

SEO-TOOLS UNTER DER LUPE

Mario Fischer

Damit hatten wir nicht gerechnet. Während unseres Kontakts zum Hersteller des SEO-Tools Screaming Frog ließ dieser wenige Tage vor dem Go-live einer neuen, stark erweiterten Version 10 keinen Ton darüber verlauten. Kurz nach Redaktionsschluss und unmittelbar vor dem Druck gab man dann bekannt, dass eine neue Version erscheint. Leider konnten wir an den Test des Tools in der letzten Ausgabe nur noch am Ende einen kleinen Kasten mit einem Hinweis auf die neuen Funktionen anhängen – und dort versprechen, alles Wichtige hier in dieser Ausgabe nachzureichen. Dieses Versprechen lösen wir hiermit ein, und weil es eine ganze Menge Nützliches und Neues in der Version 10 gibt, entschieden wir uns, gleich einen umfassenden Teil 2 zu liefern. Sie werden sehen.

Wer den letzten Testbericht in der Ausgabe 52 gelesen hat, wird erkennen, dass es sich wirklich lohnt, über die neuen Funktionen ausführlicher zu sprechen. Insbesondere die jetzt eingebauten Möglichkeiten der internen Berechnung der Linkpower bzw. -verteilung und die innovativen Visualisierungen haben es tatsächlich in sich. Auf alle bereits in der letzten Ausgabe besprochenen Funktionen des Tools gehen wir hier aus Platzgründen nicht mehr ein.

Bisher in der SEO-Tool-Serie erschienen:

Sistrix Toolbox:	Ausgabe #42
LinkResearchTools:	Ausgabe #43
SEO-Tools für Excel:	Ausgabe #44
XOVI SEO-Tool:	Ausgabe #45
SEO-Diver:	Ausgabe #46
linkbird:	Ausgabe #47
Audisto:	Ausgabe #48
SEMrush:	Ausgabe #49
Screaming Frog (SEO Spider):	Ausgabe #52
Screaming Frog V 10 - Update:	Ausgabe #53

Version 10: Was ist neu dazu gekommen?

Auf eine Art Autostart-Möglichkeit haben wohl viele Nutzer des Screaming Frog gewartet. Damit lassen sich zeitgesteuert Crawls starten und die gewonnenen Daten automatisch sehr flexibel als Dateien für spätere Analysen ablegen. Damit kommt man einer professionelleren Überwachung einer oder mehrerer Domains schon recht nahe. Man muss nicht mehr daran denken, eine Analyse zu starten, und auf die Ergebnisse warten. Auch die Möglichkeit, verschieden definierte Crawls zu unterschiedlichen Startterminen laufen zu lassen, hat einen deutlichen Mehrwert. Einen weiteren Vorteil hat die Zeitsteuerung: Meist arbeitet man ja tagsüber, sollte wegen der möglichen

Serverbelastung solche Datenerhebungen aber besser nachts laufen lassen. Auch das ist mit dem Autostart jetzt machbar, weil man nicht mehr selbst vor dem Rechner sitzen und klicken muss.

Der Screaming Frog erhebt ja in Summe eine Menge an Metriken. Bisher blieb es bei der statischen Wiedergabe der Daten in der Oberfläche. Wer Berechnungen bezüglich der Linkstruktur vornehmen wollte, musste dies anschließend mit eigenen Programmen oder z. B. Excel tun. Hier hatte der Anbieter ein Einsehen und führt in der neuen Version die Möglichkeit der Berechnung weiterer Kennmetriken, wie zum Beispiel eines Link Score ein, welcher der Berechnung des (internen) PageRanks nachempfunden wurde. Auch dazu im Verlauf des Beitrags mehr.

TEIL 10: SCREAMING FROG

UPDATE

Was nützen die schönsten Tabellen mit Daten, wenn man keine vernünftige Visualisierung hat? Diese Frage beantwortet das Tool endlich mit einer völlig neuen, dynamischen und filterbaren Darstellung der Beziehungen zwischen Seiten und deren Verlinkung. Neben einer strukturellen Übersicht lassen sich nun auch Darstellungsarten wählen, die man bis dato aufwendig in das Visualisierungstool Gephi importieren und dort justieren musste. Die beim Im-/Export nötigen Aufräumarbeiten in den Daten entfallen damit und Erkenntnisse für die Optimierung lassen sich nun praktisch live durch Veränderung von Darstellungsparametern per Mausklick gewinnen.

Eine leicht übersehbare Ergänzung ist die neue Spalte „Indexability“. Zusammen mit der Folgespalte „Indexability Status“ bekommt man nun einen einfacheren Einblick. „Indexable“ bedeutet, dass die URL gecrawlt werden kann, einen 200er-Statuscode zurückliefert und im Google-Index zu finden sein dürfte. „Non-Indexable“ als Eintrag bedeutet entsprechend das Gegenteil. In der Folgespalte wird dann bei solchen nicht indexierbaren URLs der Grund angegeben. Dieser kann z. B. in einer Blockierung per robots.txt oder dem Metatag „noindex“, aber auch in einem Redirect, einem Canonical-Tag

oder Serverfehlern liegen. Auch wenn eine URL intern ausschließlich per „nofollow“ angelinkt wird, bekommt sie nun leicht erkennbar den „Non-Indexable“-Status. Der Reiter „Pagination“ fördert in diesem Zusammenhang auch mehr Transparenz, weil er neben den Adressen für ein hinterlegtes rel=“next“ und „prev“ den Zusammenhang mit (falschen) Kanonisierungen oder eben noindex sichtbar macht.

Weiterhin neu ist ein sog. Command Line Interface für Windows, Mac und Linux. Damit lässt sich der Frog kommandozeilenorientiert steuern, ohne das Programm zu öffnen. Eine Übersicht über nutzbare Befehle bzw. wie das generell funktioniert, findet man bei Bedarf unter <http://einfach.st/frogline>.

Der Frog hüpfet jetzt von alleine!

Eine wirklich nützliche Neuerung findet man unscheinbar unter dem Menüpunkt File/Scheduling. Dort öffnet sich eine Maske, über die man den Screaming Frog automatisch zeitgesteuert mit definierbaren Parametern starten kann. Was auf den ersten Blick nach Nerd-Funktion aussieht, entpuppt sich bei längerem Nachdenken als äußerst hilfreich. Damit lassen sich nämlich unbeaufsichtigt auf dem einen Rechner

„Auf den zweiten Blick extrem nützlich: der Scheduler.“

(muss natürlich eingeschaltet bleiben) oder einem zweiten PC bestimmte Crawlvorgänge und deren Archivierung bewerkstelligen. Doch wofür braucht man das?

