

Karsten Buth

So gelingen Website-Audits mit dem Screaming Frog SEO Spider

Wer im Bereich Suchmaschinenoptimierung arbeitet, stolpert früher oder später über den Screaming Frog, den Quasistandard für technische SEO-Prüfungen. Während SEO-Experten das Tool schon lange in ihr Herz geschlossen haben, fühlen sich Anfänger oft von den komplexen Möglichkeiten erschlagen. Weniger ist dann mehr: Mit der richtigen Konfiguration und dem Blick aufs Wesentliche können auch Einsteiger nützliche Erkenntnisse für ihre SEO-Arbeit gewinnen. Karsten Buth zeigt Ihnen Schritt für Schritt den Start mit dem beliebten Tool.

Der Screaming Frog SEO Spider gilt als Standard für technische SEO-Prüfungen. Nicht selten wird in Stellenausschreibungen für Online-Marketing- und SEO-Manager ein versierter Umgang mit diesem Tool vorausgesetzt.

Mehr als ein Bauchgefühl: warum Primärdaten so wertvoll sind

Wer bereits über SEO-Erfahrung verfügt, weiß, wie wichtig es ist, eine Website kontinuierlich zu optimieren, um ein besseres Ranking in den Suchmaschinenergebnissen zu erzielen. Dabei hilft es auch zu wissen, welche technischen Aspekte einer Website zu verändern sind. Idealerweise wird die Entscheidung, was zu ändern ist, nicht aus dem Bauch heraus, sondern anhand von Daten getroffen – im besten Fall anhand von Primärdaten.

„SEO-Entscheidungen sollten immer anhand von Daten getroffen werden.“

Als Primärdaten werden Daten bezeichnet, die direkt aus der Quelle stammen und dort direkt erhoben werden. Im Gegensatz dazu beziehen sich Sekundärdaten auf Informationen, die bereits auf anderem Weg gesammelt und veröffentlicht wurden.

Es gibt eine Reihe von Tools, die anbieten, die technischen SEO-Aspekte einer Website zu analysieren. Sie greifen in Echtzeit auf Websites

zu und durchleuchten die Inhalte. In den meisten Fällen stellen sie aber nicht die erhobenen Rohdaten zur Verfügung, sondern bieten Zusammenfassungen inklusive Interpretationen an: „Diese fünf Title-Tags behindern ein gutes Ranking“ oder „Vermeide die Verwendung mehrerer H1 auf einer Seite“.

Für ein SEO-Audit sind Primärdaten so wertvoll, weil sie einen vollständigen und ungefilterten Einblick in die technischen Aspekte einer Website ermöglichen.

Um Missverständnissen vorzubeugen: Der Screaming Frog SEO Spider nimmt keine Veränderungen oder gar Optimierungen an einer Website vor. Das Programm sammelt Daten, stellt sie übersichtlich dar und gibt damit eine Hilfestellung für die nachfolgenden Umsetzungen.

„Der Screaming Frog kann SEO-Expertise und gegebenenfalls Entwicklerkapazitäten nicht ersetzen.“

Der Screaming Frog SEO Spider ist damit ein wichtiges Werkzeug bei der SEO-Arbeit, kann aber SEO-Expertise und gegebenenfalls Entwicklerkapazitäten nicht ersetzen.

Seine Arbeitsweise macht das Programm so besonders: Ausgehend von einem Startpunkt (meist die Homepage einer Website) arbeitet sich der Spider über jeden von dort aus erreich-

DER AUTOR



Karsten Buth ist Head of SEO bei moebel.de und selbstständiger Digitalstratege. Unter <https://frogger.training/> hat er ein kostenfreies Training zum Screaming Frog veröffentlicht.



Abb. 1: Auch Google Trends verrät: In den letzten Jahren hat das Interesse am Screaming Frog SEO Spider stetig zugenommen.

baren Link durch den Internetauftritt und protokolliert dabei seine Beobachtungen. So entsteht ein vollständiges Bild, das schonungslos alle Fehler und Auffälligkeiten festhält und sich nicht nur auf eine kleine Stichprobe oder gar die Homepage selbst beschränkt.

Eine weitere Besonderheit des Screaming Frog ist seine Distributionsform. Es handelt sich nicht um eine webbasierte Anwendung, sondern um ein dediziert zu installierendes Programm. Unterstützt werden dabei die Betriebssysteme Windows (ab Windows 7), MacOS und Ubuntu Linux.

Weil das Programm direkt auf einem Computer installiert wird, hat es in der Regel auch die gleichen Zugriffsrechte wie andere Programme (inklusive Webbrowser) dieses PCs. Das bedeutet, dass der Screaming Frog nicht nur öffentlich erreichbare Websites crawlen, sondern auch für lokale Entwicklungsumgebungen oder VPN-gesicherte Webanwendungen verwendet werden kann.

Die Crawling-Methode (auch Spider-Methode genannt) hält mitunter auch Überraschungen bereit. Denn nur wenn eine Datei (nachfolgend allgemein als URL bezeichnet) über direkte oder weiterführende Links erreicht werden kann, kann sie das Programm untersuchen. Das bedeutet in der Praxis, dass eine im Auftritt nicht verlinkte Landingpage (beispielsweise für eine Google-Ads-Kampagne) nicht erreicht

und analysiert wird, wenn sie zusätzlich auch nicht in der Google-Sitemap aufgeführt ist.

Kostenlos vs. kostenpflichtig: Darf es ein bisschen mehr sein?

Wer überlegt, ob das Programm für die eigenen Anwendungszwecke geeignet ist, darf sich zum Ausprobieren ermutigt fühlen. Neben einer kostenpflichtigen Version mit allen Features

wird auch eine kostenlose Version des Screaming Frog angeboten, die bereits viele Funktionen beinhaltet (Abbildung 2).

Wer die Software für sein Betriebssystem herunterlädt und installiert, aber nachfolgend keinen gültigen Lizenzschlüssel einträgt, nutzt dann diese kostenlose Basisversion. Für kleinere Auftritte ist das mitunter ausreichend und ermöglicht in jedem Fall erste Analysen. So lässt sich ganz ohne

TIPP

Das SEO-Tool gibt es seit 13 Jahren und erfreut sich seither steigender Beliebtheit. Es handelt sich um eine downloadbare Software, die auf dem eigenen Windows, Mac oder Linux läuft. Der Haupteinsatzzweck ist das sogenannte Crawlen einer Webpräsenz oder eines Teils davon. Dabei spielt es technisch keine Rolle, ob man die eigene oder eine beliebige andere Site analysiert. Der Crawler geht dabei so vor, dass er die anzugebende URL (meist die Startseite) aufruft und analysiert. Jede Datei (HTML, PHP, CSS, JS etc.) von der Anfangsadresse und jedes dort verlinkte Bild werden vorgemerkt und anschließend ebenfalls aufgerufen. So crawlt (krabbelt) das Tool eine Site komplett durch und speichert dabei alles, was gefunden wurde – was durchaus einige Zeit in Anspruch nehmen kann. Anschließend setzt das Tool alles nach gewissen Regeln zusammen und ermittelt, ob und welche Probleme vorhanden sind. Beispielsweise werden tote Links auf 404-Seiten gefunden oder Seiten mit zu langen Seitentiteln. Lässt man den

