

CRAWLS DURCHFÜHREN UND ANALYSIEREN

ALLGEMEINE EINSTELLUNGEN

Rendering/HTML (1)

- > Der Frog kann Seiten auch rendern
(HTML + JS + Bilder + Fonts ...)
- > Ähnlich wie „Abruf wie durch Googlebot“ (Search Console)
- > Aktivieren:
 - > Configuration > Spider > Rendering
 - > Rendering: „Text only“ (aus) oder „JavaScript“ (ein)
 - > Device und Größe auswählen (Desktop/Smartphone)
- > Warnung: Rendern braucht mehr Zeit, Speicher und CPU

Rendering/HTML (1)

- > Option „Enable Rendered Page Screen Shots“
 - > Legt einen Screenshot von jeder Seite an
 - > Exportierbar über „Bulk Export > Screenshots“

Rendering/HTML (1)

The screenshot shows the Screaming Frog SEO Spider tool interface. At the top, there is a table with two rows of data:

1	https://www.zdf.de/	text/html; charset=UTF-8	200	OK	Indexab
2	https://www.zdf.de/static/0.45.3800/css/styles.css	text/css	200	OK	Indexab

Below the table, there is a large red arrow pointing from the top-left towards the right side of the interface. On the right side, there is a preview window titled "Zoom: 50%" showing a dark-themed ZDF news page about Angela Merkel. The page title is "Politik | maybrit illner". The main headline reads "Streit um Merkels Erbe – die CDU sucht Kurs". Below the headline, there is a snippet of text: "Angela Merkel macht den Weg frei: Im Dezember für eine neue CDU-Spitze und später für einen neuen Regierungschef ...". At the bottom of the preview, it says "65 min | 01.11.2018".

At the bottom of the interface, there is a navigation bar with several tabs: "URL Info", "Inlinks", "Outlinks", "Image Info", "Resources", "SERP Snippet", "Rendered Page", and "View Source". The "Rendered Page" tab is highlighted with a green background and white text.

Rendering/HTML (1)

- > HTML-Code (vor/nach Rendern) kann gespeichert werden
 - > Configuration > Spider > Advanced
 - > „Store HTML“ (vor dem Rendern)
 - > „Store Rendered HTML“ (nach dem Rendern)
- > Auch hier: Braucht Speicher!

Rendering/HTML (1)

The screenshot shows the Screaming Frog SEO Spider tool interface. At the top, there is a table with two rows of data:

1 https://www.zdf.de/	text/html; charset=UTF-8	200	OK	Indexable
2 https://www.zdf.de/static/0.45.3800/css/styles.css	text/css	200	OK	Indexable

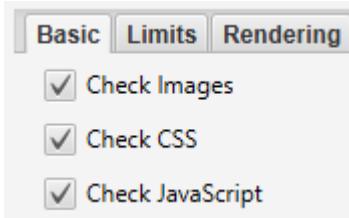
Below the table are two code editors side-by-side:

- Original HTML:** Shows the raw HTML code with various meta tags and conditional comments for IE compatibility.
- Rendered HTML:** Shows the final rendered HTML code after the browser has applied styles and scripts.

At the bottom of the interface, there is a navigation bar with several tabs: URL Info, Inlinks, Outlinks, Image Info, Resources, SERP Snippet, Rendered Page, and View Source. The 'View Source' tab is highlighted with a red arrow pointing to it.

Rendering/HTML (1)

- > Eng damit verbunden:
Sollen auch Bilder, CSS und JS-Dateien geladen werden?
- > Rendern funktioniert nur, wenn alles angeklickt wird!
- > Configuration > Spider:



Speicher (2)

- > Früher: hakelig
- > Jetzt: bequem einstellbar
 - > Configuration > System > Memory (RAM-Speicher)
 - > Configuration > System > Storage (Speicher oder Festplatte)
- > Gerade für große Crawls (>10k Seiten): Festplatte
 - > SSD, kein Netzwerk-Speicher

Speed (3)

- > Configuration > Speed
- > Hilfreich bei langsamem Servern
- > Einstellungen:
 - > Threads (gleichzeitige Verbindungen)
 - > URLs/sec

Authentication (4)

- > Configuration > Authentication
- > Braucht man, wenn die Website per Passwort geschützt ist
(vor allem bei Relaunch-Projekten)
- > Zwei Methoden:
 - > „Standards Based“ (HTTP-Login)
 - > „Forms Based“ (Formulare – EXTREM gefährlich!)

HTTP-Header (5)

- > Configuration > HTTP Header
- > Braucht man faktisch nie
- > Möglichkeiten:
 - > „Accept-Language“ =
Wie reagiert die Website auf Nutzer mit Sprache „Englisch“
 - > ...

User-Agent (6)

- > Configuration > User-Agent
- > Als was soll sich der Frog dem Server gegenüber ausgeben?
 - > Verschiedene Googlebots
 - > Bingbot
 - > ...
- > Braucht man faktisch kaum

Limits (7)

- > Configuration > Spider > Limits
- > Viele Beschränkungen für den Crawler:
 - > Limit Crawl Total (URLs insgesamt)
 - > Limit Crawl Depth (Klicktiefe)
 - > Limit Max Folder Depth (Ordnertiefe in URL)
 - > Limit Number of Query Strings (Anzahl URL-Parameter)
 - > Max Redirects to Follow (wie vielen Umleitungen folgen)

DER CRAWL KANN STARTEN...

So einfach ist ein Crawl

- > Mode = Spider auswählen
- > Oben die Start-URL eingeben
- > Auf „Start“ drücken
- > Was passiert?

Crawl von blog.bloofusion.de

The screenshot shows the homepage of the Bloofusion website. At the top, there's a navigation bar with links for BLOG, BERATUNG, VERANSTALTUNGEN, INFOS, BLOOFUSION, and KONTAKT. Below the navigation is a header with the Bloofusion logo and a search bar. The main content area features a section titled "Das Online-Marketing-Blog von Bloofusion" with a post about a Google Analytics seminar. A large, dark blue callout box is overlaid on the page, containing the text: "Ein Link, den der Frog noch nicht besucht hat – also ab in die Pipeline". At the bottom left is the Bloofusion footer, and at the bottom right is a "Screaming Frog SEO Spider > [17]" watermark.

BLOOFUSION
die suchmaschinen-experten

BLOG BERATUNG VERANSTALTUNGEN INFOS BLOOFUSION KONTAKT

Startseite Suche

Das Online-Marketing-Blog von Bloofusion

[Endlich Durchblick: Google Analytics-Seminar bei OMCampus](#)

Endlich Durchblick: Google Analytics-Seminar bei OMCampus

12. November 2018 | Von [Helen Ramm](#) in [Google Analytics](#), [OMCampus / BLOO:CON](#) | Keine Kommentare

Was passiert da eigentlich wenn Traffic kommt? Diese und viele andere Fragen werden im Seminar „Google Analytics für Marketing“ beantwortet. Wenn du weißt, woher der Traffic kommt, dann kann du besser auf ihn reagieren – denn dann ist das Marketing zielführend.

Ein Link, den der Frog noch nicht besucht hat – also ab in die Pipeline

<https://blog.bloofusion.de/google-analytics-seminar/>

Newsletter

Ihre E-Mail-Adresse

Anmelden

Unsere Webinare

15. November 2018, 16–17 Uhr
Die häufigsten SEO-Fehler von B2B-Unternehmen (und wie man sie vermeidet!)

Also...

- > Unkonfiguriert würde der Frog erstmal allen Links folgen
- > Das muss ja nicht schlimm sein –
bei großen Websites aber schon:
 - > Server-Last
 - > Frog: Speicherbedarf, Laufzeit
- > Welchen Einfluss hat man darauf,
welche Links sich der Frog anschaut?