Wer sich länger mit SEO beschäftigt, dem ist folgendes Problem sicherlich schon öfter begegnet: Rankings ändern sich nennenswert ins Negative und es ist kein Google-Update in Sicht, dem man die Schuld geben könnte. Sind irgendwelche Veränderungen am Shop bzw. der Website die Ursache? Hat sich Content verändert, die interne Verlinkung oder ist vielleicht die neue Navigationsebene die Ursache für eine Neubewertung seitens der Suchmaschinen? Die meisten Unternehmen fischen hier im Trüben, niemand weiß bei einer „lebenden“ Website mit beständigen Änderungen, wie die Metriken der vorigen Woche, im letzten Monat oder vor einem halben Jahr waren. Hier setzt das Scheduling-Modul an. Scans bzw. alle relevanten Site-Daten werden automatisch und in definierten Zeitintervallen erhoben, gespeichert und bei Bedarf analysiert.

2018	gefilterte URL	Size	Word Count	Crawl Depth	Link Score	Unique Inlinks	Title	H1-1
Jan	https://www.domain.de/produkt4711.html	69.931	1.391	2	68	74	Produkt 4711 günstig online kaufen	Die Vorteile von Produkt 4711 si
Feb	https://www.domain.de/produkt4711.html	69.931	1.391	2	68	74	Produkt 4711 günstig online kaufen	Die Vorteile von Produkt 4711 si
Mrz	https://www.domain.de/produkt4711.html	69.931	1.391	2	52	74	Produkt 4711 günstig online kaufen	Die Vorteile von Produkt 4711 si
Apr	https://www.domain.de/produkt4711.html	69.931	1.391	2	32	16	Produkt 4711 günstig online kaufen	Die Vorteile von Produkt 4711 si
Mai	https://www.domain.de/produkt4711.html	69.931	1.391	2	30	16	Produkt 4711 günstig online kaufen	Die Vorteile von Produkt 4711 si
Jun	https://www.domain.de/produkt4711.html	69.931	1.391	2	31	18	Produkt 4711 günstig online kaufen	Die Vorteile von Produkt 4711 si
Jul	https://www.domain.de/produkt4711.html	69.931	1.391	2	32	18	Produkt 4711 günstig online kaufen	Die Vorteile von Produkt 4711 si
Aug	https://www.domain.de/produkt4711.html	54.837	821	2	27	14	Firma X - Produkt 4711 günstig kau	
Sep	https://www.domain.de/produkt4711.html	54.837	821	2	28	14	Firma X - Produkt 4711 günstig kau	
Okt	https://www.domain.de/produkt4711.html	54.837	821	2	26	14	Firma X - Produkt 4711 günstig kau	

Abb. 1: Was hat sich z. B. bei einer bestimmten URL im Lauf der Zeit verändert?

Über eine so aufgebaute Historie kann praktisch jede Veränderung zu jedem gescannten Zeitpunkt gegenübergestellt werden! Hat sich die Anzahl der eingehenden Links oder der berechnete Link Score verändert? Welchen Einfluss auf die Verlinkung hat die neu eingeführte Blätternavigation? Was hat sich an den Metriken für SeiteA.html verändert, die sich im Ranking beständig verschlechtert? Solche und sehr viel mehr Fragen lassen sich jetzt mit wenigen Mausclicks bzw. ein paar Excelfiltern und -sortierungen beantworten.

Ein Beispiel zeigt Abbildung 1. Angenommen, das Ranking der URL *www.domain.de/produkt4711.html* hätte sich im Lauf des Jahres spürbar verschlechtert. Extrahiert man aus den automatisch abgelegten Dateien die jeweilige Zeile mit dieser URL und legt sie zusammen mit den in den Spalten stehenden Metriken monatsweise manuell oder per Formel zeilenweise ab, erkennt man recht schnell, was hier passiert ist. Im (fiktiven) Beispiel sind die Zellen mit Datenänderung der besseren Erklärung halber orange eingefärbt.

Man kann unschwer erkennen, dass sich bereits im März der Link Score (der intern berechnete PageRank von Screaming Frog) veränderte, von vormals 68 auf 52. Das heißt, die interne Linkpower floss wahrscheinlich durch zusätzlich neue Links auf andere URLs, da die Zahl der eingehenden Links stabil blieb. Im April sieht man, dass sich dann auch die Anzahl der eingehenden Links zusätzlich verringerte, nämlich von 74 vorher auf nun nur noch 16. Der Link Score fiel folgerichtig erneut massiv auf 32. Im August hatte jemand den Title der URL bearbeitet und den Firmennamen nach vorne gezogen – und damit die Wortwertung verändert. Gleichzeitig wurde die als wichtig erachtete Überschrift H1 entfernt. Das alles können, müssen aber natürlich nicht die Gründe für eine stetige Rankingverschlechterung sein. Wie immer sollte man über SEO-Wissen

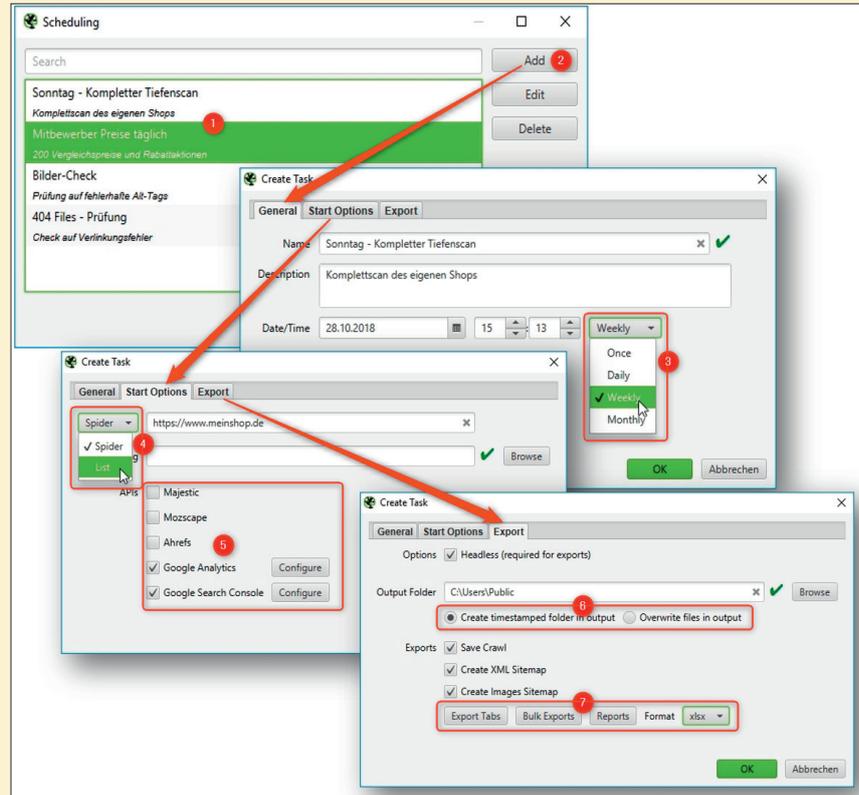


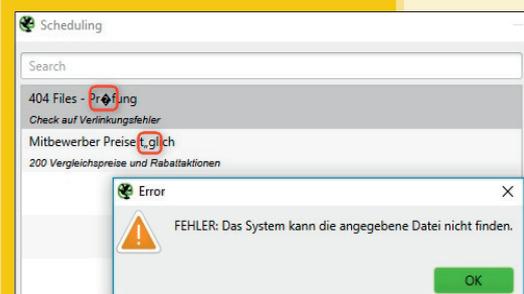
Abb. 2: Das neue Automatikmodul lässt praktisch keine Wünsche offen

