BLOOFUSION

# SCREAMING FROG FROM ZERO TO HERO

## Der Ablaufplan für den perfekten Crawl

24.11.2023

Markus Hövener

# Heute: die kombinierte Erfahrung aus 1000+ Crawls mit dem Frog

Die Konfiguration

Analyse im Frog

Was fehlt dann noch?

BLOOFUSION

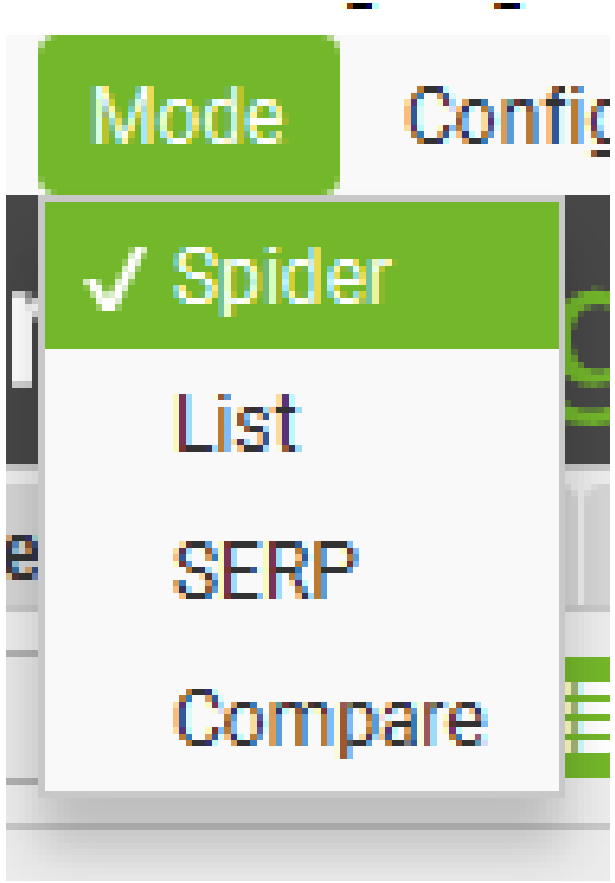
# DIE RICHTIGE KONFIGURATION

## … ist schon die halbe Miete!

# „Was sind die wichtigsten Konfigurationen für den Frog?“

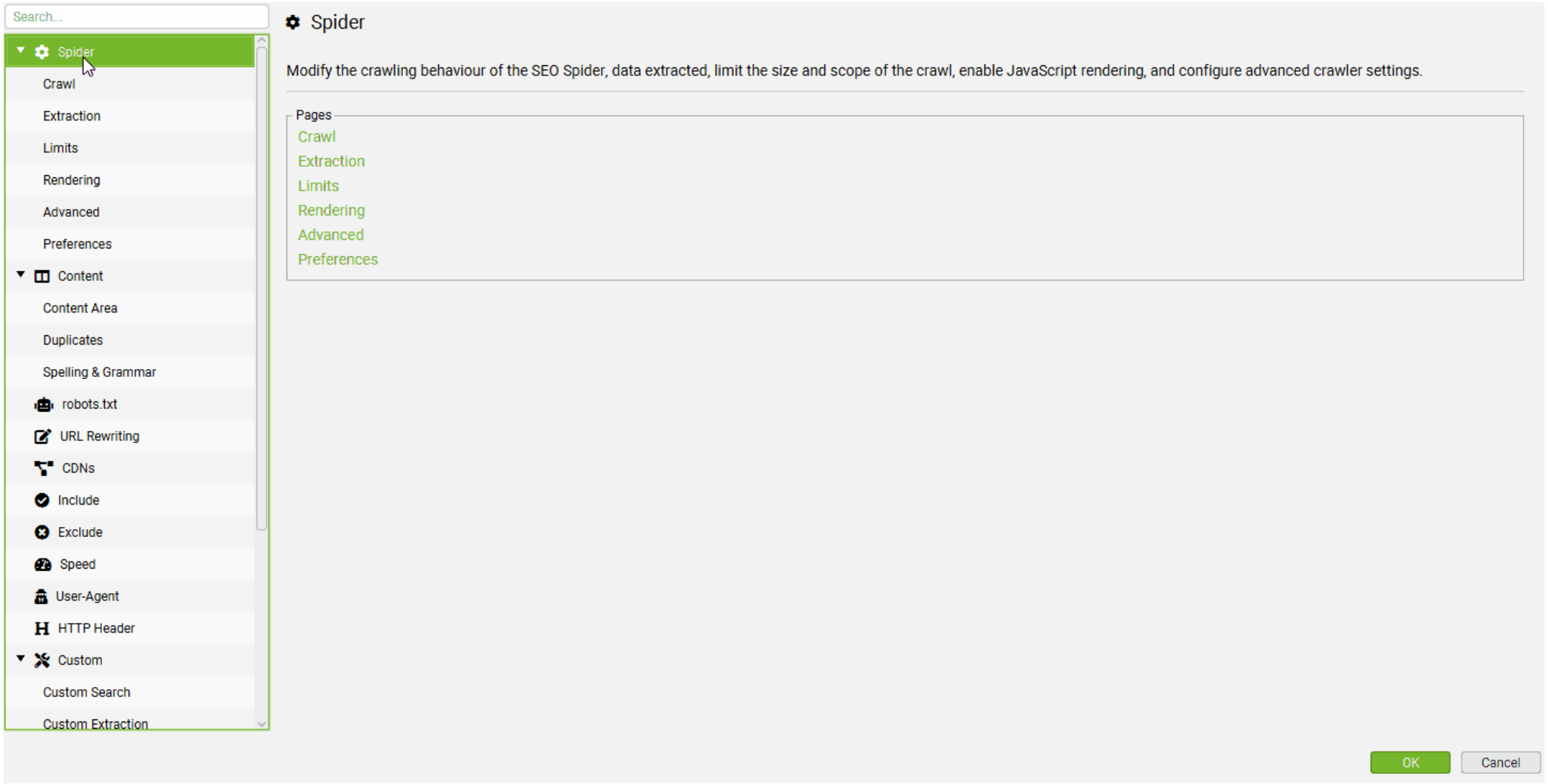
Letztens, in einem Seminar

# DIE WICHTIGSTE OPTION ...



# ALSO …

- Der ganz normale Crawl ohne …
  - … Automatisierung/Scheduling
  - … Dashboards in Looker Studio
  - … Data Analysis (ex Code Interpreter)

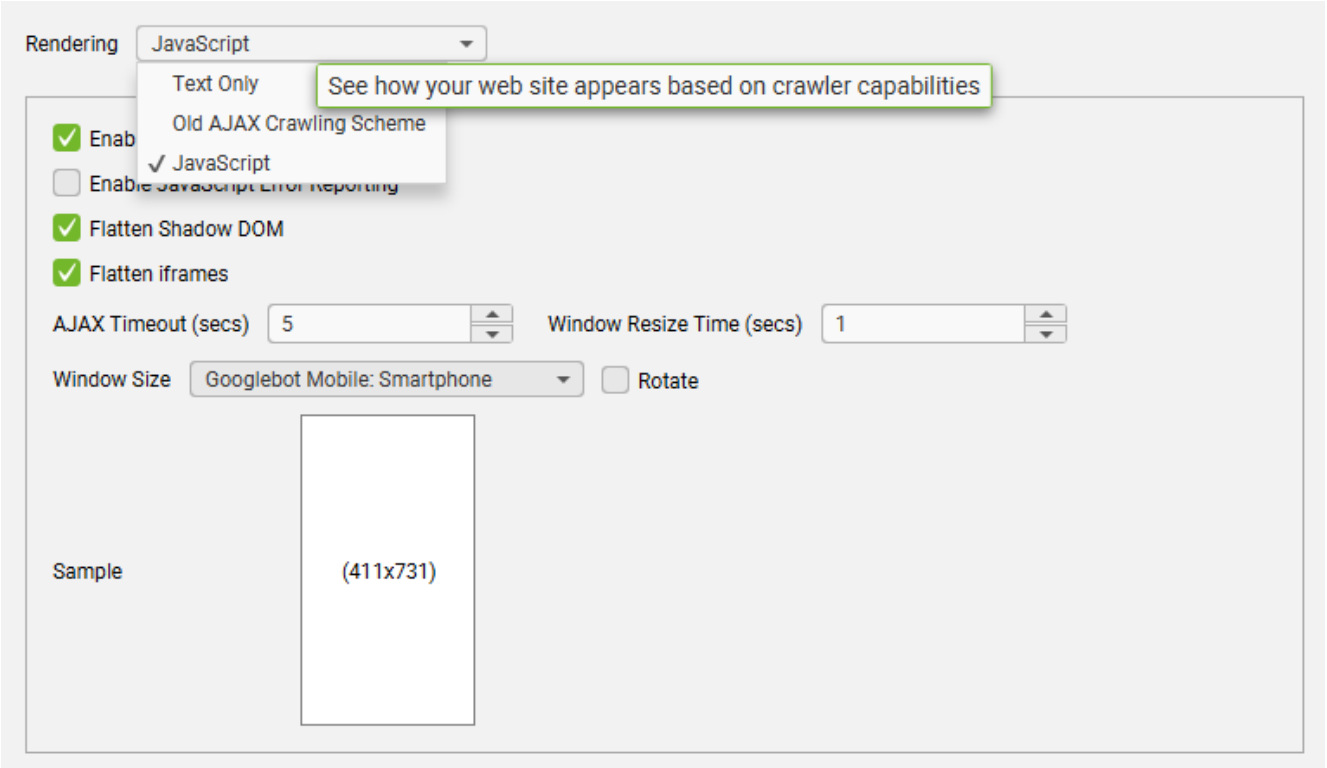

Search...
Spider
Crawl
Extraction
Limits
Rendering
Advanced
Preferences
Content
Content Area
Duplicates
Spelling & Grammar
robots.txt
URL Rewriting
CDNs
Include
Exclude
Speed
User-Agent
HTTP Header
Custom
Custom Search
Custom Extraction
Spider
Modify the crawling behaviour of the SEO Spider, data extracted, limit the size and scope of the crawl, enable JavaScript rendering, and configure advanced crawler settings.
Pages
Crawl
Extraction
Limits
Rendering
Advanced
Preferences
OK
Cancel

# „HÄNGT DAVON AB“

- Der beste Standard:
  - Spider > Crawl > Page Links = alles anwählen
  - Spider > Extraction = alles anwählen
  - Spider > Limits = passt schon
  - Spider > Advanced > „Always follow …“ = anwählen
  - Spider > Preferences = passt schon

# #1: RENDERN ODER NICHT?

- Warum drüber nachdenken?
  - Frisst Zeit
  - Frisst Ressourcen
- Vor allem wichtig für:
  - JavaScript-lastige Websites
  - Optische Kontrolle