INCLUDES/EXCLUDES

Includes und Excludes

- > Configuration > Include
- > Configuration > Exclude
- > Wenn man etwas als Include definiert:
 - > URL wird geladen, wenn mind. eine Include-Regel darauf passt
- > Wenn man etwas als Exclude definiert:
 - > URL wird nicht geladen, wenn mind. eine Exclude-Regel darauf passt

Wichtig: Reguläre Ausdrücke

- > Für Includes/Excludes gelten reguläre Ausdrücke
- > Siehe: Cheat Sheet für reguläre Ausdrücke
- > Der häufigste Fall:
 - > Man will eine URL anhand eines Bestandteils erkennen, z. B. druckversion
 - > Regulärer Ausdruck: .*druckversion.*

Beispiel (1)

- > Ich will von einer mehrsprachigen Website nur den Ordner /de/ herunterladen
- > Beispiel: <https://www.hotelzurpost.de/> hat zwei Ordner /de/ und /nl/

Path	Address
▼ https/	
▼ www.hotelzurpost.de/	
► de/	https://www.hotelzurpost.de/de/
► nl/	https://www.hotelzurpost.de/nl/

Beispiel (1)

- > Wie löse ich das Problem?
 - > Lösung #1:
Include: .*/de/.*
 - > Lösung #2:
Exclude: .*/nl/.*

Beispiel (2)

- > Eine Website hat ganz viele PDF-Dateien verlinkt
- > Also:
 - > Exclude: .*pdf

Das Cheat Sheet

Cheat Sheet:

Reguläre Ausdrücke (für „Includes“/„Excludes“)
Für den Bereich der „Includes“ und „Excludes“ müssen reguläre Ausdrücke angegeben werden, für die dann geprüft, ob diese auf die jeweilige URL „machen“.

Bausteine
Ein regulärer Ausdruck setzt sich aus verschiedenen Bestandteilen zusammen. Das folgende ist dabei nur ein kleiner Ausschnitt der Möglichkeiten:

Was soll matchen?	Wie schreibt man das?
Irgendwas (0 – unendlich viele Zeichen)	,*
Irgendwas (1 – unendlich viele Zeichen)	,
Eine einzelne Zahl	[0-9]
Eine oder mehrere Zahlen	[0-9]+
Ein Buchstabe in Kleinschreibung	[a-z]
Ein Buchstabe in Klein- oder Großschreibung	[a-zA-Z]
Bestimmte Zeichen (vor allem: * + ?)	\^ \? \{ [vor den Zeichen einen Backslash platzieren]

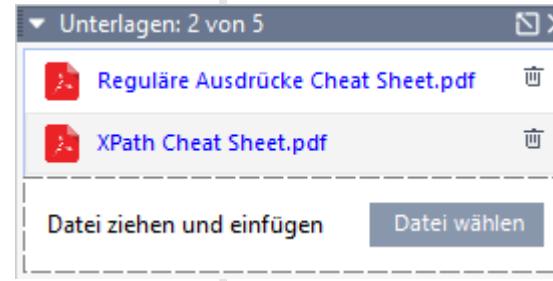
Beispiel
Gesucht:
Ein regulärer Ausdruck für alle URLs, die auf / enden

Zusammenbauen:

- Erst kommt irgendwas (also: *)
- Dann kommt ein Slash (also: /)
- Danach kommt nichts mehr
- Also: */

Weitere Beispiele

Beispiel	Regulärer Ausdruck
Alle URLs, die ein /de/ im Pfad haben	^/de/,*
Alle URLs, die auf .html enden	.*\.html
Alle URLs, die auf .html enden und davor eine Zahlenkombination haben	^([0-9]+)\.html
Alle URLs, die ein .? enthalten	.?\?



SUBDOMAINS UND START FOLDER

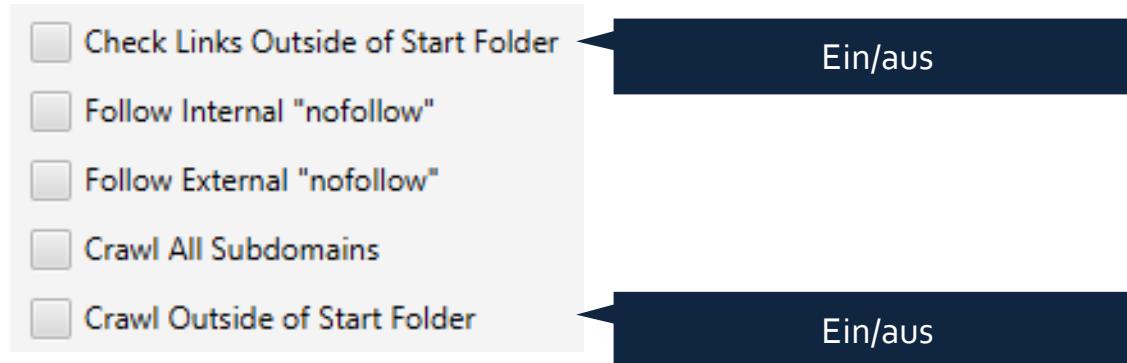
Subdomains

- > Soll der Frog einem Link von www.bloofusion.de auf blog.bloofusion.de folgen?
- > Configuration > Spider > Basic



Start Folder

- > Manche starten den Crawl auf einem Ordner, z. B.
<https://www.hotelzurpost.de/de/>
- > Darf der Frog dann auch /nl/ crawlen?



NOFOLLOW

Noch zwei Besonderheiten

- > <meta name="robots" content="nofollow"> bedeutet:
 - > Der Crawler wird Links auf einer Seite nicht folgen
- > bedeutet:
 - > Der Crawler „traut“ dem Link nicht
 - > Er kann ihm folgen, muss er aber nicht

Optionen

- Follow Internal "nofollow"
- Follow External "nofollow"

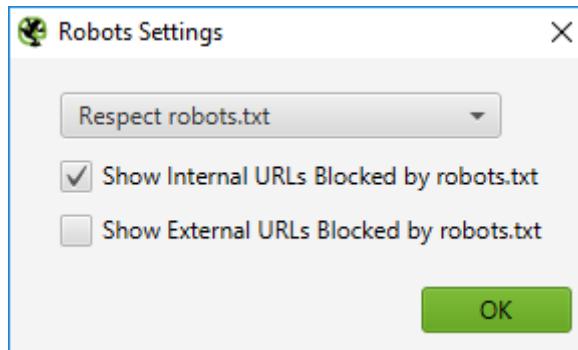
Für interne Links

Für externe Links

ROBOTS.TXT

robots.txt

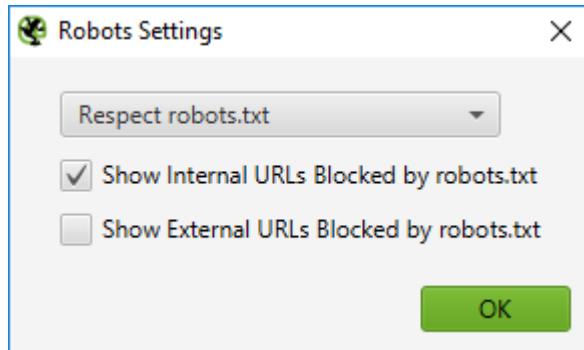
- > Standardmäßig hält sich der Frog an die robots.txt
- > Configuration > robots.txt > Settings



Kann man abschalten

Anzeigen oder nicht?

	Address	Content	Status Code	Status	Indexability	Indexability Status	
130	https://www.schnullireich.de/	text/html; charset=UTF-8	200	OK	Indexable		S
131	https://www.schnullireich.de/checkout/cart		0	Blocked by robots.txt	Non-Indexable	Blocked by robots.txt	



Sollen URLs angezeigt werden?

robots.txt

- > Der Frog lädt grundsätzlich die „reale“ robots.txt
- > Man kann ihm aber auch eine andere vorgeben:
 - > Configuration > robots.txt > Custom

Custom robots.txt

Custom Robots Configuration

Per subdomain robots.txt configuration. You can write your own robots.txt file to be used in a crawl, or download and modify the existing one for the site.