verfügen, um aus Daten bzw. Metriken die richtigen Schlüsse zu ziehen, besser: die richtigen Spuren für mögliche Optimierungsmaßnahmen zu finden. Automatisch und auf Knopfdruck geht auch beim Screaming Frog nichts. Findet man keine Auffälligkeiten in der internen Struktur, weiß man zumindest relativ sicher, dass man die Gründe woanders suchen muss (z. B. mögliche Keyword- oder URL-Strafen, ein Google-Update, der Wegfall externer Backlinks oder stark verbesserte Mitbewerber).

Damit man genau die Metriken bekommt, die man braucht, lässt sich der Scheduler sehr flexibel konfigurieren (Abbildung 2). Über „Add“ (Ziffer 2) können einzelne Crawls angelegt werden, die dann in der Übersicht (Ziffer 1) zu sehen sind. Zunächst hinterlegt man Startdatum und ob bzw. wann der Crawl später automatisch starten soll. Im Reiter „Start Options“ lässt sich statt einer Start-URL auch eine Liste mit definierten URLs hinterlegen (Ziffer 4). Man muss also nicht immer eine gesamte Domain durchleuchten, es können auch Daten verschiedener Domains bzw.

TIPP: ACHTUNG BUG!

Mit deutschen Umlauten hat der Frog noch ein paar Probleme. Vermeiden Sie deutsche Sonderzeichen wie ä, ü, ö und ß bei eigenen Einträgen im Scheduler. Man kann diese zwar eintippen, aber beim nächsten Öffnen macht das Tool unschöne Symbole daraus. Schlimmer ist jedoch, dass man diese Einträge dann nicht mehr editieren oder löschen kann, da der Frog beim Lösch- oder erneutem Speicher-Versuch eine Fehlermeldung erzeugt. Erstaunlich, dass dieser gravierende Bug selbst in der Version 10.3 noch nicht gemeldet bzw. entdeckt wurde. Die meisten Anwender nutzen Tools meist wohl doch nicht so intensiv?



URLs von dort abgearbeitet werden.

Zu den URLs lassen sich über entsprechende API-Zugänge jeweils weitere Daten von SEO-Tools (aktuell Majestic,

Mozscape und Ahrefs) dazuholen, sofern man dort über einen entsprechenden Zugang verfügt (Ziffer 5). Wichtig und nützlich sind vor allem natürlich Google Analytics und die Search Console, um zu jeder (eigenen) URL z. B. Page-Impressions, Bounce-Rate, Conversion-Rate etc. und aktuelle Rankingdaten zu erhalten.

Im Reiter „Export“ legt man dann fest, wo die Daten gespeichert werden sollen, und vor allem, ob sie mit Datum nacheinander archiviert oder einfach nur überschrieben werden sollen (Ziffer 6). Am Ende gilt es noch zu definieren, welche Daten und Berichte man tatsächlich haben möchte. Die im Frog im manuellen Betrieb abrufbaren Berichte (Reports) und Exporte (Bulk Exports) (Ziffer 7) lassen sich ebenso dazuwählen wie einzelne Spaltensektionen, die man unter „Export Tabs“ auswählen kann. Es lässt sich also bei Bedarf gleich eine ganze Batterie an Auswertungen speichern (im CSV- oder Excel-Format). So kann man z. B. ein 404- (broken link report) oder ein hreflang-Fehlerreport direkt an andere Personen weitergeben, ohne weiteres Zutun bzw. ohne diese erst aus einer Gesamttabelle filtern zu müssen.

Wer sich hier wichtige Reports anlegt, wird diese Art der Archivierung und den damit möglichen schnellen Zugriff auf die Vergangenheit zu Vergleichszwecken sicher bald zu schätzen wissen!

Am Ende sei noch erwähnt, dass sich bei der Aktivierung des Renderings für einen Crawlvorgang auf diese Art und Weise auch Screenshots von Webseiten mit ablegen lassen. Ob man damit einzelne wichtige Seiten von definierten Mitbewerbern ablegt und/oder die eigenen, bleibt den Anforderungen und der Fantasie überlassen. Der Effekt ist auch hier der gleiche: Man kann jeweils bei Bedarf einen Blick in die Vergangenheit werfen, Preisnachlässe oder Sonderaktionen etc. recherchieren oder einfach auch nachvollziehen, wie

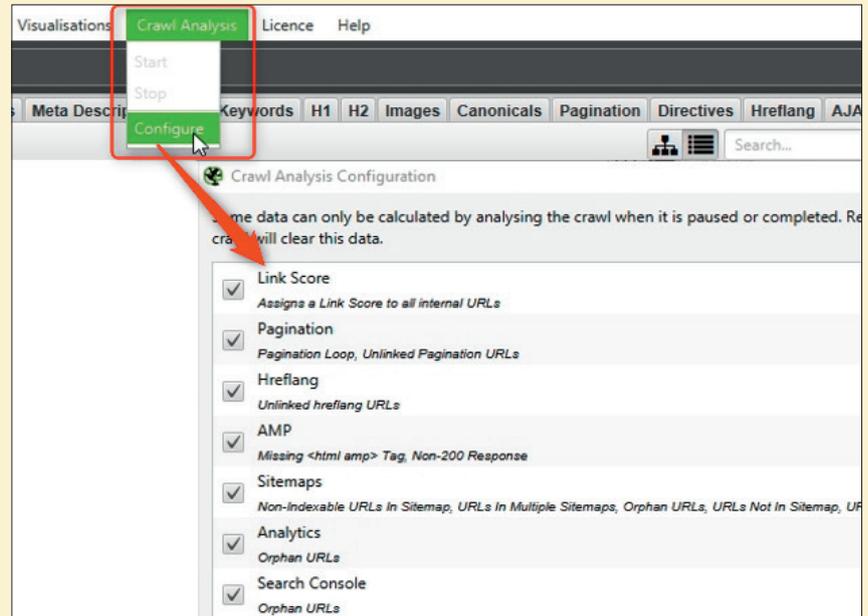


Abb. 3: Welche Berechnungen soll der Screaming Frog durchführen?