# EINFACH AUSPROBIEREN

## – X URLs rendern (aus GSC, SISTRIX & Co.)

| JavaScript | | |
|---|---|---|
| All | 17 | 100% |
| Pages with Blocked Resources | 0 | 0% |
| Contains JavaScript Links | 3 | 17,65% |
| Contains JavaScript Content | 3 | 17,65% |
| Noindex Only in Original HTML | 0 | 0% |
| Nofollow Only in Original HTML | 0 | 0% |
| Canonical Only in Rendered HTML | 0 | 0% |
| Canonical Mismatch | 0 | 0% |
| Page Title Only in Rendered HTML | 0 | 0% |
| Page Title Updated by JavaScript | 0 | 0% |
| Meta Description Only in Rendered HTML | 0 | 0% |
| Meta Description Updated by JavaScript | 0 | 0% |
| H1 Only in Rendered HTML | 3 | 17,65% |
| H1 Updated by JavaScript | 0 | 0% |
| Uses Old AJAX Crawling Scheme URLs | 0 | 0% |
| Uses Old AJAX Crawling Scheme Meta Fragment Tag | 0 | 0% |
| Pages with JavaScript Errors | 0 | 0% |
| Pages with JavaScript Warnings | 0 | 0% |
| Pages with Chrome Issues | 0 | 0% |

Contains JavaScript Content | Export

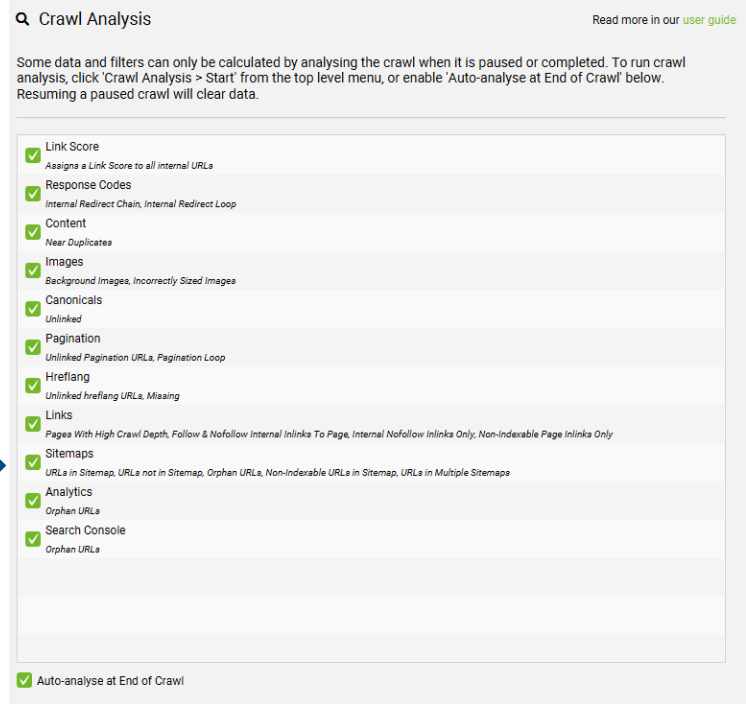
| | Address | HTML Word Count | Rendered HTML Word Count | Word Count Change |
|---|---|---|---|---|
| 1 | https://www.dresdner-essenz.com/de/feste-duschpflege | 5 | 28 | 23 ▲ |
| 2 | https://www.dresdner-essenz.com/de/cremeduschen | 5 | 68 | 63 ▲ |
| 3 | https://www.dresdner-essenz.com/de/duschpeelings | 5 | 32 | 27 ▲ |

| | Address | HTML H1 | Rendered HTML H1 |
|---|---|---|---|
| 1 | https://www.dresdner-essenz.com/de/feste-duschpflege | | Feste Duschen |
| 2 | https://www.dresdner-essenz.com/de/cremeduschen | | Reichhaltige Cre... |
| 3 | https://www.dresdner-essenz.com/de/duschpeelings | | Duschpeelings |

# Vorsicht! Immer erst die Content Areas (siehe #3) definieren!

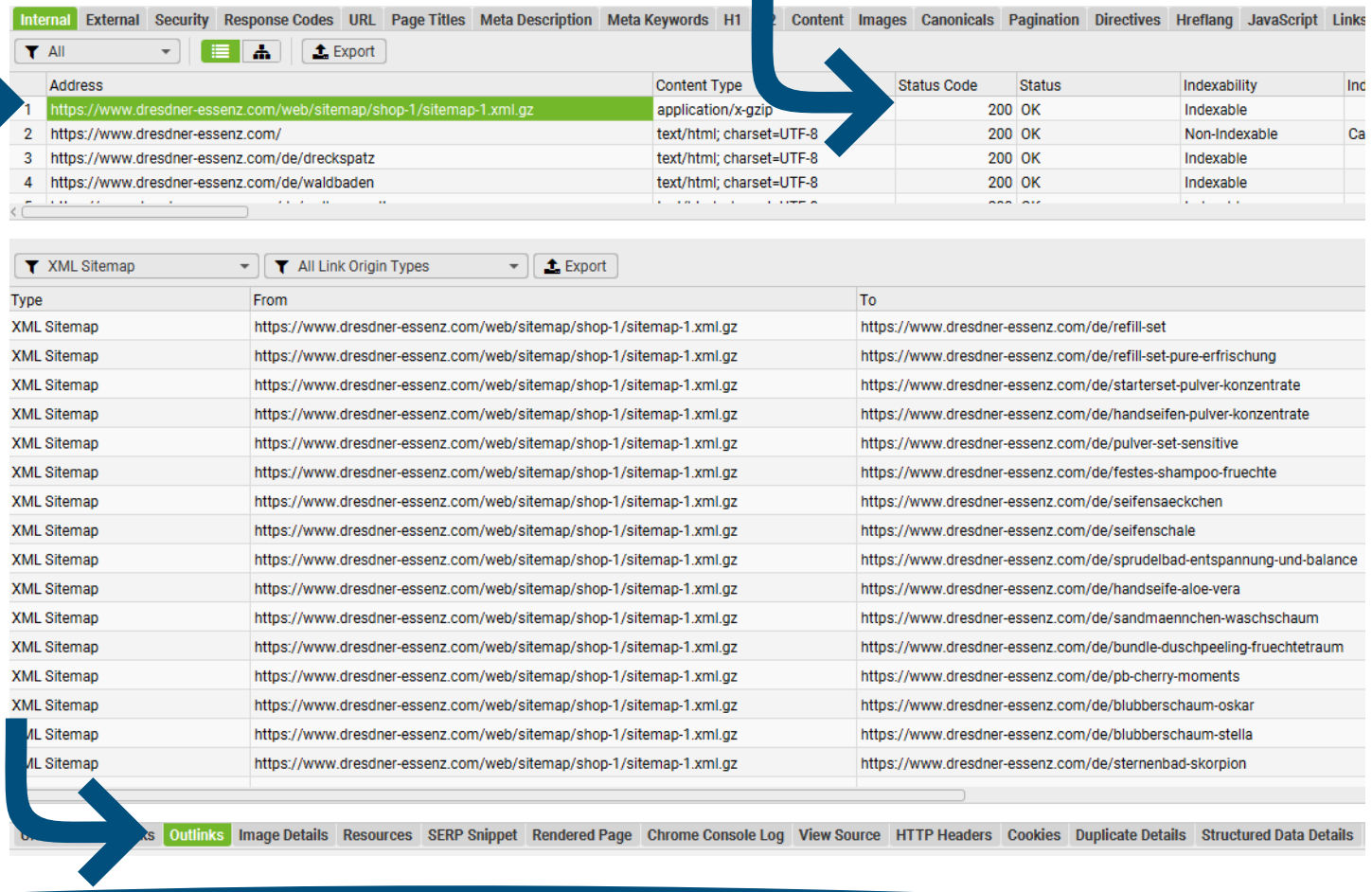
# #2: XML-SITEMAPS

## – Immer hinterlegen → interne Verlinkung prüfen



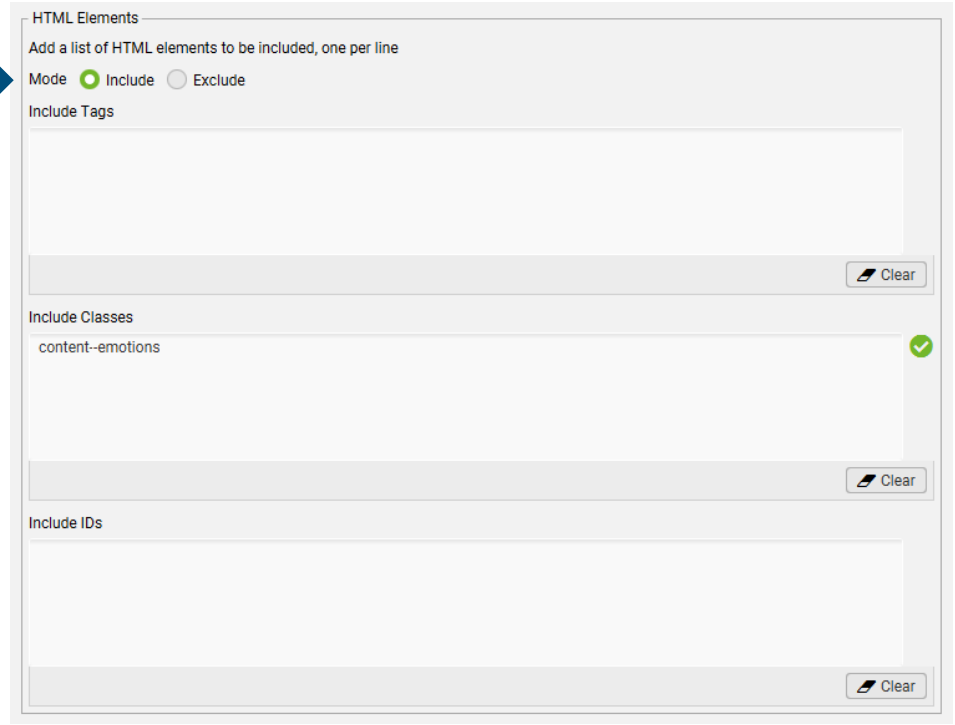
# FEHLER BEI DEN SITEMAPS? HMMM …

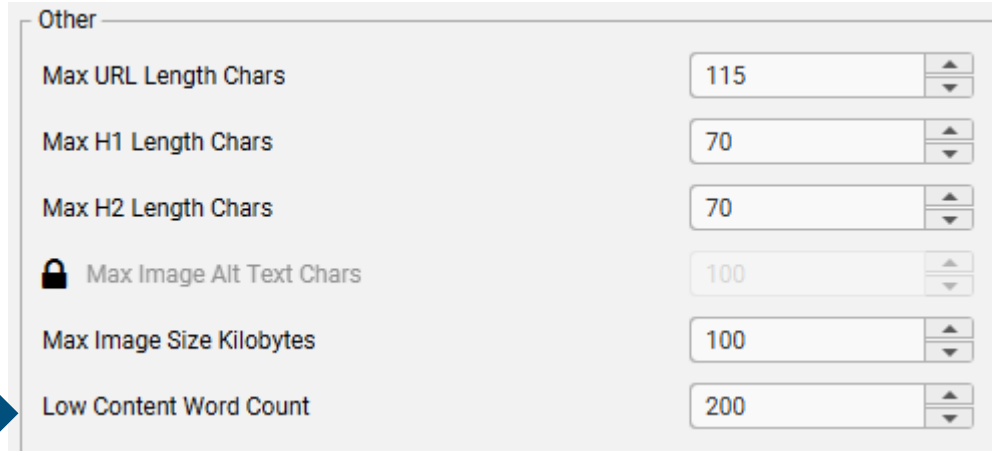
- Wenn was schief läuft, erfahren wir es erst am Ende des Crawls.
- Also:
  - Crawl starten
  - Kurz warten
  - Auf Pause