Subdomains

- https://www.schnullireich.de
Override Enabled

```
1 User-agent: *
2
3
4
5 Disallow: /compare
6 Disallow: /checkout
7 Disallow: /register
8 Disallow: /account
9 Disallow: /address
10 Disallow: /note
11 Disallow: /widgets
12 Disallow: /listing
13 Disallow: /ticket
14
15 Disallow: /probierichmalaus
16
17 Allow: /widgets/emotion
18
19
20 Sitemap: https://www.schnullireich.de/sitemap.xml
21
22
23
24 Sitemap: https://www.schnullireich.de/sitemapMobile.xml
25
26
27
```

CANONICAL UND NOINDEX

Welche Sichtweise möchte ich haben? (1)

- > Was macht der Crawl, was passiert im Index?
- > Canonical Tag:
 - > 100 Dubletten, alle mit demselben Canonical Tag
 - > Crawler lädt 100 Seiten
 - > Im Index erscheint 1 Seite

Welche Sichtweise möchte ich haben? (2)

- > Meta-Robots-Tag „noindex“:
 - > Seiten werden vom Crawler heruntergeladen
 - > Erscheinen dann nicht im Index
- > In beiden Fällen:
 - > Der Crawler kann seeeeeeeehr beschäftigt sein

Zwei Optionen

> Configuration > Spider > Advanced

Respect noindex

Aus = Noindex-Seiten erscheinen im Crawl
Ein = Noindex-Seiten erscheinen nicht

Respect Canonical

Aus = Dubletten erscheinen im Crawl
Ein = Dubletten erscheinen nicht

An oder aus?

> Respect Canonical = Aus

Address	Indexability Status	Indexability
https://www.schnullireich.de/kindergeschenke-mit-namen/109/kindermuetze-mit-namen-fleece		Indexable
https://www.schnullireich.de/kindergeschenke-mit-namen/109/kindermuetze-mit-namen-fleece?c=11	Canonicalised	Non-Indexable

> Respect Noindex = Aus

https://www.schnullireich.de/krabbelchuhe-mit-namen/?p=3	noindex
---	---------

PAGINIERUNG (REL-PREV-NEXT)

Rel-prev-next

- > Manche Website nutzen Paginierung
 - > /kategorie/
 - > /kategorie/?page=2
 - > /kategorie/?page=3
 - > ...
- > rel-prev-next verzahnt diese Komponentenseite miteinander

Beispiel:

- > In /kategorie/:
 - > <link rel="next" href="/kategorie/?page=2"
- > In /kategorie/?page=2:
 - > <link rel="prev" href="/kategorie/"
 - > <link rel="next" href="/kategorie/?page=3"
- > Effekt:
 - > Google zeigt in den Suchergebnissen nur die erste Komponentenseite an (in diesem Fall: /kategorie/)

Überprüfen

- > Möchten Sie rel-prev-next überprüfen?
 - > Configuration > Spider > Basics



Crawl Next/Prev

Dran halten?

- > Soll im Ergebnis nur die erste Komponentenseite erscheinen?
 - > Configuration > Spider > Advanced



Respect Next/Prev

UMLEITUNGEN (REDIRECTS)

Redirects

- > Was passiert, wenn der Frog auf eine URL stößt, die auf etwas umleitet?
- > Configuration > Spider > Advanced



Beispiel: Ein

	Address	Indexability Status	Indexability	Content	Status Code	Status
1	https://www.bloofusion.de/scripts/redirect-n.cgi?n=1	Redirected	Non-Indexable	text/html; charset=iso-8859-1	302	Found
2	https://www.bloofusion.de/scripts/redirect-n.cgi?n=2	Redirected	Non-Indexable	text/html; charset=iso-8859-1	302	Found
3	https://www.bloofusion.de/scripts/redirect-n.cgi?n=3	Redirected	Non-Indexable	text/html; charset=iso-8859-1	302	Found
4	https://www.bloofusion.de/scripts/redirect-n.cgi?n=4	Redirected	Non-Indexable	text/html; charset=iso-8859-1	302	Found
5	https://www.bloofusion.de/scripts/redirect-n.cgi?n=5	Redirected	Non-Indexable	text/html; charset=iso-8859-1	302	Found
6	https://www.bloofusion.de/scripts/redirect-n.cgi?n=6	Client Error	Non-Indexable		404	Too Many Redirects

Beispiel: Aus

	Address	Indexability Status	Indexability	Content	Status Code	Status
1	https://www.bloofusion.de/scripts/redirect-n.cgi?n=1	Redirected	Non-Indexable	text/html; charset=iso-8859-1	302	Found

Wie viele Redirects?

- > Configuration > Spider > Limits:
 - > Max Redirects to Follow
- > Standardwert: 5
- > Passt zum Verhalten von Google

UND NOCH EIN PAAR OPTIONEN

hreflang

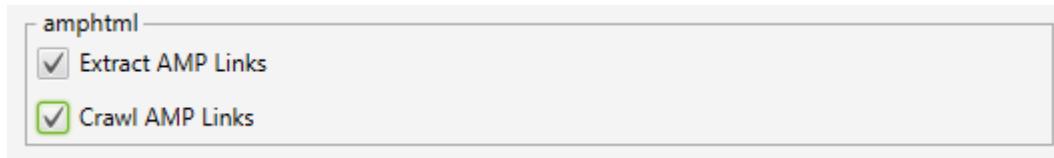
- > Möchten Sie hreflang-Tags überprüfen?
 - > Configuration > Spider > Basic



AMP

> Haben Sie AMP?

> Configuration > Spider > Basics



ERGEBNIS-ANALYSE

Viele Reports

Internal External Protocol Response Codes URI Page Titles Meta Description Meta Keywords H1 H2 Images Canonicals Pagination Directives Hreflang AJAX AMP Sitemaps Custom Analytics Search Console Link Metrics

Später

Abkürzung/Überblick

The screenshot shows the Screaming Frog SEO Spider software interface. The main window displays a list of crawl errors under the 'All' filter. A red arrow points from the 'All' filter dropdown to the error list. Another red arrow points from the 'Hreflang' tab in the top navigation bar to the 'Hreflang' category in the error list. A third red arrow points from the 'Missing' link in the error list to the 'Missing' category in the detailed error list.

Filter: All

Hreflang

- All (0) (0.00%)
- Non-200 hreflang URLs (0) (0.00%)
- Unlinked hreflang URLs (Crawl Analysis Required) (0) (0.00%)
- Missing Confirmation Links (0) (0.00%)
- Inconsistent Language Confirmation Links (0) (0.00%)
- Non Canonical Confirmation Links (0) (0.00%)
- Noindex Confirmation Links (0) (0.00%)
- Incorrect Language Codes (0) (0.00%)
- Multiple Entries (0) (0.00%)
- Missing Self Reference (0) (0.00%)
- Not Using Canonical (0) (0.00%)
- Missing X-Default (0) (0.00%)
- Missing (0) (0.00%)

AJAX

- All (0) (0.00%)
- With hash fragment (0) (0.00%)
- Without hash fragment (0) (0.00%)

AMP

- All (0) (0.00%)
- Non-200 Response (Crawl Analysis Required) (0) (0.00%)
- Non-Confirming Canonical (0) (0.00%)
- Missing Non-AMP Canonical (0) (0.00%)
- Non-Indexable Canonical (0) (0.00%)
- Indexable (0) (0.00%)

No Data

ERGEBNIS-ANALYSE (1): INTERNAL/EXTERNAL

Internal/External

> Liste aller internen/externen URLs:

	Internal	External	Prot
Filter:	All		1
	✓ All		
1	h	c.de/	
2	h	c.de/	
3	h	c.de/	
4	h	c.de/	
5	h	d.bre	
6	h	c.de/	
7	h	d.bre	
8	h	d.bre	
9	h	d.bre	
10	h	d.bre	
11	h	d.bre	
12	h	c.de/	
13	https://cms.bnstc.de/		

ERGEBNIS-ANALYSE (2): PROTOKOLL

HTTP oder HTTPS?

> Überblick über URLs beider Protokolle:

Screaming Frog SEO Spider						
	Address	Content	Status Code	Status	Indexability	Indexability Status
1	http://blog.bloofusion.de/	text/html; charset=iso-8859-1	301	Moved Permanently	Non-Indexable	Redirected
2	http://twitter.com/CoraRothenba		301	Moved Permanently	Non-Indexable	Redirected

ERGEBNIS-ANALYSE (3): RESPONSE CODES

HTTP-Codes

> Übersicht aller Codes bzw. der jeweiligen URLs:

The screenshot shows the 'Response Codes' tab of the Screaming Frog SEO Spider tool. The filter dropdown is set to 'Client Error (4xx)'. The results table lists three URLs, each with its status code (403 or 404), the URL itself, and a brief description of the error type. The descriptions include 'Blocked by Robots.txt', 'Blocked Resource', and 'No Response'.