Abb. 4: Ein neuer Fortschrittsbalken gibt Auskunft, wie lange die Berechnung noch dauert

und wann sich die eigenen Webseiten optisch ändern. Das Rendering ging zwar schon in früheren Versionen, aber mit der neuen Möglichkeit der unüberwachten bzw. nicht aktiv angestoßenen Durchläufe bekommt dies deutlich mehr Nützlichkeit.

Wie immer erkennt man den wahren analytischen Wert solcher Daten erst, wenn man sie denn tatsächlich im Fall des Falles zur Verfügung hat.

Echte Datenauswertungen

Unter dem neuen Menüpunkt „Crawl Analysis“ lässt sich nach einem Crawlvorgang eine Berechnung für interne Auswertungen anstoßen. Je nach Konfiguration werden nun die ermittelten Daten für weitere Auswertungen und Checks aufbereitet (Abbildung 3). Möchte man die Berechnung nicht jedes Mal manuell anstoßen, kann man unten (Ziffer 1) auswählen, dass dies automatisch nach jedem Crawlvorgang geschieht. Bei umfangreichen Daten kann dies jedoch je nach Rechnerpower ein klein wenig dauern (Abbildung 4). Die sieben auswählbaren Analysefunktionen sind in Abbildung 3 zu sehen.

Link Score

Die von den beiden Google-Gründern patentierte PageRank-Berechnung wird nach Aussagen von Mitarbeitern nach wie vor regelmäßig durchgeführt. Bekanntlich steckt dahinter ein mathematisches System zur Vererbung von „Bedeutung“ für ein Dokument, das iterativ über eingehende Links berechnet wird. Das heißt, ein Dokument hat einen initialen PageRank-Wert und gibt diesen um einen notwendigen Dämpfungsfaktor gekürzt über ausgehende Links an andere Dokumente weiter, die diesen „auf sammeln“. Dabei spielt es keine Rolle, ob diese Links inner- oder außerhalb einer Domain gesetzt werden. So weit, so gut.

Wenn aber die Wichtigkeit eines Dokuments, sprich einer Webseite, – ganz grob und vereinfacht gesagt – durch die Anzahl eingehender Links bestimmt wird, ist es hilfreich, diesen Wert zumindest ansatzweise zu erkennen. Hierzu gibt es verschiedene Möglichkeiten, wie u. a. im Titelbeitrag der Ausgabe #44 („Es lebe der PageRank“) ausführlich beschrieben wurde. Die Macher des Screaming Frog

Crawl Depth	Link Score	Inlinks	Unique Inlinks	% of Total
1	100	619	412	71.78
2	63	207	207	36.06
2	63	207	207	36.06
2	63	210	207	36.06
2	63	415	207	36.06

Abb. 5: Neu ist die Berechnung eines Link Scores, der dem PageRank nachempfunden ist

haben nun in der neuen Version einen Berechnungsalgorithmus integriert. Damit wird es auf einfache Art und Weise möglich, die Wichtigkeit, die man per Linksetzung möglicherweise auch unbeabsichtigt erzeugt, besser transparent zu machen. Nachdem man, wie oben beschrieben, die Berechnung nach einem Crawl angestoßen hat, erscheint eine neue Spalte mit „Link Score“, in der pro URL ein Wert von 0-100 ausgegeben wird. Es ist wichtig zu verstehen, dass es sich hier nur um einen internen Wert handelt, der nur erkennbare Links innerhalb einer Domain repräsentiert und Links von außen ignoriert. Insofern handelt es sich natürlich nicht um einen echten PageRank-Wert – er ist nur im Vergleich mit allen anderen Seiten zu sehen. Trotzdem ist diese strukturelle Betrachtungsweise sehr hilfreich, um zu sehen, ob die wirklich wichtigen Seiten auch genügend interne Linkpower abbekommen.

Pagination, Hreflang, AMP und Sitemaps

Die neuen Analysefunktionen der gecrawlten Daten ermöglichen nun auch einen wesentlich tieferen und aufschlussreicheren Einblick in die oft kritische bzw. fehlerbehaftete Umsetzung von Blätternavigationen (Pagination) und die hinterlegten Tags für Sprach- und Länderversionen (Hreflang). So findet man jetzt z. B. ganz einfacher Mauseklick Blätter-URLs, die nicht korrekt in Tags von Vorgängern oder Nachfolgern hinterlegt wurden oder die fälschlich eine Schleife bilden (Pagination Loop). Ebenso einfach lassen sich fehlende und fehlend rückverlinkende

hreflang-Tags aufstöbern. Zwar zählt dies eher zum sog. „Housekeeping“ beim SEO, also zu den Dingen, die man routinemäßig in Ordnung oder, um im Bild zu bleiben, sauber halten sollte. In der Praxis aber hat sich gezeigt, dass diese Art struktureller Analyse wegen der Datenkomplexität ohne leistungsfähige Tools oft nicht mit vertretbarem Aufwand manuell zu machen ist.

Wer die Technik der sog. Accelerated Mobile Pages (AMP) bereits nutzt, wird sich über die tiefere Integration im Screaming Frog freuen. So ist es ab der Version 10 möglich, die AMP-Adressen automatisch zu extrahieren, mit zu crawlen und am Ende auch zu validieren.

Nach einem Crawl lässt sich übrigens auch eine gültig formatierte Sitemap im XML-Format erzeugen sowie auch eine Image-Sitemap für Bilder. Sollten sich URLs auf der aktuellen Website, aber nicht in der Sitemap finden – und umgekehrt –, ist auch das einfach über Filter oder einen speziellen Bericht aufzuzeigen.

Analytics- und Search-Console-Daten werden effizienter genutzt

Auch die Integration der von Google über die API abrufbaren Daten aus Google Analytics und der Search Console wurde weiter verbessert. Es ist jetzt möglich, neue URLs, die von der Search Console oder Analytics übertragen werden, gleich automatisch mit zu crawlen. In der Vorversion musste man diese erst umständlich über den sog. „Orphan File“-Bericht abfischen und manuell integrieren, um einen vollständigen Überblick zu bekommen. Diese

<p>▼ Canonicals</p> <p>All (1128) (100,00%)</p> <p>Contains Canonical (1089) (96,54%)</p> <p>Self Referencing (1089) (96,54%)</p> <p>Canonicalised (0) (0,00%)</p> <p>Missing (39) (3,46%)</p> <p>Multiple (0) (0,00%)</p> <p>Non-Indexable Canonical (0) (0,00%)</p> <p>▼ Pagination</p> <p>All (1128) (100,00%)</p> <p>Contains Pagination (4) (0,35%)</p> <p>First Page (4) (0,35%)</p> <p>Paginated 2+ Pages (0) (0,00%)</p> <p>Pagination URL Not In Anchor Tag (0) (0,00%)</p> <p>Non-200 Pagination URLs (0) (0,00%)</p> <p>Unlinked Pagination URLs (Crawl Analysis Required)</p> <p>Non-Indexable (0) (0,00%)</p> <p>Multiple Pagination URLs (0) (0,00%)</p> <p>Pagination Loop (Crawl Analysis Required)</p> <p>Sequence Error (0) (0,00%)</p>

