

Pascal Horn

Indexierung in Suchmaschinen: robots.txt, Nofollow und Noindex im Praxistest

DER AUTOR



Pascal Horn ist leidenschaftlicher SEO-Blogger und Tester. Als SEO-Manager ist er vor allem für Datenerhebung und Datenauswertung zuständig.

Noindex, Nofollow und robots.txt – mit welchen Methoden hält man eine Seite zuverlässig aus dem Index von Suchmaschinen heraus? Und wie halten sich eigentlich die großen drei Suchmaschinen Google, Bing und Yahoo! an die üblichen Standards? Der SEO-Experte Pascal Horn hat für die Website-Boosting-Leser ausführliche Tests gefahren und den aktuellen Stand analysiert und dokumentiert.



robots.txt Test: Blogartikel	Artikel 1	Artikel 2	Artikel 3	Artikel 4	Artikel 5	Artikel 6	Artikel 7	Artikel 8
per robots.txt gesperrt	✓	✓	✓	✓	✓	✓	✓	✓
Title # Artikelüberschrift	✓	✓	✓	✓	✓	✓	✓	✓
Im Blog verlinkt?	✓	✓	✓	✓	✓	✓	✓	✓
Google Analytics Trackingcode?	✓	✓	✓	✓	✓	✓	✓	✓
In XML Sitemap übertragen?	✗	✗	✗	✗	✓	✓	✓	✓
Manuell versuchter Crawl Google WMT	✗	✗	✓	✗	✗	✓	✓	✓
Meta Robots: Noindex	✗	✓	✗	✓	✗	✓	✗	✓
Im Google Index gelandet?	✓	✓	✓	✓	✓	✓	✓	✓
Google: Noindex berücksichtigt?	✗	✗	✗	✗	✗	✗	✗	✗
Google: Title verwendet?	✗	✗	✗	✗	✗	✗	✗	✗
Google: Beschreibung = Seiteninhalt?	✗	✗	✗	✗	✗	✗	✗	✗
Im Bing Index gelandet?	✓	✓	✓	✓	✓	✗	✓	✓
Bing: Title verwendet?	✗	✗	✓	✗	✗	✗	✓	✗
Bing: Beschreibung = Seiteninhalt?	✗	✗	✗	✗	✗	✗	✓	✗
Im Yahoo! Index gelandet?	✓	✓	✓	✓	✓	✓	✓	✓
Yahoo: Title verwendet?	✗	✗	✓	✗	✗	✗	✓	✗
Yahoo: Beschreibung = Seiteninhalt?	✗	✗	✗	✗	✗	✗	✗	✗
Facebook: Title ausgelesen?	✓	✓	✓	✓	✓	✓	✓	✓
Facebook: Beschreibung ausgelesen?	✓	✓	✓	✓	✓	✓	✓	✓
Facebook: Bild verwendet?	✓	✓	✓	✓	✓	✓	✓	✓
Twitter: Title ausgelesen?	✗	✗	✗	✗	✗	✗	✗	✗
Twitter: Beschreibung ausgelesen?	✗	✗	✗	✗	✗	✗	✗	✗
Twitter: Bild verwendet?	✗	✗	✗	✗	✗	✗	✗	✗
Google+: Title ausgelesen?	✓	✓	✓	✓	✓	✓	✓	✓
Google+: Beschreibung ausgelesen?	✗	✗	✗	✗	✗	✗	✗	✗
Google+: Bild verwendet?	✓	✓	✓	✓	✓	✓	✓	✓