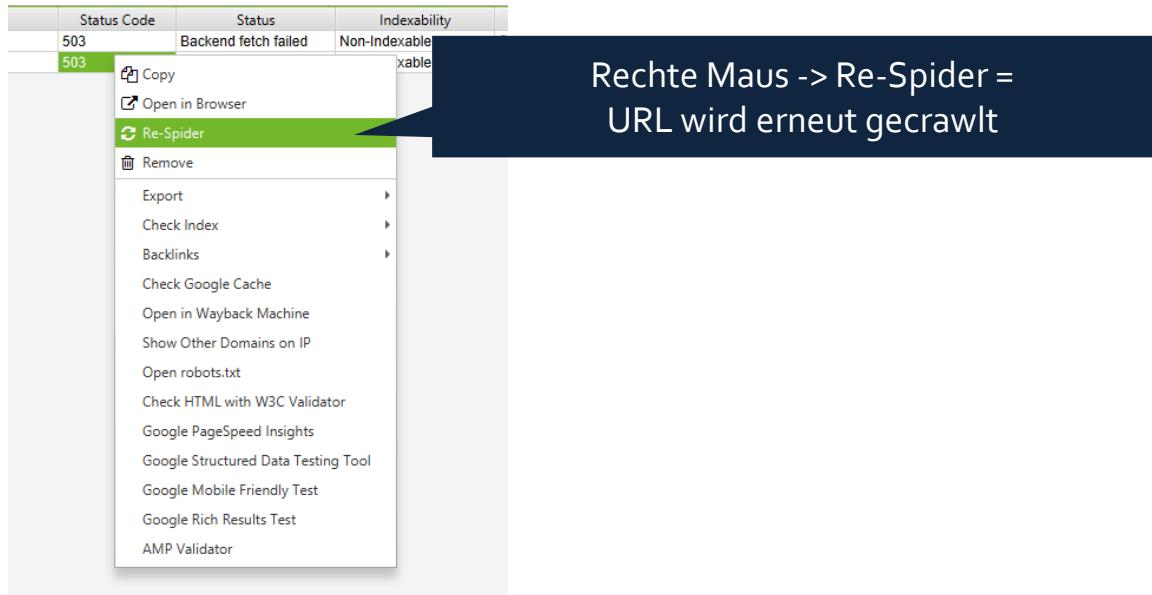
Code	URL	Description
403	http://www.einsteiger-teil-3-keywords.de/hy-cam.-the-gif-camera/	Blocked by Robots.txt
404	http://www.einsteiger-teil-3-keywords.de/hy-cam.-the-gif-camera/	Blocked Resource
404	http://www.einsteiger-teil-3-keywords.de/hy-cam.-the-gif-camera/	No Response

Erklärung

The screenshot shows the Screaming Frog SEO Spider software interface. At the top, there is a navigation bar with tabs: Internal, External, Protocol, Response Codes, URI, and Page. Below the navigation bar is a filter dropdown labeled "Filter: Client Error (4xx)" with a dropdown menu open. The dropdown menu contains the following options: All, Blocked by Robots.txt, Blocked Resource, No Response, Success (2xx), Redirection (3xx), Redirection (JavaScript), Redirection (Meta Refresh), Client Error (4xx) (which is checked with a checked icon), and Server Error (5xx). To the right of the filter dropdown, there is a button labeled "Export". In the main content area, a dark blue callout box highlights the "Blocked by Robots.txt" option from the dropdown menu. The text inside the callout box reads: "Bilder, JS-Dateien, Fonts ... , die per robots.txt gesperrt sind".

Tipp

- > Wenn es mal einen temporären Fehler gibt:



ERGEBNIS-ANALYSE (4): URI

Liste aller URLs

	Internal	External	Protocol	Response Codes	URI	Page T
Filter:	All					
1	h	Non ASCII Characters			-content/uploads/2018/10/3D-suchi	
2	h	Underscores			-content/plugins/bbspoiler/inc/bbsp	
3	h	Uppercase			-includes/js/jquery/jquery-migrate.n	
4	h	Duplicate			-content/plugins/wp-code-highlightj	
5	h	Parameters			-includes/js/jquery/jquery.js?ver=1.	
6	h	Over 115 Characters			-content/themes/bloofusion-blog/jst	
7	h				oren/julia-leutloff/	
8	h				oren/ricarda-overbeck/	
9	h					
10	https://blog.bloofusion.de/autoren/kathrin-schroepfer/					
11	https://bloq.bloofusion.de/wp-content/plugins/wp-code-highlightj					

Was machen wir damit?

- > „Non ASCII Characters“:
 - > Z. B. Umlaute
 - > Ideal: Vermeiden (also z. B. Umlaute auflösen)
 - > Google kommt damit trotzdem oft gut zurecht

Was machen wir damit?

- > „Underscores“:
 - > Wörter werden per Unterstrich getrennt (z. B. seite_a.html)
 - > Besser wäre ein Bindestrich (also seite-a.html)
 - > Google kommt auch mit dem Unterstrich zurecht

Was machen wir damit?

- > „Uppercase“:
 - > Großbuchstaben in der URL (URL ist kontextsensitiv)
 - > Nur dann suboptimal, wenn Groß- und Kleinschreibung gemischt genutzt werden (Dubletten)
 - > Standard-Empfehlung: konsequent Kleinschreibung

Was machen wir damit?

> „Duplicate“:

- > Frog berechnet für jede Seite einen Hash-Wert
- > Angezeigt werden URLs mit identischen Hash-Werten
- > Also: Exakte Dublicetten (HTML-Code 1:1)

Hash
3ceff5a66167a79945073209b659a0f0

Was machen wir damit?

- > „Parameters“:
 - > In der URL kommen Parameter vor (also ...?v1=w1&v2=w)
 - > Muss nicht schlimm sein
 - > Manchmal eine Quelle von Dubletten (a.html und a.html?c=5)
 - > URL-Parameter kritisch prüfen

Was machen wir damit?

- > „Over 115 characters“:
 - > Recht willkürliche Grenze
 - > URLs sollten möglichst konzis sein
(keine irrelevanten Bestandteile enthalten)

Filter:	Over 115 Characters	Export
	Address	
1	https://blog.bloofusion.de/2018/03/14/google-tag-manager-tutorial-iv-adwords-remarketing-und-conversion-tag-einbinden/	
2	https://blog.bloofusion.de/2016/12/15/6-fragen-an-gregor-hilden-warum-und-wie-er-videos-fuer-seine-webseite-nutzt-ein-kleiner-vorgesmack-auf-die-blocon-2017/	
3	https://blog.bloofusion.de/webinar-mitschnitt-15-minuten-praxis-webinar-bessere-e-mails-fuer-die-kontaktaufnahme-bei-content-marketing-online-pr-und-linkaufbau/...	
4	https://blog.bloofusion.de/unser-webinar-am-freitag-die-interne-verlinkung-optimieren-mehr-seo-boost-fuer-die-wichtigen-seiten/amp/	
5	https://blog.bloofusion.de/webinar-mitschnitt-google-analytics-e-commerce-berichte-nutzen-und-richtig-auswerten/amp/	
6	https://blog.bloofusion.de/2017/08/15/search-camp-episode-32-das-grosse-backlink-faq-20-fragen-rund-um-linkaufbaubacklinks/	

ERGEBNIS-ANALYSE (5): SEITENTITEL

Seitentitel

		Internal	External	Protocol	Response Codes	URI	Page Titles	Me
		Filter: All						
		<input checked="" type="checkbox"/> All						
1	h	Missing						
2	h	Duplicate						
3	h	Over 65 Characters						
4	h	Below 30 Characters						
5	h	Over 568 Pixels						
6	h	Below 200 Pixels						
7	h	Same as H1						
8	h	Multiple						
9	https://bloa.bloofusion.de/categories/seo/							
10	https://bloa.bloofusion.de/categories/seo/							
11	https://bloa.bloofusion.de/categories/seo/							
12	https://bloa.bloofusion.de/categories/seo/							
13	https://bloa.bloofusion.de/categories/seo/							

Grundregeln

- > Wichtig für Rankings + Klicks =
Was sieht der Nutzer als erstes im Suchergebnis?
- > Zeichen: 50-60 (inkl. Satz- und Leerzeichen)
- > Jeder Seitentitel sollte unique sein
- > Der Seitentitel sollte zum Inhalt passen
- > Keine Auflistung von Suchbegriffen

Was machen wir damit?