Abb. 6: Die neuen Berechnungsalgorithmen bringen mehr Transparenz in Berichte und Übersichten

Art der Analyse ist besonders nützlich, denn ein Crawler kann ja nur alle Seiten aufstöbern, die von einer bestimmten Startseite (meist der Homepage) aus verlinkt sind. Allenfalls können dabei verwaiste URLs gefunden werden, wenn sie in einer hinterlegten Sitemap verzeichnet sind, aber eben nicht mehr angelinkt werden. Nimmt man jedoch die URLs dazu, die Klicks über Google Analytics bekommen oder via Suche bei Google gefunden werden, was ja die Search Console meldet, bekommt man eben auch die Adressen, die man selbst oft nicht mehr auf dem Schirm hat, weil man sie – vermeintlich – gelöscht hat. Nicht jedes CMS löscht bekanntlich Seiten auch physikalisch vom Webserver, sondern nimmt sie nur aus der Navigationsstruktur. Bei Analysen findet man oft Hunderte oder gar Tausende solcher Seiten, die noch im Netz herumschwirren und im Lauf der Zeit immer mehr 404-Fehler produzieren, weil z. B. die dort verlinkten Bilder oder abgehenden internen Links im Lauf der Monate oder Jahre nicht mehr existieren. Experten empfehlen daher, gerade auch solche verwaisten Seiten (Orphan Pages) im Auge zu behalten, damit bei Suchmaschinen nicht die Zahl der 404-Fehler ansteigt und die Vertrauenswürdigkeit einer Domain sinkt. Zuletzt ist es ja

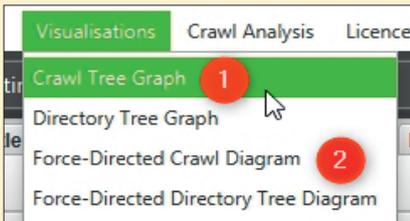


Abb. 7: Im Menü „Visualisations“ gibt es eine Darstellung in Baumstrukturen oder als dynamisches Crawl-Diagramm

auch keine gute Nutzererfahrung, eine Seite bei Google zu finden, die dann ohne die mittlerweile fehlenden Bilder dargestellt wird, und wenn weiterführende Links in die Irre führen.

Optische Liveanalysen – Visualisierungsmöglichkeiten

Dieser neue Menüpunkt hat es in sich und er war auch der Hauptgrund, den Screaming Frog erneut in die Tool-Liste zu nehmen. Prinzipiell gibt es zwei unterschiedliche Ansichten. Die beiden Funktionen „Crawl Tree Graph“ und „Directory Tree Graph“ (Abbildung 7, Ziffer 1) zeigen in einer Baumstruktur die analysierte Website an. Dabei kann man auswählen, ob man die Ansicht lieber horizontal oder vertikal haben möchte. Ein Dreh am Mousrad zoomt dabei. Mit gedrückter Maustaste lässt sich das Baumdiagramm einfach

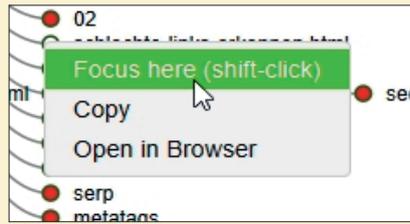


Abb. 8: Mit der rechten Maustaste kann man auf das ausgewählte Dokument zentrieren, die URL kopieren oder einfach im Browser öffnen

verschieben. Je nach Wunsch kann man die Ansicht dann über alle Dokumente oder eben nur nach Verzeichnissen (Directory Tree Graph) generieren.

Wählt man mit der Maus einen Knoten (Dokument oder Verzeichnis) aus und klickt mit der rechten Maustaste, kann man diesen ins Zentrum stellen bzw. darauf fokussieren. Das hilft vor allem bei größeren Websites, bei denen man der Übersichtlichkeit halber zunächst die Anzahl der angezeigten Knoten einschränkt (oben rechts mit dem Zahnradsymbol, dann „Expand to Level“) und dann gezielt „tiefer“ geht. Der Baum baut sich dann sofort in Echtzeit erneut auf. Wie in Abbildung 8 zu sehen ist, lässt sich die URL auch einfach einzeln kopieren oder im Browser zur Ansicht öffnen.

Das Aussehen des Baumdiagramms lässt sich nach Farben, danach, welche Knoten (Nodes) wie angezeigt werden,

und nach den Abständen zwischen den Ebenen und den Knoten gut an die persönlichen Bedürfnisse anpassen.

Streich man mit dem Mauszeiger über einen Knoten, erscheint eine grüne Hinweisbox mit den zu dem Knoten gehörigen Daten (Abbildung 9). Besonders interessant sind hier natürlich die schon beim Crawl ggf. angereicherten Daten der diversen API-Schnittstellen. Im Beispiel wurden über einen Connect zu Google Analytics und der Searchconsole zu jeder URL Kennmetriken gezogen, die man nun in der grünen Box sehen kann. Der markierte Knoten bzw. die URL hatte 8.428 Sessions im letzten Monat und 9.507 Klicks in den organischen Suchergebnissen. Die gewünschte Zeitspanne stellt man ganz einfach bereits beim Crawl ein. Hat man von anderen SEO-Tool-Schnittstellen Daten mit abgerufen, bekommt man diese hier mit angezeigt. Links oben neben dem URL-Feld kann man über den Pfeile-Kreis oder den Zurück-Button jederzeit wieder auf die Startdarstellung zurückschalten oder einen Refresh auslösen. Über den Speicherbutton lassen sich die Abbildungen praktischerweise direkt im skalierbaren SVG-Format abspeichern.