Abb. 1: Wordpress-Artikel, die per robots.txt blockiert wurden

robots.txt Test: Statische HTML-Seiten	Seite 1	Seite 2	Seite 3	Seite 4	Seite 5	Seite 6	Seite 7	Seite 8
per robots.txt gesperrt	✓	✓	✓	✓	✓	✓	✓	✓
Title # Artikelüberschrift	✓	✓	✓	✓	✓	✓	✓	✓
Im Blog verlinkt?	✗	✓	✗	✓	✗	✗	✗	✓
Google Analytics Trackingcode?	✗	✗	✗	✗	✗	✗	✗	✗
In XML Sitemap übertragen?	✗	✗	✗	✗	✓	✓	✓	✓
Manuell versuchter Crawl Google WMT	✗	✗	✓	✓	✗	✗	✓	✓
Meta Robots: Noindex	✗	✗	✗	✗	✗	✗	✗	✗
Im Google Index gelandet?	✗	✓	✗	✓	✗	✓	URL*	✓
Google: Noindex berücksichtigt?	✗	✗	✗	✗	✗	✗	✗	✗
Google: Title verwendet?	✗	✗	✗	✗	✗	✗	✗	✗
Google: Beschreibung = Seiteninhalt?	✗	✗	✗	✗	✗	✗	✗	✗
Im Bing Index gelandet?	✗	✗	✗	✗	✗	✗	✗	✗
Bing: Title verwendet?	✗	✗	✗	✗	✗	✗	✗	✗
Bing: Beschreibung = Seiteninhalt?	✗	✗	✗	✗	✗	✗	✗	✗
Im Yahoo! Index gelandet?	✗	✗	✗	✗	✗	✗	✗	✗
Yahoo: Title verwendet?	✗	✗	✗	✗	✗	✗	✗	✗
Yahoo: Beschreibung = Seiteninhalt?	✗	✗	✗	✗	✗	✗	✗	✗
Facebook: Title ausgelesen?	✓	✓	✓	✓	✓	✓	✓	✓
Facebook: Beschreibung ausgelesen?	✓	✓	✓	✓	✓	✓	✓	✓
Facebook: Bild verwendet?	✗	✗	✗	✗	✗	✗	✗	✗
Twitter: Title ausgelesen?	✗	✗	✗	✗	✗	✗	✗	✗
Twitter: Beschreibung ausgelesen?	✗	✗	✗	✗	✗	✗	✗	✗
Twitter: Bild verwendet?	✗	✗	✗	✗	✗	✗	✗	✗
Google+: Title ausgelesen?	✓	✓	✓	✓	✓	✓	✓	✓
Google+: Beschreibung ausgelesen?	✗	✗	✗	✗	✗	✗	✗	✗
Google+: Bild verwendet?	✓	✓	✓	✓	✓	✓	✓	✓

Abb. 2: Statische HTML-Seiten, die per robots.txt blockiert wurden

Es kann immer Seiten geben, die man Suchmaschinen lieber nicht zeigen oder deren Inhalte man gar in den Suchergebnissen sehen möchte. Dazu gibt es drei Möglichkeiten, die unterschiedliche Stärken und Schwächen aufweisen.

Für eine aussagekräftige Analyse wurde die Wirkung der robots.txt, Nofollow-Links und die Anweisung Noindex überprüft. Werden per robots.txt blockierte Inhalte gelesen und indexiert? Folgen Suchmaschinen Nofollow-Links tatsächlich nicht? Und wie schnell kann man mit Noindex Seiten aus den Suchergebnissen nehmen?

Zunächst eine kurze Zusammenfassung der Begriffe und welchen Nutzen sie haben sollen.

Die „Standarddefinitionen“ zu robots.txt, Nofollow, Noindex

Die drei Methoden zur Beeinflussung von Suchmaschinen werden von Suchmaschinenoptimierern und den Suchmaschinen in etwa so gedeutet:

Die **robots.txt** ist eine .txt-Datei, die man im Hauptverzeichnis seiner Seite ablegt (Beispielpfad für die Datei: <http://www.domain.de/robots.txt>). Hier kann man Seiten hinterlegen, die Suchmaschinen gar nicht erst betreten sollen. Die Anwendung findet vor allem bei Seiten und Dateien statt, die für die jeweilige Verwendung eines Content-Management-Systems kritisch sind, oder bei Seiten, auf denen man in eine Art privatisierte Oberfläche kommt. Für das Crawlmanagement bedeutet das, dass wir explizit Seiten vom Crawling ausschließen und so den Fokus für Suchmaschinen auf die nicht ausgeschlossenen Seiten lenken können. Ein Ausschluss einer Seite findet pro Zeile statt, in etwa so:

```
Disallow: <URL>
```

Nofollow bedeutet nicht automatisch das, was der Name vermuten lässt. Nofollow heißt nicht, dass Suchmaschinen den Links wirklich nicht folgen. Ein mit Nofollow ausgestatteter Link signalisiert

Suchmaschinen in erster Hinsicht, dass sie der verlinkten Seite keinen Wert beimessen soll, der dieser einen Vorteil im Ranking verschaffen kann. Man setzt Nofollow vor allem dann ein, wenn man einer verlinkten Seite selbst nicht vertraut oder keine direkte Kontrolle über alle ausgehenden Links einer bestimmten „Kategorie“ hat (wie etwa bei Blogkommentaren oder Links in Foren). Google folgt nach eigenen Aussagen Nofollow-Links wirklich nicht. Wird der Nofollow anstelle innerhalb eines Links in dem Meta Tag robots eingesetzt, werden alle Links dieser Seite zu Nofollow-Links, was aber in den seltensten Fällen empfohlen wird. Ein mit Nofollow ausgestatteter Beispiellink:

```
<a href="http://www.domain.de/" rel="nofollow">Linktext</a>
```

Der **Noindex** hat den Auftrag, Suchmaschinen mitzuteilen, dass diese Seite nicht in den Suchindex geraten oder aus diesem nachträglich wieder entfernt werden soll. Nicht mehr und auch nicht

weniger. Noindex beeinträchtigt weder das Crawlverhalten von Suchmaschinen noch andere Rankingsignale innerhalb der kompletten Website. Der Noindex wird in das Meta-Tag robots geschrieben, etwa:

```
<meta name="robots" content="noindex" />
```

Jede dieser drei Methoden hat ihre ganz individuellen Vor- und Nachteile.