- > „Missing“:
 - > Seitentitel fehlt = darf einfach nicht passieren!

Was machen wir damit?

- > „Duplicate“:
 - > Nicht unique Seitentitel
 - > Gefahr, dass Google sie nicht übernimmt

Was machen wir damit?

- > „Over 65 characters“ bzw. „over 568 pixels“:
 - > Zu lang, wird in der Darstellung abgeschnitten
 - > Sollte vermieden werden
 - > Also: Seitentitel einkürzen

Was machen wir damit?

- > „Below 30 characters“ bzw. „below 200 pixels“:
 - > Zu kurz = Raum nicht optimal genutzt
 - > Z. B. „Startseite“ oder „Herren“
 - > Also: Sinnvoll erweitern

Was machen wir damit?

- > „Same as h1“:
 - > Passiert ohnehin selten
(weil an Seitentitel oft ein Suffix angehängt wird)
 - > Nicht schlimm, aber:
h1 darf gerne mehr Prosa enthalten (kein Zeichenlimit)

Was machen wir damit?

- > „Multiple“:
 - > Im HTML-Code gibt es mehrere Seitentitel
 - > Darf nicht passieren

ERGEBNIS-ANALYSE (6): META DESCRIPTION

Analog: Meta Description

> Wichtig für das Snippet

		Address
1	Missing	utoren/julia-leutloff/
2	Duplicate	utoren/ricarda-overbeck/
3	Over 155 Characters	utoren/kathrin-schroepfer/
4		category/suchradar/
5		category/omcampus-bloofusion/
6	Below 70 Characters	utoren/lisa-heidemann/
7	Over 930 Pixels	utoren/andreas-engelhardt/
8	Below 400 Pixels	category/search-camp/
9		category/content-marketing/
10	Multiple	
11	https://blog.bloofusion.de/autoren/martin-roettgerding/	
12	https://blog.bloofusion.de/category/anentur/	

Grundregeln

- > Nicht wichtig für Rankings,
wichtig für Klick-Rate =
Was sieht der Nutzer als zweites im Suchergebnis?
- > Zeichen: 150-160 (inkl. Satz- und Leerzeichen)
- > Jede Meta Description sollte unique sein
- > Zusammenfassung der Seite, Argumente für Kauf/Klick
- > Abgestimmt auf Seitentitel (gleiche Suchbegriffe)

Was machen wir damit?

- > „Missing“:
 - > Meta Description fehlt
 - > Google holt sich dann einen Textausschnitt aus dem Content (muss nicht schlecht sein)
 - > Empfehlung: Alle relevanten Seiten haben eine Meta Description

Was machen wir damit?

- > „Duplicate“:
 - > Nicht unique Meta Descriptions
 - > Gefahr, dass Google sie nicht übernimmt

Was machen wir damit?

- > „Over 155 characters“ bzw. „over 930 pixels“:
 - > Zu lang, wird in der Darstellung abgeschnitten
 - > Sollte vermieden werden
 - > Also: Meta Descriptions einkürzen

Was machen wir damit?

- > „Below 70 characters“ bzw. „below 400 pixels“:
 - > Zu kurz = Raum nicht optimal genutzt (nur eine Zeile!)
 - > Also: Sinnvoll erweitern

Was machen wir damit?

- > „Multiple“:
 - > Im HTML-Code gibt es mehrere Meta Descriptions
 - > Darf nicht passieren

ERGEBNIS-ANALYSE (7): META KEYWORDS

Meta Keywords

- > Für SEO nicht mehr relevant
- > Empfehlung: ignorieren!

ERGEBNIS-ANALYSE (8): H1

h1

		Internal	External	Protocol	Response Codes	URI	Page Titles	Meta Description	Meta Keywords	H1	H2
		Filter: All									
		<input type="button" value="Export"/>									
		<input checked="" type="checkbox"/> All									
1	h	Missing								Das	
2	h	Duplicate								Julia	
3	h	Over 70 Characters								Rica	
4	h	Multiple								Kath	
5	h									Beitr	
6	h									Beitr	
7		https://blog.bloofusion.de/autoren/lisa-heidemann/								Lisa	

Grundregeln

- > Jede relevante Seite sollte genau eine <h1> haben
- > Die h1 sollte zum Seitentitel passen (mehr Prosa)
- > Bestätigung für den Nutzer: Du bist hier richtig
 - > Vorteile kommunizieren
 - > Erwartungshaltung aufbauen
 - > ...

Was machen wir damit?

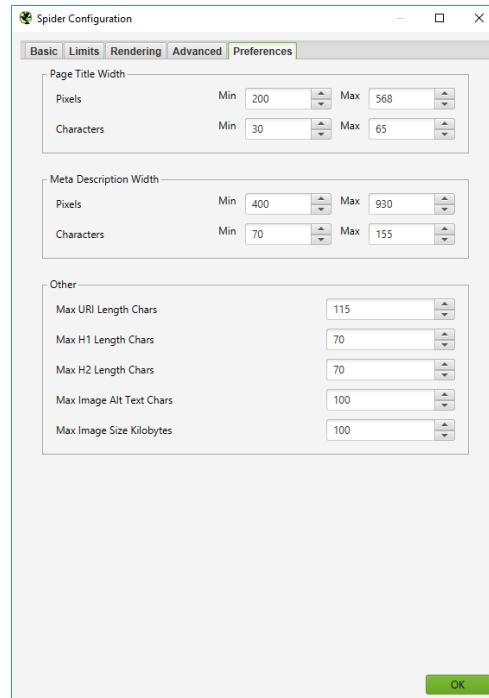
- > „Missing“:
 - > H1 fehlt
 - > Für Google nicht sonderlich relevant
 - > Empfehlung: Es sollte immer genau eine h1 geben

Was machen wir damit?

- > „Duplicate“:
 - > Es gibt mehrere Seiten mit derselben h1
 - > Faktisch nicht relevant

Was machen wir damit?

- > „Over 70 characters“:
 - > Willkürliche Grenze (anpassbar)
 - > Faktisch nicht relevant



Was machen wir damit?

- > „Multiple“:
 - > Es gibt Seiten, die mehrere h1 haben
 - > Nicht unbedingt schlimm, sollte aber vermieden werden

ERGEBNIS-ANALYSE (9): H2

H2

- > Analog: Tab „H2“
- > Grundregeln:
 - > H2..h6 sollten für Zwischenüberschriften genutzt werden
 - > Nicht für Navigationselemente nutzen
- > Frog zeigt max. zwei h2 an (h2-1 und h2-2)
- > Nicht der wichtigste Report...

ERGEBNIS-ANALYSE (10): IMAGES

Bilder auf der Seite

- > Wichtig in Bezug auf:
 - > Bildersuche
 - > Ladezeit

		Internal	External	Protocol	Response Codes	URI	Page Titles	Meta Description	Meta Keywords	H1	H2	Images	Canc
		Filter: All											
		Address											
1	h	Over 100 KB				it/uploads/2018/10/3D-suchradar-74.png							
2	h	Missing Alt Text				it/themes/bloofusion-blog/images/logo-bloofusion.svg							
3	h	Alt Text Over 100 Characters				it/uploads/2018/11/Michaela-Linhart-1024x913.jpg							
4	h					it/uploads/2018/11/Christoph-Raue-1024x884.jpg							
5	h					Jd4c2d559bf95af371196dda4093ad3d?s=80&d=mm&r=g							
6		https://bloo.bloofusion.de/wp-content/uploads/2018/11/Christian-Lorenz-1024x697.jpg											

Grundregeln

- > Sprechende URL (spanien.jpg anstelle von DSCN9922.jpg)
- > Alt-Attribut (kompakte Beschreibung des Bilders)
- > Möglichst hohe Auflösung, möglichst geringe Größe

Was machen wir damit?

- > „Over 100 KB“:
 - > Willkürliche Grenze, aber nicht ganz schlecht gewählt
 - > Prüfen: Muss das Bild so groß sein? (hilfreich für Ladezeit)

Was machen wir damit?