# #3: CONTENT AREAS

- Wichtig für:
  - Erkennung von internem Duplicate Content
  - Thin Content



# #4: LINK POSITIONS

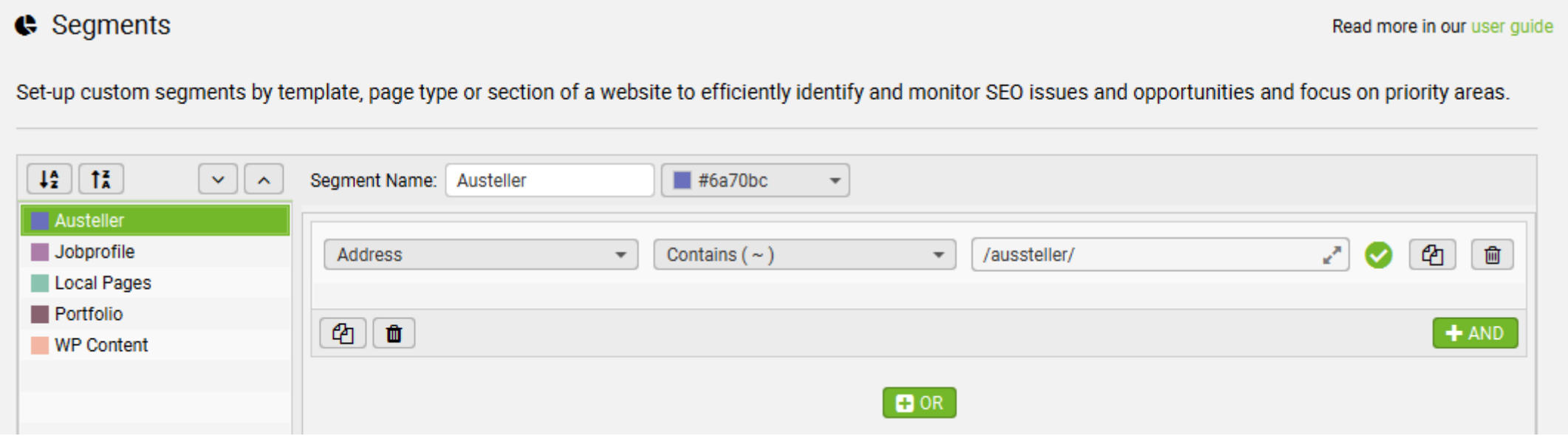
- Kann man brauchen für:
  - Tiefergehende Analysen der internen Verlinkung

| Link Path | Link Position |
|---|---|
| //body/div[@class='page-wrap']/header/div[@class='headnavi-wrap']/div[2]/nav/ul/li[1]/a | Navigation |
| //body/div[@class='page-wrap']/header/div[@class='headnavi-wrap']/div[2]/nav/ul/li[2]/div[@class='hint--bottom-left']/a | Navigation |
| //body/div[@class='page-wrap']/section/div/aside/div[@class='navigation-smartphone']/ul/li/a | Navigation |
| //body/div[@class='page-wrap']/section/div/aside/nav/ul/li[1]/a | Navigation |

# #5: SEGMENTS

– Seit V19

– Optimal für Seitentypen – aber nicht nur …

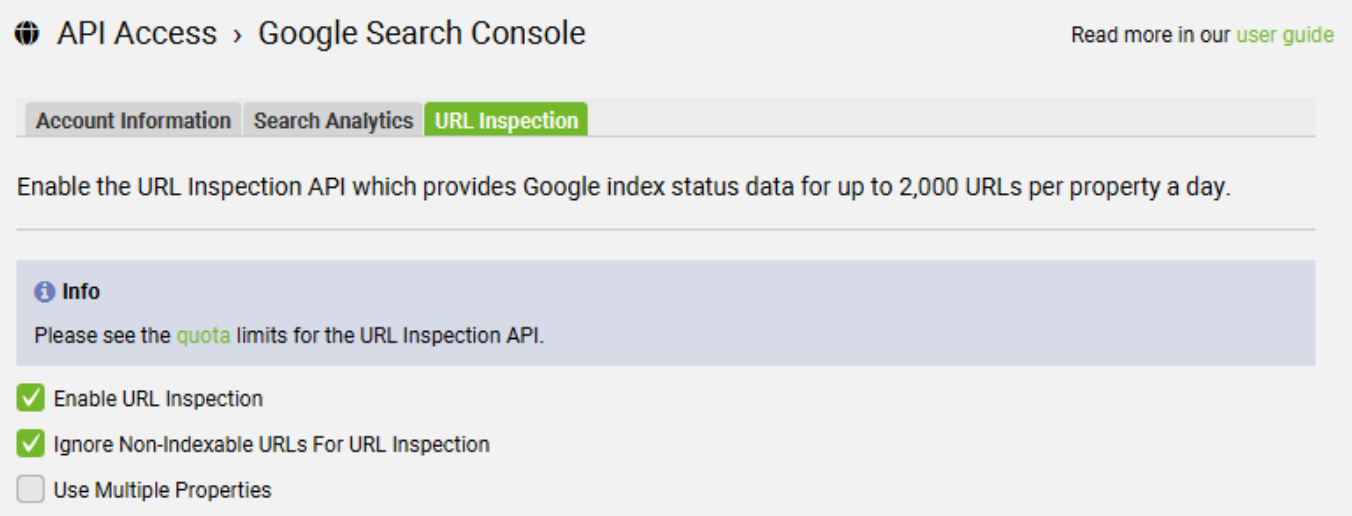


# GOOD TO KNOW …

- Die Segmentierung kann auch nach einem Crawl vorgenommen werden.
- Eine Seite kann in beliebig vielen Segmenten liegen.
- Segmente können auf allen Daten basieren (z. B. auch Custom Extraction).

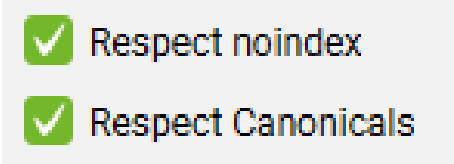
# #6: APIS

- Search Console geht immer ☺
- URL Inspection auch:
  - Max. 2k URLs/Tag/Property
  - Verzögert natürlich



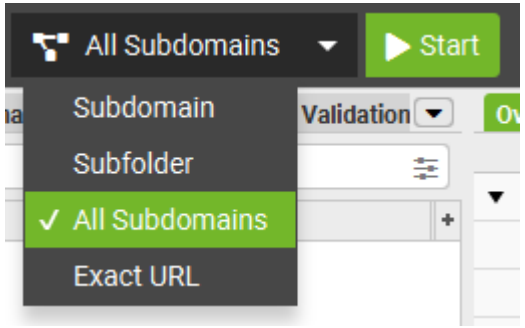
# #7: CRAWLER VS. INDEX-SICHT

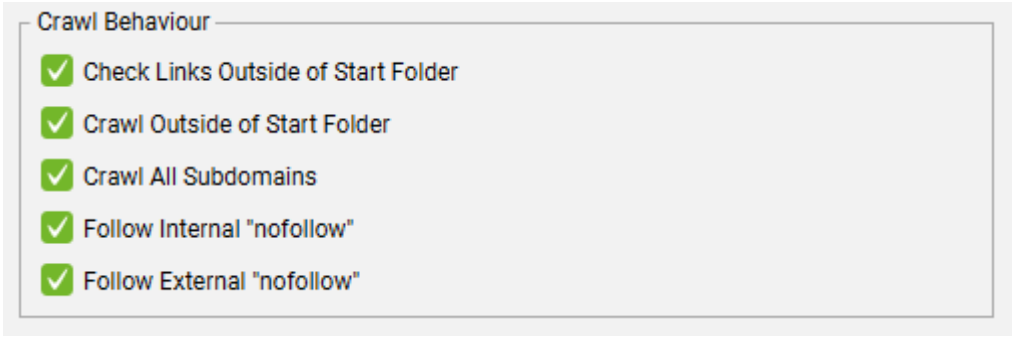
- Was will ich?
  - Z. B. Crawl-Budget optimieren → Crawler-Sicht
  - Z. B. Inhaltliche Optimierung → Index-Sicht