Frog regelmäßig auf die eigene Site los, findet man damit ohne großen Aufwand kleinere, aber auch große und schwerwiegende Fehler. Das ist perfekt für eine verantwortungsvolle Kontrolle und Überwachung. Wer andere als die eigene Website crawlt, sollte in jedem Fall die Geschwindigkeit des Tools deutlich herabsetzen, damit ein fremder Webserver nicht durch den Crawlvorgang „gestresst“ oder gar überlastet wird. Das könnte juristisch als bösartiger Angriff interpretiert werden. Wessen Webserver einem solchen Crawl nicht standhält, hat allerdings meist noch ganz andere Probleme ... Die Lizenzkosten für die Vollversion über derzeit 239 Euro fallen jährlich an. Die umgerechnet circa 20 Euro Kosten pro Monat spielt man meist zehnfach oder hundertfach bei der dann folgenden Optimierung der Site wieder ein. Jeder nicht bekannte oder behobene schwere Fehler kann wertvolle Ranking-Plätze bei Suchmaschinen kosten oder gar ein Ranking ganz verhindern.

Risiko ausprobieren, ob man mit dem Screaming Frog zurechtkommt.

Die wichtigste Einschränkung der kostenlosen Basisversion ist das Crawl-Limit: Jeder Arbeitsdurchgang ist auf maximal 500 URLs begrenzt. Außerdem stehen nicht alle Funktionen der Software zur Verfügung.

Wer größere Auftritte vollständig crawlen lassen möchte, benötigt die kostenpflichtige Version. Hier entfällt nicht nur das Crawl-Limit, zusätzlich stehen auch alle weiteren Funktionen des Programms zur Verfügung (Abbildung 3).

Der derzeitige Preis für eine Jahreslizenz liegt bei 239 Euro. Dabei gilt jede Lizenz zwölf Monate ab Kauf und explizit für eine Person. Automatische Verlängerungen gibt es nicht. Läuft die Lizenz ab, muss sie händisch erneuert werden.

Der Anfang wirkt komplex: die Benutzeroberfläche

Zugegeben, der erste Aufruf der Benutzeroberfläche lässt einen vielleicht erst einmal ratlos zurück. Aber tatsächlich gelingt es relativ schnell, zu durchschauen, wofür die unterschiedlichen Fenster (Abbildung 4) verwendet werden. Oder anders ausgedrückt: Schlimmer wird es nicht!

Im Groben teilt sich die Oberfläche in vier Fensterbereiche: das URL-Übersicht-Fenster oben links, das URL-Details-Fenster direkt darunter, das Datenanalyse-Fenster oben rechts und das Info-Fenster unten rechts.

URL-Übersicht

In der URL-Übersicht werden während eines Crawls die Daten eingefüllt. Nach und nach werden hier in Echtzeit die URLs ergänzt, die bereits erfasst wurden. Das Fenster listet die URLs mit verschiedenen Parametern in Spalten auf: Adresse, Inhaltstyp, Status-Code, Indexierbarkeit und viele weitere.

Funktion	Basisversion
Broken Links, Errors & Redirects finden	✓
Seitentitel & Meta-Angaben analysieren	✓
Meta Robots & Direktiven überprüfen	✓
hreflang-Attribute prüfen	✓
Exakte Seitenkopien entdecken	✓
XML Sitemaps erstellen	✓
Visualisierung der Website-Struktur	✓

Abb. 2: Bereits mit der kostenlosen Basisversion lassen sich wichtige Erkenntnisse gewinnen.

Crawls terminieren & planen	✓
Crawls detailliert konfigurieren	✓
Crawls speichern und erneut hochladen	✓
JavaScript Rendering aktivieren	✓
Crawls vergleichen	✓
Near Duplicate Content finden	✓
robots.txt individualisieren	✓
AMP-Seiten crawlen und validieren	✓
Structured Data finden und überprüfen	✓
Grammatik und Rechtschreibung prüfen	✓
Quellcode durchsuchen	✓
Custom Extractions	✓
Google Analytics Integration	✓
Google Search Console Integration	✓
Google PageSpeed Insights Integration	✓
Linkmetriken Integration	✓
Formularbasierte Authentifizierung nutzen	✓
Gerendertes & Quell-HTML speichern und anzeigen	✓
Crawls an Google Looker Studio exportieren	✓
Kostenloser technischer Support	✓

Abb. 3: Mit diesem Funktionsumfang wird die Vollversion des Screaming Frog SEO Spider beworben.

