

Autore

RAFFAELE VISINTIN

PREFAZIONE DI

DAN SHARP

SEO AUDIT AVANZATO

KILL SEO

VOLUME 1

Come ottimizzare siti web sui motori di ricerca con



**SCREAMING
FROG.CLUB**

SOMMARIO

AUTORE	01
PREFAZIONE: DAN SHARP	02
INTRODUZIONE	03
ISTRUZIONI PER L'USO	04
URL DETAILS	04
OVERVIEW	04
CAPITOLO 1: COMANDI BASE DI SCREAMING FROG	05
MENU PRINCIPALE	05
FILE	05
MODE	06
CONFIGURATION	06
BULK EXPORT	07
REPORT	07
VISUALISATION	07
CRAWL ANALYSIS	08
LICENCE/HELP	08
TAB SUPERIORI	08
FINESTRA SUPERIORE	08
TAB INFERIORI	08
SIDEBAR LATERALE	08
CAPITOLO 2: I PRIMI PASSI CON SCREAMING FROG	09
IL SEO SPIDER	09
LE BASI DEL "CRAWLING"	09
FARE LA SCANSIONE DI UNA SOTTOCARTELLA	10
CRAWLER DI UN ELENCO DI URL	11
REDIRECTS E MIGRAZIONE	14
CRAWLER SITI WEB DI GRANDI DIMENSIONI	16
COME SALVARE UN AUDIT SEO	24
CONFIGURAZIONI PERSONALI CON IL SEO SPIDER	26
SCHEDULING	27
ESPORTAZIONE DEI DATI DI UN CRAWL	30
IL FILE ROBOTS.TXT	43
SITEMAP	47
VISUALISATION	51
CRAWL ANALYSIS	60
CAPITOLO 3: Reports	63
CRAWL OVERVIEW REPORT	63
REDIRECTS	63
REPORTS CANONICAL	64
PAGINATION REPORTS	67
HREFLANG REPORTS	67
ORPHAN PAGES REPORT	74
REPORT INSECURE CONTENT	79
REPORT SUI DATI STRUTTURATI	80
REPORT PAGESPEED REPORTS	81
REPORT HTTP HEADER SUMMARY	82
COOKIE SUMMARY	82
CRAWL PATH REPORT	83
SEARCH FUNCTION	84
CAPITOLO 4: LE FUNZIONI DEL SEO SPIDER	86
SCHEDA INTERNAL	86
EXTERNAL TAB	91
SECURITY TAB	91
RESPONSE CODE TAB	94
TAB URL	96
TAB PAGE TITLES	97
TAB META DESCRIPTION	98
TAB META KEYWORDS	99
TAB HEADING H1	99
TAB HEADING H2	100
TAB CONTENT	100
TAB IMAGES	102
TAB CANONICAL	107
TAB PAGINATION	108
TAB DIRECTIVES	112
TAB HREFLANG	114
TAB AMP	116
TAB STRUCTURED DATA	119
TAB SITEMAP	120

TAB PAGE SPEED	120
TAB CUSTOM SEARCH	127
TAB CUSTOM EXTRACTION	127
TAB ANALYTICS	127
TAB SEARCH CONSOLE	128
TAB LINK METRICS	128
URL DETAILS	131
INLINKS	131
OUTLINKS	132
IMAGE DETAIL	132
DUPLICATE DETAILS	132
TAB RESOURCES	133
TAB SERP	133
RENDERED PAGE	134
VIEW SOURCE	134
TAB HTTP HEADERS	135
TAB COOKIES	135
STRUCTURED DATA	136
PAGESPEED DETAILS	137
TAB SPELLING & GRAMMAR DETAILS	137
OVERVIEW	138
SITE STRUCTURE	138
LATERALE RESPONSE TIMES	139
API	139
SPELLING & GRAMMAR	139
INTRODUZIONE: GOOGLE E JAVASCRIPT	185
JAVASCRIPT, IL SEO SPIDER E ALCUNE CONTROINDICAZIONI	185
JAVASCRIPT LATO SERVER E LATO CLIENT	186
I TEMPI DI CRAWLING DI GOOGLE	187
COME IDENTIFICARE JAVASCRIPT	187
AUDIT CON SCREAMING FROG DI SITI JAVASCRIPT	191
CONCLUSIONI	242
CREDITS:	243



[raffaelevisintin](#)

WEB: www.screamingfrog.club

EMAIL: info@screamingfrog.club

AUTORE

Ciao, sono **Raffaele Visintin** l'autore di **KILL SEO**, il manuale operativo dedicato al mondo della Search Engine Optimization e ti do il benvenuto in **Screaming Frog Club**, l'Hub di condivisione degli appassionati e professionisti Seo.

Laureato presso la facoltà degli studi di Trieste con indirizzo economico/statistico sono un dinosauro della SEO e del Search Marketing.

Negli anni mi sono confrontato con i migliori esponenti del settore sia in ambito nazionale che internazionale che hanno arricchito il mio background e permesso di consolidare sul campo le conoscenze acquisite.

Dopo anni da Consulente Seo freelance, dal 2016 sono Seo Specialist presso Webtek, società di comunicazione di Milano.

Lavorare nella Seo non è solo un lavoro ma una vera e propria passione che coltivo giorno dopo giorno cercando di aggiornarmi costantemente per stare al passo con le continue evoluzioni dei Motori di Ricerca.

Giornalmente mi diletto a fare l'investigatore e scovare anche il più piccolo dettaglio per smascherare ogni delitto perfetto che potrebbe ostacolare il successo di un sito web o ecommerce sulle serp di Google, Yahoo, Bing, Yandex e Baidu.

Questa passione per i Seo Audit mi ha portato a testare diversi tool ma quello che prediligo è sicuramente Screaming Frog, un SEO Spider che mi accompagna in tutte le analisi, anche le più complesse.

Screaming Frog mi è entrato nel cuore già dal primo utilizzo e la vera differenza rispetto a tool simili è la sua completezza e i continui aggiornamenti che Dan Sharp e il suo staff rilasciano regolarmente.

L'intento della guida è quello di accompagnarti nelle analisi di siti web, e-commerce e portali e dare degli spunti ed approfondimenti per la redazione di Seo Audit tecnici sempre più avanzati utilizzando il Seo Spider.

Ringrazio Dan per aver accettato di scrivere la prefazione della guida.

PREFAZIONE: DAN SHARP

SCREAMING FROG DIRECTOR



The Screaming Frog SEO Spider is a tool we launched over a decade ago, to automate website data collection and analysis of SEO to help SEOs work more efficiently.

Since that time, our mission has been pretty simple - Listen to the SEO community and help make their lives easier by solving their day-to-day problems in technical SEO. This has always felt quite natural to us, as SEOs ourselves who experience many of the same pain points.

The SEO community has been a big part of the continued growth of the Screaming Frog SEO Spider, providing incredible feedback, suggestions and testing over the years. This has led to the continued evolution of the tool, which we relentlessly improve every day.

Raffaele Visintin is part of this amazing community, and has helped us in beta testing of new versions of the software.

Raffaele approached us wanting to help Italian speakers and the Italian SEO community learn more about the tool, and has worked tirelessly on this guide with input from our team.

KILL SEO is an excellent guide to using the Screaming Frog SEO Spider for Italian SEOs.

Raffaele has produced a hugely comprehensive and practical guide, covering the basics, to more advanced features and functions of the tool.

Every tab, filter and configuration is described thoroughly and more importantly, how they can help you in your technical SEO auditing.

Whether you're a beginner, or more advanced and just want to level up your SEO skills using the tool, I highly recommend reading.

INTRODUZIONE

Il mondo web è in continua evoluzione e rappresenta un'occasione di business unica per ogni azienda.

Ogni giorno nascono milioni di siti web e altrettanti scompaiono rimettendo nel cassetto le proprie speranze di successo. Il mercato online non ha confini e permette una vetrina senza precedenti per ogni tipologia di business. Ma in questo "Paese dei Balocchi" internazionale ci sono un numero immenso di marionette ma un solo burattinaio: Google.

Il Motore di Ricerca cannibalizza oltre il 98% del traffico mondiale lasciando le briciole agli altri Motori di Ricerca come Yandex, Bing o Baidu. Di conseguenza le direttive di Google diventano un mantra ed essere indicizzati e posizionati bene nei suoi risultati può significare aumentare il traffico e l'authorship che successivamente si trasformano in conversioni e maggiore fatturato.

La guida ti accompagnerà passo passo a conoscere tutti gli aspetti tecnici di un Seo Audit avanzato ed affrontare tutte le possibili criticità che un sito web, portale o ecommerce potrebbe presentare.

Alla fine di questo viaggio sarai in grado di scansionare, analizzare e prendere le giuste decisioni strategiche per ottimizzare al meglio il tuo progetto e al contempo migliorare la crescita organica sul Motore di Ricerca.

Per il nostro Seo Audit utilizzeremo Screaming Frog, un Web Spider che emula il comportamento di Google, un vero e proprio alleato nella diagnosi dei siti internet.

Il Seo Spider sarà il nostro compagno di viaggio e ci fornirà tutte le informazioni e dati necessari per mettere a punto delle analisi avanzate.

Questa guida vuole essere un punto di riferimento pratico per maneggiare con destrezza il tool, ottimizzare al meglio le tue attività di crawling e mettere a punto degli Audit Seo professionali.

NOTA BENE:

Questo manuale è stato inteso come supporto pratico per gli Audit Seo.

Per questo motivo alcuni contenuti sono stati ripetuti volutamente diverse volte nei vari capitoli per permetterti di non perdere il focus ed avere per ogni argomento tutte le nozioni necessarie a portata di mano.



KILL SEO è un progetto editoriale personale e no profit e viene distribuito gratuitamente sul sito web screamingfrog.club e su tutti i principali Marketplace.

Se vuoi sostenere **KILL SEO** e i suoi futuri aggiornamenti ti chiedo **5 minuti del tuo tempo per lasciare una recensione e condividere l'iniziativa sui Social Network.**

Un piccolo gesto che garantisce un pieno di energia per i progetti futuri.

ISTRUZIONI PER L'USO

Prima di immergerti nello studio di Screaming Frog e nel mondo del “crawling” ti voglio dare alcune indicazioni sulla simbologia utilizzata nella guida per una lettura più semplificata:

1. Percorso da seguire per abilitare/disabilitare delle funzioni nel Seo Spider.

Config > Mode> Spider

2. Le sezioni con il colore blu identificano degli esempi pratici applicativi delle funzionalità di Screaming Frog o alcune best practice nel mondo seo.

ESEMPIO

3. “HOW TO”: tutorial completi e approfondimenti su tipi particolari di crawl.

HOW TO

4. Codice: identifica una porzione di codice sorgente HTML, sintassi RegEx o altri input utilizzati sul Seo Spider.

<h1> Screaming Frog </h1>

5. FAST & CRAWL: brevi tutorial per scansioni particolari con Screaming Frog.

FAST & CRAWL

6. TAB: Identifica le Tab della finestra inferiore del Seo Spider.

URL DETAILS

7. TAB: Identifica le Tab della sidebar a destra del Seo Spider.

OVERVIEW

CAPITOLO 1: COMANDI BASE DI SCREAMING FROG

Iniziamo il nostro viaggio nell'ispezione di Screaming Frog e sottoponiamo ai raggi X il Seo Spider analizzando le parti principali che lo compongono.

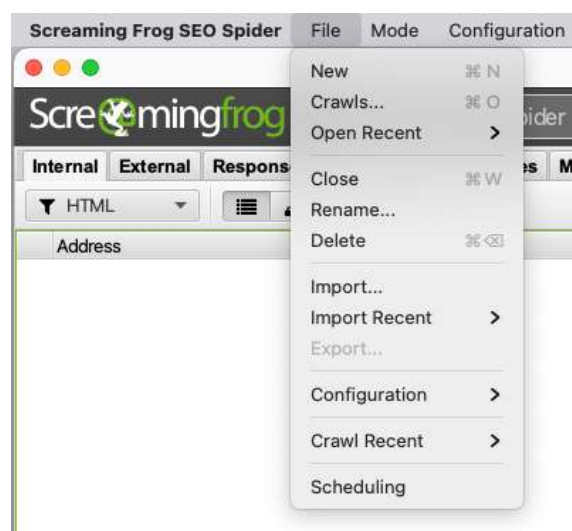


MENU PRINCIPALE

Rappresenta il cuore della console e ti permette di gestire le impostazioni più rilevanti che influiranno sul tuo Seo Audit tenico.