- > „Missing alt text“:
 - > Alt-Attribut fehlt
 - > Alle relevanten Bilder auf relevanten Seiten sollten ein individuelles Alt-Attribut haben

Was machen wir damit?

- > „Alt text over 100 characters“:
 - > Willkürliche Grenze, aber nicht ganz schlecht gewählt
 - > Alt-Text sollte das Bild beschreiben, aber in max. 10 Wörtern
 - > Empfehlung: einkürzen

ERGEBNIS-ANALYSE (11): CANONICALS

Canonical Tag

> Sehr wichtiger Teil (Crawling/Indexierung)

	Internal	External	Protocol	Response Codes	URI	Page Titles	Meta Description	Meta Keywords	H1	H2	Images	Canonicals	P
	Filter: All												
	Address Occurrences Indexability												
1	h	Contains Canonical											
2	h	Self Referencing											
3	h	Canonicalised											
4	h	Missing											
5	h	Multiple											
6	h	Non-Indexable Canonical											
10	https://www.euroncar.de/FRF/module/render/Airline-Partners/RoyalAirMarocSafarFlyer												1

Grundregeln

- > Seiten mit gleichem Canonical Tag:
 - > Ein Stellvertreter erscheint in den Suchergebnissen
(die URL aus dem Canonical Tag)
 - > Signale werden zusammengezogen
- > Canonical Tag beseitigt die Negativeffekte von internem Duplicate Content
- > Gut, aber der Crawler muss die Dubletten immer noch crawlern

Was machen wir damit?

- > „Contains Canonical“, „Missing“:
 - > Seite enthält oder enthält nicht Canonical Tag
 - > Keine Aussage, ob das Canonical Tag sinnvoll ist
- > Alle HTML-Seiten sollten ein Canonical Tag haben
- > Bei Nicht-HTML-Seiten möglich, aber schwierig

Was machen wir damit?

- > „Self-Referencing“:
 - > Positivfall = Canonical-Tag stimmt mit URL überein
 - > Theoretisch könnte man das Canonical Tag dann weglassen
(ist aber absolut nicht notwendig)

Was machen wir damit?

> „Canonicalised“:

- > Der interessante Fall: Gecrawlte URL ≠ Canonical Tag
- > An sich nicht schlimm = dafür haben wir ja das Canonical Tag
- > Bei hoher Anzahl: Kann man das auch anders realisieren
(also z. B. auf URL-Parameter verzichten)?

Filter:	Canonicalised	Address	Occurrences	Indexability	Indexability Status	Canonical Link Element 1
13		https://www.europcar.de/autovermietung-angebote/auto-mieten-in-spanien?fromHP=true&...	1	Non-Indexable	Canonicalised	https://www.europcar.de/autovermietung-angebote/auto-mieten-in-span...
14		https://www.europcar.de/angebote/transporter-stundenmiete-guenstig-mieten?fromHP=tru...	1	Non-Indexable	Canonicalised	https://www.europcar.de/angebote/transporter-stundenmiete-guenstig...
15		https://www.europcar.de/business/services/schlüsselautomat	1	Non-Indexable	Canonicalised	https://www.europcar.de/cms/render/live/de_DE/sites/Europcar.de/temp...
16		https://www.europcar.de/mietwagen-notfall-panne?noredirection=true	1	Non-Indexable	Canonicalised	https://www.europcar.de/mietwagen-notfall-panne
17		https://www.europcar.de/unfallersatz?noredirection=true	1	Non-Indexable	Canonicalised	https://www.europcar.de/unfallersatz
18		https://www.europcar.de/versicherungsschutz-schulzpakete-details/frankreich-deutschland...	1	Non-Indexable	Canonicalised	http://www.europcar.de/angebote/mietwagen-weltweit

Was machen wir damit?

- > „Multiple“:
 - > Seite enthält mehrere Canonical Tags
 - > Darf es nicht geben!

Was machen wir damit?

> „Non-Indexable Canonical“:

- > Seite zeigt per Canonical Tag auf Seite, die nicht indexierbar ist
(primär: gesperrt per robots.txt, noindex)
- > Darf es nicht geben!

Address	Occurrences	Indexability	Indexability Status	Canonical Link Element 1
1. https://www.europcar.de/business/travel	1	Non-indexable	Canonicalised	https://www.europcar.de/cms/render/live/de_DE/sites/Europcar.de/templates/simple.html
2. https://www.europcar.de/business/firmenkunden	1	Non-indexable	Canonicalised	https://www.europcar.de/cms/render/live/de_DE/sites/Europcar.de/templates/simple.html
3. https://www.europcar.de/business/greenworldraten	1	Non-indexable	Canonicalised	https://www.europcar.de/cms/render/live/de_DE/sites/Europcar.de/templates/simple.html
4. https://www.europcar.de/business/services/schlusselfaktoren	1	Non-indexable	Canonicalised	https://www.europcar.de/cms/render/live/de_DE/sites/Europcar.de/templates/simple.html

Name	Value
Address	https://www.europcar.de/business/firmenkunden
URL Encoded Address	https://www.europcar.de/business/firmenkunden
Content	text/html;charset=UTF-8
Status Code	200
Status	OK
Indexability	Non-indexable
Indexability Status	Canonicalised
Size	156378
Title 1	Firmenkunden Business Car Rental Europcar
Title 1 Length	45
Meta Description 1	Discover all Europcar business solutions tailor made for your needs. We have special offers for SME, tour operators, travel agents and corporate accounts.
Meta Description 1 Length	156
Meta Description 1 Width	900
H2-1	FIRMENTHEMEN
H2-1 Length	12
H2-2	Kleinunternehmen und Mittelstand
H2-2 Length	35
Canonical Link Element 1	https://www.europcar.de/cms/render/live/de_DE/sites/Europcar.de/templates/simple.html
Canonical Link Element 1 Indexability	Non-indexable
Canonical Link Status 1 Indexability Status	Blocked by robots.txt

Ein wichtiger Hinweis

- > Eine Seite, die ein Canonical Tag hat,
sollte kein „noindex“ haben
- > Dann einfach das Canonical Tag weglassen!

ERGEBNIS-ANALYSE (12): PAGINATION

Pagination

> Pagination = rel-prev-next

The screenshot shows the Screaming Frog SEO Spider interface. The top navigation bar includes tabs for Internal, External, Protocol, Response Codes, URI, Page Titles, Meta Description, Meta Keywords, H1, H2, Images, Canonicals, Pagination, and Di. A filter dropdown is set to "Contains Pagination". An "Export" button is visible. The main content area displays a table with three columns: Address, Indexability, and Indexability Status. There are two rows of data:

Address	Indexability	Indexability Status
thon-produkte-und-events/	Indexable	https:
thon-produkte-und-events/	Indexable	https:

To the left of the table, a sidebar lists various pagination-related filters:

- All
- Contains Pagination
- First Page
- Paginated 2+ Pages
- Pagination URL Not In Anchor Tag
- Non-200 Pagination URLs
- Unlinked Pagination URLs
- Non-Indexable
- Multiple Pagination URLs
- Pagination Loop
- Sequence Error

Grundregeln

- > Wenn eine Seite einen Vorgänger hat:
`<link rel="prev" href="URL" />` (im `<head>!`)
- > Wenn eine Seite einen Nachfolger hat:
`<link rel="next" href="URL" />` (im `<head>!`)
- > Komponentenseiten sollten intern verlinkt werden
- > Indexierbar (HTTP 200, nicht noindex, nicht robots.txt)

ERGEBNIS-ANALYSE (13): DIRECTIVES

Direktiven

> Direktiven können wichtig sein – vor allem noindex!

Noindex

- > Noindex = Seite erscheint nicht im Index
- > Verbraucht aber trotzdem Crawl-Budget
- > Also prüfen:
 - > Gibt es sehr viele Noindex-Seiten?
 - > Kann man die auch per robots.txt sperren?

Zweimal Vorsicht!

