Risiken einer dauerhaften Verwendung von noindex, follow: Isolation von HTML-Dokumenten



Risiken einer dauerhaften Verwendung von noindex, follow: Isolation von HTML-Dokumenten – websiteboosting.com

Wenn Suchmaschinen Dokumente plötzlich anders bewerten, kann dies zu schwerwiegenden Folgen für Webseitenbetreiber führen. Auch wenn dies mit vorheriger Ansage – wie beim Langzeiteinsatz von noindex, follow – passiert, ignorieren einige SEOs die möglichen Konsequenzen. Zwei ausgewiesene…
Wenn Suchmaschinen Dokumente plötzlich anders bewerten, kann dies zu schwerwiegenden Folgen für Webseitenbetreiber führen. Auch wenn dies mit vorheriger Ansage – wie beim Langzeiteinsatz von noindex, follow – passiert, ignorieren einige SEOs die möglichen Konsequenzen. Zwei ausgewiesene SEO-Experten, Tobias Schwarz und Sören Bendig, erklären anhand eingängiger Beispiele, wie und warum Seiten vom Link-Graphen isoliert werden können.

Um das Internet crawlen zu können, müssen Suchmaschinen eine sehr große Menge an URLs erfassen. Damit dies sinnvoll und wirtschaftlich erfolgen kann, bedienen sich moderne Suchmaschinen einer Vielzahl an Techniken, um wichtige, weniger wichtige und unwichtige Dokumente effektiv zu crawlen. Sie passen diesen Prozess kontinuierlich an und versuchen dabei, ihre eigenen Ressourcen möglichst effizient einzusetzen.

Wenn Google noindex, follow nachträglich als noindex, nofollow interpretiert

Aus Sicht der Suchmaschinen ist das Crawlen von Dokumenten, die von der Indizierung ausgeschlossen sind, ineffektiv. Ende 2017 erklärte John Mueller (Google Webmaster Trends Analyst) in einem Google Webmaster Hangout (http://einfach.st/jmu43), dass sich die Bewertung von langfristig als noindex markierten Seiten im Laufe der Zeit ändern kann und somit Auswirkungen auf den Link-Graphen hat. Google interpretiert diese Seiten dann als noindex, nofollow und wertet die Links zu Folgeseiten in diesem Fall nicht mehr. Dies führt im schlimmsten Fall auch dazu, dass die weiterführenden Seiten nicht mehr gecrawlt werden, sofern diese keine weiteren starken Signale von außen bekommen.

Wie lange dauert es, bis Google die Direktive für sich ändert? Leider gibt es dazu keinen eindeutigen Zeitraum und es heißt wie so oft; „Es kommt darauf an …“



Abbildung 1: Quelle: https://twitter.com/JohnMu/status/945952610992967680

Tipp

Ein Webmaster hat verschiedene Möglichkeiten, Suchmaschinen beim Crawling-Prozess zu unterstützen:

Um den Crawler zu steuern:

- URLs in der robots.txt verbieten
- Nofollow-Links

Um das Crawling zu optimieren, können Änderungen angegeben

werden:

- XML-Sitemap
  - lastmod
  - changefreq

- Robots-Direktive
  - unavailable_after

Wie die Indexierbarkeit eines HTML-Dokuments andere Dokumente isoliert

Um die Problematik besser nachvollziehen zu können, werden im Folgenden ein paar einfache Beispiele mit drei Dokumenten (A, B und C) untersucht. Der Fokus liegt auf der Situation von Dokument C unter Berücksichtigung des Indexierungsstatus von B. In allen Szenarien gibt es interne Links zwischen A ⇆ B und B ⇆ C. Dokument A gilt immer als gecrawlt und hat die Robots-Direktive „index, follow“ gesetzt. Dokument C ist immer „index, follow“ und mit Dokument B verknüpft.

Normales Indexierungsverhalten: In diesem Beispiel erhält **Dokument B** eine **„index, follow“-**Direktive. Im Ergebnis können alle Dokumente gecrawlt und indiziert werden.

Abbildung 2: Normales Indexierungsverhalten
Langzeit-Noindex-Szenario: In diesem Beispiel erhält **Dokument B** eine **„noindex, follow“-**Direktive. Im kurzfristigen Ergebnis werden alle Dokumente gecrawlt und indiziert. Wenn eine Suchmaschine sich nach einiger Zeit dazu entschließt, die Direktive von **Dokument B** als **„noindex, nofollow“** zu interpretieren, dann wird Dokument C isoliert wie in Abbildung 3 zu sehen.