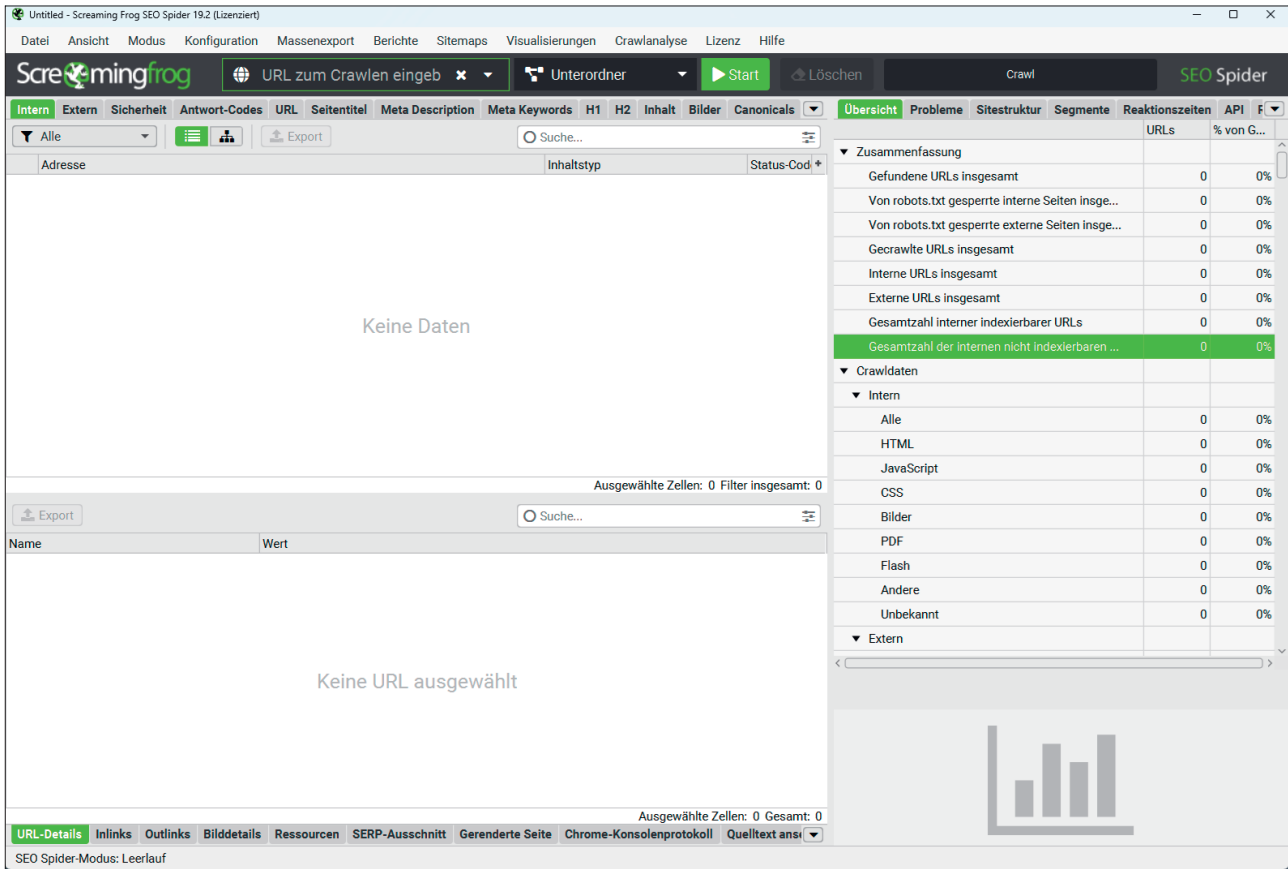


Abb. 4: So sieht die Benutzeroberfläche des Screaming Frog SEO Spider nach dem Programmstart aus.

URL-Details

Das URL-Details-Fenster korrespondiert direkt mit dem darüberliegenden URL-Übersicht-Fenster. Wird im oberen Fenster eine URL aus der Liste angeklickt, sind im unteren Fenster dann weitergehende Informationen zu dieser URL zu sehen. Die verschiedenen Parameter sind auch hier wieder über Tabs erreichbar, die unterhalb des Fensters positioniert wurden.

Datenanalyse

Das Datenanalyse-Fenster bietet Ergebnisse, die sich auf die Gesamtheit des Crawls beziehen. Anders als im URL-Details-Fenster werden hier die Werte aller URLs aufbereitet.

Infos

In Abhängigkeit vom aktiven Tab in der URL-Übersicht werden im Info-Fenster kleine grafische Veranschaulichungen präsentiert. Bei der Anzeige der internen oder externen URLs visualisiert beispielsweise ein Kreisdiagramm die Anteile der jeweiligen Dateitypen an der Gesamtanzahl der URLs.

gramm die Anteile der jeweiligen Dateitypen an der Gesamtanzahl der URLs.

Auf die Einstellung kommt es an: hilfreiche Tipps zur Konfiguration

Ein Blick in den Punkt „Konfiguration“ im Hauptmenü verrät schnell, dass sich der Screaming Frog SEO Spider sehr kleinteilig an die eigenen Bedürfnisse anpassen lässt. Im Rahmen dieser Einführung kann nicht auf alle Konfigurationspunkte eingegangen werden. Es gibt jedoch einige Einstellungen, die neue Nutzer unbedingt kennen sollten. Sie werden im Folgenden kurz vorgestellt.

Speichermodus wählen

Etwas versteckt und abseits der übrigen Einstellungen findet sich im Hauptmenü unterhalb von „Datei“ der Eintrag „Speichermodus“ (Abbildung 5).

Standardmäßig verwendet der SEO Spider den Arbeitsspeicher, um Crawl-Daten zu speichern. Der wichtigste

Vorteil in diesem RAM-Speichermodus: die Geschwindigkeit. Da Computer aber meist deutlich weniger Arbeitsspeicher als Festplattenspeicher haben, ist dieser Modus für große Websites (ab circa 500.000 URLs und mehr) nicht die richtige Wahl.

In solchen Fällen bietet sich die Datenbankspeicherung an. Hierbei werden die Daten direkt auf der Festplatte gesichert, was zwingend eine SSD-Festplatte voraussetzt, um die nötigen Schreibvorgänge während des Crawls verzögerungsfrei durchführen zu können.

„Datenbankspeicherung: Nur in diesem Modus können zwei Crawls miteinander verglichen werden.“

Die Datenbankspeicherung bietet einen weiteren Vorteil: Nur in diesem

Modus können zwei Crawls miteinander verglichen werden. Das ist beispielsweise hilfreich, um vor und nach erfolgten Änderungen an einem Auftritt die technischen SEO-Auswirkungen zu testen oder Staging- und Liveauftritte auf Unterschiede zu prüfen.

Crawl-Geschwindigkeit einstellen

Unterhalb des Hauptmenüpunkts „Konfiguration“ können im Abschnitt „Geschwindigkeit“ die Anzahl gleichzeitiger Abfragen (Threads) oder die pro Sekunde abzurufenden URLs festgelegt werden (Abbildung 6).

Der erste Impuls ist dabei natürlich, die Geschwindigkeit hochzusetzen, um den Crawling-Vorgang zu beschleunigen. Das kann aber dazu führen, dass schwächere Webserver durch die Anfragenflut überlastet werden oder zu viele Anfragen innerhalb kürzester Zeit zum automatischen Blockieren durch den Server führen.

Wer sich bei diesen Einstellungen unsicher ist, sollte vorab Rücksprache mit dem Webmaster oder Hosting-Provider halten, um Problemen vorzubeugen.

Wie soll mit der robots.txt umgegangen werden?

Standardmäßig respektiert der SEO Spider die Regeln einer vorhandenen robots.txt (Abbildung 7). Er ist dann ebenso wie der Googlebot nicht in der Lage, URLs zu crawlen, die in dieser Datei ausgeschlossen wurden.

In einigen Fällen ist es aber sinnvoll, von diesem Verhalten abzuweichen: „robots.txt ignorieren“. Wird diese Option gewählt, führt das dazu, dass der SEO Spider die robots.txt-Datei nicht einmal anguckt. Sämtliche Direktiven bleiben ohne Berücksichtigung.