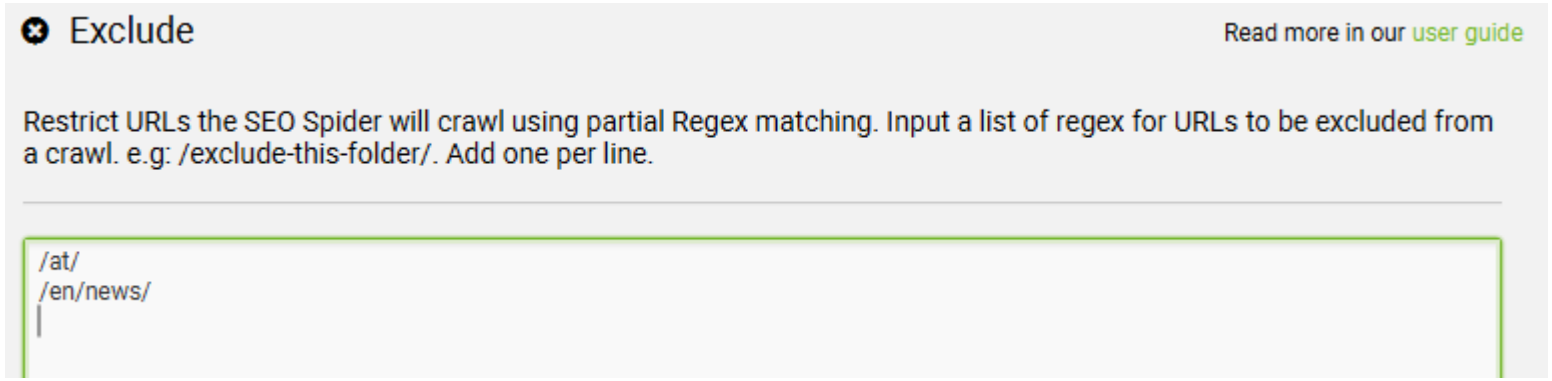


# #8: SCOPE

## – Viel besser #v19

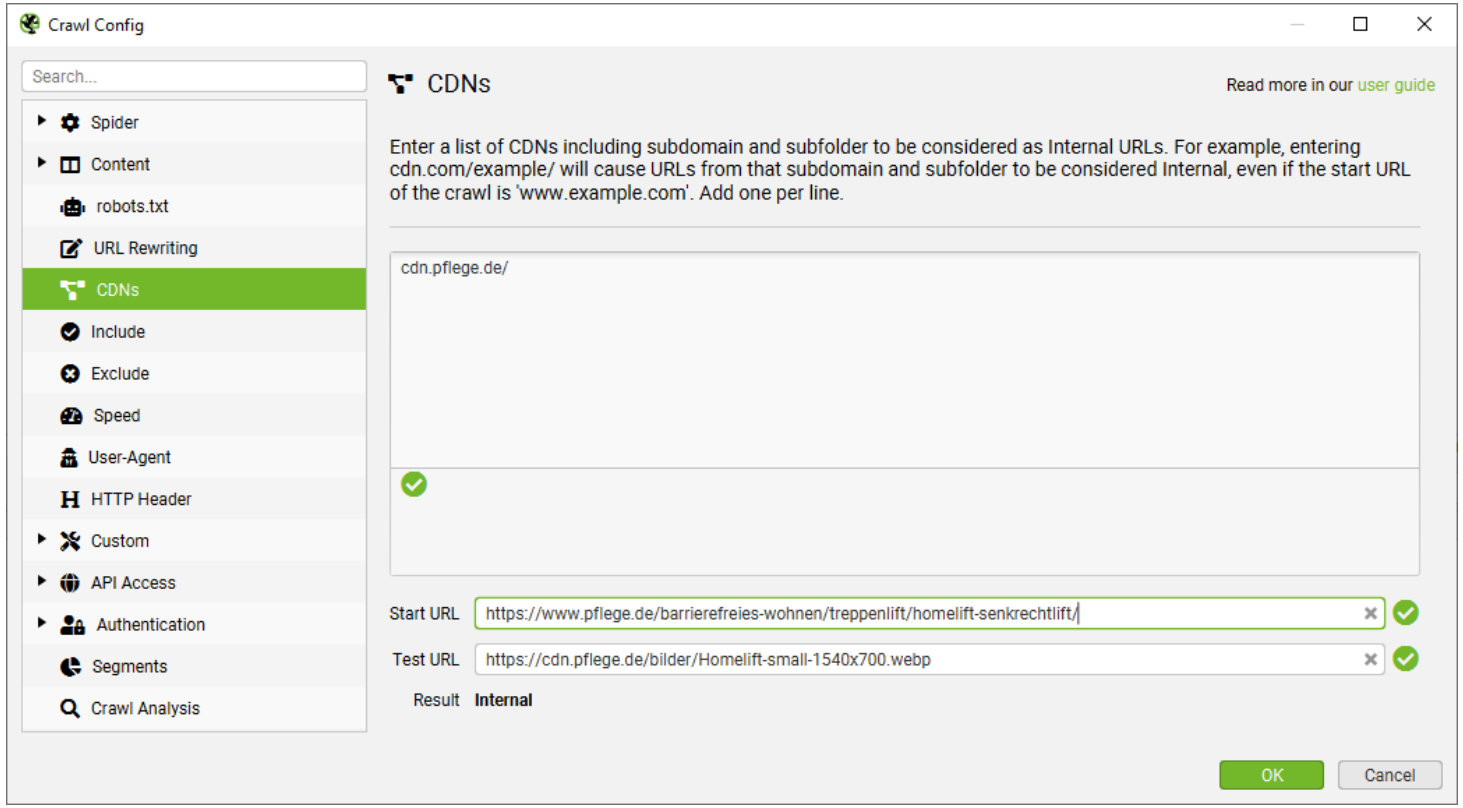


# #9: CDN

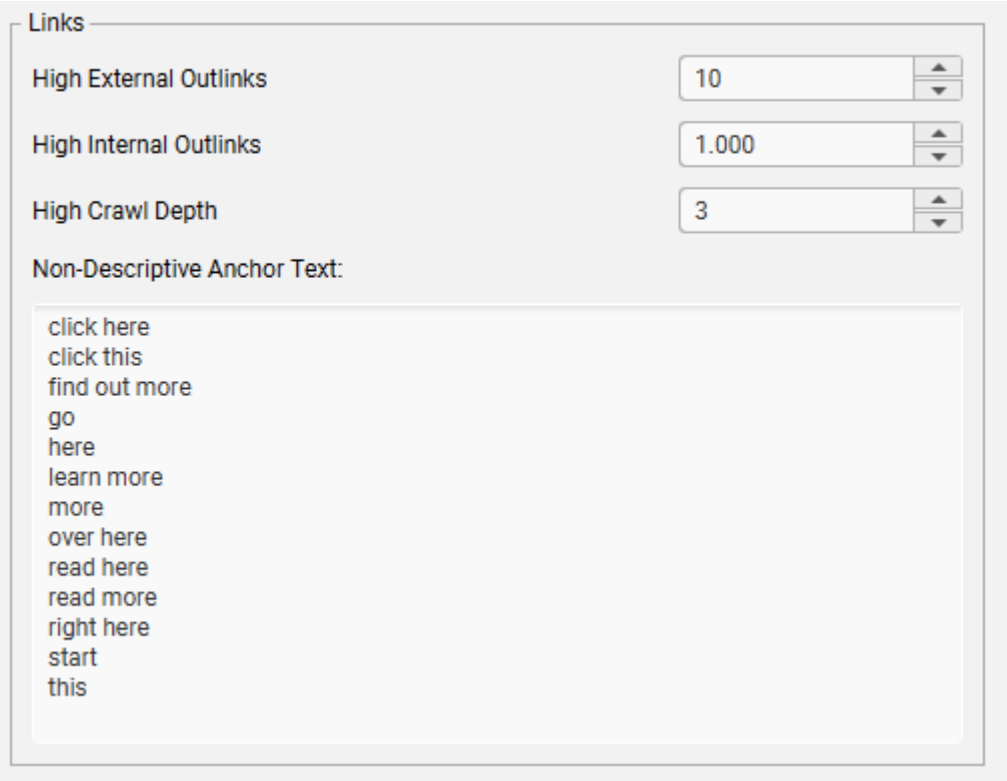
## – Wer‘s hat, sollte es angeben!



# … UND NOCH MEHR!

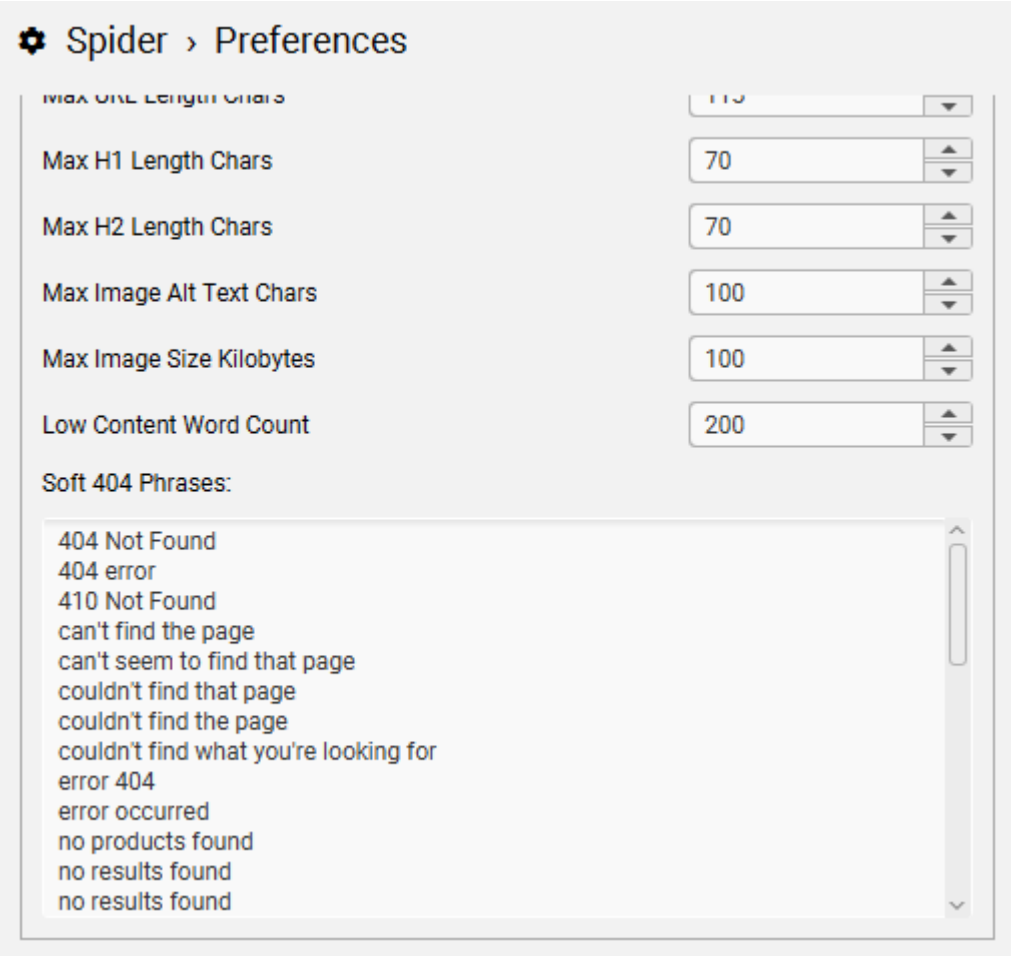
| | |
|---|---|
| – Speed | → je nach Server |
| – Authentication | → vor allem für Staging Server (niemals „Forms Based“ nutzen!) |
| – Custom Extraction/Search | → hilfreiches Universal-Werkzeug |
| – Page Speed | → eher nicht im Rahmen eines normalen Crawls (frisst Zeit!) |
| – Spelling & Grammer | → eher nicht für SEOs relevant |

# ANKERTEXTE ANPASSEN!

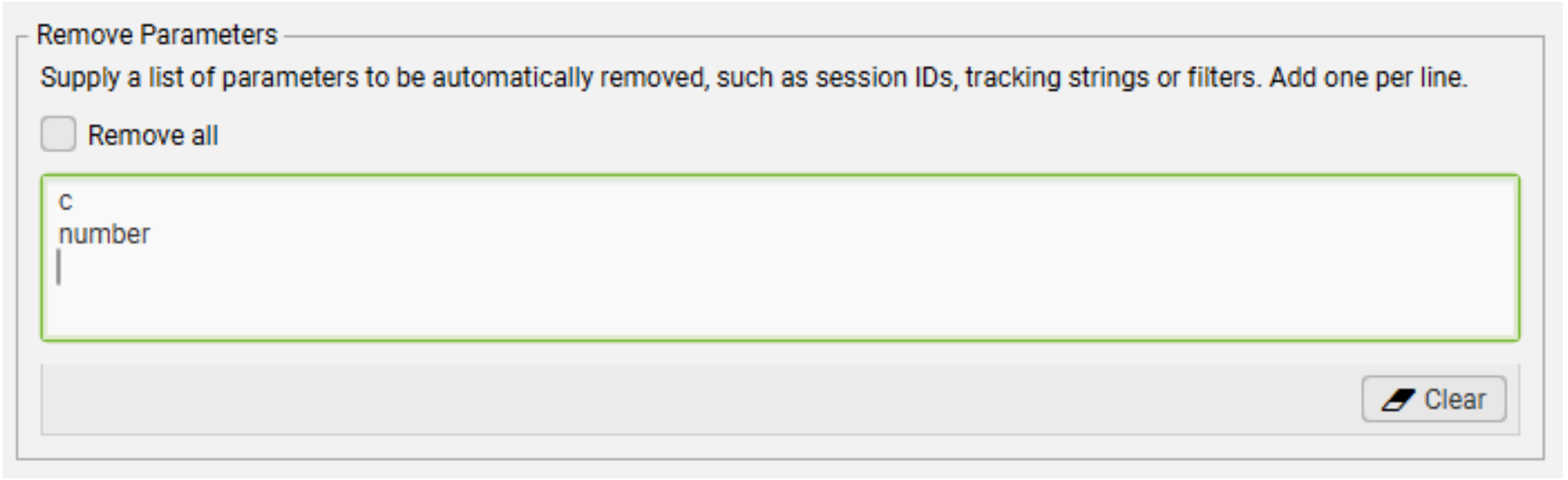



hier
klicken
mehr erfahren
...