FILE

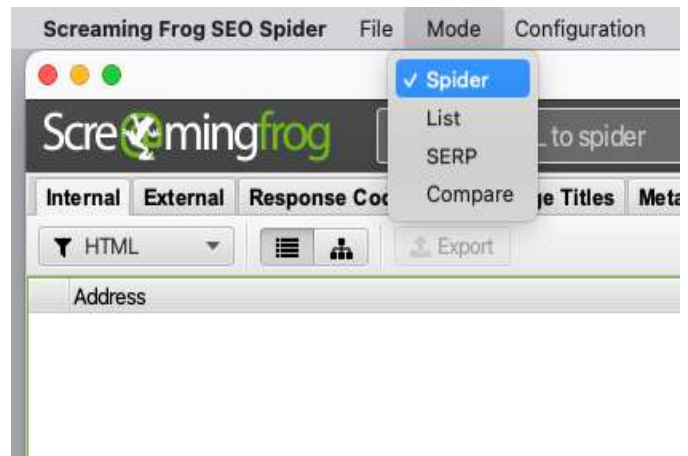
Puoi creare un nuovo progetto, aprire le scansioni precedenti nella modalità "Database Storage" (Crawl) o "Memory Storage" (Open Recent), gestire la pianificazione dei crawl (Scheduling) e modificare le configurazioni di default delle scansioni creando anche più profili personalizzati pronti all'uso.



MODE

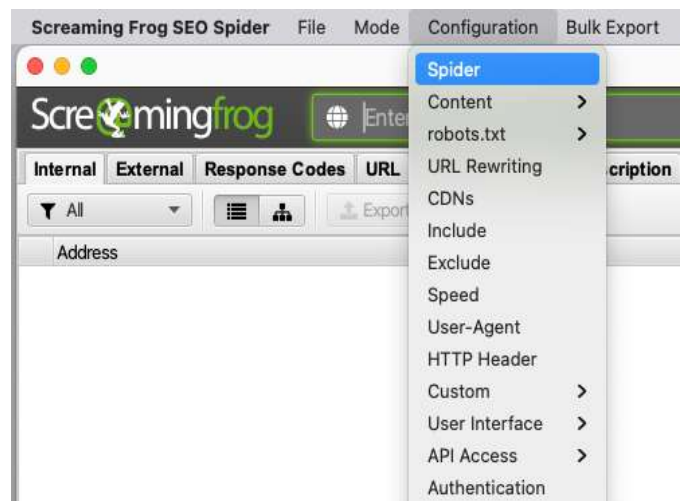
Ti permette di scegliere quale modalità di scansione utilizzare.

- **Spider:** modalità di default per una scansione completa di un sito web.
- **List:** ti permette di fare dei crawl specifici da elenchi di URL che puoi caricare manualmente o copiare nel Seo Spider.
- **Compare:** rappresenta l'ultima funzionalità aggiunta dagli sviluppatori di Screaming Frog per confrontare due scansioni precedentemente eseguite e analizzare le differenze in termini di URL e filtri (utile per siti in staging vs siti live, desktop vs versione mobile etc.).
- **Serp:** ti permette di ottenere un'anteprima degli snippet attraverso il caricamento di un file .csv o .xls tabulato.



CONFIGURATION

Questa voce di menu è tra le più importanti del Seo Spider e corrisponde al motore della console mettendo a tua disposizione tutta una serie di personalizzazioni per eseguire Seo Audit Avanzati e collegamenti a servizi esterni come Google Analytics, Search Console, Lightspeed, Moz, Ahrefs.



- **Spider:** attraverso questa opzione puoi gestire il comportamento dello Spider, controllare quali risorse scansionare, limitare le sue scorribande, definire i parametri di alcuni filtri e decidere se fare una scansione dell'HTML statico o includere anche il rendering Javascript.
- **Content:** puoi definire l'area in cui il Seo Spider deve esaminare il contenuto, impostare gli indici di tollerabilità per considerare un contenuto come duplicato e scegliere varie opzioni riguardo al vocabolario da utilizzare per il controllo grammaticale e di battitura.
- **Robots.txt:** questa opzione ti permette di definire in modo personalizzato il file Robots.txt modificandolo direttamente nel Seo Spider o quale atteggiamento debba tenere lo Spider nei confronti di questo file. In sintesi puoi decidere se il Seo Spider debba seguire le direttive o ignorarle per la scansione.
- **URL Rewriting:** ti permette di riscrivere gli URL prima di eseguire la scansione. Molto utile per evitare URL parametrati nei risultati del crawl e per altre best practice avanzate che utilizzano la sintassi RegEX.
- **Include/Exclude:** queste funzionalità ti permettono di includere o escludere degli URL dalla scansione impostando delle regole RegEX che puoi personalizzare in base al focus della tua analisi.
- **Speed:** ti consente di definire la velocità di scansione dello Spider Tool; molto utile per non sovraccaricare il server del sito internet da analizzare.

- **User Agent:** come sai Screaming Frog è un emulatore di Spider, con questa opzione puoi scegliere con quale User Agent eseguire la scansione scegliendo tra GoogleBot, GoogleBot per smartphone, Baidu per il mercato orientale, Yandex per mercato Russo, Yahoo/Bing etc. in base al tuo mercato e focus.
- **HTTP Header:** ti permette di configurare alcune opzioni dell'HTTP Header in modo personalizzato.
- **Custom:** attraverso questa voce sei in grado di eseguire scansioni di particolari elementi HTML ed estrarli a tuo piacimento per un'analisi personalizzata. Molto utile anche per definire la posizione dei Link all'interno del documento Html.
- **Api Access:** questa funzionalità ti consente di impostare delle connessioni tramite API con servizi quali Google Analytics, Search Console, Ahrefs, Majestic, PageSpeed Insight e Moz. Utilissime integrazioni per ottenere degli Audit Seo completi con dati granulari.
- **Authentication:** permette di scansionare un sito web che richiede un'autenticazione.
- **System:** permette di scegliere la modalità di archiviazione dei dati delle scansioni.

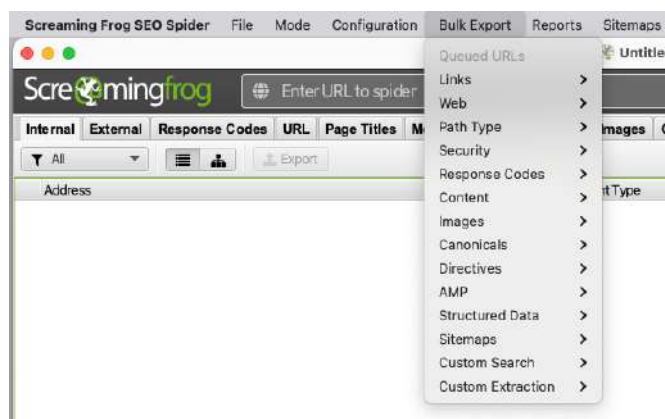
BULK EXPORT

(Esportazioni di massa): le voci incluse in questo menu ti permettono di scaricare quasi tutti i dati più rilevanti delle tue scansioni.

Questi documenti sono in formato “.csv”, “.xlsx” e “.gsheet”.

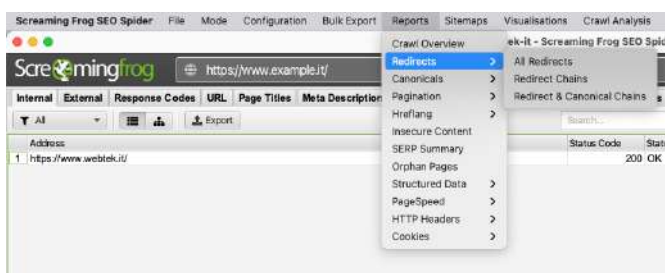
Di seguito le metriche a tua disposizione:

- Links;
- Web;
- Path Type;
- Security;
- Response Code;
- Content;
- Images;
- Canonicals;
- Directives;
- AMP;
- Structured Data;
- Sitemaps;
- Custom Search;
- Custom Extraction;



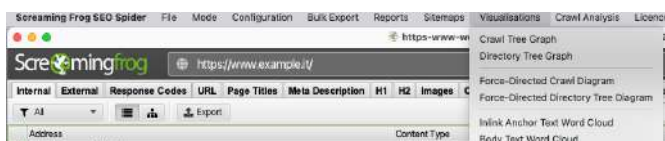
REPORT

Questa voce di menu ti permette di trovare tutta una serie di rapporti che mettono a focus i reindirizzamenti, i canonical, le paginazioni del sito web, eventuali contenuti non considerati sicuri, l'individuazione delle pagine orfane, i dati strutturati e molte altre informazioni utili da esportare e gestire durante l'ottimizzazione Seo.



VISUALISATION

Questa funzionalità ti permette di visualizzare i dati riguardanti gli URL in modo molto intuitivo ed impattante secondo diverse modalità grafiche. Le visualizzazioni sono interattive ed ideali per concentrarsi sui diversi nodi o cartelle della struttura del sito e garantire un controllo maggiormente specifico e puntuale.



CRAWL ANALISYS

La crawl analysis è una nuova funzione introdotta da Screaming Frog negli ultimi aggiornamenti del Seo Spider. Attraverso la crawl analysis sei in grado di analizzare, alla fine della scansione, metriche che richiedono un calcolo ulteriore sui dati popolati (es. "Link Metrics", alcuni aspetti della Sitemap etc.)

Dal menu puoi avviare la Crawl Analysis e/o configurare su quali elementi applicarla e/o automatizzare alla fine della scansione.



LICENCE/HELP

Corrispondono alle ultime voci del menu e sono dedicate all'inserimento della licenza di Screaming Frog e ai link per la richiesta di approfondimenti e supporto.

TAB SUPERIORI

Questa sezione include tutta una serie di schede che vengono popolate in base ai diversi intenti di analisi. Comprende ad esempio la raccolta degli URL in base al Response Code, alle Meta Description e Title, Directive etc.

Ogni scheda è stata commentata nelle sezioni successive di questa guida.

FINESTRA SUPERIORE

Corrisponde alla schermata principale del Seo Spider e viene popolata dagli URL scoperti durante la scansione con tutti i dettagli relativi. Ricorda che ogni URL può essere filtrato mediante i "Filtri dei Dati" nelle diverse tab superiori.

TAB INFERIORI

Quest'area rappresenta un approfondimento ai risultati ottenuti nella finestra superiore e viene popolata ogni qualvolta selezioni uno o più URL dalla stessa. In questa sezione trovi i dettagli riguardo gli Inlinks, gli Outlinks, dettagli sulle immagini e molto altro. Come nel caso precedente ogni singola Tab è stata documentata nel dettaglio nelle pagine seguenti.

SIDEBAR LATERALE

Quest'area del Seo Spider ti permette una visione preliminare e aggregata molto interessante sulla scansione eseguita. Nella prima scheda trovi una panoramica con dati assoluti e relativi dei singoli elementi scansionati, nella seconda puoi valutare l'architettura del sito in base al "Crawl Depth", nella terza i tempi di caricamento delle pagine aggregate per intervalli di tempo, nella quarta lo stato di avanzamento della scansione con API a servizi terzi e nell'ultima un riassunto degli errori riscontrati nel contenuto.

CAPITOLO 2: I PRIMI PASSI CON SCREAMING FROG

IL SEO SPIDER

Screaming Frog lo puoi trovare in due versioni: la prima gratuita che permette di scansionare fino a 500 Url e la seconda a pagamento senza limitazioni e con accesso a opzioni, configurazioni avanzata e integrazioni tramite API a servizi di terze parti come Google Analytics, Search Console, Page Speed Insight, Ahrefs, Moz e Majestic Seo al costo di 149 sterline.

Un confronto fra le due versioni lo puoi trovare qui: <https://www.screamingfrog.co.uk/seo-spider/pricing/>

LE BASI DEL “CRAWLING”

Di base Screaming Frog è configurato con la modalità di scansione “Spider” che permette di eseguire un crawling completo di un sito web, e-commerce o portale.



Ricordati però che il Seo Tool di default considera solamente il sottodominio che hai inserito e se ci sono degli altri li valuta come fossero dei collegamenti esterni (“External Links” - Link esterni).

ESEMPIO:

Il sito web di Screaming Frog è composto da 3 sottodomini:

1. <https://www.screamingfrog.co.uk>;
2. <https://cdn.screamingfrog.co.uk>;
3. <https://images.screamingfrog.co.uk>

Se utilizzi la configurazione base del Seo Spider e scansioni ad esempio il sottodominio A gli altri due “cdn.” e “images.” vengono trattati come link esterni e li puoi visualizzare nella tab “External Links” della finestra inferiore del Seo Spider.