- > „noindex, follow“ wird zu „noindex, nofollow“
- > „none“ = „noindex, nofollow“

ERGEBNIS-ANALYSE (14): HREFLANG

hreflang

> Wichtig für das korrekte Ausliefern internationaler Inhalte

Category	Page URL	Title	Occurrence
All		Title 1	19
Non-200 hreflang URLs		Autovermietung - Auto, Transporter und LKW weltweit mieten Europcar	19
Unlinked hreflang URLs		Times Car Rental & Europcar Partnerschaft	1
Missing Confirmation Links		Firmenkunden Business Car Rental Europcar	1
Inconsistent Language Confirmation Links		Replacement Business Car Rental Europcar	1
Non Canonical Confirmation Links		Advantage & Europcar Partnerschaft	1
Noindex Confirmation Links		Business Services Business Car Rental Europcar	1
Incorrect Language Codes		Important Information	1
Multiple Entries		Zubehör für Mietwagen mieten bei Europcar	1
Missing Self Reference		Selection Mietwagen Europcar Autovermietung	1
Not Using Canonical		Mietwagen ab 18 Jahren Europcar Autovermietung	1
Missing X-Default		Autovermietung Mailand	15
Missing		Werbung bei Europcar Europcar Autovermietung	1
		Übersicht Business Car Rental Europcar	1
		Europcar mobil - Jetzt App downloaden und Mietwagen buchen	1
		Extras Europcar Autovermietung	1
		Privileg Loyaltätsprogramm Europcar Autovermietung	1
		Kontakt per E-Mail Europcar Autovermietung	1
		Discount & Europcar Partnerschaft	1

Beispiel

```
<link rel="alternate" hreflang="en-NZ" href="https://www.europcar.co.nz/" />
<link rel="alternate" hreflang="en-GB" href="https://www.europcar.co.uk/" />
<link rel="alternate" hreflang="fr-FR" href="https://www.europcar.fr/" />
<link rel="alternate" hreflang="pt-PT" href="https://www.europcar.pt/" />
<link rel="alternate" hreflang="it-IT" href="https://www.europcar.it/" />
<link rel="alternate" hreflang="es-ES" href="https://www.europcar.es/" />
<link rel="alternate" hreflang="en-AU" href="https://www.europcar.com.au/" />
<link rel="alternate" hreflang="en-BE" href="https://www.europcar.be/en" />
<link rel="alternate" hreflang="nl-BE" href="https://www.europcar.be/nl" />
<link rel="alternate" hreflang="fr-BE" href="https://www.europcar.be/fr" />
<link rel="alternate" hreflang="de-AT" href="https://www.europcar.at/de" />
<link rel="alternate" hreflang="en-AT" href="https://www.europcar.at/en" />
<link rel="alternate" hreflang="no-NO" href="https://www.europcar.no/" />
<link rel="alternate" hreflang="fr-CH" href="https://www.europcar.ch/fr" />
<link rel="alternate" hreflang="de-CH" href="https://www.europcar.ch/de" />
<link rel="alternate" hreflang="it-CH" href="https://www.europcar.ch/it" />
<link rel="alternate" hreflang="en-CH" href="https://www.europcar.ch/en" />
```

Grundregeln (1)

- > Jede Seite, von der es Sprachversionen gibt, sollte hreflang-Tags haben
- > In allen Sprachversionen sollten dieselben Tag vorhanden sein („Return Tags“)
- > Genannte URLs:
 - > HTTP-Code 200
 - > Nicht auf noindex, nicht per robots.txt gesperrt
 - > URLs in hreflang-Tags = Canonical Tag

Grundregeln (2)

- > Falls sinnvoll sollte es auch ein x-default-Tag geben
- > URL, wenn keine andere Regel greift

```
<link rel="alternate" hreflang="en-NZ" href="https://www.europcar.co.nz/" />
<link rel="alternate" hreflang="en-GB" href="https://www.europcar.co.uk/" />
<link rel="alternate" hreflang="fr-FR" href="https://www.europcar.fr/" />
<link rel="alternate" hreflang="pt-PT" href="https://www.europcar.pt/" />
<link rel="alternate" hreflang="it-IT" href="https://www.europcar.it/" />
<link rel="alternate" hreflang="es-ES" href="https://www.europcar.es/" />
<link rel="alternate" hreflang="en-AU" href="https://www.europcar.com.au/" />
<link rel="alternate" hreflang="en-BE" href="https://www.europcar.be/en" />
<link rel="alternate" hreflang="nl-BE" href="https://www.europcar.be/nl" />
<link rel="alternate" hreflang="fr-BE" href="https://www.europcar.be/fr" />
<link rel="alternate" hreflang="de-AT" href="https://www.europcar.at/de" />
<link rel="alternate" hreflang="en-AT" href="https://www.europcar.at/en" />
<link rel="alternate" hreflang="no-NO" href="https://www.europcar.no/" />
<link rel="alternate" hreflang="fr-CH" href="https://www.europcar.ch/fr" />
<link rel="alternate" hreflang="de-CH" href="https://www.europcar.ch/de" />
<link rel="alternate" hreflang="it-CH" href="https://www.europcar.ch/it" />
<link rel="alternate" hreflang="en-CH" href="https://www.europcar.ch/en" />
```

Diagnose (1)

Filter	Was tun?
Non-200 hreflang URLs	URL liefert einen HTTP-Code ≠ 200
Unlinked hreflang URLs	In hreflang-Tags genannte URLs werden intern nicht verlinkt (sollte nicht passieren!)
Missing confirmation links	Nicht alle Sprachversionen von A verweisen in den hreflang-Tags zurück auf A
Inconsistent Language Confirmation Links	Es gibt unterschiedliche Sprach-Angaben für eine bestimmte Seite

Diagnose (2)

Filter	Was tun?
Non Canonical Confirmation Links	Im hreflangTag einer Sprachversion wird auf eine URL ≠ Canonical Tag verwiesen
Noindex Confirmation Links	Hreflang-Tag verweist auf per noindex gesperrte Seite
Incorrect languages Codes	Sprachangaben passen nicht zu ISO 639-1 und ISO 3166-1 alpha-2 (also z. B. „eu“)

Diagnose (3)

Filter	Was tun?
Multiple Entries	Auf eine Seite wird mit unterschiedlichen Sprachen verwiesen
Missing Self Reference	In Seite A fehlt das hreflang-Tag für A
Not Using Canonicals	URL im hreflang-Tag stimmt nicht mit dem Canonical-Tag der Seite überein
Missing X-Default	Kein x-default vorhanden (nicht so schlimm)
Missing	Keine hreflang-Tags in Seite vorhanden

Grundsätzlich

- > Hreflang-Tags sind immer fehlerträchtig
- > Fehler = dann wird evtl. eine unpassende Seite ausgeliefert
- > Keine Ranking-Nachteile
- > Für das Debugging einzelner Seite:
 - > Hreflang.org (kostenlos, kein Login)

ERGEBNIS-ANALYSE (15): AJAX

Ganz kurz

- > Es geht um das AJAX crawling scheme
- > Ignorieren wir, denn Google sagt:
(<https://developers.google.com/search/docs/ajax-crawling/docs/getting-started>)



The syntax described by this page is no longer required, and will not be supported after the summer of 2018.

ERGEBNIS-ANALYSE (16): AMP

AMP

- > AMP = „eingeschränktes HTML“
- > Separate Seiten/URLs:
 - > <https://blog.bloofusion.de/die-5-groessten-remarketing-fehler/>
 - > <https://blog.bloofusion.de/die-5-groessten-remarketing-fehler/amp/>

Anforderungen

- > HTML-Seite verweist auf AMP-Seite
 - > <link rel="amphtml" href="https://blog.bloofusion.de/die-5-groessten-remarketing-fehler/amp/" />
- > AMP-Seite kanonisiert auf HTML-Seite
 - > <link rel="canonical" href=" https://blog.bloofusion.de/die-5-groessten-remarketing-fehler/" />