Für die Crawlingsteuerung eignen sich die robots.txt-Datei und teilweise auch Nofollow-Links, während Noindex keinen Einfluss darauf hat. Im Gegenzug dazu ist nur der Noindex ein zuverlässiges Hilfsmittel, um Seiten tatsächlich aus den Suchergebnissen auszuschließen. In den einzelnen Tests und deren Auswertungen wird sich herauskristalisieren, warum das so ist.

Für alle drei Bereiche wurden seit März 2015 zeitversetzt Tests angelegt, um zu prüfen, wie vor allem die drei größten Suchmaschinen Google, Bing und Yahoo! auf gesperrte Inhalte, Noindex-Seiten und Nofollow-Links reagieren. Um die detaillierten Auswertungen und deren Testgrundlage lesen zu können, wird auf die entsprechenden Seiten verwiesen. Hier konzentrieren wir uns hauptsächlich auf deren Interpretationen und Anwendungsmöglichkeiten.

robots.txt-Test: Indexieren Suchmaschinen gesperrte Seiten?

Die einfache Frage lautet: Indexieren Suchmaschinen auch per robots.txt gesperrte Inhalte? Und wenn ja, wie?

Kurz gesagt:

- » Die robots.txt-Datei eignet sich NICHT für die Nicht-Aufnahme in Suchmaschinen.
- » Sie eignet sich aber durchaus für die Crawling-Steuerung.

Beim robots.txt-Test wurden 18 Testseiten online gestellt, die alle für Suchmaschinen-Bots gesperrt wurden. Laut oben genannter „Standarddefinition“ darf kein Bot diese Seiten betreten und crawl-

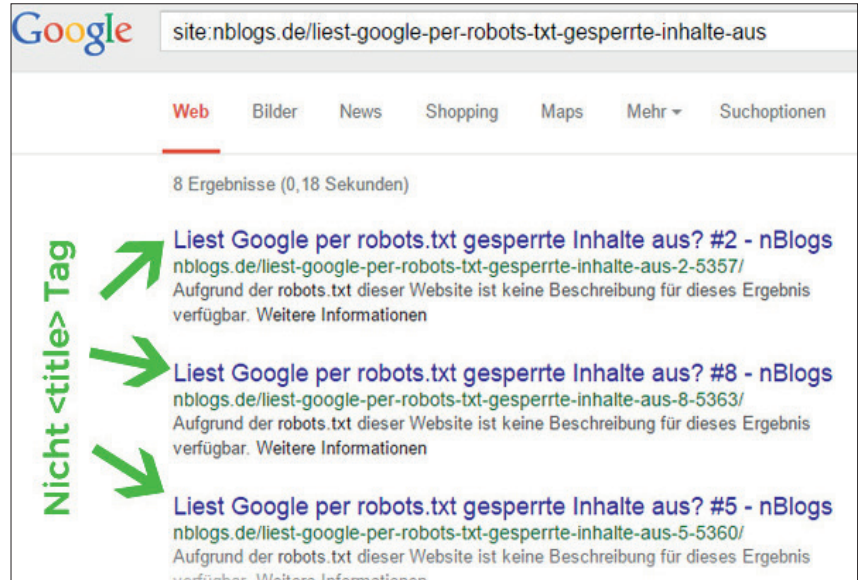


Abb. 3: Suchergebnisse zu den per robots.txt gesperrten Artikeln in WordPress: Google verwendet dabei nicht den Title der Artikel, sondern den Verlinkungstext (zum Beispiel von der Startseite)

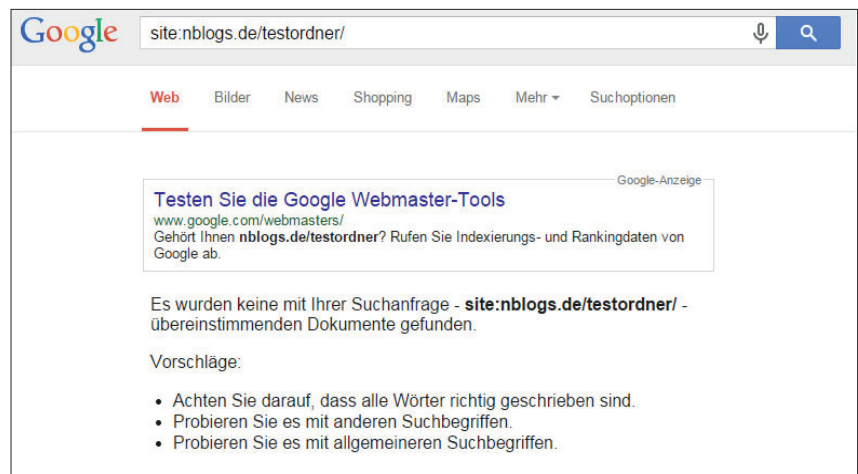


Abb. 4: Google nahm keine Testseite in den Index auf

len. Untersucht wurde hier das Verhalten der Suchmaschinen Google, Bing und Yahoo! sowie als Zusatz die drei großen Social-Media-Plattformen Facebook, Twitter und Google+. Der Test fand über sechs Tage statt.