Den User-Agent ändern

Mit der Einstellung zum User-Agent kann festgelegt werden, wie sich der SEO Spider dem Webserver gegenüber

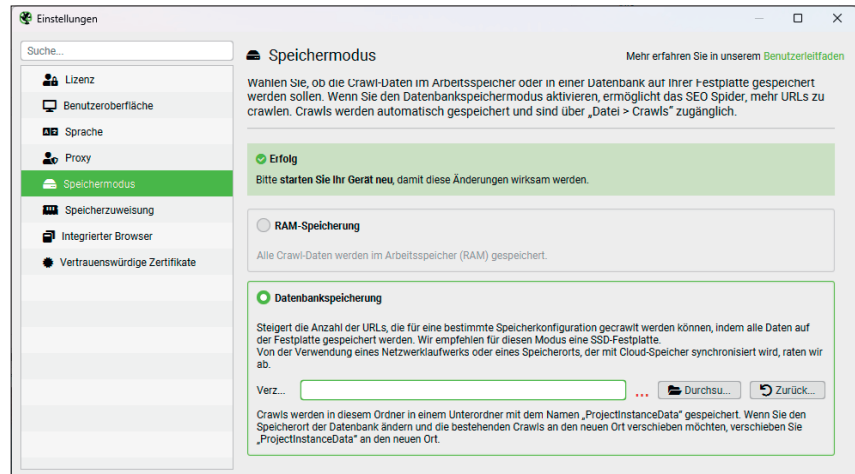


Abb. 5: Etwas versteckt, aber dennoch wichtig: die Einstellungen zum Speichermodus

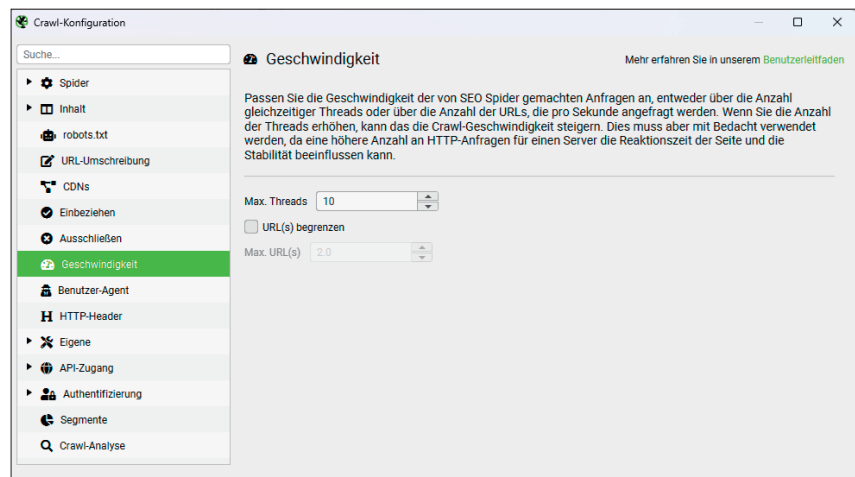


Abb. 6: Auf die richtige Balance kommt es an: Welches Tempo lässt der Webserver zu?

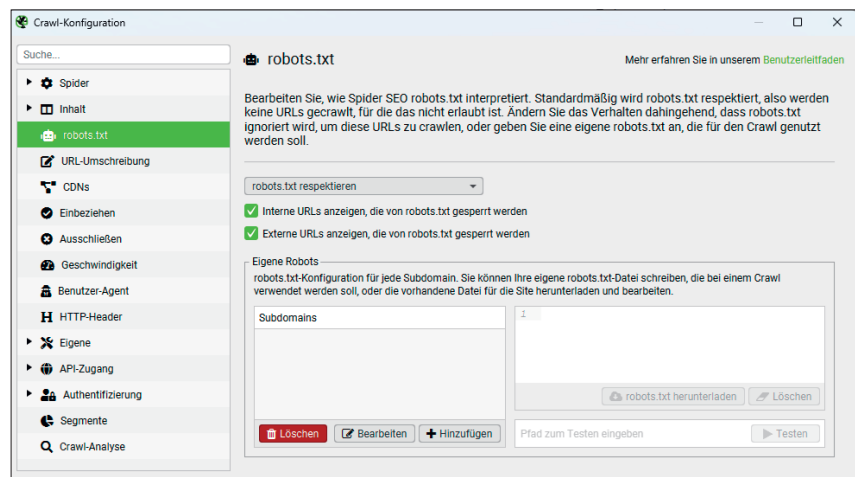


Abb. 7: Was wäre wenn? In der Crawl-Konfiguration lässt sich auch das Ignorieren der robots.txt einstellen.

identifiziert. Standardmäßig stellt der SEO Spider Anfragen mit seinem eigenen User-Agent: Screaming Frog SEO Spider (Abbildung 8).

Manche Webmaster sperren den Screaming Frog explizit von ihren

Seiten aus, um die Analyse durch Wettbewerber zu erschweren. In anderen Fällen ist es einfach interessant zu wissen, wie der Google- oder Bingbot einen Auftritt wahrnimmt.

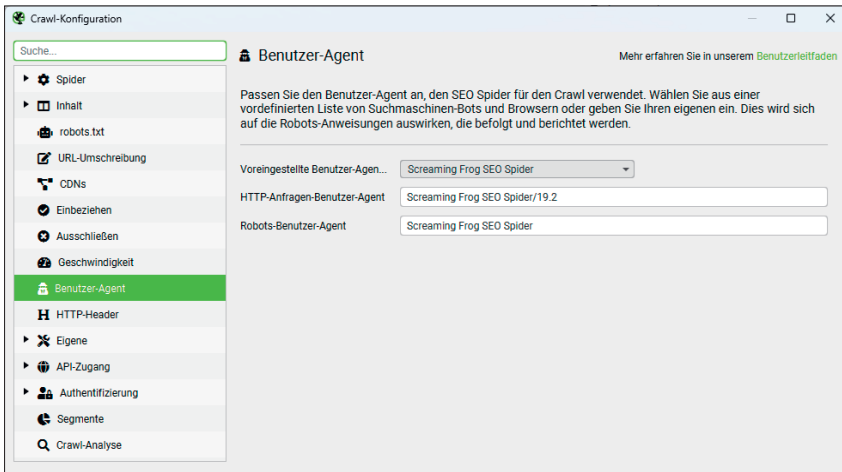


Abb. 8: Klassisch oder in geheimer Mission: Wie soll sich der Crawler den Webservern gegenüber vorstellen?

Fehler finden: von Einzelfällen bis zu strukturellen Schwächen

Wenn die ersten Anpassungen in der Konfiguration vorgenommen wurden, kann es losgehen. Der erste Website-Crawl wird gestartet.

In das Feld „URL zum Crawlen“ wird dazu die Homepage der zu analysierenden Website eingetragen und der Crawl anschließend mit Klick auf „Start“ begonnen. Die eingegebene URL und alle URLs, die über Hyperlinks im HTML-Code von Seiten derselben Subdomain gefunden werden, werden nun gecrawlt und deren Daten zusammengetragen.

Schon während der Laufzeit des Crawls kann damit begonnen werden, erfasste Daten zu analysieren. Im Fenster URL-Übersicht wächst die Liste beständig an. Nachfolgend finden sich erste Untersuchungsansätze.

Title- und Meta-Tags

Für die Verwendung von Titles lassen sich grundlegende Regeln aufstellen. Jede Seite einer Website verfügt idealerweise über genau einen Title, der innerhalb des HTML-Headers platziert

wird und der dabei gleichzeitig einzigartig ist, sich also innerhalb der Website nicht an anderer Stelle als Title wiederholt.

Von der Länge her gelten Titles als optimal, die zwischen 30 und 60 Zeichen lang sind und bei der Darstellung in den Suchergebnissen nicht weniger als 200 und nicht mehr als 560 Pixel breit sind.