# ÜBERSIEHT MAN GERNE



# WENN MAN SICH SICHER IST: URL REWRITING



BLOOFUSION

# DIE ANALYSE DER ERGEBNISSE

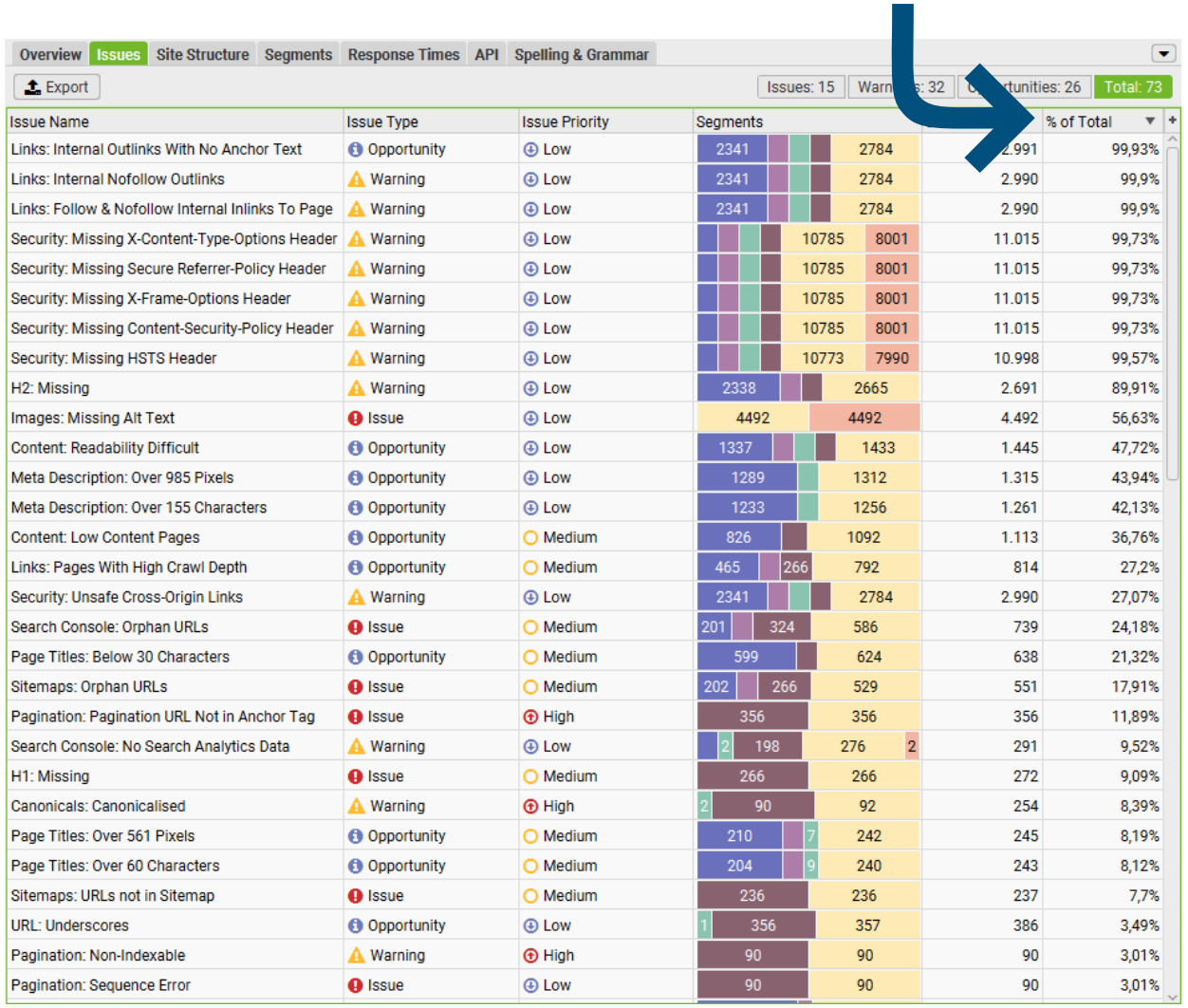
2

# EMPFEHLUNGEN

1. Empfehlungen für gesamte Website
1.1 robots.txt
1.2 Page Speed
1.3 BreadcrumbList Markup
...

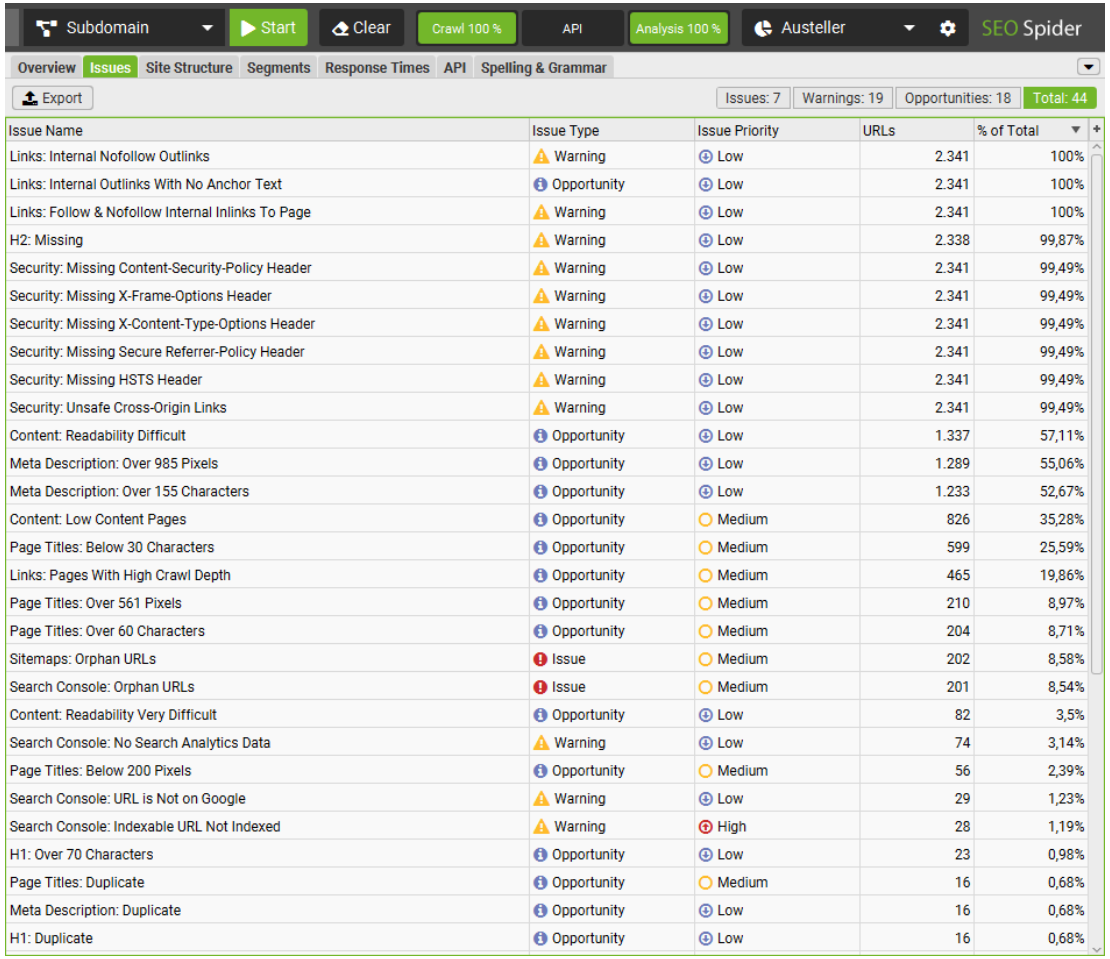
2. Empfehlungen pro Seitentyp
2.1 Startseite
2.2 Kategorieseiten
2.3 Detailseiten
...

# WAS GILT SITE-WIDE?




| Issue Name | Issue Type | Issue Priority | Segments | | | | % of Total |
|---|---|---|---|---|---|---|---|
| Links: Internal Outlinks With No Anchor Text | Opportunity | Low | 2341 | 2784 | 2.991 | 99,93% |
| Links: Internal Nofollow Outlinks | Warning | Low | 2341 | 2784 | 2.990 | 99,9% |
| Links: Follow & Nofollow Internal Inlinks To Page | Warning | Low | 2341 | 2784 | 2.990 | 99,9% |
| Security: Missing X-Content-Type-Options Header | Warning | Low | 10785 | 8001 | 11.015 | 99,73% |
| Security: Missing Secure Referrer-Policy Header | Warning | Low | 10785 | 8001 | 11.015 | 99,73% |
| Security: Missing X-Frame-Options Header | Warning | Low | 10785 | 8001 | 11.015 | 99,73% |
| Security: Missing Content-Security-Policy Header | Warning | Low | 10785 | 8001 | 11.015 | 99,73% |
| Security: Missing HSTS Header | Warning | Low | 10773 | 7990 | 10.998 | 99,57% |
| H2: Missing | Warning | Low | 2338 | 2665 | 2.691 | 89,91% |
| Images: Missing Alt Text | Issue | Low | 4492 | 4492 | 4.492 | 56,63% |
| Content: Readability Difficult | Opportunity | Low | 1337 | 1433 | 1.445 | 47,72% |
| Meta Description: Over 985 Pixels | Opportunity | Low | 1289 | 1312 | 1.315 | 43,94% |
| Meta Description: Over 155 Characters | Opportunity | Low | 1233 | 1256 | 1.261 | 42,13% |
| Content: Low Content Pages | Opportunity | Medium | 826 | 1092 | 1.113 | 36,76% |
| Links: Pages With High Crawl Depth | Opportunity | Medium | 465 | 266 | 792 | 814 | 27,2% |
| Security: Unsafe Cross-Origin Links | Warning | Low | 2341 | 2784 | 2.990 | 27,07% |
| Search Console: Orphan URLs | Issue | Medium | 201 | 324 | 586 | 739 | 24,18% |
| Page Titles: Below 30 Characters | Opportunity | Medium | 599 | 624 | 638 | 21,32% |
| Sitemaps: Orphan URLs | Issue | Medium | 202 | 266 | 529 | 551 | 17,91% |
| Pagination: Pagination URL Not in Anchor Tag | Issue | High | 356 | 356 | 356 | 11,89% |
| Search Console: No Search Analytics Data | Warning | Low | 2 | 198 | 276 | 2 | 291 | 9,52% |
| H1: Missing | Issue | Medium | 266 | 266 | 272 | 9,09% |
| Canonicals: Canonicalised | Warning | High | 2 | 90 | 92 | 254 | 8,39% |
| Page Titles: Over 561 Pixels | Opportunity | Medium | 210 | 7 | 242 | 245 | 8,19% |
| Page Titles: Over 60 Characters | Opportunity | Medium | 204 | 9 | 240 | 243 | 8,12% |
| Sitemaps: URLs not in Sitemap | Issue | Medium | 236 | 236 | 237 | 7,7% |
| URL: Underscores | Opportunity | Low | 1 | 356 | 357 | 386 | 3,49% |
| Pagination: Non-Indexable | Warning | High | 90 | 90 | 90 | 3,01% |
| Pagination: Sequence Error | Issue | Low | 90 | 90 | 90 | 3,01% |