Innerhalb der eigenen Website wurden Links zu diesen Seiten so präsent platziert, dass die Suchmaschinen sie auf jeden Fall erfassen mussten. Und es wurden die Suchmaschinen per Webmaster-Tools (bei Google heute Search Console) auf die Links aufmerksam gemacht.

Was dabei herauskam, ist äußerst interessant:

- » Google nahm alle 18 Testseiten in den Index auf, sie wurden aber wohl nicht untersucht.

- » Bing nahm 17 (von 18) Testseiten in den Index auf und verwendete zweimal den Title.
- » Yahoo! nahm alle 18 Testseiten in den Index auf und verwendete zweimal den Title, davon einmal den Inhalt.

Genau hier ist es aber Zeit, eine Lanze für Bing und Yahoo! zu brechen, denn bei dem Test wurde explizit auf Google Rücksicht genommen. In den (damaligen) Webmaster-Tools von Google konnte man genau feststellen, wann eine robots.txt-Datei aktualisiert wurde. Bei Bing und Yahoo! wurde darauf keine Rücksicht genommen. Bei Bing und Yahoo! ist das nicht so einfach möglich und entsprechend wurde darauf kein Fokus gelegt. Es

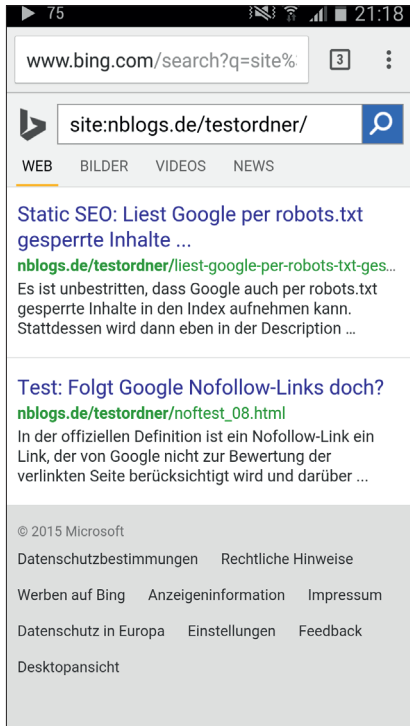


Abb. 5: Eine per Nofollow verlinkte Seite wurde indexiert (unteres Suchergebnis)

kann also sein, dass Bing und Yahoo! die Seiten unter einer veralteten robots.txt untersuchten und indexierten.

Dennoch stellt sich folgende Problematik: Eine Seite wurde vom Crawling ausgesperrt. Und Bing und Yahoo! waren nicht in der Lage, das Stoppschild zu sehen. Selbst wenn das Stoppschild ganz neu aufgestellt wurde, so wären sie bei einem Unfall auf der Straße verantwortlich zu machen. Kurz gesagt: Bing und Yahoo! reagierten nicht wie erwünscht und lasen trotzdem Inhalte.

Das bringt uns zu folgendem Fazit: Die robots.txt-Datei anzupassen, um Inhalte vom Crawling auszuschließen, eignet sich nicht für den Per-sofort-Effekt.

Suchmaschinen nehmen gesperrte Inhalte auf!

Trotzdem kann man einen Punkt für alle Suchmaschinen feststellen: Alle drei Suchmaschinen (Google, Bing und Yahoo!) nahmen nahezu alle Seiten in den Index auf. Das ist merkwürdig, wenn man berücksichtigt, dass sie die Seiten nicht mal untersuchen dürften.

Bei der Ausgabe einer dieser Seiten in einem Suchergebnis passiert Folgen-

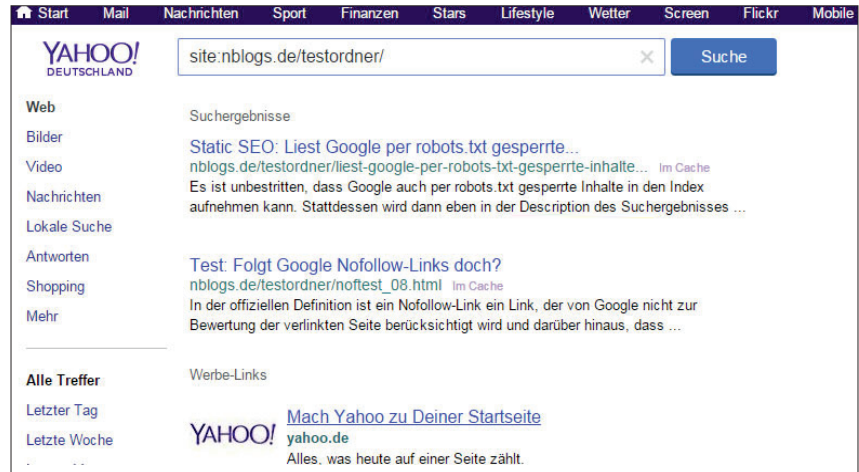


Abb. 6: Auch Yahoo! indexierte eine Testseite (unteres Suchergebnis)