Per fare un crawl completo di tutti i sottodomini hai a disposizione principalmente due alternative:

1. Abilitare l'opzione “Crawl All Subdomains”

Configuration > Spider > Crawl > Crawl All Subdomains

2. Scansionare il domain root di un sito (Es. <https://screamingfrog.co.uk>)

FARE LA SCANSIONE DI UNA SOTTOCARTELLA

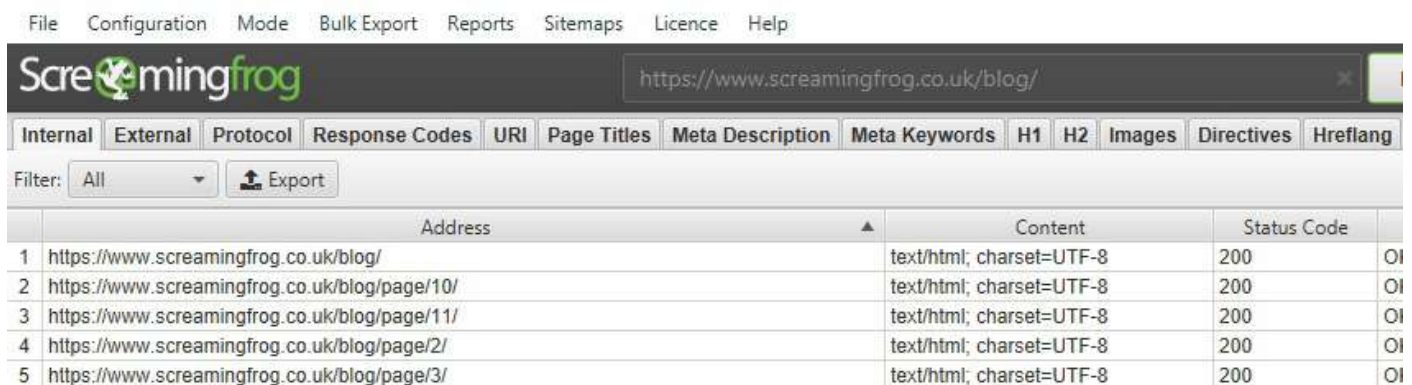
Per impostazione predefinita il Seo Spider esegue una scansione completa del “root domain” o sotto-dominio ma può capitare per esigenze di tempo o per rispondere ad un focus preciso che tu abbia la necessità di fare una scansione solo parziale di un sito web.

Vediamo insieme come ottenere un crawl di una sottocartella specifica.

ESEMPIO:

Vuoi analizzare gli errori e criticità presenti solamente nell'area blog del sito web di Screaming Frog:
<https://www.screamingfrog.co.uk/blog/>

Per questa scansione parziale è sufficiente inserire l'URL della sottocartella, avviare il Seo Spider e lasciare concludere il crawl. In questo modo troverai popolata la finestra principale con gli URL della sottocartella, i rispettivi filtri e metriche.



The screenshot shows the Screaming Frog SEO Spider application. The address bar displays <https://www.screamingfrog.co.uk/blog/>. The 'Internal' tab is selected in the top navigation bar. Below the navigation bar, there are tabs for 'Internal', 'External', 'Protocol', 'Response Codes', 'URI', 'Page Titles', 'Meta Description', 'Meta Keywords', 'H1', 'H2', 'Images', 'Directives', and 'Hreflang'. The 'Filter' dropdown is set to 'All'. An 'Export' button is visible. The main table displays the following data:

	Address	Content	Status Code	
1	https://www.screamingfrog.co.uk/blog/	text/html; charset=UTF-8	200	OI
2	https://www.screamingfrog.co.uk/blog/page/10/	text/html; charset=UTF-8	200	OI
3	https://www.screamingfrog.co.uk/blog/page/11/	text/html; charset=UTF-8	200	OI
4	https://www.screamingfrog.co.uk/blog/page/2/	text/html; charset=UTF-8	200	OI
5	https://www.screamingfrog.co.uk/blog/page/3/	text/html; charset=UTF-8	200	OI

Come potrai constatare durante i tuoi test, oltre agli URL della “subfolder”, saranno presenti anche ulteriori URL che non appartengono all'area che avevi limitato.

Questo comportamento anomalo però non è un bug del sistema ma rappresenta una funzionalità intrinseca del Seo Spider.

Infatti il crawler considera in prima istanza la sottocartella ma segue contestualmente anche tutti i link presenti proponendo i dati degli URL collegati.

Questo accorgimento è vitale per un'analisi più accurata e ti aiuta a scoprire eventuali collegamenti interrotti che partono dalla sottocartella verso URL esterni.

Configuration > Spider > Crawl Behaviour > Check Links Outside of Start Folder

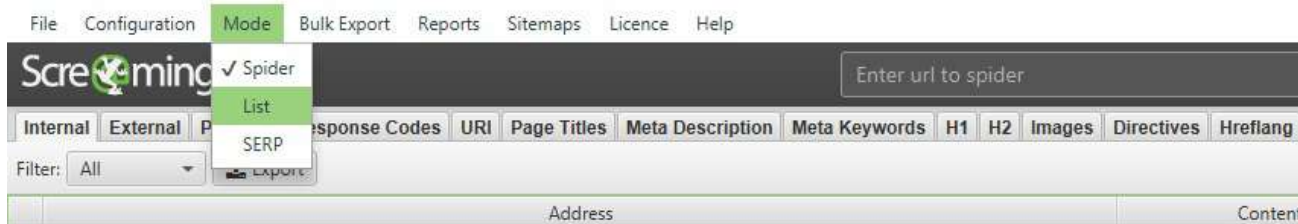
Nota Bene: il Seo Spider è in grado di scansionare una sottocartella solo con questa sintassi “./nome-cartella/”; se dimentichi lo “/” finale Screaming Frog non la riconosce. La medesima situazione si verifica quando la versione con “/” finale presenta un redirect verso la versione senza.

Un'ulteriore alternativa per scansionare una sottocartella è quella di utilizzare la funzione “include” sfruttando le regole RegEX.

Configuration > Include > ./blog.*

CRAWLER DI UN ELENCO DI URL

Fino a questo momento hai visto solamente la modalità di scansione “Spider” ma il Seo tool prevede ulteriori 3 tipologie di crawl: List, Compare e Serp. Iniziamo l’analisi ponendo l’attenzione sulla modalità “List - elenco” con la quale sei in grado di ottenere dati e metriche da una selezione specifica di URL.



Per utilizzare questa opzione puoi inserire manualmente gli URL, copiarli e incollarli nel Seo Spiders, fare un “Upload” di un documento in “.csv” o scegliere di far scaricare dal Seo Spider una sitemap xml.

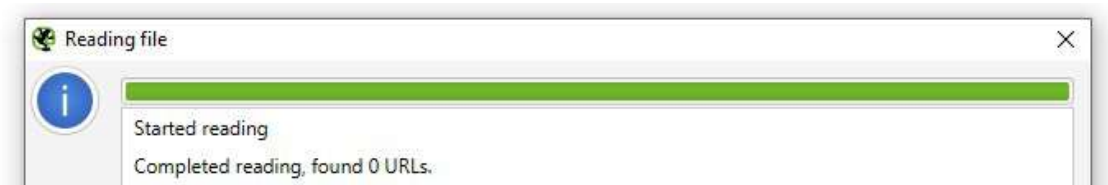
Le differenze sostanziali rispetto alla modalità “Spider” sono essenzialmente due:

1. la “Address Bar” è sostituita da un pulsante per eseguire l’upload.
2. La profondità di scansione nella modalità “List” è impostata a zero per default; Screaming Frog non esamina le immagini e i link esterni ma si concentra solamente sugli URL aggiunti.

Questa modalità di scansione è particolarmente utile durante la fase di migrazione per verificare la correttezza dei reindirizzamenti ed evitare fastidiose “Redirect Chain” o destinazioni con status code diverse da 200 che andrebbero a sprecare Crawling Budget.

Una volta conclusa la scansione puoi fare l’esportazione dei dati e delle metriche su un foglio di calcolo per gestire al meglio le criticità rilevate.

Nota Bene: il Seo Spider esamina solamente gli URL che presentano il protocollo http/https, nel caso mancasse restituisce un avviso di errore.



Quando scegli la modalità “List” non devi preoccuparti di avere URL duplicati nel tuo elenco perché vengono normalizzati direttamente da Screaming Frog.

ESEMPIO:

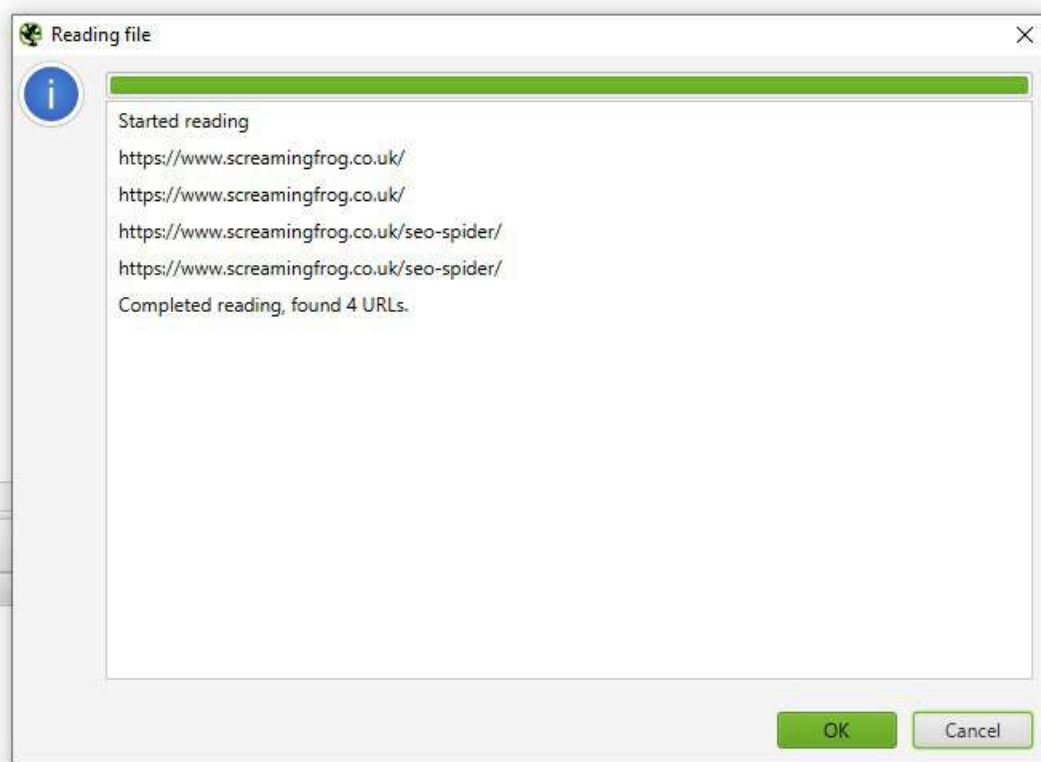
Aggiungiamo i seguenti 4 URL con la modalità “List”:

1. <https://www.screamingfrog.co.uk/>
2. <https://www.screamingfrog.co.uk/#fragment-this-wont-be-crawled>
3. <https://www.screamingfrog.co.uk/seo-spider/>
4. <https://www.screamingfrog.co.uk/seo-spider/>

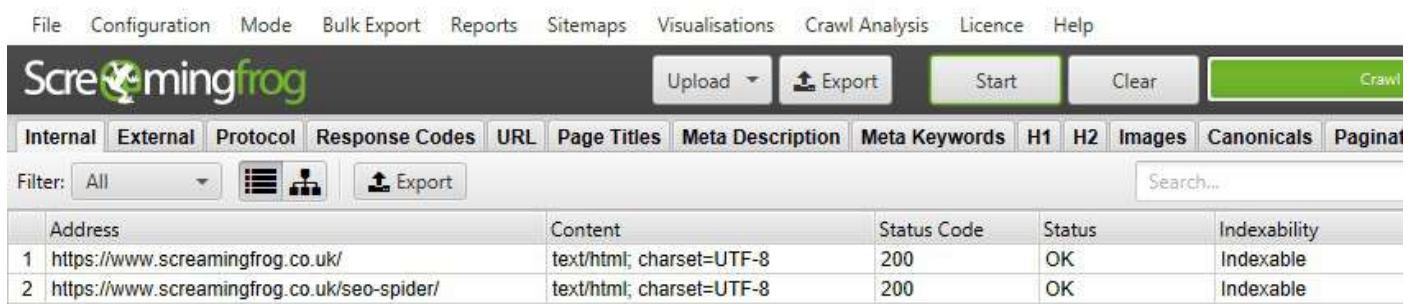
L’elenco presenta 4 elementi ma solamente due saranno processati dal Seo Spider.

Se guardi con attenzione il terzo e il quarto sono duplicati mentre il primo e il secondo differiscono solamente per un “fragment” (#) che Google non considera. Naturalmente in una lista così limitata le duplicazioni sono molto semplici da individuare ma se consideri delle liste con migliaia di URL questa normalizzazione da parte del tool diviene imprescindibile.

Vediamo nel dettaglio come si comporta Screaming Frog:



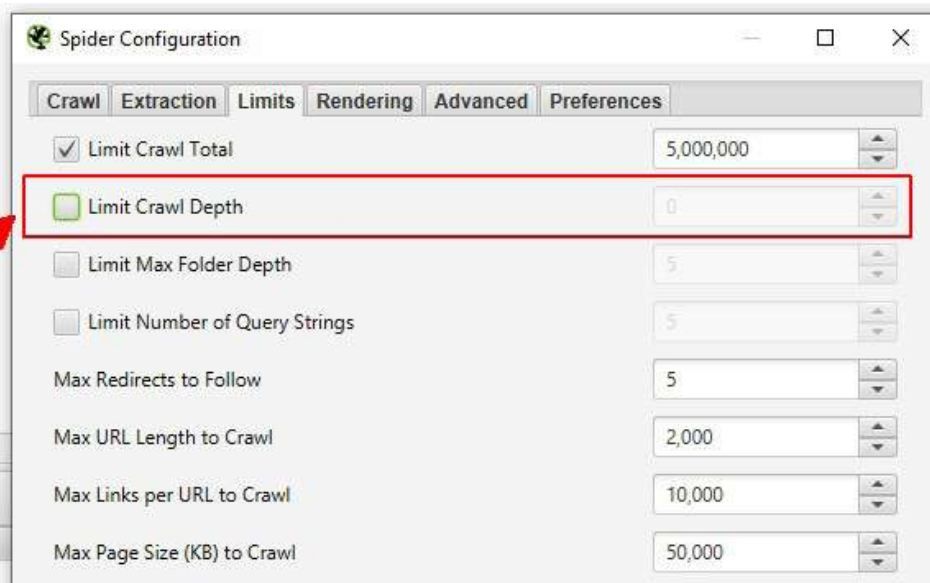
Il Seo Spider legge tutti e quattro gli URL aggiunti ma come vedrai in fase di crawling considera solamente quelli unici.