# WAS GILT PRO SEITENTYP?

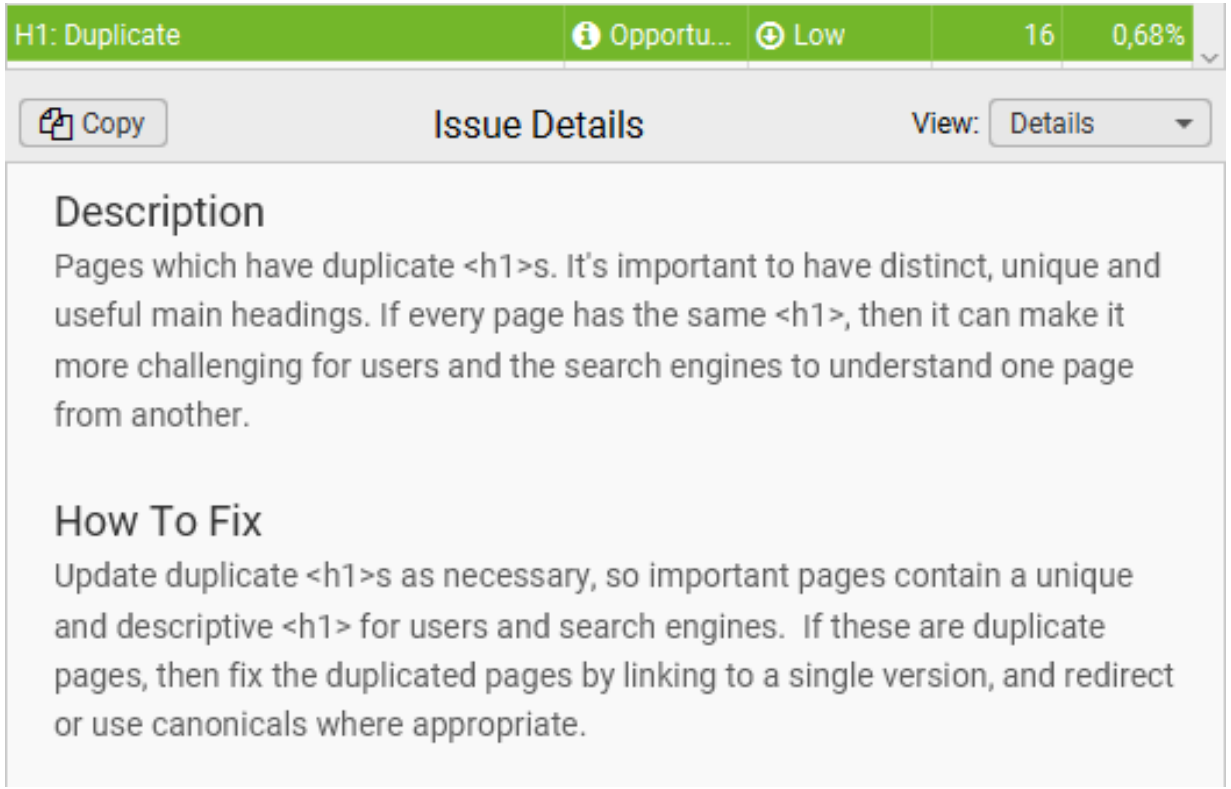



| Issue Name | Issue Type | Issue Priority | URLs | % of Total |
|---|---|---|---|---|
| Links: Internal Nofollow Outlinks | Warning | Low | 2.341 | 100% |
| Links: Internal Outlinks With No Anchor Text | Opportunity | Low | 2.341 | 100% |
| Links: Follow & Nofollow Internal Inlinks To Page | Warning | Low | 2.341 | 100% |
| H2: Missing | Warning | Low | 2.338 | 99,87% |
| Security: Missing Content-Security-Policy Header | Warning | Low | 2.341 | 99,49% |
| Security: Missing X-Frame-Options Header | Warning | Low | 2.341 | 99,49% |
| Security: Missing X-Content-Type-Options Header | Warning | Low | 2.341 | 99,49% |
| Security: Missing Secure Referrer-Policy Header | Warning | Low | 2.341 | 99,49% |
| Security: Missing HSTS Header | Warning | Low | 2.341 | 99,49% |
| Security: Unsafe Cross-Origin Links | Warning | Low | 2.341 | 99,49% |
| Content: Readability Difficult | Opportunity | Low | 1.337 | 57,11% |
| Meta Description: Over 985 Pixels | Opportunity | Low | 1.289 | 55,06% |
| Meta Description: Over 155 Characters | Opportunity | Low | 1.233 | 52,67% |
| Content: Low Content Pages | Opportunity | Medium | 826 | 35,28% |
| Page Titles: Below 30 Characters | Opportunity | Medium | 599 | 25,59% |
| Links: Pages With High Crawl Depth | Opportunity | Medium | 465 | 19,86% |
| Page Titles: Over 561 Pixels | Opportunity | Medium | 210 | 8,97% |
| Page Titles: Over 60 Characters | Opportunity | Medium | 204 | 8,71% |
| Sitemaps: Orphan URLs | Issue | Medium | 202 | 8,58% |
| Search Console: Orphan URLs | Issue | Medium | 201 | 8,54% |
| Content: Readability Very Difficult | Opportunity | Low | 82 | 3,5% |
| Search Console: No Search Analytics Data | Warning | Low | 74 | 3,14% |
| Page Titles: Below 200 Pixels | Opportunity | Medium | 56 | 2,39% |
| Search Console: URL is Not on Google | Warning | Low | 29 | 1,23% |
| Search Console: Indexable URL Not Indexed | Warning | High | 28 | 1,19% |
| H1: Over 70 Characters | Opportunity | Low | 23 | 0,98% |
| Page Titles: Duplicate | Opportunity | Medium | 16 | 0,68% |
| Meta Description: Duplicate | Opportunity | Low | 16 | 0,68% |
| H1: Duplicate | Opportunity | Low | 16 | 0,68% |

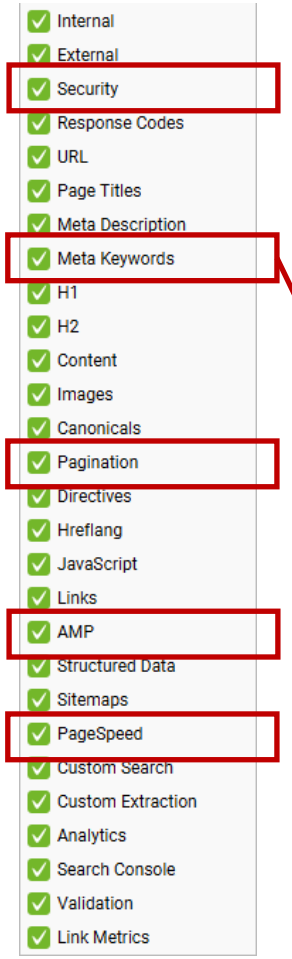
# Es bleibt ein altes, großes Problem: Was ist nur ein Hint, was ist ein echter Hebel?

# Fixe nur etwas, wenn Du genau verstehst, worum es geht!

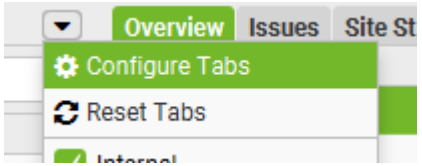
# WITH A LITTLE HELP FROM MY FROG ...



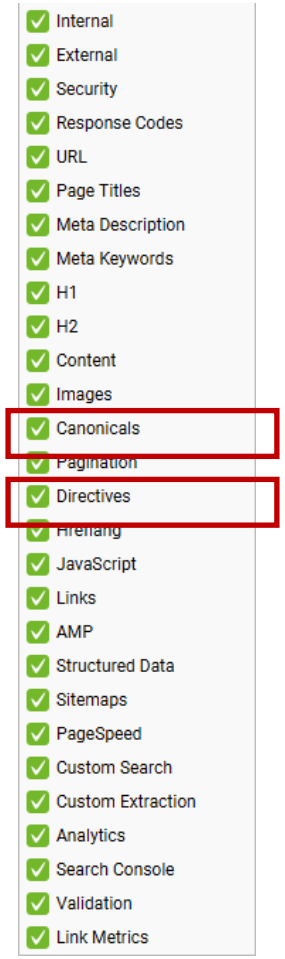
# ÜBERWIEGEND NICHT SEO-RELEVANT




- Page Speed:
  - GSC für Symptome
  - PSI oder Lighthouse für Diagnose

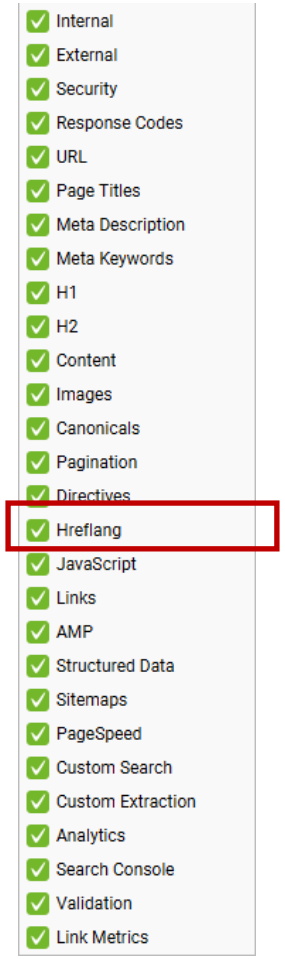


# WICHTIG, ABER FAST NIE PROBLEME



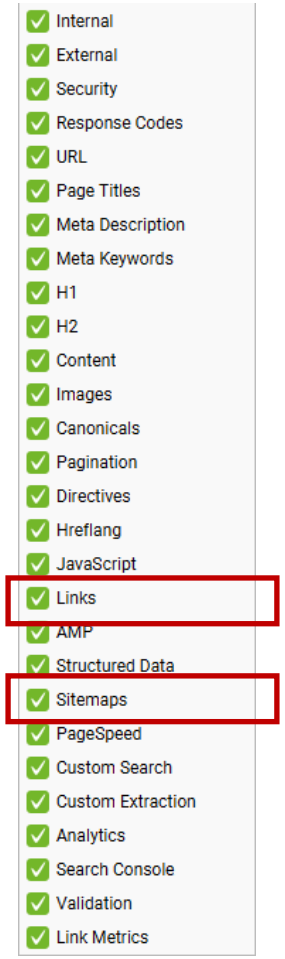

- Oft fehlt:
  - max-snippet
  - max-image-preview
  - max-video-preview
- Diskussionsthema:
  - Noindex bei paginierten Seiten