Auf das Einhalten genau dieser Regeln kann nun geprüft werden. Nach dem Start eines Crawls wird dazu in der URL-Übersicht der Tab „Seitentitel“ ausgewählt. Hier stehen die folgenden, selbsterklärenden Filter zur Verfügung:

- » Fehlende: Auf der Seite wurde kein Title-Tag gefunden.
- » Duplikate: Auf der Seite wurde ein Title-Tag gefunden, das auch auf einer anderen Seite der Website verwendet wird.
- » Über 60 Zeichen
- » Unter 30 Zeichen
- » Über 561 Pixel
- » Unter 200 Pixel
- » Gleich wie H1: Das Title-Tag ist mit einer/der H1-Überschrift der Seite identisch.

» Mehrere: Es sind fälschlicherweise mehrere Title-Tags auf der Seite eingebunden.

Wird in der URL-Übersicht auf eine URL geklickt, kann nachfolgend im unteren URL-Details-Fenster der Tab „SERP-Ausschnitt“ gewählt werden. Hier wird dargestellt, wie möglicherweise ein Suchergebnis für diese URL in den organischen Google-Ergebnissen aussieht (Abbildung 9). Es gibt außerdem die Möglichkeit, den Seitentitel und die Meta-Beschreibung zu bearbeiten, um eine Vorschau für eine veränderte Darstellung des Suchergebnisses zu erhalten. Nicht vergessen: Diese Änderungen werden nicht gespeichert, also auch nicht für die Website übernommen.

Das Überprüfen von Meta-Beschreibungen erfolgt analog zum Überprüfen der Title-Tags. Es gelten die gleichen Regeln zur auftrittsweiten Einmaligkeit des Inhalts und zum einmaligen Platzieren innerhalb des HTML-Headers einer Seite. Lediglich die Vorgaben zur Länge weichen in Zeichen und Pixel ab. Standardmäßig prüft der SEO Spider, ob die Beschreibungstexte nicht weniger als 70 und nicht mehr als 155 Zeichen lang sind und dabei eine Breite zwischen 400 und 985 Pixeln einhalten.

In der URL-Übersicht gibt es dazu den Tab „Meta Description“. Die Liste lässt sich wie folgt filtern:

- » Fehlende: Auf der Seite wurde keine Beschreibung gefunden.
- » Duplikate: Auf der Seite wurde eine Beschreibung gefunden, die auch auf einer anderen Seite der Website verwendet wird.
- » Über 155 Zeichen
- » Unter 70 Zeichen

missinglink.de
https://missinglink.de

Digitalstrategie & SEO aus Hamburg - missing link - Karsten ...

Ich unterstütze Unternehmen, im Internet erfolgreicher zu werden: Durch Digitalstrategie, Website-Audits, SEO und Workshops.

Element	Zeichen			Pixel		
	Länge	Angezeigt	Abgeschnitten	Länge	Verfügbar	Verblei
Titel	64	59	5	575	561	-14
Beschreibung	124	124	0	777	985	208

Abb. 9: Die Vorschau eines möglichen Suchergebnisses im Tab „SERP-Ausschnitt“ im URL-Details-Fenster

- » Über 985 Pixel
- » Unter 400 Pixel
- » Mehrere: Es sind fälschlicherweise mehrere Beschreibungen auf der Seite eingebunden.

Broken Links

Wenn auf eine URL verlinkt wird, die auf dem Webserver nicht existiert, wird bei korrekter Konfiguration eine Fehlermeldung in Form eines Status-codes zurückgeschickt. Der klassische Fehlerfall ist „404 – not found“.

Wenn der Crawl einer Website gestartet wurde, kann bereits in der URL-Übersicht mitverfolgt werden, ob der SEO Spider fehlerhaft verlinkte URLs findet. Dazu wird der Tab „Antwort-Codes“ ausgewählt und die Ansicht dort auf Client-Fehler (4xx) gefiltert.

Zu beachten ist dabei noch eine weitere Filterbox, mit der bestimmt wird, ob nur interne, nur externe oder beide Arten von URLs angezeigt werden. Auf diese Art und Weise können nicht nur fehlerhaft verlinkte URLs auf der eigenen Website, sondern auch Verlinkungsfehler zu externen Seiten und Dokumenten erkannt werden.

In der Liste erscheinen also die URLs, die es auf dem Webserver nicht gibt. Die spannende Frage für die Problemlösung ist nun aber, wo sich diese fehlerhaften Links darauf befinden. Wenn dazu einer der Einträge in der URL-Übersicht angeklickt wird, kann im URL-Details-Fenster unten der Tab „Inlinks“ gewählt werden. Hier bekommt man entsprechend alle Seiten gezeigt, auf denen sich die fehlerhafte Verlinkung findet (Abbildung 10).

Bei größeren Websites kann die Liste nicht existierender URLs und auch

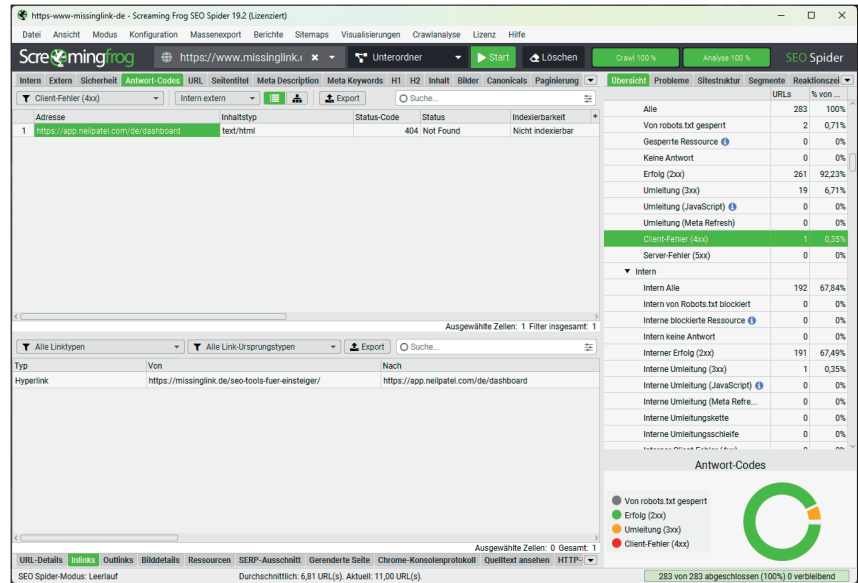


Abb. 10: Eine externe URL liefert eine Fehlermeldung zurück. Der Tab „Inlinks“ im URL-Details-Fenster verrät, wo die Verlinkung auf diese URL gefunden wurde.