des: Der blaue Verlinkungstext in Suchergebnissen (der in der Regel durch den Title präsentiert wird) wird durch den Verlinkungstext einer anderen Webseite dargestellt. Die schwarze Schrift (eigentlich die Meta-Description) wird durch den Hinweis ersetzt: „Aufgrund der robots.txt dieser Website ist keine Beschreibung für dieses Ergebnis verfügbar. Weitere Informationen.“

Und es kann noch „schlimmer“ kommen. Tatsächlich erreichen diese Seiten nicht nur den Index der Suchmaschinen als solche, sondern schaffen es sogar auch ins Ranking. Auch wenn in dem Test die Position 90 bei Google nie überboten wurde, so ist das ein starkes Anzeichen dafür, dass die gesperrten Seiten ohne Crawling es durch andere Rankingeinflüsse in die Top 100 von Google schaffen.

Die robots.txt gilt nicht für Social-Media-Plattformen

In diesem Test wurden auch explizit die „Sharability“ von Facebook, Twitter und Google+ untersucht, also würden diese Plattformen die Inhalte auslesen, obwohl sie per robots.txt blockiert wurden.

Die Ergebnisse sind relativ ernüchternd:

- » Facebook und Google+ lasen die Inhalte voll aus.
- » Twitter las sie nicht aus. Allerdings gab es keine expliziten Twitter-Meta-Angaben auf den Seiten, die Twitter hätte auslesen können. Es ist durchaus mög-

lich, das Twitter die Angaben abseits der robots.txt las. Das gab der Test leider nicht her.

Stellt man ein Resultat zusammen, ergibt sich also Folgendes: Mindestens die Social-Media-Plattformen Facebook und Google+ halten sich auch sechs Tage nach der Testphase nicht an die aktuelle robots.txt, die jederzeit ganz easy abrufbar wäre.

Das ist eigentlich als enttäuschend anzusehen. Denn sperrt man eine Seite aus, möchte man in der Regel nicht, dass diese auf anderen Seiten ausgelesen wird, was aber passiert.

Fazit: Die robots.txt-Datei und das Aussperren der Inhalte eignen sich nicht für den Ausschluss der Seiten in Suchmaschinen.

Die Testumgebung und Ergebnisse können auf <http://nblogs.de/robots-txt-test/> nachgelesen und diskutiert werden.

Nofollow-Test: Folgen Suchmaschinen Nofollow-Links?

Es hält sich hartnäckig der Eindruck, dass Suchmaschinen den Links tatsächlich nicht folgen. Und Google gibt in der Search-Console-Hilfe an, dass sie diesen Links tatsächlich nicht folgen.

Das gilt aber nicht verbindlich für andere Suchmaschinen. Bing und Yahoo! ist es genau wie Google freigestellt, ob sie Nofollow-Links folgen oder nicht. Ein Test sollte also klären, ob diese Suchmaschinen Nofollow-Links nicht doch folgen.

Der Testaufbau ist einfach: Es wurden 16 statische HTML-Testseiten angelegt, die alle per Nofollow verlinkt wurden, teilweise in Sidebar und Footer, teilweise in Contentbereichen und auch von externen Websites. Jede der verlinkenden Seiten wurde in den Google- und Bing-Webmaster-Tools angewiesen, diese zum Beispiel unter „Abruf wie durch Google“ zu untersuchen und, im Fall von Google, auch die direkten Links zu crawlen. Der Test lief über sechs Tage. Wenn eine der verlinkten Testseiten es in den Index der Suchmaschinen schaffte, wäre klar, dass diesen Links gefolgt wurde.

Und tatsächlich bestätigt der Test die Aussagen in der Google-Search-Console-Hilfe:

- » Google nahm keine Testseite in den Index auf.
- » Bing und Yahoo! nahmen jeweils zwei Testseiten auf.

Das bestätigt, dass sich Nofollow nicht dafür eignet, Seiten nicht in den Index von Suchmaschinen aufnehmen zu lassen. Wird eine Seite zudem von einer anderen Herkunft normal verlinkt oder wird die Seite per XML-Sitemap an Suchmaschinen übermittelt, werden diese ganz normal aufgenommen.

Interne Anwendungsmöglichkeiten für Nofollow-Links

Auch wenn der Einsatz von Nofollow-Links innerhalb der eigenen Website in den meisten Fällen nicht empfohlen wird, weil eben keine Linkpower den verlinkten Seiten übermittelt wird, so ergeben sich doch Anwendungsmöglichkeiten, um speziell das Crawlbudget von Google besser steuern zu können.

