

Mario Träger

Screaming Frog – Anwendungsmöglichkeiten in der Praxis

Screaming Frog ist ein beliebtes Tool, um die eigene Webseite im Bereich der Suchmaschinenoptimierung zu verbessern. Somit ist es bspw. sehr einfach, die Verwendung doppelter Title und Descriptions oder aber auch nicht gepflegte Links auf Fehlerseiten ausfindig zu machen. Neben diesen Funktionen steckt jedoch noch viel mehr in diesem Tool. Insbesondere neue Funktionen vergrößern die Anwendungsmöglichkeiten innerhalb der Suchmaschinenoptimierung um ein Vielfaches. In diesem Leitfaden erhalten Sie verschiedene Anwendungsbeispiele aus der Praxis, die zeigen sollen, was mit Screaming Frog möglich ist, sowie Denkanstöße, um eigene Verwendungsmöglichkeiten auszutesten.

Screaming Frog ist ein sogenanntes Crawling-Tool, welches sich als Programm bequem auf Windows und Mac installieren lässt. Durch seine einfache Bedienung zählt es zu den beliebtesten Tools im Bereich der Suchmaschinenoptimierung. Insbesondere die Neuerungen 2015 haben die Möglichkeiten mit Screaming Frog und die Beliebtheit des Programms deutlich gesteigert. Man kann komplette Content- und Technical-Audits der eigenen Webseite ohne viel Aufwand durchführen. In der Basisversion lassen sich 500 URLs kostenlos analysieren, für größere Webseiten sollte auf die kostengünstige Pro-Version zurückgegriffen werden.

Screaming Frog – Neuerungen 2015

Die Entwickler von Screaming Frog waren 2015 mehr als fleißig. Immer wieder gab es größere Updates mit neuen Funktionen, die es wirklich in sich haben. Folgende drei Funktionen sollen auf jeden Fall genannt und näher vorgestellt werden:

» **Custom Search** – Die Custom Search erlaubt es, innerhalb des Quellcodes nach beliebigen Inhalten zu suchen. In einem speziellen Reiter namens Filter kann man sich sowohl Seiten ausgeben lassen, welche einen Inhalt im Quellcode enthalten, als auch Seiten, welche

den Inhalt nicht enthalten. Somit lässt sich bspw. in Form eines Content-Audits einfach ermitteln, ob in einem Online-Shop auf den Kategorien ein Text vorhanden ist.

» **Custom Extraction** – Die Custom Extraction ermöglicht es bspw. mithilfe von CSS Path, Inhalte aus einer Webseite auslesen zu lassen. In einer zusätzlichen Tabelle kann man somit bspw. die Anzahl an Kommentaren zu sämtlichen Blogartikeln oder komplette Inhalte, welche sich stets im Quellcode an der gleichen Stelle befinden, extrahieren.

» **Google Search Console & Google Analytics API-Anbindung** – Die Möglichkeit, Trafficdaten aus der Google Search Console oder aus Google Analytics direkt in Screaming Frog einlaufen zu lassen, schafft neue Optionen – insbesondere in der Priorisierung von Inhalten. So kann bspw. nach Seiten sortiert werden, welche den meisten Traffic generieren, um diese gezielt weiter zu verbessern bzw. durch interne Verlinkung zu stärken oder umgekehrt Potenziale ausfindig zu machen.

Die drei näher beschriebenen Funktionen sollten nicht unterschätzt werden. Sie eröffnen neue Anwendungsmöglichkeiten und lassen sich beliebig miteinander kombinieren.

DER AUTOR



Mario Träger ist Partner der Agentur Webworks (www.webworks-agentur.de). Er unterstützt große und mittelständische Online-Shops im Bereich SEO und PageSpeed. Zuletzt war er bis 2014 Head of SEO der Witt-Gruppe (member of the otto group).

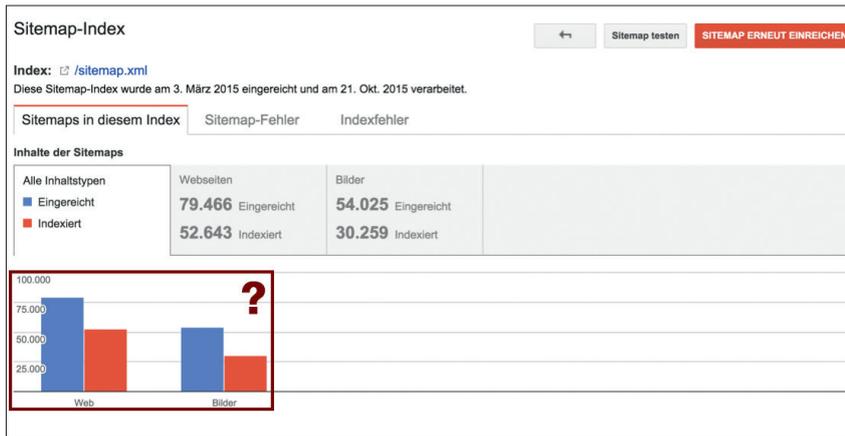


Abb. 1: Auswertung Sitemap-Index über die Google Search Console

Response Codes	URI	Page Titles	Meta Description	Meta Keywords	H1	H2	Images	Dir...
Blocked by Robots.txt								
No Response								
Success (2xx)								
Redirection (3xx)								
Client Error (4xx)								
Server Error (5xx)								