	Address	Content	Status Code	Status	Indexability
1	https://www.screamingfrog.co.uk/	text/html; charset=UTF-8	200	OK	Indexable
2	https://www.screamingfrog.co.uk/seo-spider/	text/html; charset=UTF-8	200	OK	Indexable

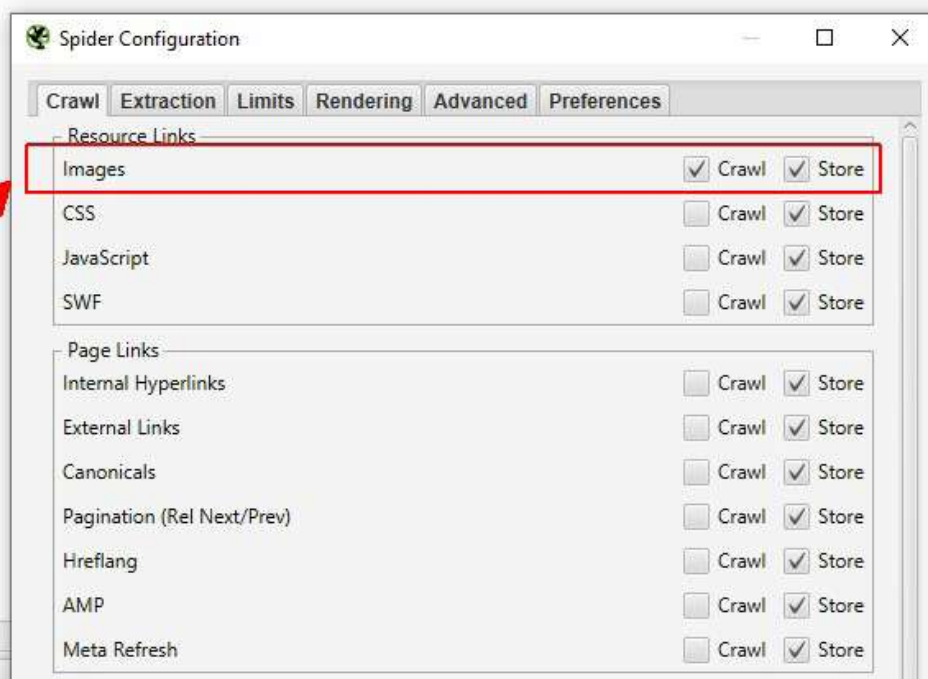
Questa modalità di scansione prevede la diagnosi solamente degli URL aggiunti ma come devi fare se necessiti di conoscere elementi supplementari come gli "external links", gli attributi "hreflang" o altro?

In questo caso l'unica cosa che devi modificare è il "Limit Crawl Depth - Limite di profondità di scansione" che è impostato a zero. E' sufficiente deselezionare questa opzione per eseguire un crawl inclusivo degli URL in lista e di tutti quelli scoperti durante la scansione attraverso i link presenti che si trovano nello stesso sottodominio.



Anche con la modalità “List” puoi decidere il livello di granularità della tua analisi scegliendo quali metriche ed elementi includere e quali non considerare dalla scansione gestendo le opzioni “Resource Links” e “Page Links”.

Config > Spider > Crawl



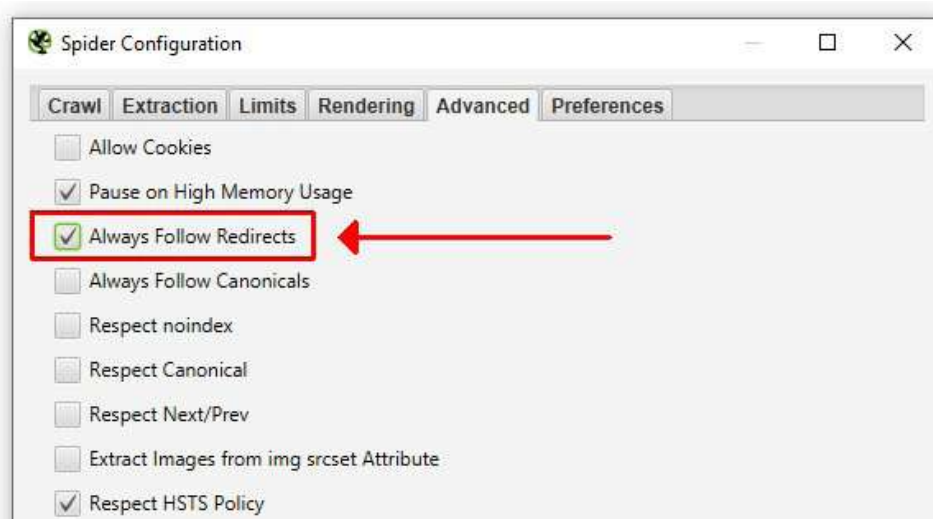
REDIRECTS E MIGRAZIONE

La modalità di scansione “List” è particolarmente utile nella gestione dei reindirizzamenti 3xx delle migrazioni di siti web in cui varia l'architettura o cambia il nome del dominio perché Google, attraverso i “Link Signals”, trasmette il ranking tra i diversi URL.

Mediante l'utilizzo del Seo Spider sei in grado di diagnosticare tutte le criticità relative ai “redirect”, valutare lo “Status Code” degli URL di destinazione ed evitare “Redirect Chain” o loop che consumano inesorabilmente il crawling budget del Motore di Ricerca.

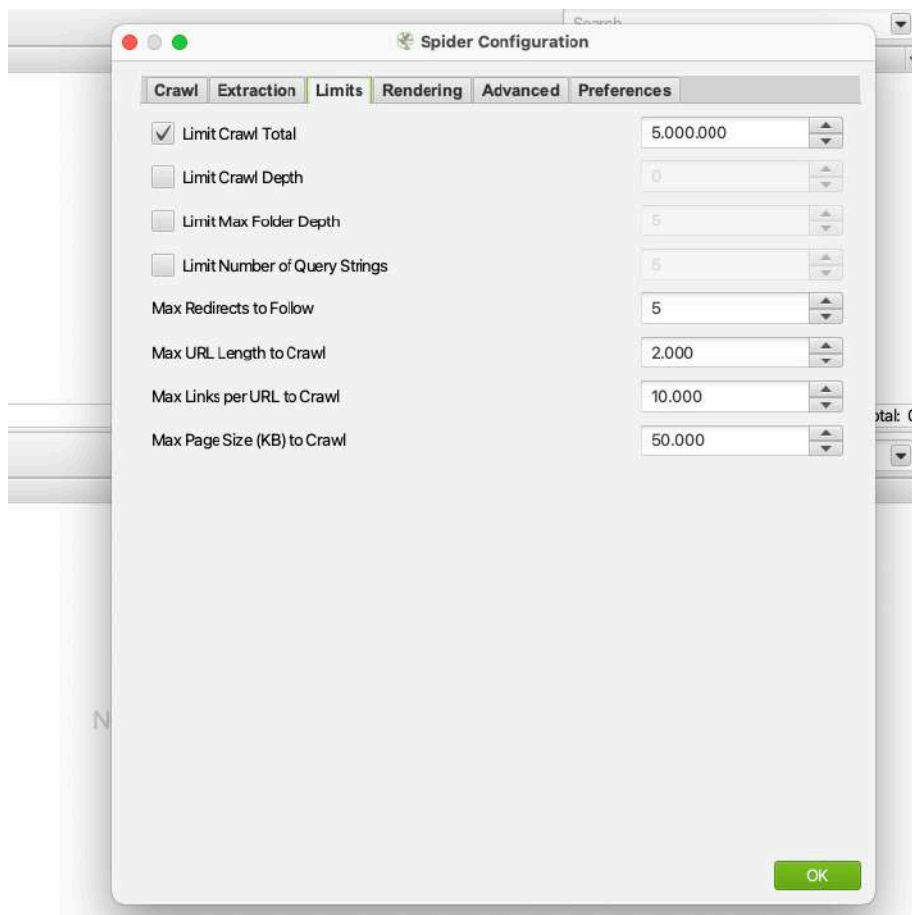
Per ovviare a questa problematica è sufficiente selezionare l'opzione “Always Follow Redirects” nelle impostazioni.

Config > Spider > Avanzate > Always Follow Redirects



Abilitando la funzione “Always Follow Redirects” il Seo Spider segue il reindirizzamento fino al primo URL che risponde con uno status code diverso da 3xx. Per default il limite massimo è impostato a 5 ma puoi modificare questa soglia intervenendo su “Max Redirect to Follow”.

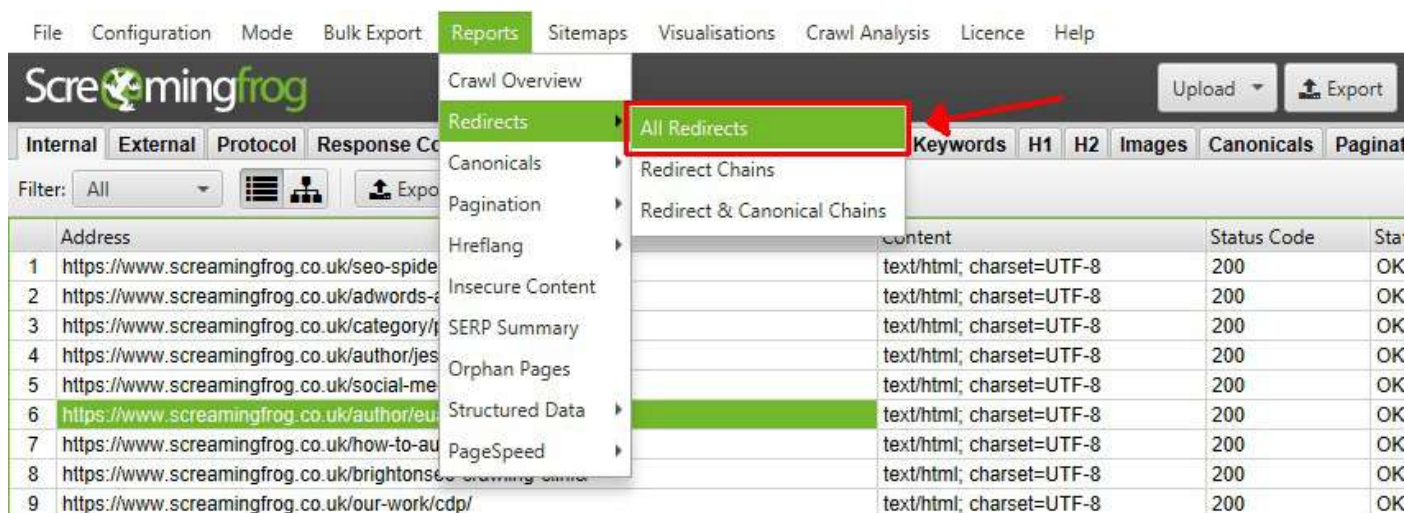
Config > Spider > “Max redirect to Follow”



Definito il setup iniziale puoi eseguire la scansione.

Conclusa la fase di crawling hai a disposizione un report con tutti i dettagli dei reindirizzamenti da poter esportare.

Reports > All Redirects



Il rapporto "All Redirects" è un documento molto articolato e completo che ti permette di valutare in modo semplice ed intuitivo lo status dei tuoi reindirizzamenti.

Il documento è composto dai seguenti campi:

1. **Chain Type:** specifica la tipologia del reindirizzamento (es. Http/Javascript/Meta Refresh).
2. **Number of Redirects:** rappresenta il numero di nodi presenti nella catena dei redirect prima di trovare un URL con uno status code differente da 3xx.
3. **Redirect Loop:** identifica se ci sono dei redirect circolari e li definisce con True/False secondo regole booleane.
4. **Temp redirect in Chain:** segnala se esiste un reindirizzamento temporaneo nella catena (es. 302).
5. **Address:** l'indirizzo dell'URL di partenza in cui è presente un redirect 3xx.
6. **Final Address:** la destinazione finale del reindirizzamento.
7. **Indexability:** identifica se l'URL di destinazione (Final Address) è indicizzabile o non indicizzabile. Questo campo è molto utile per comprendere se durante la migrazione hai linkato un vecchio URL verso una destinazione preclusa ai Motori di Ricerca.
Se una risorsa l'hai gestita con un reindirizzamento è perché volevi passare il trust verso una nuova pagina ma se questa presenta l'attributo "Non Indexable" il lavoro precedente non porterebbe alcun effetto.
8. **Indexability Status:** identifica la ragione per cui l'indirizzo di destinazione non è indicizzabile, per esempio presenta il tag "No Index".
9. **Final Content:** mostra la tipologia del contenuto di destinazione (esempio: html).
10. **Final Status Code:** offre lo Status Code della destinazione finale e può essere 200, 4xx, 1xx, 5xx o "No Response". Anche questo campo è molto importante da valutare per evitare di sprecare il trust del reindirizzamento verso una risorsa che presenta errori di server, pagine non trovate o altro.
In sintesi se non vedi lo status code 200 devi fermarti ed intervenire.