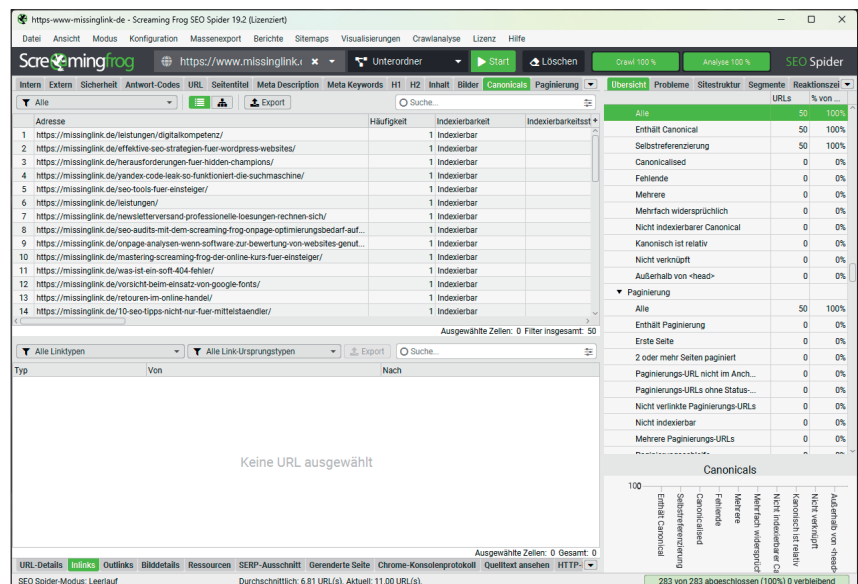


Abb. 11: Alle URLs mit Canonical-Tags finden sich in dieser Liste in der URL-Übersicht. Rechts im Datenanalyse-Fenster gibt es eine praktische Zusammenfassung.

die Anzahl der falsch gesetzten Links lang sein. Die vollständige Auflistung lässt sich daher auch bequem als CSV-, Gsheet- oder Excel-Datei exportieren. Der Bulk Report (etwas unglücklich als „Massenreport“ übersetzt) für alle internen und externen Clientfehler findet sich im Hauptmenü unter „Massenexport > Antwort-Codes > intern und extern > Clientfehler (4xx)-Inlinks“.

Canonical-Tags

Das Canonical-Tag wird verwendet, um Suchmaschinen mitzuteilen, welche

Version einer Seite als die Hauptversion betrachtet werden soll, wenn es mehrere Versionen dieser Seite unter verschiedenen URLs gibt. Das ist hilfreich, um Problemen mit doppelten Inhalten (Duplicate Content) vorzubeugen. Den Suchmaschinen wird so mitgeteilt, welche Version einer Seite die bevorzugte Version (das Original) ist, die in den Suchergebnissen erscheinen soll.

Auch hier hilft die URL-Übersicht in Echtzeit während des Crawls weiter. Im Tab „Canonicals“ wächst die Liste mit den URLs an, bei denen ein Canonical-

Fehler	Beschreibung
Fehlende hreflang-Tags	Ohne das hreflang-Tag kann die Suchmaschine die korrekte Version der Seite für die Zielgruppe nicht identifizieren.
Falsche Syntax	Falsch geschriebene oder inkorrekte hreflang-Tag-Syntax kann dazu führen, dass die Suchmaschine die Indizierung der Seite ignoriert oder fehlerhaft durchführt.
Inkonsistente Verwendung	Ein inkonsistentes hreflang-Setup auf verschiedenen Seiten der Website kann zu Verwirrung führen und die Suchmaschine daran hindern, die korrekte Version der Seite für die Zielgruppe zu identifizieren.
Verwendung von veralteten oder falschen Tags	Veraltete oder falsche hreflang-Tags können ebenfalls dazu führen, dass die Suchmaschine die Indizierung der Seite ignoriert oder fehlerhaft durchführt.
Falsche Zuordnung von Sprache oder Land	Wenn die Sprache oder das Land der Seite nicht korrekt identifiziert ist oder falsch zugeordnet wurde, kann dies zu Fehlern bei der Indizierung führen.

Abb. 12: Typische Implementierungsfehler bei der Verwendung von hreflang

Tag gefunden wurde. Dort wird in der Spalte „Canonical-Linkelement 1“ die URL aufgeführt, die für jede Seite als Originalversion über das Canonical-Tag benannt wird.

Der Tab „Canonicals“ verfügt über sechs Filter, die allesamt hilfreich sind, um typische Probleme beim Einsatz von Canonical-Tags zu identifizieren. Rechts im Datenanalyse-Fenster findet sich außerdem eine Zusammenfassung der für jeden dieser Filter enthaltenen Daten (Abbildung 11):

- » Enthält Canonical
- » Selbstreferenzierung
- » Canonicalised
- » Fehlende
- » Mehrere
- » Mehrfach widersprüchlich
- » Nicht indexierbarer Canonical
- » Kanonisch ist relativ

hreflang-Set-up

Ein hreflang-Set-up wird verwendet, um sicherzustellen, dass Google eine Website in mehreren Sprach- oder Länderversionen korrekt indiziert. Es ist besonders wichtig für Websites, die ähnliche oder gar gleiche Inhalte unter verschiedenen URLs in verschiedenen Regionen und Sprachen anbieten. Das Ziel: die richtige Version einer Seite an die richtige Zielgruppe ausliefern.

Im Grunde ist die hreflang-Logik ein recht einfaches Konzept. Dennoch zeigen sich in der Praxis oft Fehler bei der Implementierung, gerade bei umfangreichen Websites wie Online-Shops mit

vielen Kategorien und Produktdetailseiten.

Die Ergebnisse der hreflang-Überprüfung sind im Tab „Hreflang“ der URL-Übersicht zu finden. Im Tab werden alle bereits untersuchten URLs aufgelistet und (falls vorhanden) die hreflang-Angaben angezeigt. Die Spalte „Häufigkeit“ gibt Aufschluss darüber, wie viele alternative Versionen einer Seite durch hreflang-Angaben für jede URL aufgeführt werden. Liegt der Wert hier bei 0, wurden keine hreflang-Verweise gefunden.

Ein kurzer Ausblick: Auch das kann der Screaming Frog

Zugegeben, die bis hierhin vorgestellten Analyseoptionen bilden nur einen kleinen Teil der Möglichkeiten des Screaming Frog ab. Es soll deswegen nicht unerwähnt bleiben, dass noch sehr viel mehr Analyseoptionen mit dem Programm möglich sind.

Oft vergessen wird beispielsweise die Crawl-Analyse. Sie kann nach dem Abschluss eines Crawl-Vorgangs angewandt werden und geht die erhobenen Daten noch einmal unter einem ganzheitlichen Blickwinkel durch. So lässt sich auch der Link Score (ein proprietärer Wert des Screaming Frog) für jede Seite ermitteln, der bei der Einschätzung hilft, ob die wichtigsten Inhalte einer Website auch am stärksten intern verlinkt sind.

„Eine äußerst schlechte Idee wäre es, den Screaming Frog in einem Content-Management-System loszulassen.“

Erwähnenswert ist außerdem, dass der Screaming Frog auch genutzt werden kann, um passwortgeschützte Umgebungen zu crawlen. Dann können dem Programm schon vor dem Crawl Benutzernamen und Passwort mitgegeben werden, die dann in einem Session-Cookie gespeichert sind.