Abb. 2: Auswertung aller Statuscodes unter dem Reiter (Response Codes)

Anwendungsmöglichkeiten in der Praxis, die man kennen sollte

Die wichtigsten Funktionen innerhalb von Screaming Frog wurden bereits vorgestellt, nun geht es an die Umsetzung und die Vorstellung einiger Anwendungsmöglichkeiten „out of the box“. Die nachfolgend vorgestellten Anwendungen können direkt für die eigene Webseite übernommen werden und sollen zugleich Denkanstöße für eigene Anwendungsideen mit Screaming Frog liefern.

Indexierungskontrolle mit Screaming Frog

Die Indexkontrolle lässt sich als ein elementarer Baustein einer guten Suchmaschinenoptimierung bezeichnen. Nur wenn man die Kontrolle über die Indexierung seiner eigenen Webseite erlangt, hat man Einfluss, welche Seiten Google zur Bewertung der Webseite heranzieht, und hat überhaupt die Möglichkeit, mit neuen Landingpages zu ranken und Umsatz zu erzielen.

Die einfachste Möglichkeit, Kontrolle über die Indexierung zu erhalten,

ist der Sitemap-Index über die Google Search Console. Man sendet XML-Sitemaps mit allen wichtigen bzw. umsatzrelevanten Seiten an Google und bittet die Suchmaschine, diese Seiten in den Index aufzunehmen. Doch genau dort entstehen die ersten Probleme: Nicht selten ist es so, dass Google nur einen Bruchteil der geforderten Inhalte in den Index aufnimmt.

Dies ist vor allem bei Online-Shops ein großer Umsatzkiller. **Jede Seite eines Online-Shops, die sich im Index befindet, hat die Möglichkeit, Traffic zu generieren, der wiederum Umsatz schafft.** Sind nun 20 % der Seiten eines Online-Shops nicht indexiert, lässt sich schnell errechnen, welches Umsatzpotenzial an dieser Stelle verloren geht. Daher sollte das Hauptziel sein, alle umsatzrelevanten Seiten in den Index zu bekommen.

Nach dem Hochladen einer XML-Sitemap in die Google Search Console zeigt Google an, wie viele Unterseiten zur Indexierung eingereicht wurden und wie viele Seiten Google davon tatsächlich indexiert hat (siehe Abb. 1).

Anhand des oberen Beispiels lässt sich gut erkennen, dass von 79.466

Seiten nur 52.643 Seiten indexiert wurden. Diese Differenz von eingereichten zu nicht indexierten Seiten ist keine Seltenheit und bietet ein großes Optimierungspotenzial.

Warum ist diese Differenz überhaupt gegeben und was kann man dagegen tun? Es gibt zwei Gründe für den Unterschied von eingereichten zu indexierten Seiten. Der Hauptgrund liegt meist an der XML-Sitemap direkt:

Fehler innerhalb der XML-Sitemap mit Screaming Frog finden

Nur wenige Webseiten können von sich behaupten, dass die Sitemap zu 100 % sauber ist. Nicht selten werden mithilfe der Sitemap bei Google Seiten eingereicht, welche am Ende überhaupt nicht indexiert werden können. Die Gründe dafür können bspw. Folgende sein:

- » Unterseiten sind überhaupt nicht mehr erreichbar, da bspw. Inhalte gelöscht wurden oder Probleme mit dem Server bestehen – trotzdem sind diese Seiten noch häufig in XML-Sitemaps zu finden.
- » Unterseiten wurden auf Noindex gesetzt und dadurch von der Indexierung ausgeschlossen – trotzdem sind diese Seiten noch häufig in XML-Sitemaps zu finden.
- » Unterseiten enthalten ein Canonical-Tag, welches auf eine andere Seite zeigt. Dadurch wird im Normalfall nur die Seite indexiert, auf welche das Canonical-Tag zeigt – trotzdem sind diese Seiten noch häufig in XML-Sitemaps zu finden.

Um die oben genannten Punkte zu vermeiden, ist es erforderlich, die eigene XML-Sitemap zu kontrollieren und grundlegend zu säubern. Genau hier kann Screaming Frog ansetzen und große Unterstützung leisten. Mithilfe der Funktion Upload List können wir unsere XML-Sitemap direkt in Screa-

ming Frog importieren und sofort mit dem Crawling jeder einzelnen Unterseite beginnen. Je nach Größe der Sitemap erhält man in Kürze eine komplette Auswertung zu jeder Unterseite, welche in der Sitemap enthalten ist.

Welche Unterseiten sind für Google überhaupt erreichbar? Innerhalb des Reiters Response Codes erhält man zu jeder Seite den dazugehörigen HTTP-Statuscode (siehe Abb. 2).

