URLs indexiert als erwartet:

- Domain wurde erst kürzlich online gestellt
- Fehlende (interne) Verlinkung mancher Unterseiten, sogenannte verwaiste Seiten
- Einzelseiten stehen ungewollt auf noindex und werden von der Indexierung ausgeschlossen
- Durch die robots.txt ungewollt blockierte Ordner und Dateien
- fehlerhafte Verwendung der [Canonical Tags](#)
- Duplicate Content oder Content ohne Mehrwert für den Besucher
- Menge an Unterseiten ist zu groß und wurde vom Google-Bot nicht vollständig gecrawlt

Warum ist eine Kontrolle der Indexierung wichtig?

Bei der strukturellen Gestaltung einer Website gilt der Grundsatz: Nutzen Sie so viele URLs wie nötig und achten Sie darauf die Website nicht unnötig „aufzublähen“. Diese Regel wird umso wichtiger, wenn es um die Indexierung von Webseiten geht. Denn aus SEO-Sicht ist selten jede URL so relevant, dass sie über eine Suchmaschine wie Google gefunden werden müsste. Stattdessen sollte es das Ziel sein, dass ein von der Websuche kommender Nutzer direkt auf der Webseite einsteigt, die für ihn am relevantesten ist.

Zu diesen nicht-relevanten Webseiten zählen vor allem jene, die nicht für ein bestimmtes Keyword optimiert sind.

Dazu gehören:

- nicht themenrelevante Seiten (betrifft alle Websites)
- Kategorienseiten mit Paginierung (betrifft Onlineshops)
- Checkout (betrifft Onlineshops)
- dynamische Seiten (z.B. seiteninterne Suche)

Alle nicht themenrelevanten Seiten können prinzipiell auf noindex gesetzt werden. Beispiele hierfür wären Impressum, AGB, Widerrufsrecht, Datenschutzerklärung etc.

Dazu kommen bei Shops Paginierungsseiten. Durch die Paginierung (Blätter-Funktion) im Onlineshop entstehen mehrere, beinahe identische Kategorienseiten (Seite 1, Seite 2 ...). Damit diese Seiten im Ranking nicht miteinander konkurrieren, sollten diese SEO-konform eingebunden werden, um [Duplicate Content](#) zu verhindern.

Einfache Inhaltsseiten, die nur wenig hilfreiche Informationen bereitstellen, haben keine Daseinsberechtigung und sollten gelöscht oder aufgewertet werden.

Am wichtigsten ist es, dass Website-Betreiber den Indexierungsstatus regelmäßig auf seine Entwicklung hin überprüfen. Auf diese Weise kann schnell eingegriffen werden, wenn Google viele neue Website-Inhalte crawlt, ohne dass diese für den Nutzer neue Informationen bereitstellen. Zum anderen können Angriffe aus dem [Black Hat SEO](#)-Bereich schneller erkannt und unschädlich gemacht werden.

In dem Lexikonbeitrag Index/Indexierung erfahren Sie weiterführend, wie Sie eine Webseite indexieren lassen, wie Sie eine Indexierung verhindern und wie Sie mit von Google entfernten Webseiten in den Google-Index zurück gelangen.

Weiterführende Informationen:

<https://www.google.de/insidesearch/howsearchworks/thestory/index.html>

https://developers.google.com/search/reference/robots_txt