Dazu eine kleine Warnung: Eine äußerst schlechte Idee wäre es, den Screaming Frog in einem Content-Management-System loszulassen. Weil der Spider jedem klickbaren Link folgt, würde er dann gegebenenfalls auch Schaltflächen wie „Hinzufügen“, „Bearbeiten“ oder gar „Löschen“ auslösen.

Mehr Daten = mehr Informationen

Besonders mächtig werden die Crawls im Screaming Frog, wenn man sie um Daten von Drittanbietern ergänzt. Ahrefs, Majestic und Moz bieten hierfür (größtenteils kostenpflichtig) Schnittstellen zum Screaming Frog an.

Ohne zusätzliche Kosten lassen sich Daten aus den Google-Diensten PageSpeed Insights, Search Console und Analytics (Universal Analytics und GA4) anbinden.

Automatisierung von Crawls

Eine Besonderheit unter den möglichen Verknüpfungen stellt die Verbindung zum Google Looker Studio (ehemals Google Data Studio) dar. Über diese Verbindung bekommt der Screaming Frog nämlich keine zusätzlichen Daten, sondern wird mit der Möglichkeit versehen, die eigenen Crawl-Daten

dorthin zu exportieren und (bei mehreren Crawls) im zeitlichen Verlauf zu präsentieren.

Für einen schnellen Start wird im Looker Studio bereits ein Template zur Verfügung gestellt (Abbildung 13). Zusätzlich ist es mit dieser Verknüpfung möglich, Crawls für zukünftige Durchführungen zu terminieren und automatisch im Hintergrund ablaufen zu lassen. Die Daten werden dann direkt im Report des Looker Studios ergänzt.

In manchen Fällen ist es wünschenswert, die Struktur und Zusammenhänge einer Website nicht mehr nur tabellarisch darzustellen. Dann bieten sich die Visualisierungsoptionen des Programms an.

Die Visualisierungen helfen, interne Verlinkungen abzubilden, ein Verständnis für die URL-Struktur zu schaffen oder darzustellen, wie oft ein Wort für die Verlinkung zu oder im Text einer bestimmten Seite verwendet wird. Mit ihrer Hilfe lassen sich Linkmuster analysieren, Sackgassen identifizieren und nachvollziehen, wie sich Crawler und Suchmaschinen-Bots durch eine Website hangeln.

Seit Version 19.0 ist bei den Visualisierungsoptionen außerdem eine ansprechende 3-D-Variante hinzugekommen, die im Browser geöffnet wird, sodass nach Herzenslust gedreht und gezoomt werden kann (Abbildung 14).

Web Scraping

Nicht mehr nur SEO-relevant ist die Möglichkeit, den Screaming Frog auch für das Web Scraping zu nutzen. Web Scraping ist der automatisierte Prozess des Extrahierens von Daten aus Webseiten. Die Seiten werden dabei nicht händisch von einem echten Besucher nach den relevanten Informationen durchsucht, sondern Programme oder Skripte versuchen, anhand vordefinierter Regeln die gewünschten Inhalte zu finden und zu speichern. Zum Einsatz kommt das Verfahren in der Marktfor-

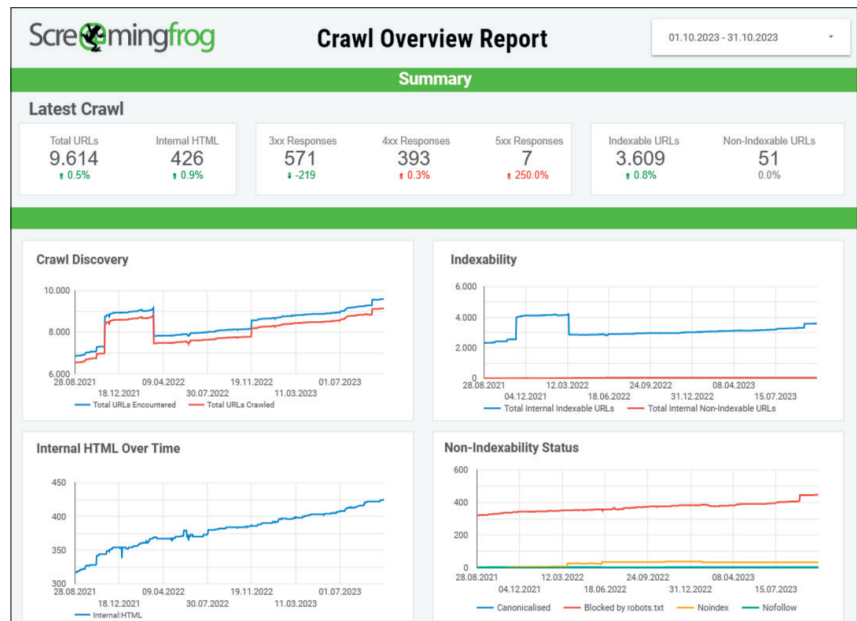


Abb. 13: Das Screaming-Frog-Template im Google Looker Studio

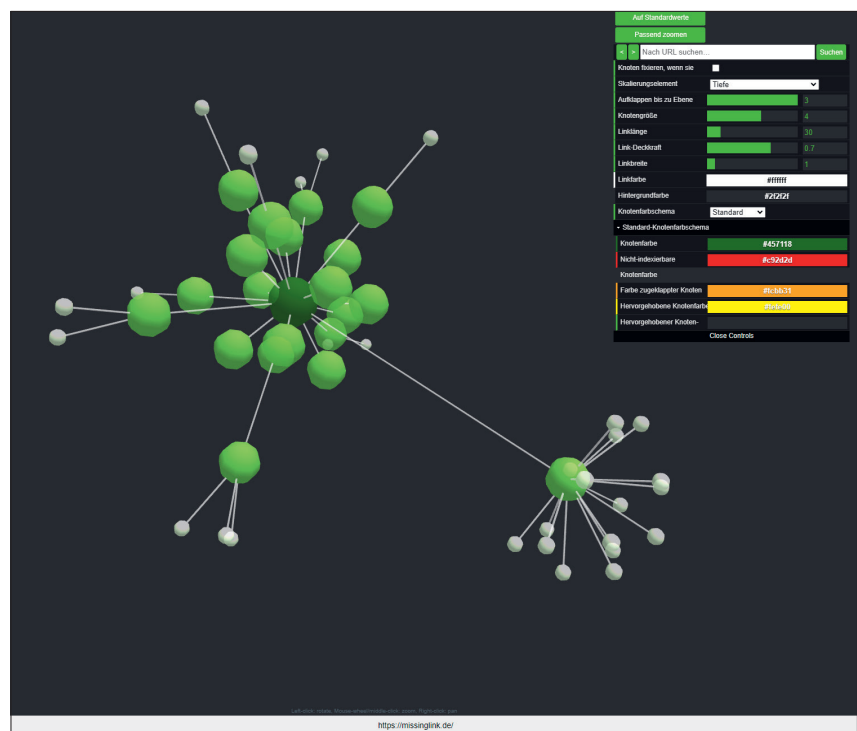


Abb. 14: Ein echter Hingucker: die Visualisierung einer Website als 3-D-Crawl-Diagramm

schung, bei der Überwachung von Wettbewerbern oder beim Extrahieren von Kontaktinformationen.