In Summe erhält man eine Auflistung folgender Informationen:

- » Blocked by Robots.txt – Welche Seiten werden zum aktuellen Zeitpunkt per robots.txt gesperrt und können somit von Google nicht indexiert werden?
- » No Response – Welche Seiten geben überhaupt keinen Statuscode zurück und können somit von Google nicht indexiert werden?
- » Success (2xx) – Welche Seiten geben einen 200-Statuscode zurück und sind somit vorhanden und von Google indexierbar?
- » Redirection (3xx) – Welche Seiten leiten bspw. per 301-Weiterleitung auf eine andere Seite weiter und können somit von Google nicht indexiert werden?
- » Client Error (4xx) – Welche Seiten geben bspw. einen 404-Statuscode aus und sind somit nicht mehr erreichbar?
- » Server Error (5xx) – Welche Seiten geben bspw. aufgrund eines Serverfehlers einen 500-Statuscode aus und können somit von Google nicht indexiert werden?

Nach der Erhebung aller Statuscodes sollte man demnach Schritte einleiten, genau die Unterseiten, welche nicht mehr erreichbar sind, dauerhaft aus der Sitemap zu entfernen.

Welche Seiten dürfen durch die Noindex-Anweisung von Google überhaupt nicht indexiert werden? Innerhalb des Reiters Directives kann

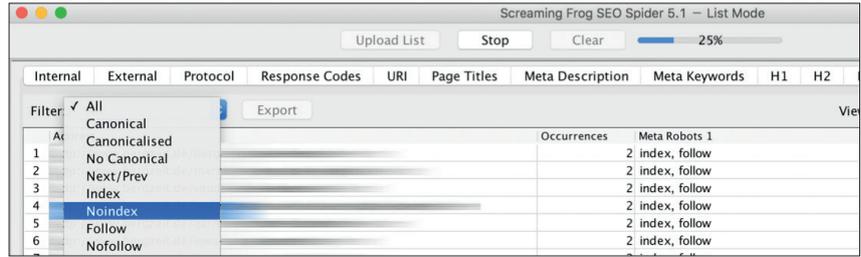


Abb. 3: Auswertung aller Noindex-Seiten unter dem Reiter (Directives)

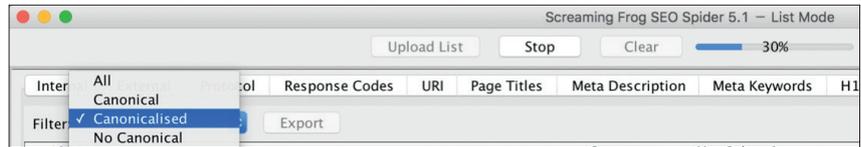


Abb. 4: Auswertung aller Seiten mit Canonical-Tag auf eine andere Seite unter dem Reiter (Directives)

man den Filter Noindex auswählen und erhält alle Seiten, welche auf Noindex gesetzt wurden (siehe Abb. 3).

Auch hier macht es keinen Sinn, Google über die Sitemap Inhalte zu zeigen, die nicht indexiert werden dürfen. Versuchen Sie, diese Unterseiten dauerhaft aus der Sitemap zu entfernen.

Welche Seiten verweisen per Canonical-Tag auf eine andere Seite?

– Innerhalb des Reiters Directives kann man den Filter Canonicalised auswählen und erhält alle Seiten, welche per Canonical-Tag auf eine andere Seite verweisen (siehe Abb. 4).

Immer wenn von einer Seite ein Canonical auf eine andere Seite und nicht auf sich selbst zeigt, sollte diese Seite dauerhaft aus der XML-Sitemap entfernt werden. Ansonsten würde man Google nur Seiten zur Indexierung anbieten, welche sich gar nicht indexieren lassen.

Anhand der letzten drei Abfragen über Screaming Frog sieht man, welches hohe Fehlerpotenzial bei einer XML-Sitemap besteht, aber auch, wie einfach die Fehler innerhalb einer Sitemap genau mit diesem Tool auffindbar sind.

Konnten sämtliche Fehler innerhalb einer Sitemap ausfindig gemacht und behoben werden, ist man auf einem sehr guten Weg, künftig nur noch relevante Inhalte an Google zu senden. Dadurch wird sich automatisch auch die in der Google Search Console ange-

zeigte Differenz zwischen eingereichten und indexierten Seiten dauerhaft verringern.

Ein weiterer Grund für das Nicht-Indexieren von Seiten innerhalb der eingereichten Sitemap liegt jedoch nicht an der Sitemap direkt, sondern an mangelnden qualitativen Inhalten selbst. So kann es sein, dass Google folgende Seiten nicht indexiert:

- » Seiten mit nur **wenig einzigartigem Inhalt** (Thin Content) können von Google aus dem Index gehalten werden – Solche Seiten sollten bei Bedarf mit hochwertigem Content angereichert oder auf Noindex gesetzt werden, um zu verhindern, dass negative Signale an Google gesendet werden.
- » Seiten, welche überhaupt nicht oder **nur sehr wenig intern verlinkt werden** – Hier kann trotz Sitemap die Indexierung sehr lange dauern oder überhaupt nicht stattfinden. Sind solche Seiten trotzdem wichtig, sollte die Seitenstruktur bzw. die interne Verlinkung verbessert werden.
- » Seiten mit einem hohen Anteil an **doppelten Inhalten** (Duplicate Content) können von Google aus dem Index gehalten werden – Solche Seiten sollten entweder mit hochwertigem Content angereichert oder vollständig vermieden bzw. von Google ferngehalten werden.

