Crawling Budget und 404 vs. 410

Zum Wochenstart eine kleine Auffrischung in Sachen Crawling Budget und 404 vs. 410. Nicht die spannendste Schlagzeile (im Gegensatz zu Panda 4.0 in der letzten Woche) aber eine häufige Quelle für SEO-Fehler, die eigentlich sehr einfach ausgemerzt werden können.

Um was geht es? Um das Crawling Budget und dessen Verbindung mit 404 bzw. 410 Fehlercodes.

Das Crawling Budget
Googles Crawler indexieren alle deine Unterseiten – aber nicht alle Unterseiten sofort. Jede Domain bekommt, je nach PageRank, ein Crawling-Budget zugewiesen. Dieses legt fest, wie viele Unterseiten gecrawlt werden. Seiten mit einem hohem PageRank bekommen mehr Budget zugewiesen. Wichtig: Das Crawling-Budget ist nicht gleich dem Index-Budget. Dieses legt fest, wie viele Seiten indexiert werden können. Logisch ist das Index-Budget dem Crawling-Budget nachgestellt: erst wird gecrawlt, dann indexiert.

Bildlich gesprochen: Im Internetland gibt es viele viele Häuser (= Domains). Google entsendet Inspekteure (= Crawler) in diese Häuser um die einzelnen Zimmer (= Unterseiten) anzuschauen und zu indexieren. Die Inspekteure gehen aber nicht wahllos von Haus zu Haus, sondern bevorzugen die bekannten, tollen Häuser (= Seiten mit einem hohem PageRank). Weniger schöne Häuser bekommen auch weniger Besuch von den Inspekteuren und diese haben dann auch weniger Zeit alle Zimmer anzuschauen (= Crawling Budget). Und es ist natürlich unschön, wenn ein Inspekteur einen Raum anschauen will, dort aber nichts zu finden ist (= 404-Fehler). Seine Zeit hätte er auch für Räume benutzen können, in denen etwas steht. Steht an der Tür aber so etwas wie „Hier ist nichts drin“ (=

SEO-Küche Internet Marketing GmbH & Co. KG
Fraunhoferstr. 6, 83059 Kolbermoor
Telefon 08031 / 2575-100
Telefax 08031 / 2575-101
E-Mail: info@seo-kueche.de

SEO-Küche Internet Marketing GmbH & Co. KG,
HRA 11167 AG Traunstein
pers. Haftende Gesellschafterin:
SEO-Küche Verwaltungs GmbH, Kolbermoor, HRB 22414 AG Traunstein
Geschäftsführer: Christian Brunnenmayer, Patrick Keller, Oliver Lindner
Ust-IdNr.: DE 286 985 708, Steuer Nr. 156/174/08500

HypoVereinsbank
IBAN DE45 700202700015260147
BIC HYVEDEMMXXX

410-Code), wird der Inspekteur direkt zur nächsten Türe gehen ohne seine Zeit zu verschwenden.

410-Code und weitere Möglichkeiten Budget zu sparen.
Ok, genug der Bildsprache. Hat eine Domain eine 404-Fehlerseite, ist das einfach ärgerlich. Bleibt dieser 404-Code bestehen, wird der Crawler immer wieder auf die Seite kommen um nachzuschauen ob sich etwas auf der Seite geändert hat. Weiß man aber, dass die Seite permanent leer bleiben wird, dann zeichnet man die Seite mit einem 410-Code aus. In Zukunft wird sich der Crawler also nicht mehr die Mühe machen, diese Seite anzuschauen – sondern seine Zeit für tatsächlich existierende Seiten aufbrauchen.

Um das Crawling-Budget effektiv einzusetzen, ist es ratsam, unwichtige Seiten wie Kontaktformulare, das Impressum (meines Erachtens) etc. mit Hilfe der robots.txt auszuschließen. Seiten die unbedingt gecrawlt werden sollen, verlinkt man intern stark und versucht für diese, Backlinks zu generieren.

Welche Seiten gecrawlt werden sollen und welche eher nicht, ist von der Art der Webseite abhängig: Im B2B-Bereich ist das Impressum häufig gut besucht, ebenso wie Kontaktformulare: Eine Sperrung mit robots.txt wäre hier nicht sinnvoll, da diese Seiten häufig Leads generieren. Ein Online-Shop hingegen, hat andere URLs als das Impressum, mit denen er Konversionen erzielt: hier würde es Sinn machen, das Impressum einfach im Footer zu verlinken, aber ansonsten zu sperren.

Eine ordentliche XML-Sitemap, in der die wichtigsten Seiten ausgezeichnet werden sowie eine möglichst flache Seitenarchitektur helfen dem Crawler, sich schneller zurecht zu finden. Bleibt die eigene Seite dauerhaft konstant, wird also nicht regelmäßig neuer Content eingestellt, Stichwort “Freshness Update”, reduziert Google den Besuch von Crawlern auf der Seite.

P.S.: Es gibt Hinweise darauf, dass der Google Browser Chrome tatsächlich ein Google Crawler-Bot ist. Ein Hinweis findet ihr hier, einen anderen hier. Ich werde mal schauen, was da dran ist und halte euch auf dem Laufenden…

SEO-Küche Internet Marketing GmbH & Co. KG
Fraunhoferstr. 6, 83059 Kolbermoor
Telefon 08031 / 2575-100
Telefax 08031 / 2575-101
E-Mail: info@seo-kueche.de

SEO-Küche Internet Marketing GmbH & Co. KG,
HRA 11167 AG Traunstein
pers. Haftende Gesellschafterin:
SEO-Küche Verwaltungs GmbH, Kolbermoor, HRB 22414 AG Traunstein
Geschäftsführer: Christian Brunnenmayer, Patrick Keller, Oliver Lindner
Ust-IdNr.: DE 286 985 708, Steuer Nr. 156/174/08500

HypoVereinsbank
IBAN DE45 700202700015260147
BIC HYVEDEMMXXX