Oltre a queste colonne fisse hai a disposizione delle colonne variabili che mappano in modo dettagliato ogni nodo delle "Redirect Chain".

Questo report è un vero e proprio Sacro Graal che ti permette di salvare il crawling budget, revisionare eventuali decisioni prese in atto di migrazione e non far penalizzare il tuo sito web.

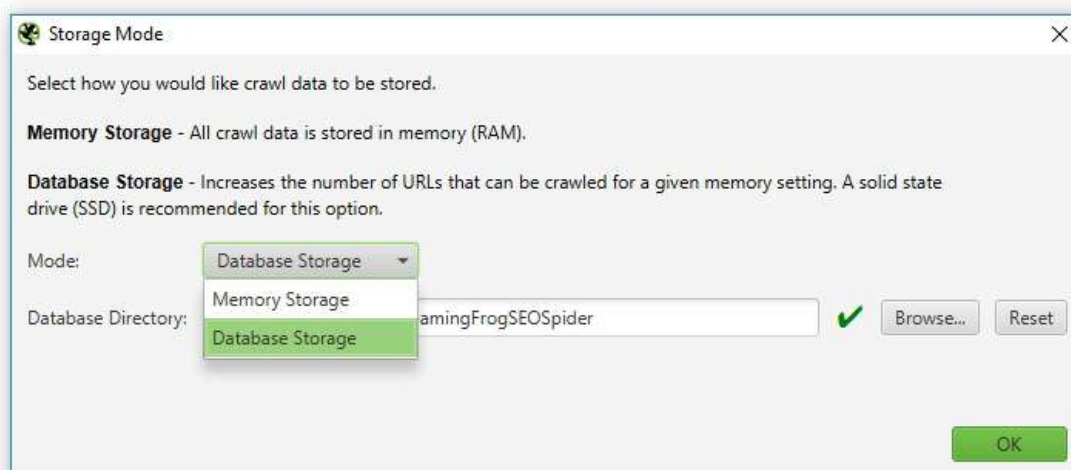
CRAWLER SITI WEB DI GRANDI DIMENSIONI

Screaming Frog rappresenta un ottimo strumento di diagnosi per tutte le taglie di siti web e rappresenta il partner ideale anche per progetti molto corposi composti da milioni di URL grazie alle diverse modalità di archiviazione dei dati.

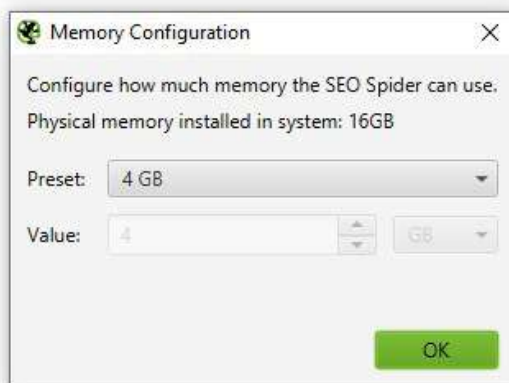
GESTIONE MODALITÀ' DI ARCHIVIAZIONE

Il Seo Spider offre due tipologie di archiviazione dei dati: "Memory Storage" e "Database Storage" che possono essere gestite dal menu di configurazione.

Config. > System > Memory Allocation



Per entrambe le modalità puoi decidere quanta RAM dedicare in base al tuo sistema. Ricorda che questa configurazione necessita del riavvio del Seo Spider per essere effettiva.



DIFFERENZE TRA MEMORY E DATABASE STORAGE

MEMORY STORAGE

La modalità “Memory Storage” prevede il salvataggio del crawl nella RAM ed è indicata per domini sotto i 500K URL.

Oltre a rappresentare un'opzione migliore per i siti web più piccoli, questa modalità di archiviazione è raccomandata per le macchine senza SSD, o dove non c'è molto spazio su disco.

Le scansioni possono essere salvate manualmente con estensione “.seospider”.

ARCHIVIAZIONE DATI SU DATABASE

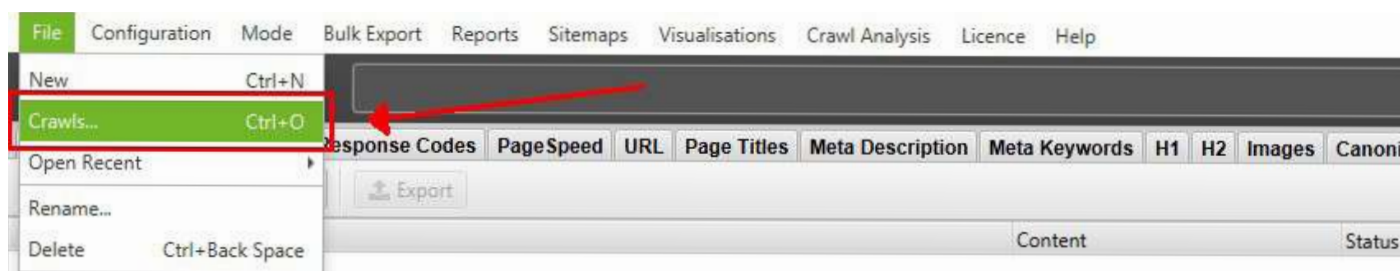
Il “Database Storage” è consigliata per crawling molto corposi e per tutti gli utenti che hanno sistemi con schede di archiviazione SSD.

Il Seo Spider è impostato di default con il limite di scansione a 5 milioni di URL, ma questa soglia può essere personalizzata in base alle diverse esigenze di progetto.

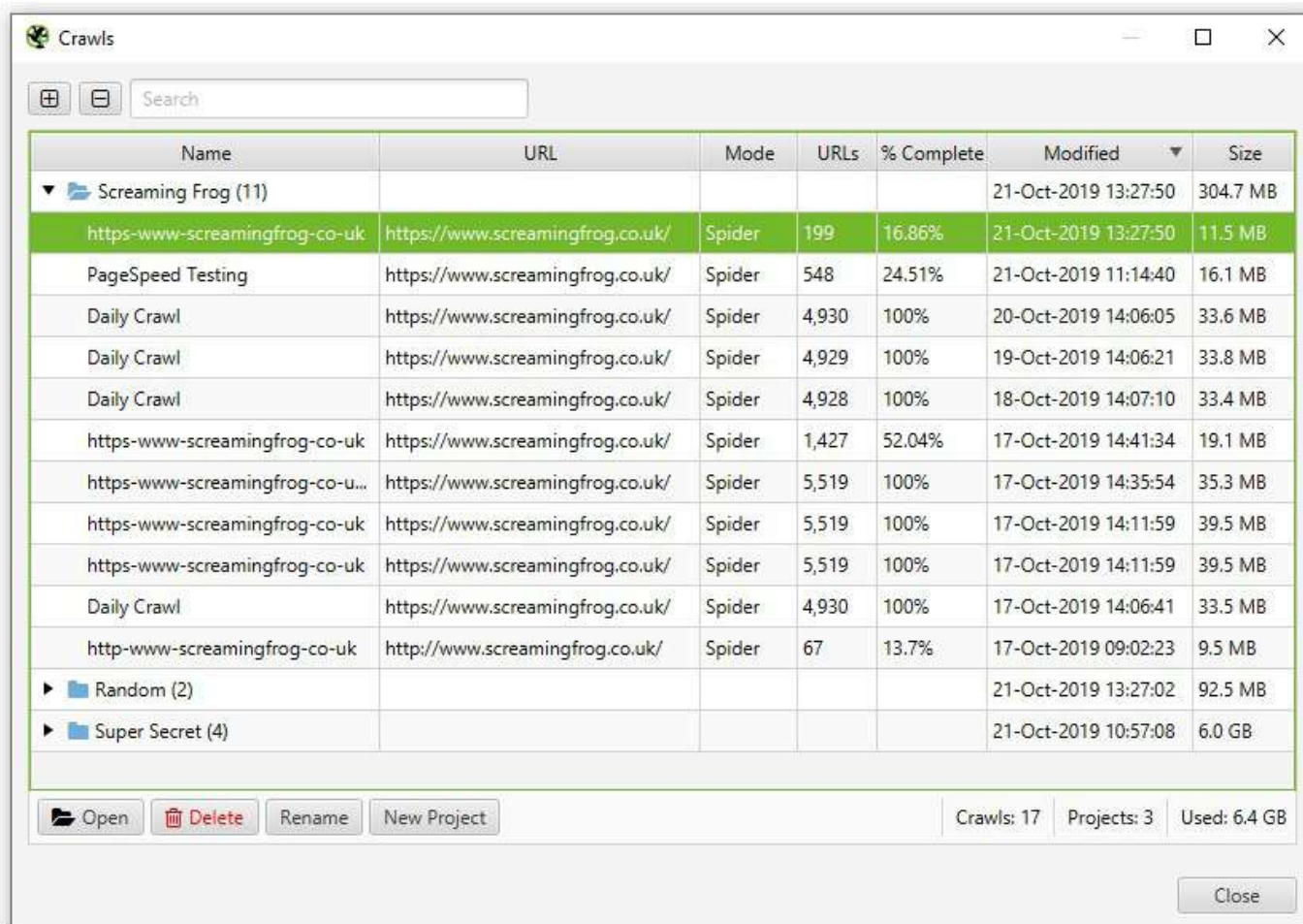
Per esempio, una macchina con un SSD da 500gb e 16gb di RAM, ti permette di scansionare fino a 10 milioni di URL circa.

Il “Database Storage” prevede il salvataggio automatico delle scansioni, molto utile in caso di interruzioni inaspettate durante il processo di crawl.

Ulteriore vantaggio del Database Storage è rappresentato dai tempi di caricamento dei “crawl” salvati; anche con progetti oversize da milioni di URL e metriche la scansione sarà richiamata in pochi secondi.



Con la modalità di archiviazione a database puoi richiedere tutte le scansioni storiche direttamente dalla voce di menu “File” selezionando l'opzione “Crawls”.



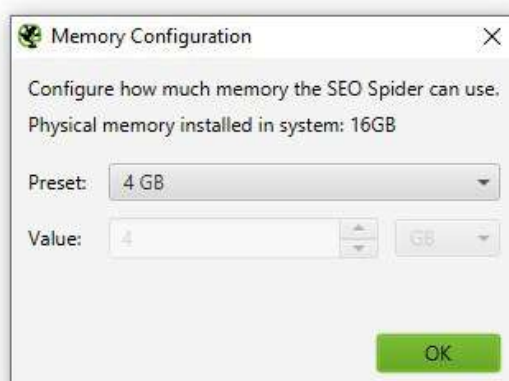
The screenshot shows the 'Crawls' window in Screaming Frog. It contains a table with columns: Name, URL, Mode, URLs, % Complete, Modified, and Size. The table lists several crawls, including 'Screaming Frog (11)' and 'Random (2)'. The 'Screaming Frog (11)' group is expanded, showing individual crawls for 'https-www-screamingfrog-co-uk' and 'Daily Crawl'. The 'Random (2)' group is also expanded, showing 'Super Secret (4)'. At the bottom, there are buttons for 'Open', 'Delete', 'Rename', and 'New Project', along with summary statistics: 'Crawls: 17', 'Projects: 3', and 'Used: 6.4 GB'. A 'Close' button is in the bottom right corner.

Name	URL	Mode	URLs	% Complete	Modified	Size
▼ Screaming Frog (11)					21-Oct-2019 13:27:50	304.7 MB
https-www-screamingfrog-co-uk	https://www.screamingfrog.co.uk/	Spider	199	16.86%	21-Oct-2019 13:27:50	11.5 MB
PageSpeed Testing	https://www.screamingfrog.co.uk/	Spider	548	24.51%	21-Oct-2019 11:14:40	16.1 MB
Daily Crawl	https://www.screamingfrog.co.uk/	Spider	4,930	100%	20-Oct-2019 14:06:05	33.6 MB
Daily Crawl	https://www.screamingfrog.co.uk/	Spider	4,929	100%	19-Oct-2019 14:06:21	33.8 MB
Daily Crawl	https://www.screamingfrog.co.uk/	Spider	4,928	100%	18-Oct-2019 14:07:10	33.4 MB
https-www-screamingfrog-co-uk	https://www.screamingfrog.co.uk/	Spider	1,427	52.04%	17-Oct-2019 14:41:34	19.1 MB
https-www-screamingfrog-co-u...	https://www.screamingfrog.co.uk/	Spider	5,519	100%	17-Oct-2019 14:35:54	35.3 MB
https-www-screamingfrog-co-uk	https://www.screamingfrog.co.uk/	Spider	5,519	100%	17-Oct-2019 14:11:59	39.5 MB
https-www-screamingfrog-co-uk	https://www.screamingfrog.co.uk/	Spider	5,519	100%	17-Oct-2019 14:11:59	39.5 MB
Daily Crawl	https://www.screamingfrog.co.uk/	Spider	4,930	100%	17-Oct-2019 14:06:41	33.5 MB
http-www-screamingfrog-co-uk	http://www.screamingfrog.co.uk/	Spider	67	13.7%	17-Oct-2019 09:02:23	9.5 MB
► Random (2)					21-Oct-2019 13:27:02	92.5 MB
► Super Secret (4)					21-Oct-2019 10:57:08	6.0 GB

Da questa finestra sei in grado di ricercare, vedere le date e percentuale di completamento dei progetti, il numero e dimensione dei crawl e decidere di rinominare o rimuovere alcune scansioni.