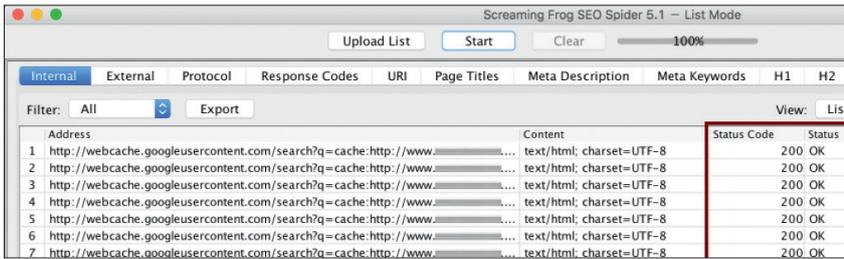


Abb. 5: Abfrage der Erreichbarkeit von Google-Webcache-Seiten

Size	Word Count	Level	Level	Inlinks	Outlinks
17950	345	1	4	1	1
17639	310	1	4	1	0
16853	257	1	4	1	1

Abb. 6: Gründe für nicht indexierte Seiten

All diese Gründe kann man mithilfe von Screaming Frog für die eigenen Sitemap-Inhalte analysieren. Zunächst sollten innerhalb von Screaming Frog alle bisher ermittelten Fehler wie bspw. Statuscodes direkt gelöscht werden (rechte Maustaste – Remove). Danach kann man diese Liste als Excel-Datei exportieren und erhält dadurch eine Liste aller Seiten, welche in Zukunft in der finalen Sitemap enthalten sein sollten.

Sitemap mit Google-Index abgleichen und Gründe für nicht indexierte Seiten finden

Interessant ist nun herauszufinden, welche Seiten innerhalb dieser Liste bei Google nicht indexiert sind, und die Gründe hierfür zu analysieren. Über die Search Console erhält man zwar den Anteil der Seiten, welche nicht indexiert wurden, man erhält jedoch nicht die Seiten selbst. Daher muss man sich zu helfen wissen. Bei kleineren Seiten kann man sich auch hier Screaming Frog zu Hilfe nehmen.

Will man herausfinden, ob eine Seite bei Google im Index enthalten ist, kann man neben einer site:-Abfrage auch die Existenz des Google Webcaches abfragen. Möchte man bspw. überprüfen, ob die Startseite der Website Boosting im Google Cache enthalten ist, genügt folgende Abfrage:

`http://webcache.googleusercontent.com/search?q=cache:https://www.websiteboosting.com/`

`com/search?q=cache:https://www.websiteboosting.com/`

Besteht zu dieser Abfrage eine URL, kann man auch davon ausgehen, dass genau diese Seite im Index enthalten ist. Wenn nicht, gelangt man auf eine Fehlerseite. Genau dieses Wissen können wir uns wiederum zunutze machen, indem wir mit Excel sämtliche Seiten innerhalb unserer Liste wie folgt verketten:

`http://webcache.googleusercontent.com/search?q=cache:[URL1]`

`http://webcache.googleusercontent.com/search?q=cache:[URL2]`

`http://webcache.googleusercontent.com/search?q=cache:[URL3]`

Daraus erhält man eine Liste, welche über die schon erwähnte Screaming-Frog-Import-Funktion in das Tool eingespielt werden kann (siehe Abb. 5).

Nach dem Crawling der einzelnen Seiten erhält man über den Statuscode Feedback, ob zu den jeweiligen Seiten eine Google-Webcache-Seite existiert (Statuscode 200) oder nicht (Statuscode 400), und kann diese Informationen direkt auf die Indexierung der Seiten bei Google übertragen.

Da Google automatisierten Abfragen nach einiger Zeit einen Riegel in Form einer Captcha-Abfrage vorschiebt, sollte man die Geschwindigkeit der maximal abgefragten URLs pro Sekunde auf den minimalsten Wert von 0,1 herabsetzen. Dies kann unter Configuration – Speed vorgenommen werden.

Bei größeren Seiten empfiehlt es sich jedoch, spezielle Tools für die Google-Indexierungsabfrage zu nutzen, welche die Verwendung unterschiedlicher Proxys ermöglichen. Hier kann bspw. auf das Tool URL Profiler zurückgegriffen werden -> <http://urlprofiler.com/>.

Am Ende sollte man somit eine Liste aller nicht indexierten Seiten erhalten und ganz genau wissen, welche Seiten aus der XML-Sitemap bisher noch nicht im Google-Index zu finden sind. Wenn man nun diese Liste mit einem Crawl der Site abgleicht, erhält man wertvolle Informationen über diese Seiten (siehe Abb. 6).