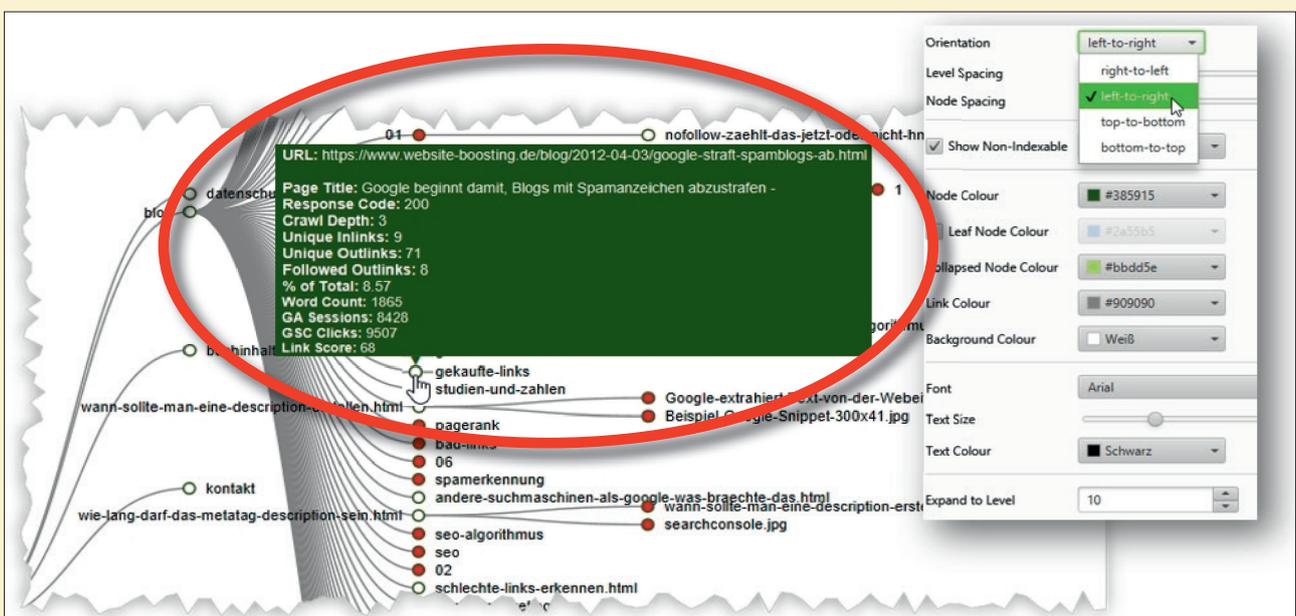


Abb. 9: Ein Baumgraph zeigt die wichtige Darstellung nach Klickleveln – nicht nach Webseitenmüs

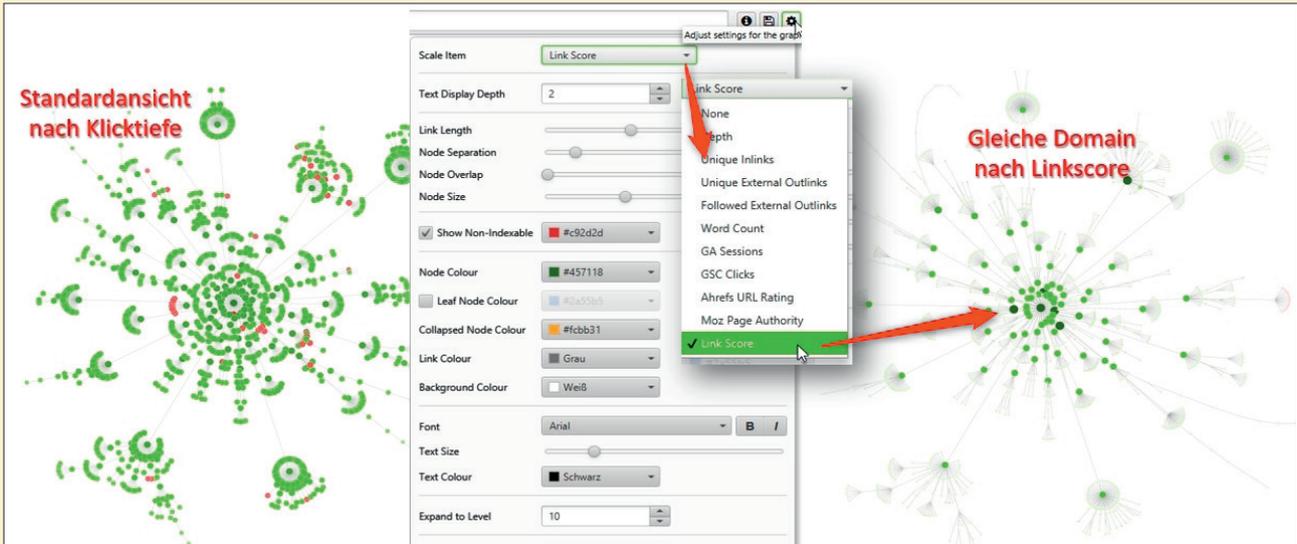


Abb. 10: Eine Domain – unterschiedliche Erkenntnisziele

Während die beiden Tree-Graphen für strukturelle Übersicht sorgen, erlauben die beiden Forced-Directed-Diagramme (Abbildung 7) konkrete SEO-Analysen bzw. Ansichten. Über sie kann man den „Gesundheitszustand“ einer Domain beurteilen. Ist die Startseite wirklich das eigentliche Zentrum oder sind andere Seiten durch eine ungeschickte bzw. unbedachte interne Verlinkung wichtiger geworden? Gerade bei der Integration eines Blogsystems kann so etwas recht leicht passieren, auch bei „explodierenden“ Websites, an die immer einfach wieder etwas angeflanscht wurde, ohne dass man von der Navigation (Verlinkung) her die nötigen Aufräumarbeiten durchführte. Entsprechende visuelle Beispiele findet man recht gut dargestellt bei Bedarf bei Portent.com unter <http://einfach.st/forceddirected>, wo allerdings mit Gephi gearbeitet wurde. Abbildung 11 zeigt ein solches Beispiel von Portent. Man sieht auf den ersten Blick, dass hier etwas ganz und gar nicht stimmt. Der Blog wurde aus Verlinkungssicht abgespalten und ist nur mit einem Link von der Startseite aus erreichbar. Von den Blogartikeln führt kein einziger wieder zurück zu den an die Startseite angebondenen Produktseiten. Der Downloadbereich ist nur über die Startseite des Blogs erreichbar. Noch schlimmer könnte sich am Ende auswirken, dass die Startseite

des Blogs durch die interne Verlinkung deutlich stärker ist (das Blau des Kreises ist kräftiger) als die eigentliche Startseite. Hier spürt man sofort die Vorteile einer derartigen visuellen Darstellung gegenüber einer zahlenorientierten.

Abbildung 10 zeigt recht gut, wie unterschiedlich eine solche Darstellung für eine Domain sein kann. Links erkennt man die einfache Darstellung nach Klicktiefe, rechts werden die Nodes dann über den Link Score angezeigt. Man sieht sofort, wo in Anlehnung an die PageRank-Berechnung die „starken“ Seiten sind. Dabei muss man natürlich immer im Kopf behalten, dass hier streng nach dem Linkgraph gearbeitet wird. Websitebetreiber „denken“ nicht selten nach ihrer Navigation. Suchmaschinen werten aber jeden Link, auch z. B. die in Texten, als Ausgangstür. Das bedeutet, dass eine Seite, die nach Navigation eigentlich erst in Ebene 4 liegt, durch einen Direktlink von der Startseite auf Ebene 1 rückt, zusammen mit den angelinkten Seiten via Navigation! Genau dies kann man hier dann sehr gut erkennen. Ebenso, ob eine eher unwichtige Seite durch zu viele Links von anderen Seiten viel zu wichtig erscheint. Diese (visuelle) Erkenntnis erlaubt entsprechende Aufräumarbeiten.