# OFT FEHLER, ABER KEIN ECHTER HEBEL



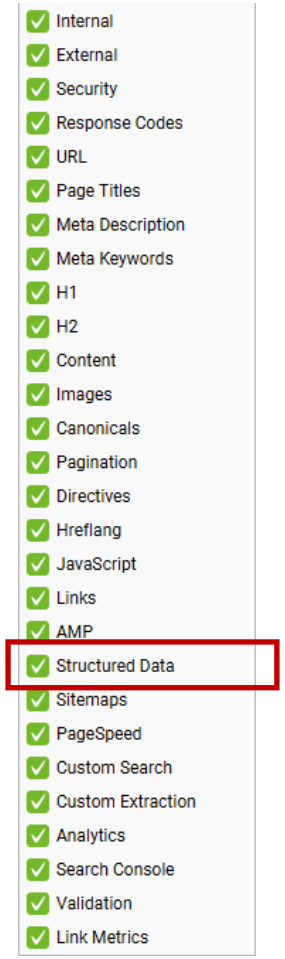

- Häufig:
  - Hreflang ≠ Canonical URL
  - X-default fehlt
  - Länder/Sprachen, die es nicht gibt
  - …
- Bringt aber alles nicht mehr Traffic!

# INTERNE VERLINKUNG: OFT UNTERSCHÄTZT



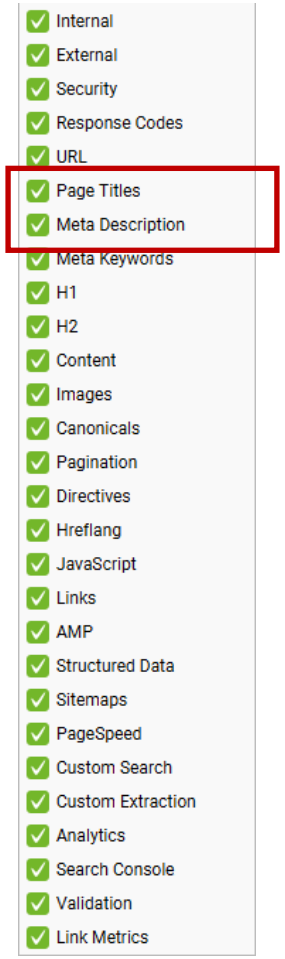

- Oft:
  - Inhalte haben zu hohe Klicktiefe
  - Inhalte werden gar nicht intern verlinkt (Orphans)
  - Nicht-deskriptive Ankertexte

# AUFPASSEN



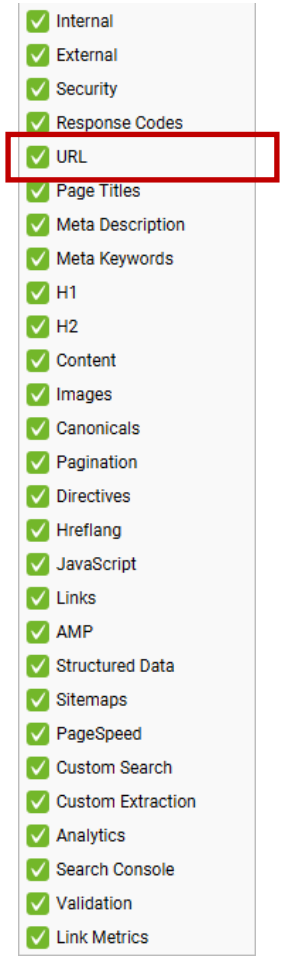

- Google kann das immer anders sehen
  - GSC, Rich Results Tester
- Obacht:
  - Die GSC zeigt nicht alle Google-relevanten Markups an

# SUPERWICHTIG




- Gary von Google spricht:
  - „You want to rely on 10 blue links“
    https://searchengineland.com/google-expect-more-simplification-of-search-results-432305
- Aber:
  - Der Frog sieht nicht, was Google ausspielt → SISTRIX & Co.

# SEHR UNWICHTIG

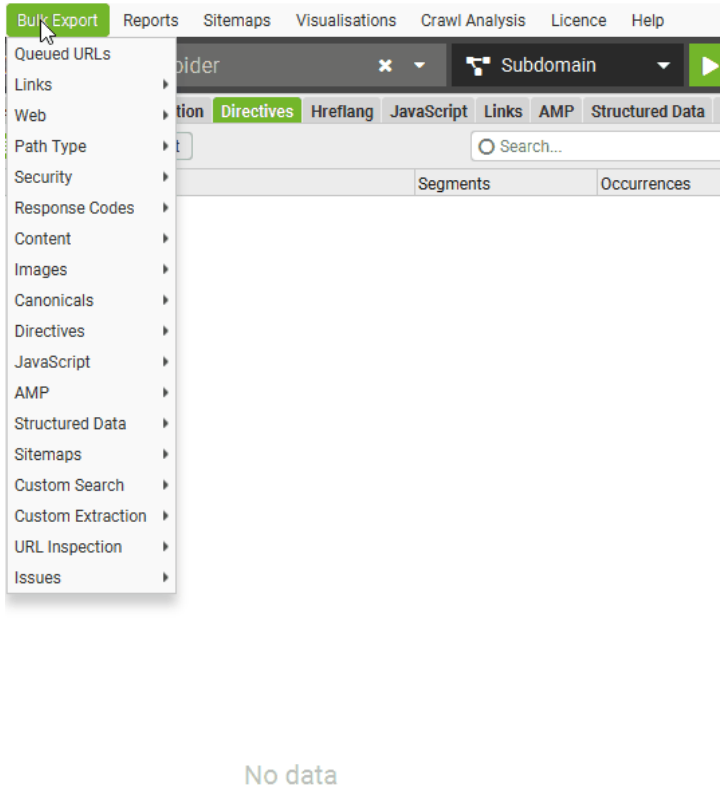


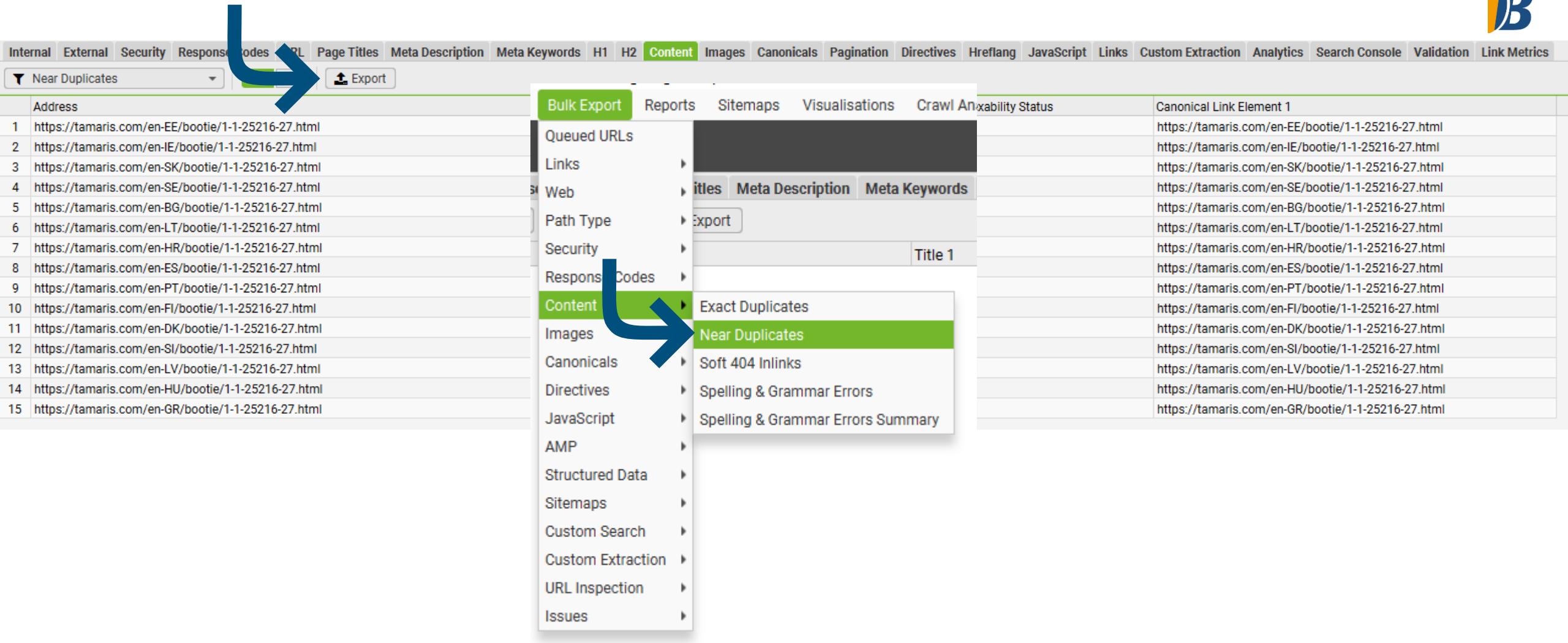

- Häufig:
  - Umlaute in URLs
  - Repetitive Path
  - URL zu lang
- Lohnt oft nicht + stellt ein Risiko dar.
- Steht bald eh ein URL-Update an?