Folgende Gründe für eine Nicht-Indexierung kann man bspw. aus einem Screaming Frog Crawl herauslesen:

- » In der Spalte *Word Count* kann man ablesen, welche Seiten nur sehr wenig Content enthalten. Dabei werden alle Wörter auf der jeweiligen Seite zusammengezählt. Sieht man sich das dazugehörige Beispiel an, sind 125, aber auch 345 Wörter sehr wenig Gesamtcontent auf einer Seite.
- » In den Spalten *Level* und *Inlinks* kann man Seiten identifizieren, welche nur sehr selten oder überhaupt nicht intern verlinkt werden. Die Spalte *Level* beschreibt die Anzahl an Klicks, welche die jeweilige Unterseite von der Startseite aus entfernt ist. Die Spalte *Inlinks* beschreibt die Anzahl an internen Links, welche auf die jeweilige Unterseite zeigen. Eine Seite, welche insgesamt vier Klicks von der Startseite entfernt ist und zudem nur über einen Link erreichbar ist, hat für Google nur eine geringe Relevanz.
- » Auch Seiten, welche Duplicate Content verursachen, sind über Screaming Frog auffindbar. Da es hierbei jedoch mehrere Möglichkeiten gibt, wird dieser Punkt im nächsten Kapitel besprochen.

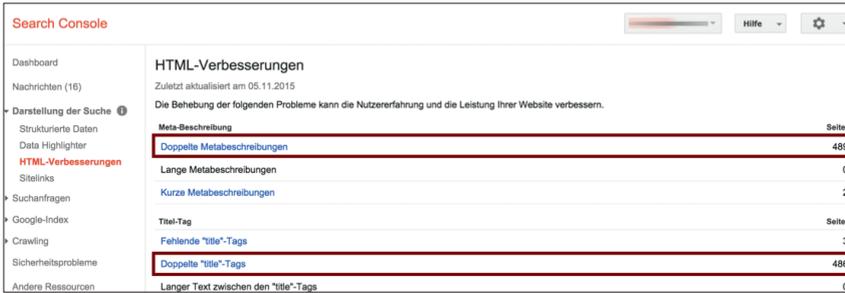


Abb. 7: Auffinden von Duplicate Content über die Search Console

Im ersten Schritt wurde gezeigt, wie man die eigene Sitemap zunächst auf Funktionalität überprüft und Seiten identifiziert, welche innerhalb der Sitemap nichts zu suchen haben. Im zweiten Schritt wurde aufgezeigt, wie man die Seiten ermittelt, welche zwar indexiert werden können, zum aktuellen Zeitpunkt aber nicht im Index enthalten sind. Im letzten Schritt wurde verdeutlicht, wie man die genauen Ursachen für die nicht indexierten Seiten herausfindet. Nur wenn man diese kennt, kann man daran arbeiten, diese Seiten entweder zu verbessern (und so indexierbar zu machen) oder aber dauerhaft aus der Sitemap zu entfernen.

Duplicate Content finden mit Screaming Frog

Bei Duplicate Content handelt es sich um gleiche bzw. sehr ähnliche Inhalte, die unter verschiedenen URLs zu finden sind. Grob kann man dabei zwischen externem und internem Duplicate Content unterscheiden.

- » **Externer Duplicate Content** – Bei Webseiten, die sich stark ähneln oder gar gleiche Inhalte auf ihren Seiten enthalten, wird die Bewertung der unterschiedlichen Webseiten für die Suchmaschine erschwert. Externer Duplicate Content entsteht bspw. durch die Verwendung von Herstellertexten oder durch die Weitergabe von Produktdaten an Affiliates und Preissuchmaschinen.
- » **Interner Duplicate Content** – Deutlich häufiger tritt der Fall auf, dass doppelte Inhalte auf der eigenen Domain zu finden sind. In diesem

Fall findet Google (und auch der User) identischen Inhalt auf Seiten mit unterschiedlichen URLs. Ein Problem dabei ist, dass Google selbst entscheidet, welche der beiden Seiten gezeigt und welche unterdrückt wird. Zusätzlich verteilt sich der Linkjuice über zwei Seiten, anstatt auf einer URL konzentriert zu werden. Interner Duplicate Content kann bspw. durch indexierbare Filterseiten oder durch Paginierungsseiten entstehen.

Um Duplicate Content zu umgehen, kann man das *Meta-Robots-Tag Noindex*, das *Canonical-Tag* oder aber auch die *robots.txt*-Datei einsetzen. Zugleich besteht die Möglichkeit, die betroffenen Seiten so zu ändern, dass Google diese als einzigartig wahrnimmt (beispielsweise durch einzigartigen Content).

Schwieriger gestaltet sich jedoch erst einmal das Auffinden von Duplicate Content auf der eigenen Seite. Ein Indiz für doppelte Inhalte sind doppelte Title und Descriptions. An diese Informationen kommt man bspw. über die Search Console im Menü HTML-Verbesserungen (siehe Abb. 7).

Das **Problem** bei der Search Console ist jedoch, dass die Daten zum einen nicht vollständig und zum anderen nicht aktuell sind. Um einen vollständigen Stand aller doppelten Meta-Tags zu erhalten, sollte man daher Screaming Frog bzw. ein Crawling-Tool nutzen.