Abbildung 12 zeigt die Daten eines Crawls mit unterschiedlichen „Scale

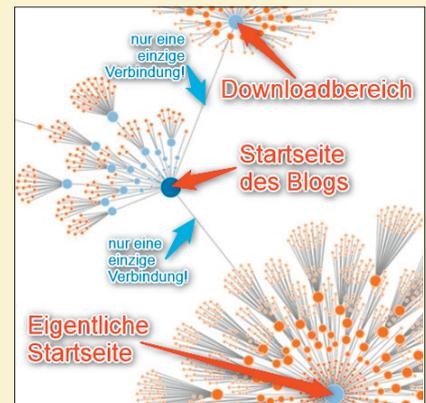


Abb. 11: Keine gute Verlinkungsstruktur (Quelle: Portent.com)

Items“ (siehe das Pull-down-Menü in Abbildung 10). Der Screaming Frog sieht noch weitere Möglichkeiten der Zuordnung von Nodes (den Kreisen), Farben und Größe vor, aber diese sechs beispielhaft Ausgewählten verdeutlichen gut, wie unterschiedlich die Diagramme wirken. Wo tummeln sich eigentlich die meisten Besucher oder wo sitzen die URLs mit besonders viel Textumfang? Gibt es nicht indexierbare URLs? Bei „Unique Inlinks“ und „Link Score“ erkennt man ebenfalls recht gut, dass es nicht mit einer einfachen Zählung eingehender Links getan ist. Das bedeutet, dass nicht automatisch die Seiten am stärksten sind, welche die meisten Eingangslinks besitzen, sondern dass ein iterativer Vererbungsalgorithmus nötig ist, der auch jeweils die Anzahl abgehender Links (mehr Links bedeutet,

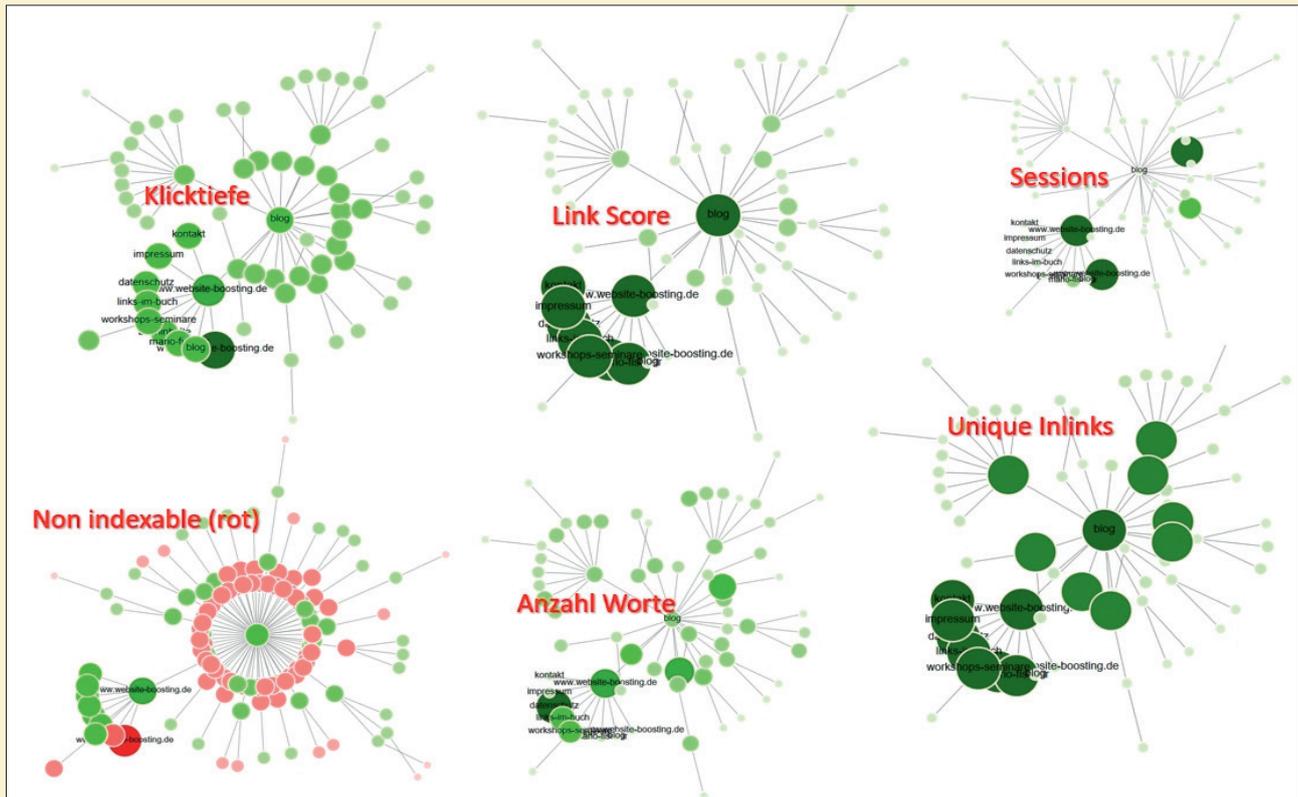


Abb. 12: Je nach Auswahl der Metriken lassen sich die Nodes farblich und von der Größe her mit unterschiedlichen Aussagen belegen

dass jeder Link weniger Power vererben kann) mit berücksichtig.

Ein Rechtsklick fördert auch in dieser Darstellung ein Kontextmenü zutage, mit dem sich bei Bedarf weitere Ebenen öffnen oder Zentrierungen realisieren lassen. Ein Tipp: Ein Klick mit gedrückter Umschalt(Shift)-Taste löst die gleiche Funktion aus und geht schneller.

Um alle Möglichkeiten der Darstellung und Filterung besser abschätzen zu können, sollte man am besten etwas experimentieren. Wesentliche Treiber sind „Text Display Depth“, die Schieberegler „Link Length“, „Node Separation“, „Node Overlap“ und „Node Size“, „Show Non-Indexable“ und „Expand to Level“. Mit diesen Parametern kann man den Graphen praktisch beliebig nach jeder Anforderung gestalten. Die Bedienung ist sehr schlüssig und man sieht Veränderungen sofort live im Diagramm und kann daher abschätzen, ob man das Richtige tut.

Wer es optisch ansprechender mag, kann auch mit den weiteren Darstellungsfarben, vor allem für den Hintergrund (z. B. in dunklem Grau) experi-

mentieren. So lassen sich durchaus vergleichsweise edel wirkende Grafiken erstellen.