- » Schlagwörter- oder Kategorieseiten, die viele Unterseiten erzeugen mit zum Beispiel Seite 1 von 80; so könnten diese paginierenden Unterseiten per Nofollow verlinkt werden.

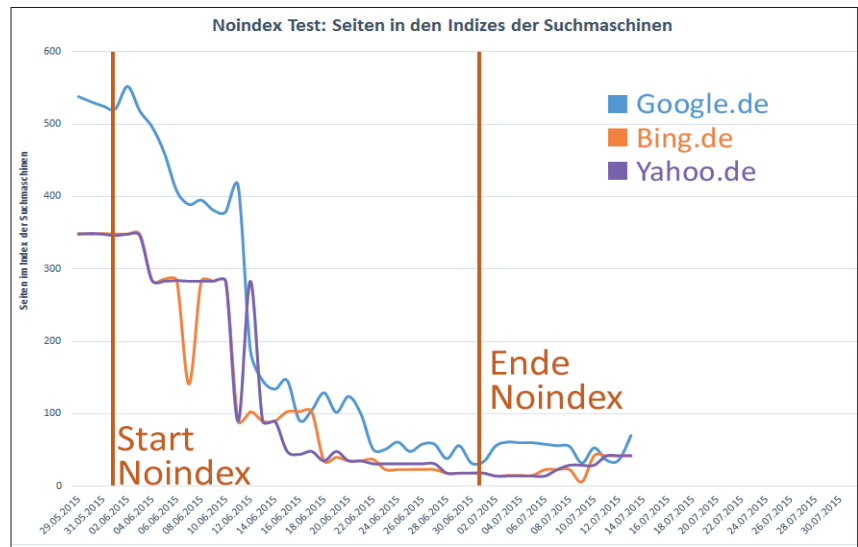


Abb. 7: Chronologische Anzahl der Suchergebnisse in Google, Bing und Yahoo!