BLOOFUSION

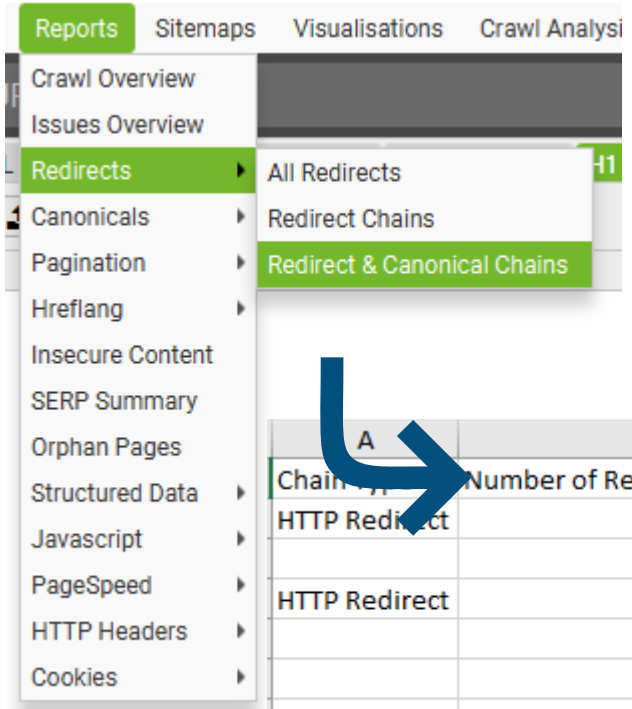
# BULK EXPORT + REPORTS

3


Bulk Export
Reports
Sitemaps
Visualisations
Crawl Analysis
Licence
Help
Queued URLs
Links
Web
Path Type
Security
Response Codes
Content
Images
Canonicals
Directives
JavaScript
AMP
Structured Data
Sitemaps
Custom Search
Custom Extraction
URL Inspection
Issues
Subdomain
Directives
Hreflang
JavaScript
Links
AMP
Structured Data
Search...
Segments
Occurrences
No data

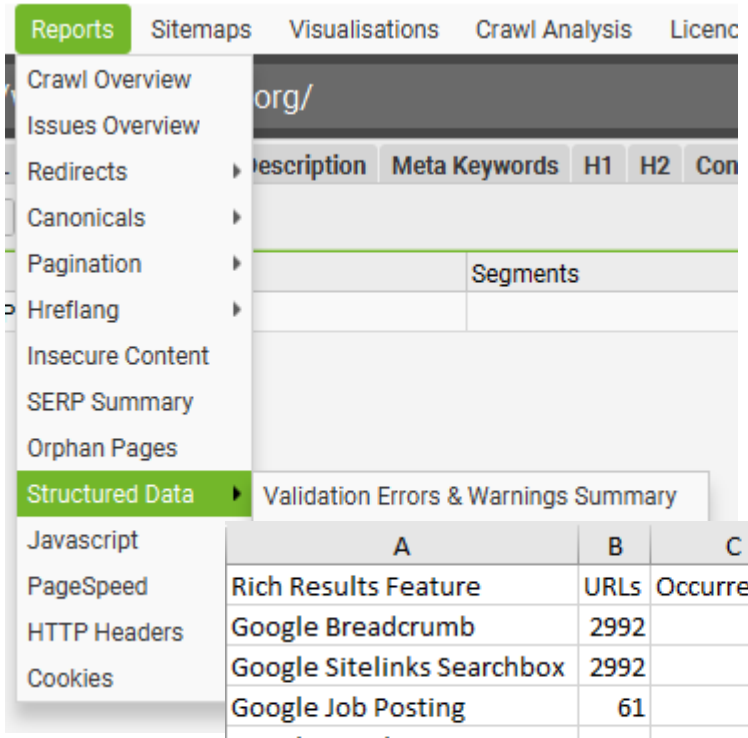

Internal External Security Page Titles Meta Description Meta Keywords H1 H2 Content Images Canonicals Pagination Directives Hreflang JavaScript Links Custom Extraction Analytics Search Console Validation Link Metrics
Near Duplicates
Export
Address
Canonical Link Element 1
https://tamaris.com/en-EE/bootie/1-1-25216-27.html
https://tamaris.com/en-IE/bootie/1-1-25216-27.html
https://tamaris.com/en-SK/bootie/1-1-25216-27.html
https://tamaris.com/en-SE/bootie/1-1-25216-27.html
https://tamaris.com/en-BG/bootie/1-1-25216-27.html
https://tamaris.com/en-LT/bootie/1-1-25216-27.html
https://tamaris.com/en-HR/bootie/1-1-25216-27.html
https://tamaris.com/en-ES/bootie/1-1-25216-27.html
https://tamaris.com/en-PT/bootie/1-1-25216-27.html
https://tamaris.com/en-FI/bootie/1-1-25216-27.html
https://tamaris.com/en-DK/bootie/1-1-25216-27.html
https://tamaris.com/en-SI/bootie/1-1-25216-27.html
https://tamaris.com/en-LV/bootie/1-1-25216-27.html
https://tamaris.com/en-HU/bootie/1-1-25216-27.html
https://tamaris.com/en-GR/bootie/1-1-25216-27.html
Bulk Export
Reports
Sitemaps
Visualisations
Queued URLs
Links
Web
Path Type
Security
Content
Images
Canonicals
Directives
JavaScript
AMP
Structured Data
Sitemaps
Custom Search
Custom Extraction
URL Inspection
Issues
Meta Description
Meta Keywords
Title 1
Exact Duplicates
Near Duplicates
Soft 404 Inlinks
Spelling & Grammar Errors
Spelling & Grammar Errors Summary

# AUSNAHME #1



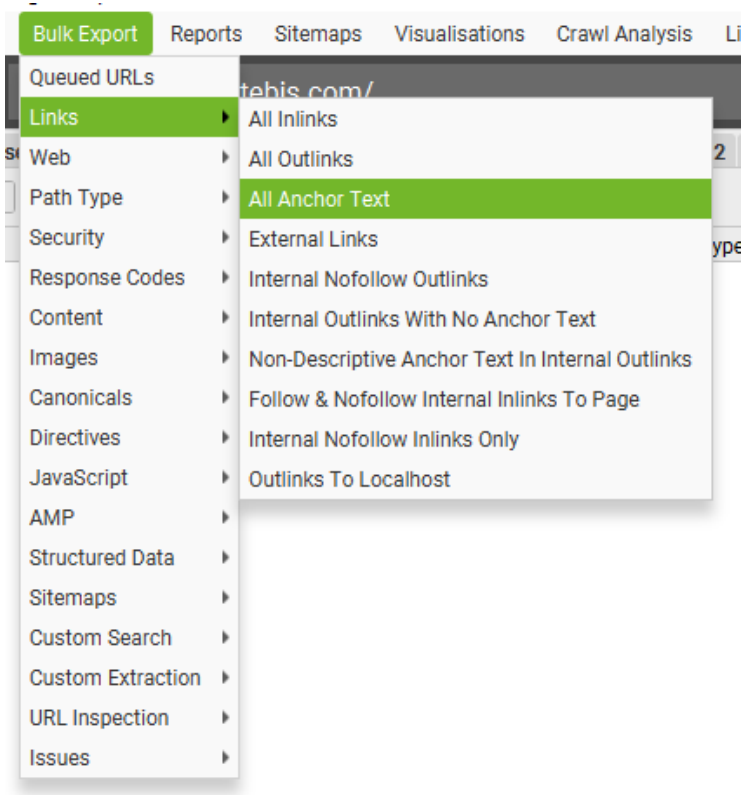

| A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|
| Chain Type | Number of Redirects/ | | Temp Redirect in Chain | Addr | | Final Indexa | | Final Content | Final Status Code |
| HTTP Redirect | 1 | FALSCH | FALSCH | https:// | https:// | Indexable | | text/html; charset=UTF-8 | 200 |
| | 0 | FALSCH | FALSCH | https:// | https:// | Indexable | | text/html; charset=UTF-8 | 200 |
| HTTP Redirect | 1 | FALSCH | FALSCH | https:// | https:// | Indexable | | text/html; charset=UTF-8 | 200 |
| | 0 | FALSCH | FALSCH | https:// | https:// | Indexable | | text/html; charset=UTF-8 | 200 |
| | 0 | FALSCH | FALSCH | https:// | https:// | Indexable | | text/html; charset=UTF-8 | 200 |
| HTTP Redirect | 1 | FALSCH | FALSCH | https:// | https:// | Indexable | | text/html; charset=UTF-8 | 200 |
| | 0 | FALSCH | FALSCH | https:// | https:// | Indexable | | text/html; charset=UTF-8 | 200 |
| HTTP Redirect | 1 | FALSCH | FALSCH | https:// | https:// | Indexable | | text/html; charset=UTF-8 | 200 |

# AUSNAHME #2




| A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Rich Results Feature | URLs | Occurrences | % Eligible | Eligible URLs | Error URLs | Warning URLs | Unique Errors | Unique Warnings | Total Errors | Total Warnings | Sample URL |
| Google Breadcrumb | 2992 | 2997 | 99 | 2990 | 2 | 0 | 1 | 0 | 4 | 0 | xxx |
| Google Sitelinks Searchbox | 2992 | 2997 | 100 | 2992 | 0 | 0 | 0 | 0 | 0 | 0 | xxx |
| Google Job Posting | 61 | 61 | 0 | 0 | 61 | 61 | 2 | 5 | 61 | 305 | xxx |
| Google Article | 8 | 109 | 100 | 8 | 0 | 8 | 0 | 8 | 0 | 377 | xxx |

# AUSNAHME #3




**Data Analysis**

**You**

all_anchor_text.xlsx
Kalkulationstabelle

Die Tabelle enthält zwei Spalten: Destination und Anchor. Ignoriere bitte alle Zeilen, bei denen der Anchor leer ist.

Erstelle mir eine Tabelle mit allen Anchor und geb in einer zweiten Spalte an, bei wie vielen Destination dieser Anchor vorkommt. Sortiere danach die Tabelle nach der zweiten Spalte absteigend.

| | Anchor | Destination_Count |
|---|---|---|
| 2 | Mehr erfahren | 162 |
| 3 | Read more | 141 |
| 4 | Más información | 117 |
| 5 | En savoir plus | 102 |
| 6 | More... | 96 |
| 7 | Mais informação | 74 |
| 8 | 詳しく見る | 72 |
| 9 | Maggiori informazioni | 68 |
| 10 | 更多..... | 66 |
| 11 | Webinar jetzt ansehen | 55 |
| 12 | Please select a desired date to book your event | 55 |
| 13 | Read full story | 46 |
| 14 | 阅读全文 | 46 |
| 15 | 了解更多 | 45 |
| 16 | Anwenderbericht lesen | 45 |
| 17 | Download | 38 |

BLOOFUSION

# BIN ICH JETZT DURCH?

4

# MIT DEM FROG: JA!

- Aber das reicht in der Regel nicht!
  - Bewertung von Content-Qualität
  - Conversions/Customer Journey vs. Sichtbarkeit/Klicks
  - Markups, die fehlen + sinnvoll sind
  - Seitentitel ohne angehängte Brand
  - Standardumleitung (http → https, non-www. → www.)
  - IP Redirect
  - Interstitials
  - …

BLOOFUSION

# FAZIT

5

# FAZIT

- Wie gesagt: Die Konfiguration ist die halbe Miete!
- Hints abarbeiten – Verständnis vorausgesetzt!
- Der Frog liefert Euch viel – aber definitiv nicht alles!

# HERZLICH WILLKOMMEN

Markus Hövener

SEO Advocate, Gründer

Kontakt: bloo.link/mh