Wichtig hierbei ist die Einstellung, dass Screaming Frog nur Seiten crawlt, welche auch von Google indexiert werden können. Seiten, welche bspw. das *Noindex-Tag* beinhalten, könnten



Abb. 8: Screaming-Frog-Einstellung, um ausschließlich indexierbare Inhalte zu crawlen

theoretisch Duplicate Content enthalten, werden jedoch von Google nicht indexiert bzw. nicht bewertet und sind daher auch kein Problem. Die Einstellung (siehe Abb. 8) ist unter *Configuration – Spider* im Reiter *Advanced* zu finden.

Nachdem diese Einstellung getätigt wurde, kann man Screaming Frog auf die Reise schicken und den indexierbaren Webseiten-Inhalt crawlen lassen. Somit erhält man umfassende Daten und kann genau ablesen, welche Title und Descriptions zum aktuellen Zeitpunkt doppelt vorkommen. Die Informationen über doppelte Title erhält man bspw. in dem Reiter *Page Titles* unter dem Filter *Duplicate* (siehe Abb. 9).

Eine weitere Möglichkeit, mithilfe von Screaming Frog Duplicate Content zu identifizieren, ist der sogenannte Hash-Wert. Screaming Frog ermittelt für jede Seite einen Hash-Wert aus dem vorhandenen Quellcode. Sind mehrere Seiten komplett identisch, so ist auch der *Hash-Wert* gleich. Für die Ermittlung doppelter Inhalte müssen in der Spalte *Hash* also nur noch die doppelten Werte ermittelt werden (siehe Abb. 10).

Durch die beiden oben genannten Methoden lassen sich Inhalte wie Filterseiten oder Paginierungsseiten innerhalb einer Webseite finden. Aufgrund immer wiederkehrender Muster sind hierbei meist schon einige Beispiele ausreichend, um diese Art von Duplicate Content aufzuspüren. Bei Online-Shops kann es zudem durch Mehrfachverwendung von Produkttexten zu zusätzlichem internem Duplicate Content kommen.

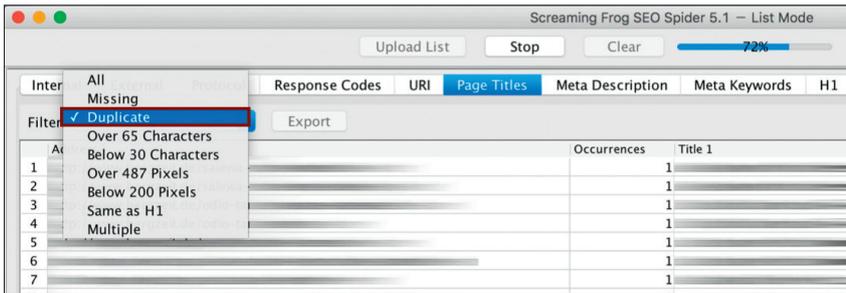


Abb. 9: Ausgeben von doppelten Title-Tags

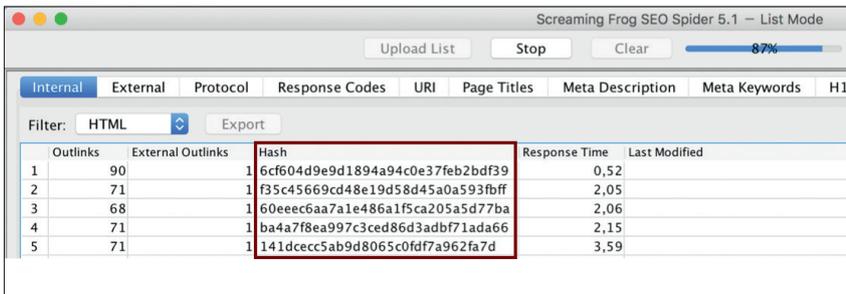


Abb. 10: Hash-Wert bei doppelten Inhalten

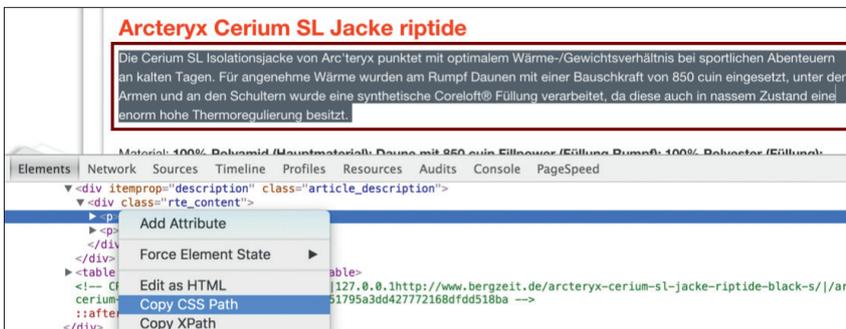


Abb. 11: CSS-Path-Code zu einem Produkttext erhalten

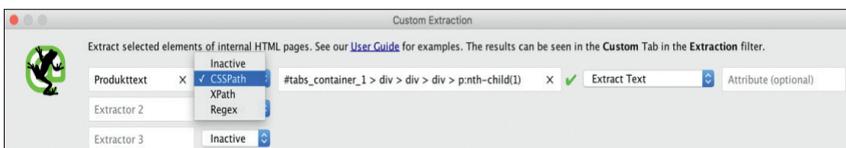


Abb. 12: CSS-Path-Code innerhalb der Custom Extraction einfügen

Mithilfe von Screaming Frog und der Funktion Custom Extraction lassen sich Inhalte aus der Webseite komplett extrahieren und im Anschluss auf Doppelungen untersuchen. Möchten man sich bspw. alle Produkttexte ausgeben lassen, um diese danach auf Dopplungen zu überprüfen, ist diese Funktion dafür wie geschaffen.