PROBLEMI DI MEMORIA ALLOCATA

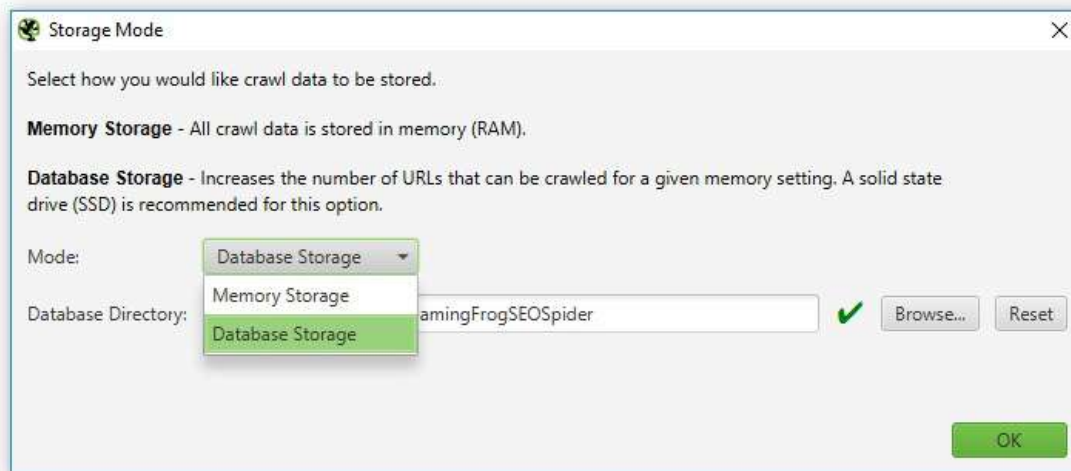
Come accennato nei paragrafi precedenti Screaming Frog ti permette di personalizzare quanta memoria allocare per le tue scansioni. Se durante un crawl la memoria è insufficiente il Seo Spider ti mostrerà un avviso di errore come questo.



COME SCANSIONARE UN SITO DI GRANDI DIMENSIONI

1. SCEGLI LA MODALITÀ DI ARCHIVIAZIONE “DATABASE STORAGE”

Per gestire la scansione di un sito da milioni di URL la modalità “Memory Storage” con archiviazione sulla RAM non è la scelta consigliata, per cui il primo passo che ti consiglio è lo switch a “Database Storage”.



Per ottenere il massimo delle performance dal Seo Spider con questa modalità dovresti prevedere una scheda di memoria SSD interna o esterna se il tuo sistema supporta la modalità UASP.

2. INCREMENTA LA “MEMORY ALLOCATION”

Scelta la modalità “Database Storage” il secondo passo è quello di incrementare la memoria allocata.

Di seguito un confronto di performance in base al numero massimo di URL che possono essere crawl-rati.

MEMORIA ALLOCATA	RAM MODE	STORAGE DATABASE
8 GB (REQUISITO MINIMO PER SITI GROSSE DIMENSIONI RAM MODE)	200 MILA URL	
4 GB		2-3 MILIONI DI URL
8 GB		5 MILIONI DI URL
16 GB		10 MILIONI DI URL

Questi requisiti sono solo dei riferimenti, parti sempre dall'idea che maggiore è la memoria allocata migliori saranno le performance del Seo Spider.

3. SCEGLI QUALI ELEMENTI E ATTRIBUTI SCANSIONARE

Non sempre è necessario scansionare un intero sito quindi è buona norma adattare il crawl agli specifici focus dell'analisi in essere. Ricorda che più dati si raccolgono e più intensa sarà la richiesta di RAM.

Per scegliere quali metriche scansionare e quali non includere puoi modificare direttamente le impostazioni del Seo Spider.

Configurazione > Spider > Crawl

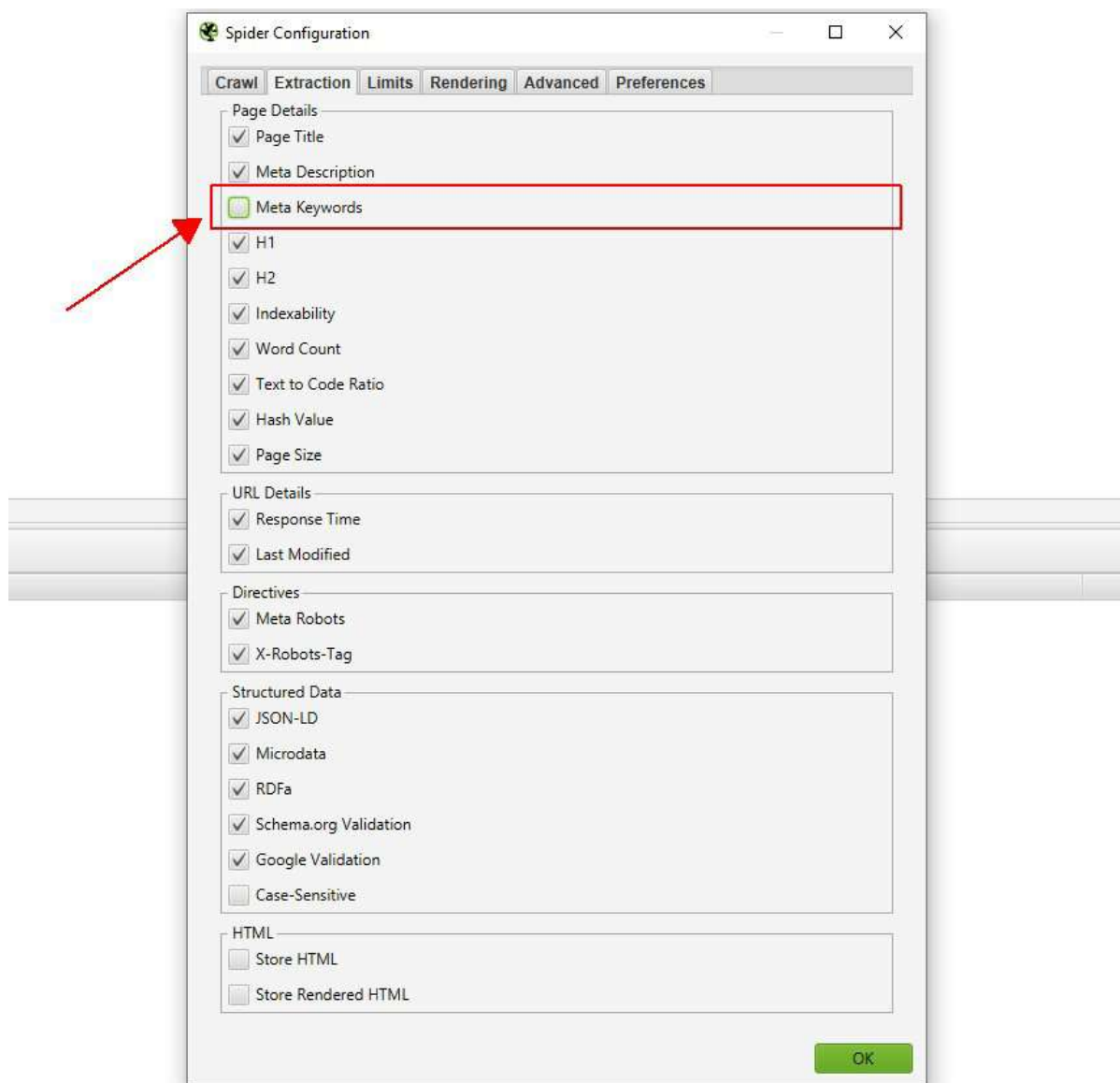
In base alla tipologia di analisi puoi scegliere se disabilitare la scansione e l'archiviazione dei seguenti elementi:

- Images*;
- CSS*;
- JavaScript*;
- SWF*.

*In termini di memoria ha molto senso disabilitare queste risorse ma nel caso utilizzi la modalità di rendering JavaScript dovrai abilitarle per non penalizzare il rendering.

- External Links;
- Canonicals;
- Pagination;
- Hreflang.
- AMP.

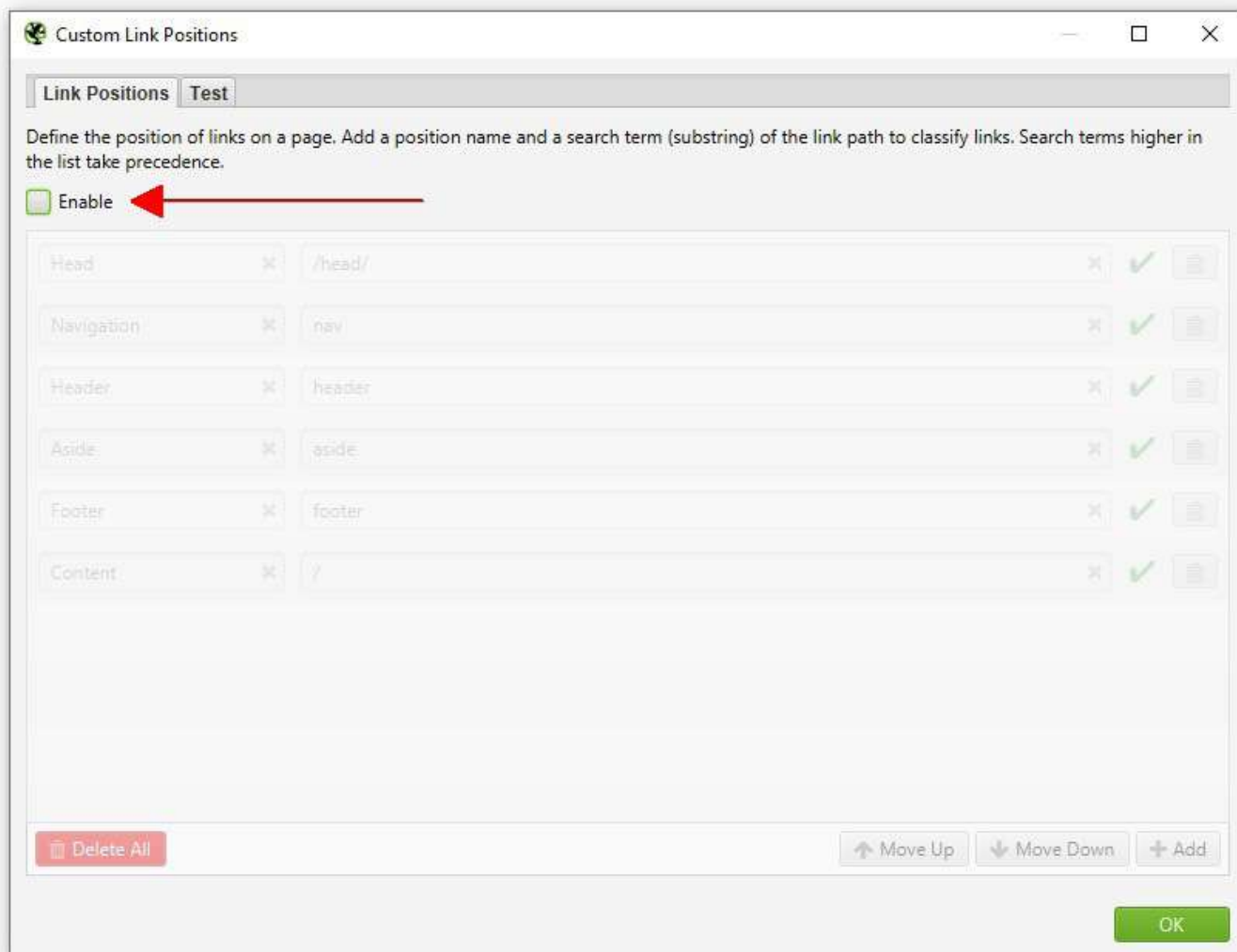
Altresì puoi disabilitare l'estrazione di alcuni elementi non essenziali come le "Meta Keyword" direttamente dalla schermata "Spider Configuration" o altri attributi che non siano strettamente necessari al momento dell'analisi scegliendo tra i "Page Details" (es. page title, word count, headings etc.), URL details, direttive o dati strutturati.



Oltre agli elementi sopra citati ti consiglio di disabilitare la funzione relativa alla posizione dei link nell'Html.