Wenn die Anzahl der Knoten über dem Darstellungslimit von derzeit 10.000 liegt, stellt Screaming Frog die weggeklappten Knoten in grauer Farbe dar. Diese lassen sich bei Bedarf per Rechtsklick (und „Explore“) öffnen.

In der üblichen Tabellenansicht des Screaming Frog lassen sich übrigens diese und weitere Visualisierungen (sog. Wordclouds nach eingehenden Linkankertexten sowie nach dem textlichen Inhalt einer Seite) direkt per Rechtsklick auslösen (Abbildung 14). Für die Wordclouds ist es allerdings nötig, beim Crawlvorgang auch das Speichern der HTML-Inhalte mit zu aktivieren (unter „Spider Configuration“, Reiter „Advanced“ und unten „Store HTML“).

Vorteile einer Visualisierung

Gerade in größeren Unternehmen ist es oft schwierig, mit Tabellen und Zahlen zu verdeutlichen, dass Optimierungsbedarf besteht. Hier kann die neue

Funktion des Screaming Frog sehr nützlich sein. Eine optische Darstellung der Domain, die z. B. zeigt, dass der integrierte Blog noch viel zu isoliert dasteht, verstehen auch Nicht-Fachleute bzw. Vorgesetzte. Gegen Änderungen, weil sie Geld kosten, besteht ja nicht selten ein gewisser interner Widerstand. Müssen wir das wirklich machen? Eine Bildkopie einer oder mehrerer Diagramme überzeugt alleine durch die faktische Erkenntnis, dass etwas nicht in Ordnung ist, viel stärker. Budgets bzw. Zeit ist leichter mit Agreement zu bekommen. Diesen oft unterschätzten Punkt sollte man nicht aus dem Auge verlieren!

Weitere nennenswerte Verbesserungen der neuen Version

Das Speicherkonzept wurde in der neuen Version nochmals optimiert und es gelang, den durchschnittlichen Speicherplatzbedarf um etwa 30 % zu reduzieren. Das ist gut, denn wenn ab jetzt durch den automatischen Start deutlich mehr Crawls gemacht werden, fallen natürlich auch sehr viel mehr

Daten an. Prinzipiell ist Speicher zwar nicht mehr teuer, aber die Praxis zeigt, dass man im Bedarfsfall doch meist zu wenig davon hat.

Bei der „Customer Extraktion“, also der individuell konfigurierten Datenerhebung von gecrawlten Seiten, wurde die Größenbeschränkung von bisher 32.000 Zeichen aufgehoben. Somit können jetzt auch umfangreichere Textbausteine gezogen werden. Wer sich viel mit Semantik für die Suchmaschinenoptimierung beschäftigt, wird es sicher begrüßen, dass gerade an dieser Stelle aufgestockt wurde.

Über sog. Regex-Bedingungen kann man sich bekanntlich komplexe Bedingungsfilter bauen. Ein einfaches Beispiel wäre, von allen Seiten eine Preisinformation zu holen, aber nicht von Seiten, die das Wort „nicht auf Lager“ enthalten oder im Pfad /info/ liegen. Der Zusammenbau solcher Regex-Anweisungen kann schnell komplex und vor allem unübersichtlich werden. Ein neuer Test-Tab hilft ab sofort, derartige Formulierungen vor dem Crawl anhand von Beispiel-URLs auf Korrektheit zu prüfen. Das erspart das früher notwendige mehrmalige Starten, weil man die Auswirkungen und mögliche Fehler immer erst danach bemerkte.

Bei der Verwendung des Listenmodus (man übergibt eine URL-Liste und fährt eben keinen Crawl mit einer Startadresse) kann man den im Quelltext der Adressen hinterlegten Canonicals per Konfiguration nun auch automatisch folgen, was die Fehlersuche enorm vereinfacht.

Preise und „Free Version“-Test-Account

Am Preis hat sich nichts geändert. Wie bereits in der letzten Ausgabe erwähnt, lässt sich der SF vorab als Free-Version ausführlich testen. Das Crawl-Limit ist bei der kostenlosen Version auf 500 URLs beschränkt, was für viele kleinere Sites bereits ausrei-



Abb. 13: Sog. Paginierungstunnel oder ULR-„Sackgassen“ lassen sich optisch meist leicht(er) erkennen (Quelle: Portent.com)

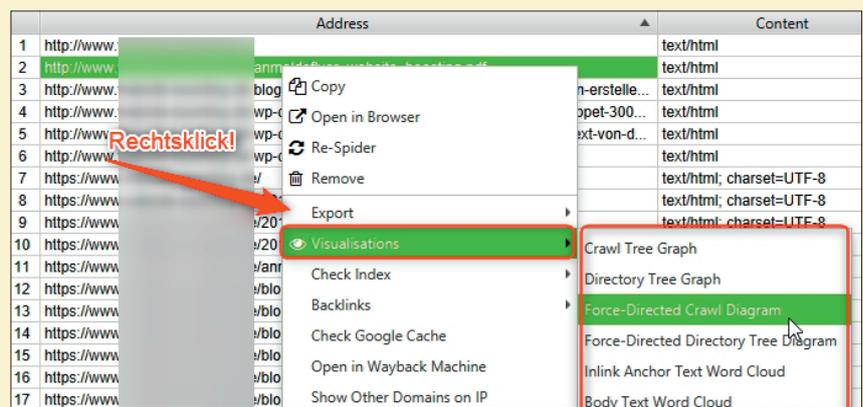


Abb. 14: Grafische Darstellung mit einem einfachen Rechtsklick

chen sollte. Trotzdem sollte man sich die funktionellen Einschränkungen der Free-Version genauer ansehen. Für ein ernsthaftes Arbeiten lohnt sich die Anschaffung der Vollversion gerade ab der Version 10 allemal.

Ab fünf Lizenzen greift eine Mengenstaffel, die den Lizenzpreis jeweils um zehn Pfund pro Lizenz vergünstigt. Ab 20 Lizenzen werden dann 134 € pro Lizenz fällig. Einen Vergleich zwischen Free-Version und der Vollversion hinsichtlich der Einschränkungen findet man auf www.screamingfrog.co.uk/seo-spider/pricing.

Fazit

Gerade nach dem Versions- bzw. Funktionssprung lohnt es sich noch mehr, den Einsatz des Screaming Frog SEO Spider für die SEO-Arbeit ernsthaft zu prüfen. Die Verwendungsmög-

lichkeiten wurden deutlich erweitert und dürften sich eigentlich bei jedem Unternehmen rechnen, das ernsthaft auf der Suche nach Potenzial beim Ranking ist. Falls Sie nur diese Ausgabe in den Händen halten, lohnt sich ggf. noch der Blick in die letzte Ausgabe #52, in der wir den Frog mit seinen Basis- und erweiterten Funktionen nebst Anwendungsmöglichkeiten und -tips umfassend beschrieben. ¶

Weitere Infos unter: www.screamingfrog.co.uk.