Abbildung 3: Langzeit-Noindex-Szenario
Ähnlich könnte es sich in einem Langzeit-Rel-Canonical-Szenario verhalten. Auch wenn John Mueller dazu in Bezug auf Google keine explizite Aussage getroffen hat, entspricht die Logik nach Effizienz und Ressourcenschonung bei einer Rel-Canonical-Direktive dem gleichen Muster. Webmaster sollten sich gut überlegen, welche Inhalte sie aus dem Index von Suchmaschinen heraushalten wollen und ob die Seitenstruktur dafür geeignet ist. Eine häufige Fehlkonzeption sind umfangreiche Kategorie- oder Paginierungsstrukturen, wo der Webmaster zwar das einzelne Produkt, aber nicht die Übersicht indizieren möchte. In Abbildung 4 ist ein Beispiel zu sehen, wo 40 % der Seiten einer Domain von einem solchen Konstrukt betroffen sind.


Abbildung 4: Freizügiger Einsatz von „noindex, follow“ kann zur möglichen Isolation großer Teile der Seite führen

Bedeutung für Paginierung und andere komplexe Szenarien

Wenn die vorherigen Beispiele auf eine komplexere Seitenstruktur wie Paginierung, Kategorien oder ähnliche HTML-Sitemaps übertragen werden, zeigt sich deutlich das mögliche Ausmaß des Problems.

Beispiel Paginierung: Im Set-up haben wir eine Kategorie C mit einer „index, follow“-Direktive und den Seiten P1, P2 und P3 mit einer „noindex, follow“-Robots-Direktive. Alle diese Seiten verweisen auf eine Reihe von Item-Seiten mit einer „index“-Direktive. Wie in Abbildung 5 zu sehen, werden kurzfristig alle Seiten indexiert bis auf die Paginierungsseiten.


Abbildung 5: Kurzfristig werden alle Einzelseiten erreicht und

indexiert
Wenn eine Suchmaschine sich nach einiger Zeit für die Neuinterpretation der Paginierungsseiten auf „noindex, nofollow“ festlegt, dann ergibt sich ein schwerwiegendes Problem (Abbildung 6), da die Einzelseiten isoliert werden. Wenn diese Seiten keine externen Signale besitzen oder neue Seiten auf bereits isolierten Paginierungsseiten verlinkt werden, werden Seiten ggf. gar nicht mehr indiziert. Insbesondere wenn regelmäßig neue Produkte oder Artikel der Webseite hinzugefügt werden, ist eine Vermeidung von Langzeit-Noindex notwendig, um eine direkte Isolierung neuer Elemente zu verhindern.


Abbildung 6: Bei einem langfristigen Einsatz von „noindex, follow“ entscheiden sich Suchmaschinen ggf. für eine „noindex, nofollow“-Interpretation

Bedeutung für komplexe Strukturen wie HTML-Sitemaps

Bei großen Portalen mit vielen unterschiedlichen Inhalten wächst historisch oft eine vielfältige Mischarchitektur. Verdeutlicht werden soll dies an einem weiteren Beispiel in Abbildung 7 und der Kombination von noindex und canonical. Im Set-up befinden sich eine Reihe von Dokumenten A bis I. Die Dokumente D und F haben eine „noindex, follow“-Direktive. Dokument B verweist per rel-canonical auf Dokument A. Die Dokumente A ⇆ I sind miteinander verlinkt.


Abbildung 7: Komplexere Struktur mit einer Mischung aus index, noindex, follow und canonical
Unter Berücksichtigung der Langzeiteffekte ändert sich die Interpretation der Dokumente D und F, daraus ergibt sich die Isolierung von Dokument E. Wenn zusätzlich Dokument B neu interpretiert wird, resultiert daraus die Isolierung von C, D

und E, wie in Abbildung 8 zu sehen. Wären die Dokumente A ⇆ I nicht miteinander verknüpft, wären auch die Dokumente G, H und I von der Isolierung betroffen. Zu beachten ist zusätzlich, dass der kürzeste Pfad zu G von A → B → G auf den längeren Weg A → I → H → G wechselt.



Abbildung 8: Bei einem langfristigen Einsatz der Struktur aus Abb. 7 werden einzelne Seiten isoliert

Die Beispiele zeigen, dass selbst einzelne Aspekte in einer komplexen Webseitenstruktur sehr ins Detail gehen können. Dies ist gleichzeitig ein Indiz dafür, dass eine Strukturanalyse und die anschließende Strukturoptimierung für große Portale unabdingbar sind. Ungeeignete Strukturen bedeuten ein erhebliches Risiko für die Interpretation durch Suchmaschinen und mitunter stark negative Auswirkungen auf die wirtschaftliche Performance der Seite.

Zu einem Strukturkonzept gehören neben einer Lösung für die Noindex-Problematik auch für die Nutzerführung relevante Themen wie Seitentiefe, Autoritätenverteilung, Schaffung geeigneter Hubs und das Thema Paginierung. Allzu oft wird bei der Optimierung leider entweder nur auf den User oder auf den Bot geschaut, aber gerade das sinnvolle Zusammenspiel bringt den größten Mehrwert. Wenn erste Analysen in Teilbereichen bereits Probleme aufzeigen, dann ist eine ausführliche Strukturanalyse notwendig und möglicherweise eine Neukonzeption der verantwortlichen Struktur die beste Lösung.