Um die Funktion Custom Extraction nutzen zu können, ist der CSS Path oder der XPath genau der Stelle notwendig, an welcher sich der Content befindet. Dabei sollte man wie folgt vorgehen: Zunächst einen beliebigen Produkttext

markieren und mit der rechten Maustaste anklicken. Danach kann man über Chrome die Funktion Copy CSS Path auswählen (siehe Abb. 11).

Schon hat man den CSS Path im Zwischenspeicher und kann diesen weiterverwenden. Unter Configuration – Custom – Extraction wählt man nun einen beliebigen Namen für die jeweilige Spalte aus. Zudem wählt man CSSPath aus und fügt den Code aus dem Zwischenspeicher in das dazugehörige Feld ein. In der darauffolgenden Auswahlfunktion wählt man mit Extract Text aus, dass man nur den Text ohne

HTML-Elemente extrahieren will (siehe Abb. 12).

Nach dem Crawl erhält man nun in einer neuen Spalte sämtliche Produkttexte des jeweiligen Online-Shops und kann diese bspw. mit Excel bequem auf doppelte Werte überprüfen.

Mithilfe von Screaming Frog lassen sich Quellen für Duplicate Content schnell und einfach ermitteln. Mit der gezielten Suche nach doppelten Title und Descriptions kann man Muster für Duplicate Content finden. Komplett identische Inhalte lassen sich über den Hash-Wert ermitteln. Produkttexte in Online-Shops lassen sich hingegen komplett extrahieren und danach auf doppelte Verwendung überprüfen.

Relaunch meistern mit Screaming Frog

Ein Webseiten-Relaunch ist aus SEO-Sicht stets mit wichtigen Überlegungen und Entscheidungen verbunden, insbesondere wenn sich die komplette URL-Struktur ändern soll. Regelmäßig kann man hierbei von Webseiten lesen, die große Sichtbarkeitsverluste hinnehmen mussten aufgrund des – aus SEO-Sicht – gescheiterten Relaunches.

Sowohl zur Vorbereitung als auch nach dem Relaunch können wir mithilfe von Screaming Frog die notwendigen Schritte vollziehen.

Vor dem Relaunch – Zunächst ist es wichtig, vor dem Relaunch einen sogenannten Redirect-Plan zu erstellen. Ändert sich bspw. die komplette URL-Struktur, müssen wir sofort bei Live-gang den Google Bot sowie den User mithilfe von 301 Redirects auf das neue Ziel weiterleiten. Ansonsten würden alte URLs ins Leere laufen und sämtliche Rankings über Google von heute auf morgen verloren gehen. Auch ärgerlich ist es, wenn Unterseiten hochwertig externe Links aufgebaut haben und bei einem Relaunch einfach komplett vergessen werden.

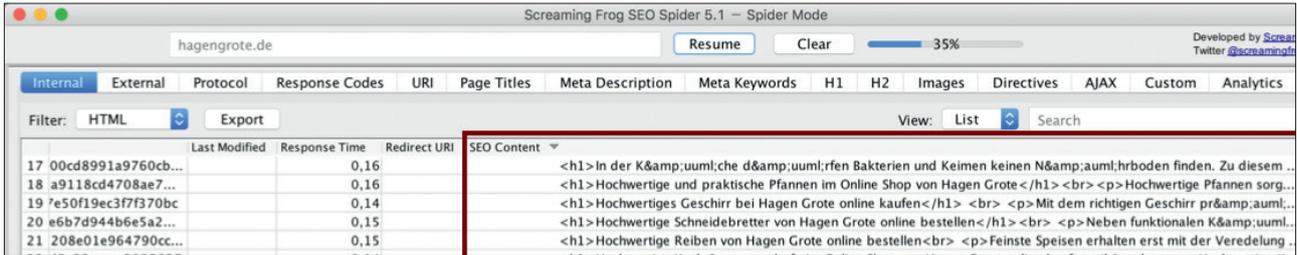


Abb. 13: SEO-Content-Back-up vor einem Relaunch

Bei einem Redirect-Plan hilft uns Screaming Frog mit einem Gesamtcrawl der Seite, alle relevanten und aktuell vorhandenen Unterseiten zu erfassen. Hier sollte wieder darauf geachtet werden, dass nur Seiten gecrawlt werden, welche auch von Google indiziert werden können (siehe Abb. 8). Nach einem Export dieser Daten haben wir damit bereits die Vorlage für unseren Redirect-Plan und können entscheiden, welche Seiten per 301-Redirect in Zukunft wohin weiterleiten sollen.