Config > Custom > Link Positions

La "Custom Link Position" memorizza l'XPath di ogni link trovato in un crawl per classificare la sua posizione (navigazione, barra laterale, piè di pagina ecc.) e richiede molta memoria per essere processata.



Altra best practice è quella di evitare la scansione di funzioni avanzate come le seguenti che possono essere incluse in una seconda fase di analisi:

- Custom Search - Ricerca personalizzata.
- Custom Extraction - Estrazione personalizzata.
- Integrazione di Google Analytics.
- Integrazione di Google Search Console.
- Integrazione di PageSpeed Insights.
- Spelling & Grammar - Ortografia e grammatica.
- Integrazione di Link Metrics (Majestic, Ahrefs e Moz).

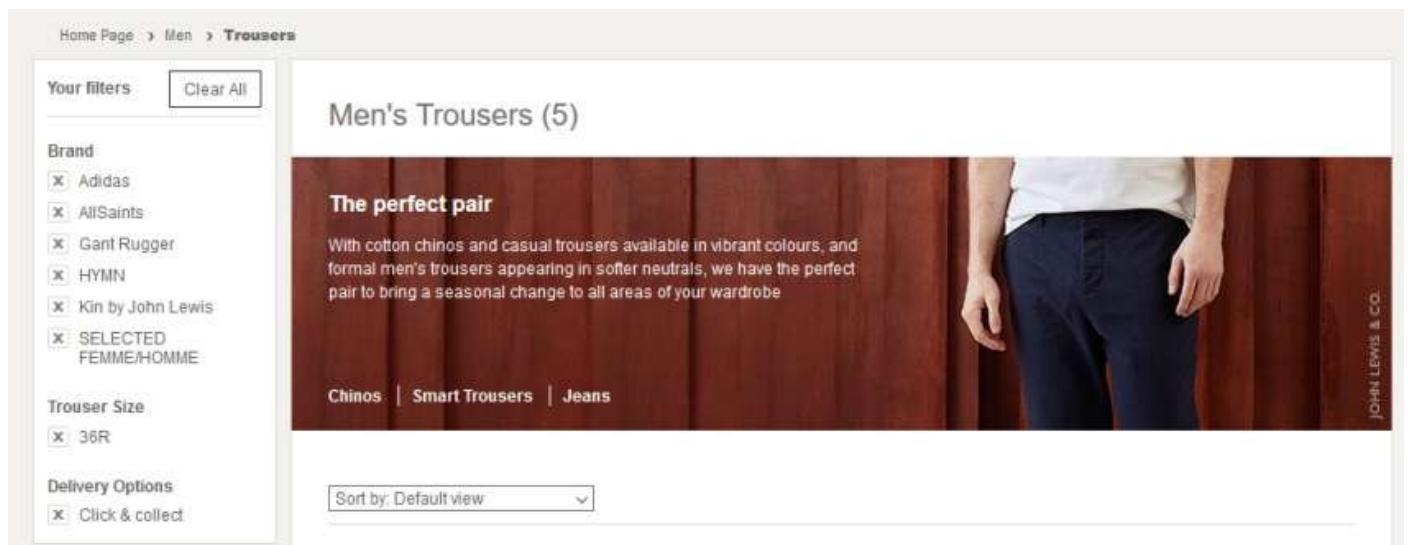
4. ESCLUDI GLI URL NON NECESSARI

Per rendere la scansione maggiormente snella e meno impattante in termini di memoria puoi limitare il crawl ad aree specifiche del sito web ed escludere elementi come gli ordinamenti generati dalle “faceted navigation”, particolari parametri presenti sugli URL o URL infiniti con directory ripetute utilizzando la funzione “exclude”.

Questa funzione attraverso espressioni RegEX ti permette di escludere gli URL che non desideri scansionare. Fai però attenzione che se gli URL esclusi contengono dei collegamenti ad altre pagine anche queste ultime non saranno crawlerate.

ESEMPIO

Per comprendere al meglio l'importanza della funzione “Exclude” prendiamo in esame uno scenario reale, come il sito web John Lewis.



Utilizzando le impostazioni di default del Seo Spider, a causa dei numerosi ordinamenti, la scansione comprenderebbe una serie infinita di URL come da esempio qui sotto:

https://www.johnlewis.com/browse/men/mens-trousers/adidas/allsaints/gant-rugger/hymn/kin-by-john-lewis/selected-femme/homme/size=36r/_/N-ebiZ1z13yvzZ1z0g0g6Z1z04nruZ1z0s0laZ1z13u9cZ-1z01kl3Z1z0swk1 .

Questa condizione è molto comune negli e-commerce che utilizzano le “faceted navigation” per le marche, per il colore, la vestibilità dei pantaloni e altro ancora creando infinite combinazioni e appesantendo inesorabilmente il crawling.

Nel caso specifico ti basterà eseguire una scansione campionaria del sito web per individuare questa tipologia di modello di URL parametrato ed escluderlo.

5. ESEGUIRE UNA SCANSIONE SU SOTTOCARTELLE O SOTTODOMINI

Se il sito web è molto consistente, puoi considerare la scansione separata dei sottodomini o ancora più granulare di particolari sottocartelle.

Per impostazione predefinita, il SEO Spider scansiona solo il sottodominio inserito, e tutti gli altri sottodomini incontrati sono trattati come link esterni aparendo nella finestra inferiore della scheda ‘External’.

È possibile scegliere di scansionare tutti i sottodomini, ma ovviamente questo richiederà più memoria.

Il Seo Spider può anche essere configurato per scansionare solamente una sottocartella inserendo l'URL della "Subfolder" con relativo percorso e disabilitando "check links outside of start folder - controlla i link fuori dalla cartella iniziale" e "crawl outside of start folder - scansiona fuori dalla cartella iniziale".

Configurazione > Spider > Check links outside of start folder | Crawl outside of start folder

Ad esempio, ipotizziamo tu voglia scansionare solamente l'area blog del sito di Screaming Frog. La prima attività è quella di identificare la sottocartella ed eseguire il crawler; che nel nostro caso sarà: <https://www.screamingfrog.co.uk/blog/>.



The screenshot shows the Screaming Frog SEO Spider application. The address bar contains <https://www.screamingfrog.co.uk/blog/>. Below the address bar, there are tabs for Internal, External, Protocol, Response Codes, URI, Page Titles, Meta Description, Meta Keywords, H1, H2, Images, Directives, and Hreflang. The 'URI' tab is selected. Below the tabs, there is a 'Filter' dropdown set to 'All' and an 'Export' button. The main table displays the following data:

	Address	Content	Status Code
1	https://www.screamingfrog.co.uk/blog/	text/html; charset=UTF-8	200
2	https://www.screamingfrog.co.uk/blog/page/10/	text/html; charset=UTF-8	200
3	https://www.screamingfrog.co.uk/blog/page/2/	text/html; charset=UTF-8	200
4	https://www.screamingfrog.co.uk/blog/page/3/	text/html; charset=UTF-8	200
5	https://www.screamingfrog.co.uk/blog/page/4/	text/html; charset=UTF-8	200

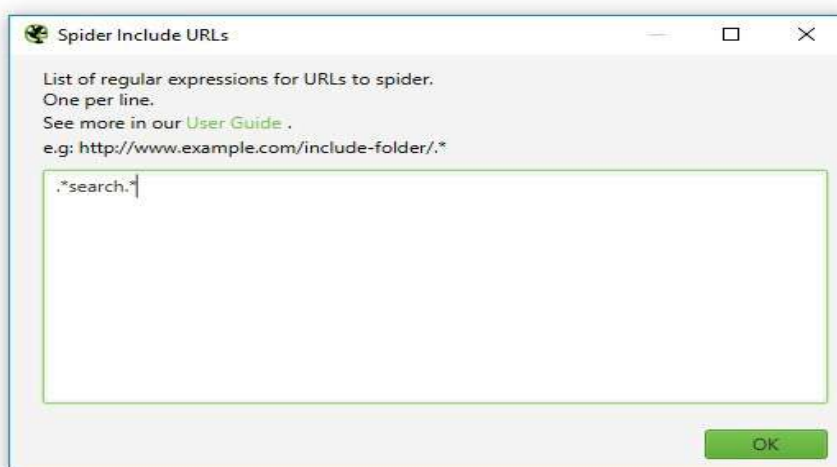
Nota Bene: per scansionare una sottocartella ricorda di aggiungere lo slash finale `"/blog/"` altrimenti il Seo Spider non la riconosce come tale. Lo stesso problema lo potresti avere se la versione con lo `"/"` viene reindirizzata verso la sua versione senza.

In questo secondo caso puoi usare la funzione "include" e utilizzare una regola RegEx per la scansione (ad esempio `.*blog.*`).

6. CIRCOSCRIVERE LA SCANSIONE UTILIZZANDO "INCLUDE"

Oltre all'esclusione di aree e/o alla disabilitazione di elementi da scansionare e archiviare puoi optare di sfruttare la funzione "Include" che ti permette di limitare gli URL da scansionare tramite le espressioni RegEx.

Per esempio, se vuoi scansionare solo le pagine che presentano "search" nella stringa dell'URL ti basta semplicemente includere la RegEx `.*search.*` nella funzione "include".



Prendendo in esame il sito di Screaming Frog otterresti i seguenti risultati

File Configuration Mode Bulk Export Reports Sitemaps Licence Help

Screamingfrog

https://www.screamingfrog.co.uk/

Internal External Protocol Response Codes URI Page Titles Meta Description Meta Keywords H1 H2 Images Directives Hreflang

Filter: HTML Export

	Address	Content	Status Code
1	https://www.screamingfrog.co.uk/	text/html; charset=UTF-8	200
2	https://www.screamingfrog.co.uk/search-engine-marketing/	text/html; charset=UTF-8	200
3	https://www.screamingfrog.co.uk/search-engine-optimisation/	text/html; charset=UTF-8	200

7. SCANSIONI E CAMPIONAMENTI

Un altro modo per risparmiare tempo e rendere la scansione meno impattante in termini di memoria è quella di lavorare su dei campionamenti del sito web attraverso:

- **Limit Crawl Total** - Limita la scansione totale : identifica la soglia massima di pagine da scansionare complessivamente.
- **Limit Crawl Depth** - Limita la profondità della scansione: limita la profondità di scansione alle pagine più prossime in termini di click dalla pagina di partenza (o la home) per ottenere un campione di tutti i modelli.
- **Limit Max URI Length To Crawl** - Lunghezza massima URL da scansionare: limita la scansione di URL fino ad un numero massimo di caratteri nella stringa evitando gli URL molto profondi.
- **Limit Max Folder Depth** - Limita la profondità massima delle cartelle: limita la scansione in base alla profondità delle cartelle.
- **Limit Number of Query Strings** - Limita il numero di stringhe di query: limita la scansione di parametri in base al numero di query string. Se imposti il limite delle stringhe di query a '1', permetti al Seo Spider di scansionare URL che presentano al massimo un singolo parametro (?=colore per esempio) nella stringa evitando combinazioni infinite.

COME SALVARE UN AUDIT SEO

Dopo la scansione è fondamentale salvare l'audit per non perdere traccia del lavoro fatto e utilizzarlo anche in un secondo momento.

Attualmente le modalità di salvataggio sono due e dipendono dal fatto che tu stia utilizzando il "Memory Storage Mode" o il "Database Storage" (consigliato).

Nel primo caso il "crawl" viene salvato con l'estensione ".seospider", tipica di Screaming Frog e ti basta cliccare "Save" dal menu principale.

File > Save

Per utilizzare il file sarà sufficiente il doppio click sul progetto oppure direttamente dal menu della console scegliendo "Apri Recenti".

File > Open Recent

File

Configuration

Mode

Bulk Export

Reports

Sitemaps

Licence

Help

Open

Ctrl+O

Open Recent

Save

Ctrl+S

Configuration

Crawl Recent

Exit

frog

https://www.screamingfrog.co.uk/

Protocol

Response Codes

URI

Page Titles

Meta Description

Meta Keywords

H1

H2

Images

Directives

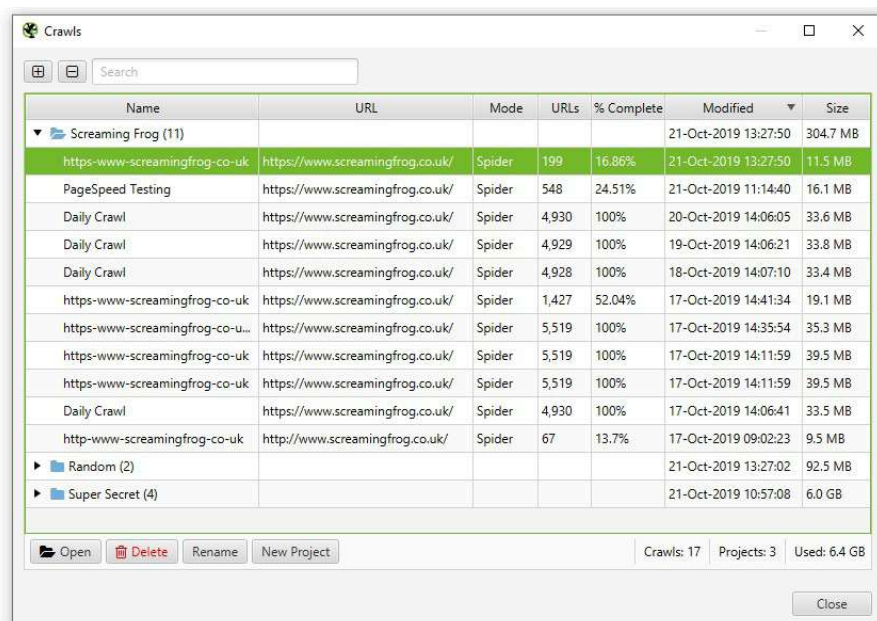
Hreflang

Export

Address	Content	Status Code
gfrog.co.uk/	text/html; charset=UTF-8	200
gfrog.co.uk/wp-content/themes/screamingfrog/public/images/frogbg1.png	image/png	200
gfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/gree...	image/png	200

Nel caso di "Database Storage" il salvataggio è automatico e la scansione la puoi trovare direttamente nel menu principale scegliendo l'opzione "Crawl".