Prüfen

ERGEBNIS-ANALYSE (17): KLICKTIEFE

Site Structure

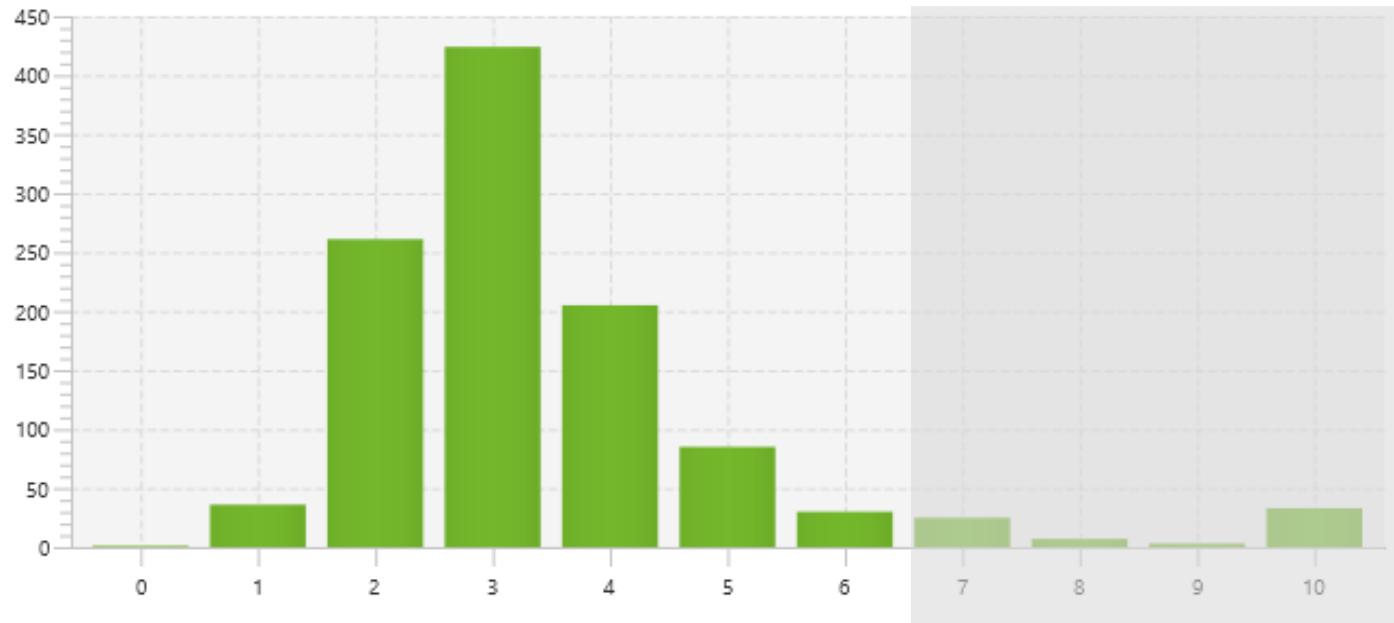
Overview						
Site Structure						
Response Times						
Unique Inlinks (Top 20 UR)						
Pos	URI	Number of URI	% of Total			
1	http://www.ints.at/	1107	99.73			
2	https://www.ints.at/grunden/start/	1041	93.78			
3	https://www.ints.at/events/	1039	93.6			
4	http://www.ints.at/startups/	1039	93.6			
5	https://www.ints.at/netzwerk/2/investoren/	1039	93.6			
6	https://www.ints.at/netzwerk/3/corporates/	1039	93.6			
7	http://www.ints.at/netzwerk/2/gesellschafter/	1038	93.51			
8	http://www.ints.at/netzwerk/2aplus/	1038	93.51			
9	http://www.ints.at/jobs/	1038	93.51			
10	http://www.ints.at/about-2uber-ints/	1037	93.42			
11	http://www.ints.at/grunden unser-angebot/	1037	93.42			
12	http://www.ints.at/netzwerk/2mentoren/	1037	93.42			
13	http://www.ints.at/about-2kontakt/	1037	93.42			
14	http://www.ints.at/news/	1037	93.42			
15	http://www.ints.at/ausichtsrat/	1037	93.42			
16	http://www.ints.at/about-2press/	1037	93.42			
17	http://www.ints.at/fragen-antworten/	1037	93.42			
18	http://www.ints.at/projektberat/	1037	93.42			
19	http://www.ints.at/startup-camp/	1037	93.42			
20	http://www.ints.at/netzwerk/2/	1037	93.42			
Depth (Clicks from Start URL)						
Depth	URI	Number of URI	% of Total			
0		1	0.09			
1		36	3.24			
2		261	23.51			
3		424	38.2			
4		205	18.47			
5		85	7.66			
6		30	2.7			
7		25	2.25			
8		7	0.63			
9		3	0.27			
10		33	2.97			
Crawl Depth						
0	500	400	300	200	100	0
1		~50	~50	~50	~50	~50
2		~250	~250	~250	~250	~250
3		~450	~450	~450	~450	~450
4		~200	~200	~200	~200	~200
5		~50	~50	~50	~50	~50
6		~20	~20	~20	~20	~20
7		~10	~10	~10	~10	~10
8		~5	~5	~5	~5	~5
9		~2	~2	~2	~2	~2
10		~1	~1	~1	~1	~1

Klicktiefe

- > Wie viele Link-Klicks braucht es, um von der Startseite auf eine bestimmte Seite zu gelangen?
- > Hohe Klicktiefe =
Geringe Chance für Crawling/Indexierung
- > Wo liegt die Grenze?

Probleme

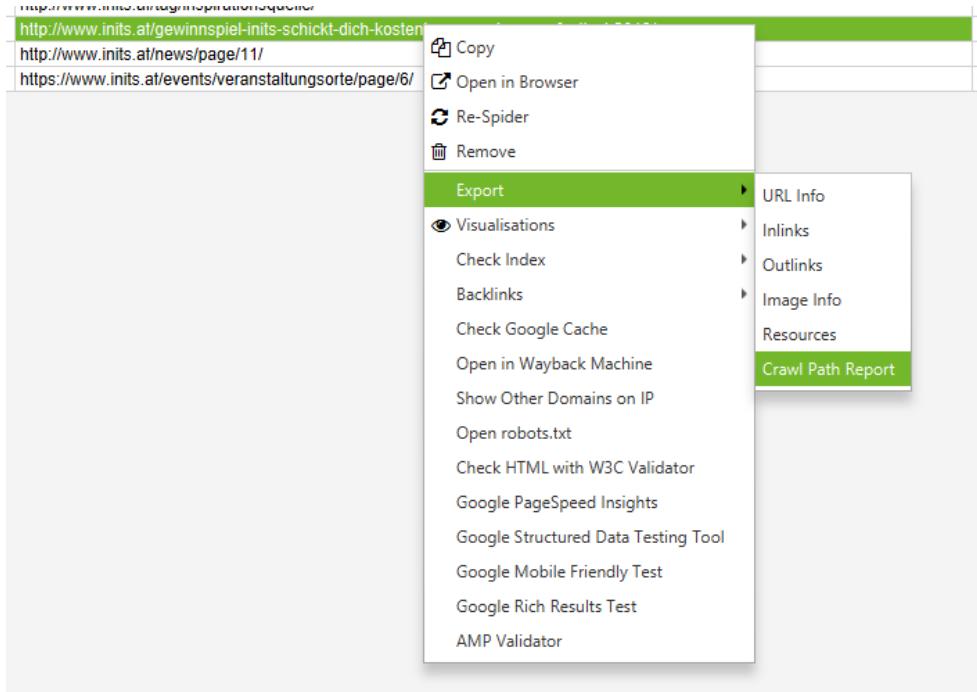
Crawl Depth



Beispiel

- > <https://www.inits.at/>
- > <http://www.inits.at/wiener-start-up-gnowsis-integriert-dropbox-in-collaboration-tool/> hat Klicktiefe 14
- > Problem hier: lineare Paginierung

Tipp: Wie wird die Seite denn verlinkt?



Beispiel

- > http://www.inits.at/ ->
- > http://www.inits.at/jobs/ ->
- > http://www.inits.at/inits-bleibt-weltspitze/ ->
- > http://www.inits.at/allgemein/ ->
- > http://www.inits.at/allgemein/page/2/ -> ..
- > http://www.inits.at/allgemein/page/11/ -> **GEFUNDEN**

Lösungen

- > Unterschiedliche Ansätze:
 - > Mehr Querverlinkungen
(Artikel A verlinkt auf ähnliche Artikel A')
 - > Paginierung anders aufbauen
(Seite N verlinkt direkt auf alle Komponentenseiten)
 - > Zeitliches Archiv
(Archiv -> 2018 -> November -> 15 -> alle Artikel vom 15. Nov 2018)
 - > ...

DAS WAR'S MIT TEIL 2