Sinnvoll vor einem Relaunch ist auch ein sogenanntes **Basic-Onpage-SEO-Back-up**. Darunter versteht man das Sichern aller relevanten Onpage-Faktoren wie bspw. Title, Descriptions, Content sowie Meta-Robots-Informationen und Canonical-Tags. Bis auf den Content haben wir bereits alle Daten innerhalb unseres Gesamtcrawls der Seite, welche wir uns beliebig exportieren können. Zudem können wir mithilfe der Funktion Custom Extraction bspw. die Kategorie-texte eines Online-Shops extrahieren und bei einem Verlust innerhalb des Relaunches sofort wieder einbauen (siehe Abb. 13).

Wie man Inhalte mithilfe der Custom Extraction extrahieren kann, wurde bereits im Bereich „Duplicate Content finden“ gezeigt (siehe Abb. 11-12). Um auch vorhandene HTML-Elemente wie H1-Tags oder auch interne Links zu erhalten, ist es notwendig, anstatt Extract Text die Auswahlfunktion Extract HTML Element zu tätigen.

Nach dem Relaunch – Direkt nach dem Livegang der neuen URL-Struktur sollte man sofort aktiv werden, um mögliche Fehlerquellen aufzuspüren und möglichst schnell zu bereinigen. Mithilfe

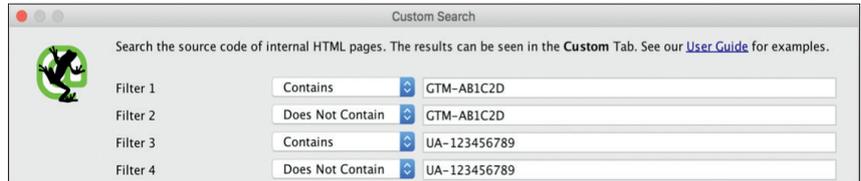


Abb. 14: Custom-Search-Einstellungen für Analytics und Tag-Manager-Code

des Redirect-Plans kann man einen Abgleich machen, ob alle Weiterleitungen korrekt eingerichtet wurden.

Durch einen Import in Screaming Frog wird das Crawling der alten URLs gestartet. Erhält man nun bei allen alten URLs unter dem Reiter *Response Codes* einen 301-Statuscode zurück, weiß man, dass alle Weiterleitungen korrekt gepflegt wurden. Erscheinen hingegen 404-Statuscodes, müssen hier die geplanten Weiterleitungen nachgepflegt werden, um keine Rankings zu verlieren.

Auch ein neuer Gesamtcrawl der Seite schadet nicht. Hier kann man insbesondere noch falsch gepflegte interne Verlinkungen aufspüren, welche bspw. auf eine 404-Fehlerseite verweisen.

Für das Webcontrolling ist es wichtig, dass nach dem Relaunch alle Daten korrekt auflaufen und keine „Tracking-Lücke“ entsteht. So eine Datenlücke kann bspw. auftreten, wenn bei einem Relaunch aus Versehen auf einzelnen Seiten oder Templates die Einbindung des Trackingcodes fehlt. Mithilfe der *Custom Search* kann man Screaming Frog auf die Reise schicken und den Auftrag mitgeben, dass das Tool alle Unterseiten auf den vorhandenen Trackingcode untersuchen soll. Somit erhält man eine detaillierte Auswertung, auf welchen Seiten denn der Trackingcode aktuell noch nicht eingebunden ist. Die Einstellung für die Custom Search befindet sich unter *Configuration – Custom – Search*.

Je nachdem, ob man bspw. auf Google Analytics oder dem Google-Tag-Manager-Code untersuchen will, sind verschiedene Codestücke möglich (siehe Abb. 13).

Der Filter 1 gibt bspw. alle Seiten aus, welche die Google-Tag-Manager-ID im Quellcode enthalten, der Filter 2 das genaue Gegenteil, nämlich alle Seiten, bei welchen diese ID nicht im Quellcode enthalten ist.

Screaming Frog unterstützt sowohl vor dem Relaunch als auch direkt nach dem Relaunch. Beide gezeigten Anwendungsmöglichkeiten sind aus SEO-Sicht essenziell für einen sauberen Relaunch und sollten niemals vernachlässigt werden.

Fazit

Die Anwendungsmöglichkeiten mit Screaming Frog sind vielfältig. Funktionen wie Custom Search und Custom Extraction haben ganz neue Möglichkeiten des Crawlings innerhalb dieses Tools mit sich gebracht. Egal ob für vollständige Content- und Technik-Audits, für die Verbesserung der XML-Sitemap, das Auffinden von Duplicate Content oder auch zur Unterstützung bei einem Relaunch – Screaming Frog ist ein zuverlässiges und kostengünstiges Tool, das darüber hinaus noch unzählige weitere Anwendungsmöglichkeiten mit sich bringt. Viel Spaß beim Ausprobieren und viel Erfolg mit dem schreienden Frosch!🐸