Wichtig: Web Scraping kann in einigen Fällen illegal sein. Es bleibt für jeden Einsatz zu prüfen, ob damit gegen Nutzungsbedingungen oder geltende Gesetze (unter anderem zum Datenschutz) verstoßen wird.

Überall Probleme: zum Umgang mit Analyseergebnissen

Am Ende eines Crawls ist im Datenanalyse-Fenster eine vollständige Auflistung der gefundenen Probleme sichtbar. Sie werden nach Problemtyp (Fehler, Warnung oder Hinweis) und nach Problempriorität (niedrig,

Problemname	Problemtyp	Problempriorität	URLs	% von Ge...
Antwort-Codes: Interner Serverfehler (5xx)	❗ Fehler	🔴 Hoch	1	0%
Hreflang: Fehlende Rückverweise	❗ Fehler	🔴 Hoch	200	2,35%
Canonicals: Nicht indexierbarer Canonical	❗ Fehler	🔴 Hoch	14	0,16%
Antwort-Codes: Interner Client-Fehler (4xx)	❗ Fehler	🔴 Hoch	2	0%
Sitemaps: Verwaiste URLs	❗ Fehler	🟡 Mittel	2	0,02%
Sitemaps: Nicht indexierbare URLs in der Sitemap	❗ Fehler	🟡 Mittel	67	0,7%
H1: Fehlende	❗ Fehler	🟡 Mittel	2	0,02%
Sitemaps: URLs nicht in der Sitemap	❗ Fehler	🟡 Mittel	30	0,31%
Bilder: Fehlendes Alt-Attribut	❗ Fehler	🟢 Niedrig	119	0,14%
URL: Enthält Leerzeichen	❗ Fehler	🟢 Niedrig	50	0,48%
Links: Nur nicht indexierbare Seiten-Inlinks	⚠️ Warnung	🔴 Hoch	42	0,49%
Links: Seiten ohne interne Outlinks	⚠️ Warnung	🔴 Hoch	2	0,02%
Direktive: Nofollow	⚠️ Warnung	🔴 Hoch	2	0,02%
Antwort-Codes: Intern von Robots.txt blockiert	⚠️ Warnung	🔴 Hoch	828	0,85%
Canonicals: Canonicalised	⚠️ Warnung	🔴 Hoch	1.692	19,88%
Direktive: Noindex	⚠️ Warnung	🔴 Hoch	306	3,3%
Antwort-Codes: Interne Umleitungskette	⚠️ Warnung	🟡 Mittel	17	0,02%
H1: Mehrere	⚠️ Warnung	🟡 Mittel	4	0,05%
Canonicals: Fehlende	⚠️ Warnung	🟡 Mittel	1	0,01%
Hreflang: Fehlende Selbstreferenz	⚠️ Warnung	🟢 Niedrig	200	2,35%
URL: Nicht-ASCII-Zeichen	⚠️ Warnung	🟢 Niedrig	28	0,27%
URL: GA-Verfolgungsparameter	⚠️ Warnung	🟢 Niedrig	3	0,03%
Antwort-Codes: Interne Umleitung (3xx)	⚠️ Warnung	🟢 Niedrig	286	0,29%

Abb. 15: Probleme, Warnungen, Hinweise: Im Datenanalyse-Fenster werden die Ergebnisse zusammengefasst.

mittel oder hoch) unterschieden (Abbildung 15).

Kann man aus dieser Liste ableiten, welche Maßnahmen vorrangig umgesetzt werden sollten? Nein, tatsächlich kann der SEO Spider die Frage nach Umsetzungsprioritäten nicht beantworten. Die Software erfasst Auffälligkeiten, die Probleme darstellen könnten, aber in der Realität nicht zwingend welche sein müssen.

Das Programm kann auch keine Aussage darüber treffen, welche konkreten SEO-Nachteile eine gefundene Auffälligkeit hat. Das liegt zum einen daran, dass es eine Vielzahl von Ranking-Faktoren bei den Suchmaschinen gibt, zum anderen aber auch daran, dass nicht alle Auffälligkeiten Korrekturen notwendig machen.

Ein Beispiel: Der SEO Spider gibt eine Warnung mit einer hohen Priorität aus, wenn er beim Crawl auf Seiten stößt, die mit einer noindex-Anweisung gekennzeichnet wurden. Das kann ein Problem sein, in einigen Fällen kann dahinter aber auch eine bewusste Entscheidung stehen.

Dateninterpretation: Wie wichtig ist ein fehlerfreier Auftritt?

„In manchen Fällen stehen Aufwand und Ertrag aber in keinem sinnvollen Verhältnis.“

Die Kategorisierung von Problemtypen und Prioritäten suggeriert, dass sich daraus auch die Wichtigkeit und somit Reihenfolge anziehender Optimierungsmaßnahmen ableiten ließe. Aber tatsächlich ist die SEO-Welt auch an dieser Stelle wieder einmal etwas komplizierter. Es gibt offensichtliche Probleme und wirkliche Fehler, deren Behebung durchaus wünschenswert wäre. In manchen Fällen stehen Aufwand und Ertrag aber in keinem sinnvollen Verhältnis mehr, weil die Behebung dieser Fehler gar keine spürbaren Auswirkungen auf die Sichtbarkeit bewirkt.

Hier ist die eigene SEO-Expertise gefordert, um Daten eines Crawls zu

interpretieren und relevante Maßnahmen zu identifizieren. Gut möglich, dass man an der einen oder anderen Stelle zu dem Ergebnis kommt, dass man mit bestimmten Fehler- und Warnhinweisen im SEO Spider leben kann, weil deren Behebung die Website nicht voranbringen würde.

Technisches SEO ist kein „Do and forget“

Fakt ist: Eine technisch optimierte Website kann dazu beitragen, dass sie sichtbarer in den Suchergebnissen erscheint. Eine Website, die schneller lädt, gut strukturiert ist und eine hohe Benutzerfreundlichkeit aufweist, hat nach wie vor Vorteile gegenüber der Website eines Wettbewerbers, die mit diesen Optimierungen nicht punkten kann.

Es ist jedoch ein Irrglaube, dass diese Arbeiten mit einem einmaligen Crawl durch den Screaming Frog abzuschließen sind. Vielmehr sollte es zur Routine werden, jede Veränderung an der eigenen Website auch parallel durch den Einsatz eines Crawling-Tools wie des Screaming Frog zu analysieren. Schon allein aus diesem Grund ist es wichtig, das dafür genutzte Programm gut zu beherrschen.

Vielleicht hilft dieser Beitrag, die Berührungängste mit dem Screaming Frog abzubauen. Warum nicht gleich einmal die eigene Website crawlen lassen? Die Ergebnisse könnten überraschen.