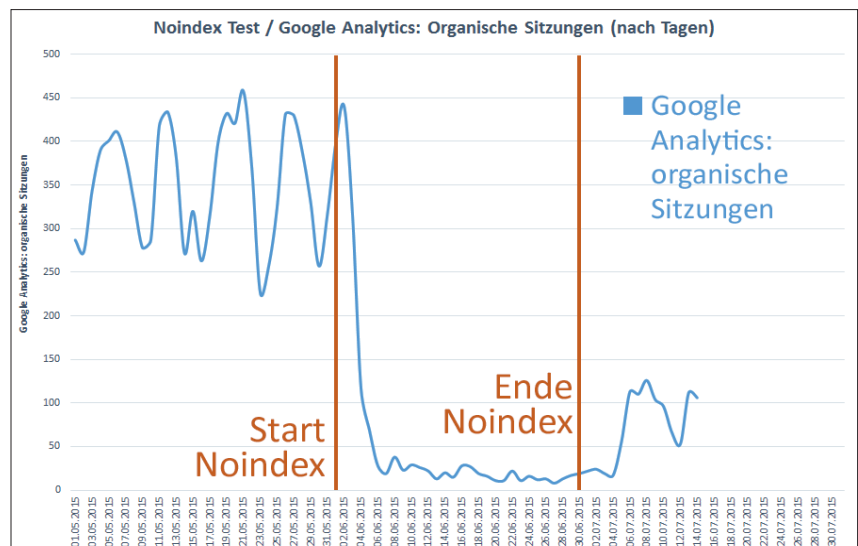


Abb. 8: Noindex-Test: So rapide bricht der organische Traffic ein

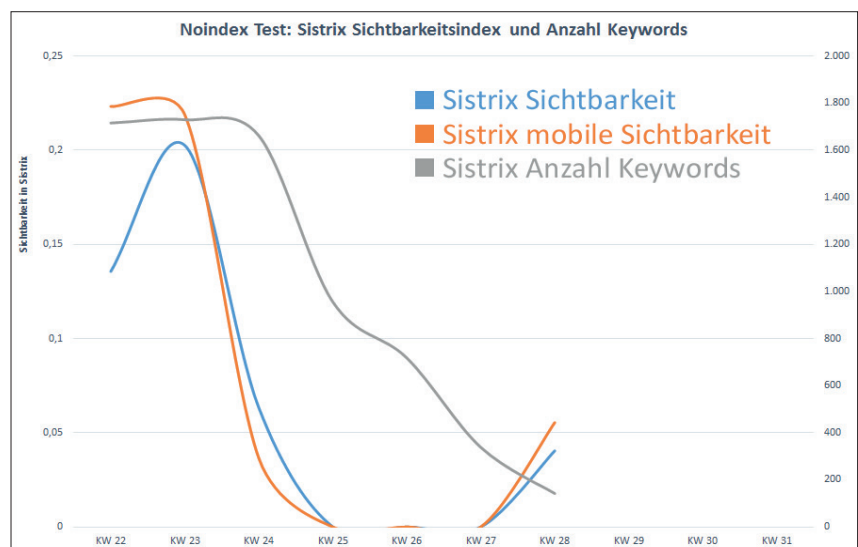


Abb. 9: Einbruch der Sistrix-Sichtbarkeit nach der Umstellung auf Noindex

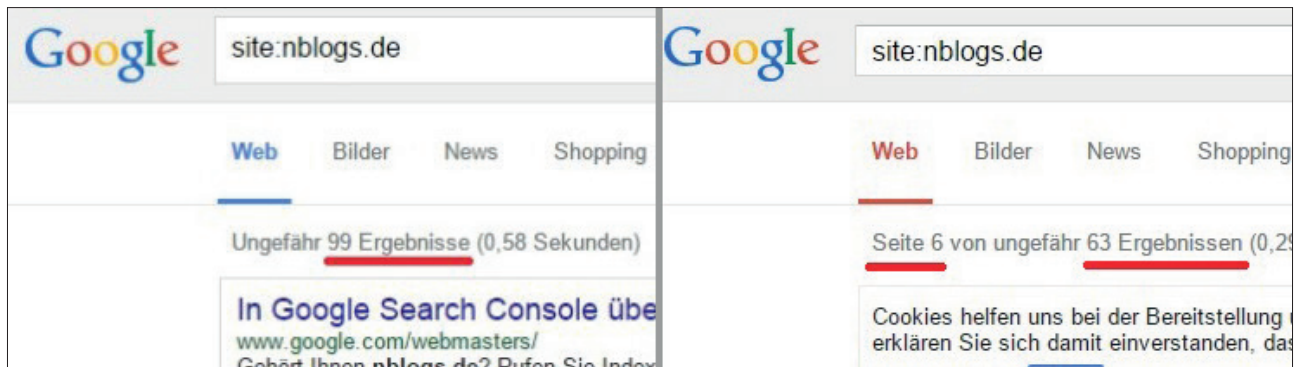


Abb. 7: Chronologische Anzahl der Suchergebnisse in Google, Bing und Yahoo!

Google muss sich dann nicht durch die Linktiefen fräsen.

» Werden viele Links auf einer Seite untergebracht, die jeweils nur auf eine einzige Unterseite verlinken, kann man diese teilweise auf Nofollow stellen. Nach einem Test von Malte Landwehr aus 2014 wird von Google ohnehin nur der erste

Text- oder Bildlink gewertet. Seinen Test findet man auf <http://einfach.st/nxpl>.

» Links zu Printversionen/PDFs eines Artikelinhalts kann man ebenfalls auf Nofollow setzen, wenn diese Versionen gegenüber der HTML-Seite keinen weiteren Vorteil bieten. Google muss diese

Inhalte nicht erfassen und indirekt schränken wir damit auch Duplicate Content ein. Was nicht heißt, dass PDF-Versionen von Inhalten nicht in den Google-Index geraten können.

Möchte man diese Seiten trotzdem im Index haben, kann man diese stattdessen auch per XML-Sitemap in der



INFORMATION SECURITY DAY

HOCHSCHULE FÜR ANGEWANDTE WISSENSCHAFTEN WÜRZBURG-SCHWEINFURT

FHWS  IDIS
Institut Design und
Informationssysteme

DER INFORMATIONSSICHERHEITSKONGRESS

HONEYPOTS – Cyberkriminelle in die Falle locken
 SOCIAL ENGINEERING – Die psychologischen Tricks der Hacker
 STEGANOGRAPHIE – Geheime Botschaften in Bildern verstecken
 ANALYSE – Sicherheit von Webanwendungen überprüfen
 AWARENESS – Mitarbeiter richtig weiterbilden



JETZT TICKETS SICHERN
ISD.FHWS.DE



12. / 13. NOVEMBER 2015



**SANDERHEINRICHSLEITENWEG 20
 97074 WÜRZBURG**

Google Search Console übermitteln. Wenn nicht, dann sollte man diese Seiten mit einem Noindex ausstatten, was uns zum nächsten Test führt:

Noindex-Test: Wie schnell deindexieren Suchmaschinen Inhalte per Noindex?

Es ist unbestritten, dass die Suchmaschinen Google, Bing und Yahoo! auf Noindex gesetzte Seiten aus dem Index nehmen oder sie gar nicht erst aufnehmen. Fragt sich nur, wie schnell Sie die Inhalte aus dem Index nehmen. Und wenn sie diese deindexiert haben, wie schnell kommen die Seiten wieder in den Index?

Im Test wurden am 1. Juni alle Seiten der Domain nblogs.de auf Noindex gestellt. Täglich wurde dann eine site:-Abfrage in den drei Suchmaschinen gestartet, um den Verlauf messbar zu machen. Dabei wurde nicht anderweitig interveniert, um die Ergebnisse möglichst nicht durch andere Faktoren zu beeinflussen. Also kein Abruf wie durch Google in der Search Console, gelöschte XML-Sitemap, keine neuen Artikel und keine freigegebenen Kommentare. Der Testzeitraum umfasste den kompletten Monat Juni.

Die einzelnen gesammelten Daten können auf nblogs.de/noindex/ nachgelesen werden. Konzentrieren wir uns hier auf die Interpretation der Ergebnisse.

Deindexierung erfolgt zeitverzögert

Es fällt auf, dass die Deindexierung in allen drei Suchmaschinen erst nach ein paar Tagen beginnt. Ein massiver Rückgang der indexierten Seiten fand erst am 12. Testtag statt. Das lässt zwei Rückschlüsse zu:

- » Entweder Suchmaschinen berücksichtigen irrtümliche Umstellungen nicht und warten ab, ob die Seite nach kurzer Zeit wieder normal auf Index steht.

- » Oder Suchmaschinen untersuchen gecrawlte Inhalte erst zeitverzögert und nehmen vom Noindex erst später Kenntnis.

Zuerst Artikel und Seiten, erst später Kategorien und Schlagwörter

Abseits der reinen Zahlen und Grafiken fällt im Test auf, dass vor allem normale WordPress-Artikel und WordPress-Seiten zuerst aus dem Index entfernt werden. Das ist logisch, denn Suchmaschinen sind auf diesen Inhalten deutlich häufiger unterwegs als auf Kategorie- und Schlagwörterseiten. Auch Anhangseiten, also eigene Unterseiten z. B. für hinterlegte Bilder, werden erst sehr spät entfernt.

In den Zahlen und Charts spiegelt sich das wider, wenn man die Anzahl der Seiten mit den organischen Besuchern vergleicht. Die am häufigsten besuchten Seiten und Artikel wurden schnell aus dem Index entfernt, was sich bald an einem Einbruch der organischen Sitzungen bemerkbar machte.

Trotzdem: Nach einem Monat befanden sich nach wie vor Inhalte in den Suchergebnissen aller drei Suchmaschinen Google, Bing und Yahoo!. Diese Ergebnisse umfassten eben nur noch paar wenige Schlagwörter-, Kategorie- und Anhangseiten.

„Ungefähr X Ergebnisse“ sind fehlerhaft

Speziell bei Google fiel auf, dass die Angaben zur ungefähren Anzahl an Suchergebnissen fehlerhaft waren. Als Beispiel wurden auf der ersten Suchergebnisseite ungefähr 99 Ergebnisse ausgeliefert. Klickt man sich bis zur letzten möglichen Seite durch, landet man nicht auf der Seite 10, sondern auf der Seite 6 mit ungefähr 63 Ergebnissen. Das bedeutet, dass die ungefähren Angaben auf der ersten Suchergebnisseite sehr ungenau sein können.

Was passiert nach der Umstellung auf „Index“?

Zum Zeitpunkt der Erstellung dieses Artikels ist der Test noch nicht abgeschlossen. Aber einen ersten Peek erlaubt der Test.

Bis zu einer Woche nach dem Entfernen des Noindex ergaben sich kaum Veränderungen in der Anzahl der indexierten Seiten, im Gegenteil. Während ein paar wenige Inhalte wieder im Index auftauchten, wurden noch mehr Inhalte weiter deindexiert. Damit ist es wahrscheinlicher, dass Google, Bing und Yahoo! den Noindex erst zeitverzögert berücksichtigen. Und auch bis zu zwei Wochen nach der Umstellung machten sich die Änderungen in der Gesamtanzahl an indexierten Seiten kaum bemerkbar.

Wenn man also nicht direkt in der Search Console bei Google bzw. den Bing-Webmaster-Tools manuell die Seiten neu crawlen lässt und keine aktualisierte XML-Sitemap hinterlegt, dauert es recht lange, bis die Inhalte wieder aufgenommen werden.

Fazit aller drei Tests

- » Die **robots.txt** ist geeignet, um Suchmaschinen von Inhalten fernzuhalten. Jedoch werden gesperrte, aber verlinkte Seiten trotzdem in den Index aufgenommen und haben das Potenzial zu ranken.
- » Google folgt **Nofollow-Links** nicht, bei Bing und Yahoo! ist es wahrscheinlich. Demnach eignet sich Nofollow nicht, um verlinkte Seiten aus dem Suchindex der einzelnen Suchmaschinen herauszuhalten.
- » **Noindex** berücksichtigen alle Suchmaschinen und damit markierte Seiten werden aus dem Index genommen. Jedoch erfolgt die Umsetzung bis zu mehreren Wochen zeitverzögert. Und Noindex beeinflusst in keiner Weise das Crawlverhalten der Suchmaschinen. ¶