Una nota positiva del “Database Storage” è rappresentata dalla velocità di apertura del crawl rispetto alla precedente modalità di salvataggio che utilizza il “Memory Storage”.



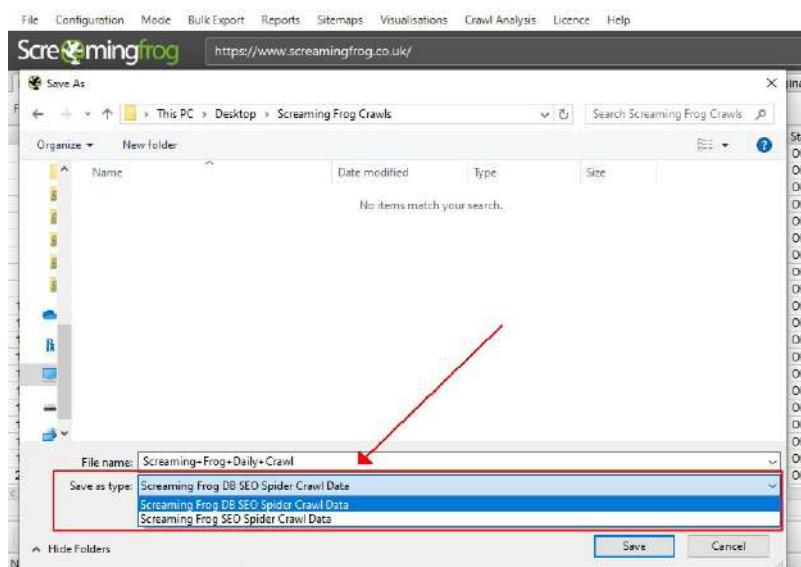
CONDIVISIONE DI UNA SCANSIONE

Anche la condivisione dei progetti risulta essere molto semplice grazie alla modalità di esportazione del Seo Spider.

A tua disposizione ci sono due formati di file che variano in base alla modalità di salvataggio prescelta:

1. ***.dbseospider**: questa opzione ti permette di importare il progetto nel Seo Spider in modalità database storage (opzione consigliata - molto veloce); non può essere importata in modalità “Memory storage”.

2. ***.seospider**: permette l'importazione sia in modalità “Memory” o “Database” Storage. L'importazione di questo crawl in modalità “Database storage” chiede un tempo discreto per la conversione automatica del file.



CONFIGURAZIONI PERSONALI CON IL SEO SPIDER

Una volta che hai preso dimestichezza con la maggior parte delle funzionalità di Screaming Frog la configurazione del crawl di default potrebbe non essere idonea per le tue esigenze specifiche.

Come sarebbe bello poter impostare dei “Crawls Profiles” esclusivi per ogni necessità da utilizzare come profili “Master”? Questa condizione farebbe risparmiare molto tempo e semplificherebbe i processi dell’Audit Seo.

Proviamo a pensare ad alcune esigenze che potrebbero sorgere in fase di crawling:

1. distinte scansioni per diversi user agent. Ad esempio per valutare il comportamento di Yandex rispetto a Bing o a Baidu sul tuo sito web.
2. Scansione dedicata a “Google per Smartphone” dato che il sito è già in fase “Mobile First Indexing”.
3. Sei in una fase contrattuale preliminare con il cliente e vuoi solo farti un’idea di massima sulla salute del sito web e non vuoi perdere tempo a scansionare immagini, css o altro che sprechi tempo e memoria.
4. Necessiti di estrapolare solo i meta tag da passare al copy senza sprecare memoria allocata per attributi come hreflang, canonical, etc.
5. Vuoi monitorare mensilmente lo stato di salute della sezione /blog/ del tuo sito web ma non vuoi continuare a mettere mano alle impostazioni di scansione.
6. Vuoi includere nella scansione solamente delle sezioni particolari del sito web.

L’elenco precedente rappresenta una parte infinitesima delle possibili occorrenze in cui puoi imbatterti giornalmente.

Ma allora che cosa puoi fare? Il Seo Spider permette di impostare e salvare una configurazione di profilo diversa per ogni esigenza di scansione che sarà richiamabile semplicemente con un click all’evenienza. In questo modo sei in grado di estrapolare immediatamente le informazioni di cui necessiti in modo sartoriale senza ulteriori modifiche delle configurazioni.

Vediamo quali sono le opzioni a disposizione per i “Crawl Profiles”:

- **Save Configuration as Default:** una volta configurate le metriche, gli attributi e le regole di scansione puoi utilizzare questa opzione per renderla di default. La modifica avrà effetto dopo la chiusura e riapertura del Seo Spider. Comunemente corrisponde al crawl generale nel Seo Audit.

File > Configuration > Save Configuration as Default.

- **Nuovo “Custom profile configuration”:** questa opzione ti permette di personalizzare gli elementi da scansionare per ogni singola esigenza specifica.
Nota Bene: non confondere il nuovo profilo con il punto precedente! Ogni qualvolta riaprirai Screaming Frog la scansione successiva si baserà sull’ultima configurazione di default impostata (punto precedente) e non sul nuovo profilo creato che dovrà essere richiamato dal menu.

File > Configuration > Save As > Nome del nuovo profilo.

Il nuovo profilo avrà un’estensione “.seospiderconfig”.

Il consiglio è quello di nominare la configurazione con testo descrittivo chiaro per renderla immediatamente riconoscibile nel tempo.

Ad esempio:

- Vuoi impostare un crawl che consideri come User Agent Google Mobile e utilizzi il Rendering JavaScript → googlemobile_javascript.seospiderconfig
- Vuoi un profilo per estrarre i dati relativi ai Meta Tag con user Agent Google Bot → metatag_googlebot.seospiderconfig

Questa funzionalità è molto utile per la condivisione immediata di una configurazione con i colleghi di un progetto e potersi confrontare anche a distanza.

Richiamare un particolare profilo di configurazione: dopo qualche tempo che smanetti con le configurazioni avrai la necessità di switchare da una all'altra in base alle nuove esigenze. Per richiamare una configurazione creata in precedenza ti basta cliccare su "Load" e scegliere dalla lista esistente.

File > Configuration > Load/Load recent

Clear Default Configuration: per ritornare alle impostazioni di default di Screaming Frog è sufficiente cliccare su Clear Default Configuration.

File > Configuration > Clear Default Configuration

SCHEDULING

Una delle funzionalità più utili del Seo Spider è quella di poter pianificare delle scansioni automatiche lasciando lavorare Screaming Frog anche in tua assenza.

Con la funzione di "Scheduling" puoi scegliere il giorno e l'ora di uno specifico crawl o calendarizzare una scansione perpetua con scadenza giornaliera, settimanale o mensile per un monitoraggio aggiornato sullo stato di salute di un sito web.

HOW TO:

COME ESEGUIRE UNA SCANSIONE PROGRAMMATA

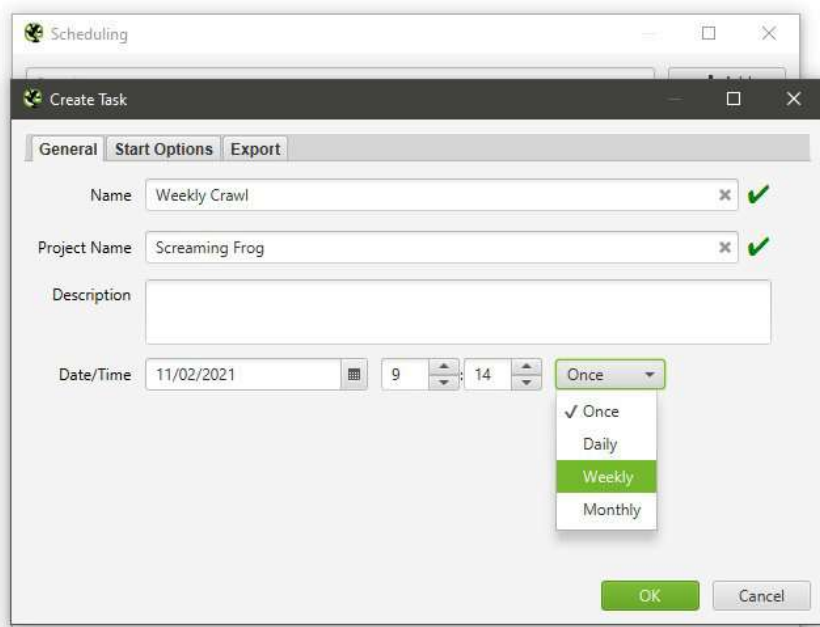
Compresa la potenzialità della funzione "Scheduling" non ti resta che configurarla seguendo alcuni step che vediamo di seguito.

File > Scheduling

1. Definire la task

In questa prima fase puoi definire alcuni aspetti generali del crawl:

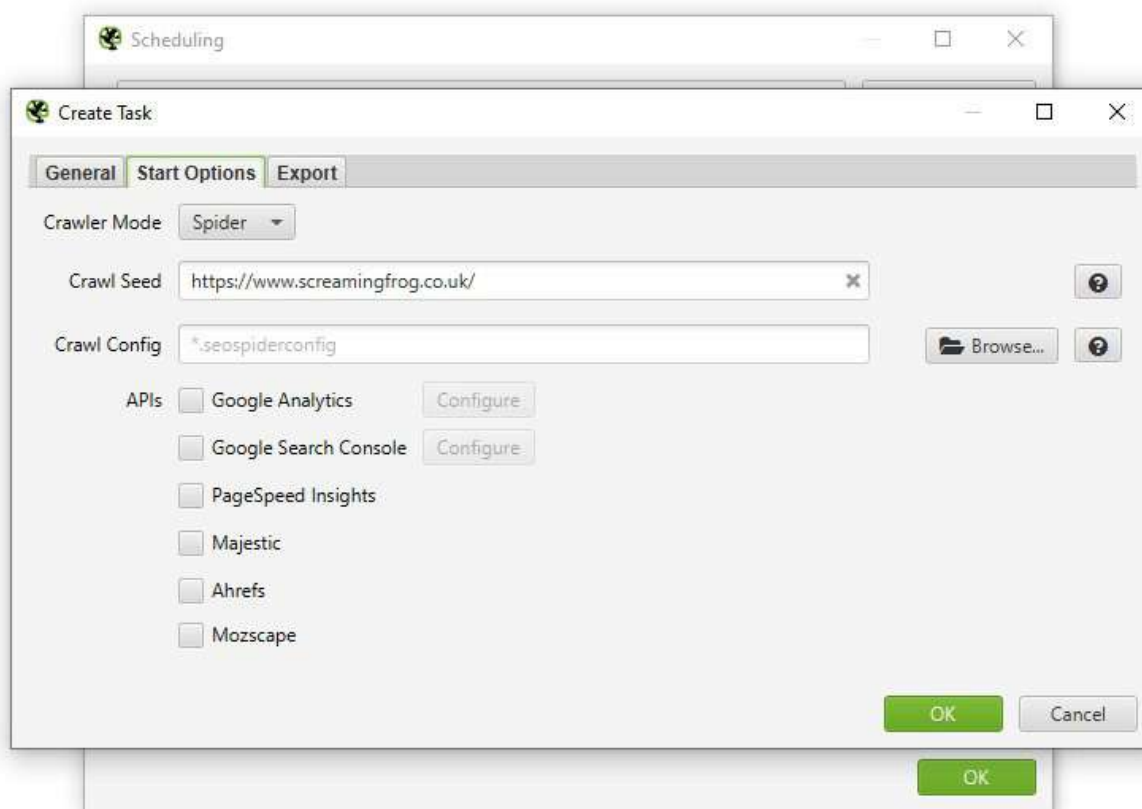
- **Name:** nome descrittivo della scansione.
- **Project Name:** nome del progetto in cui inserire la scansione.
- **Description:** molto utile per definire testualmente tutte le peculiarità e opzioni che caratterizzeranno il crawl.
- **Date/Time:** la data, l'ora della scansione e la sua periodicità (una tantum, giornaliera, settimanale o mensile).



2. Definire la configurazione del Seo Spider

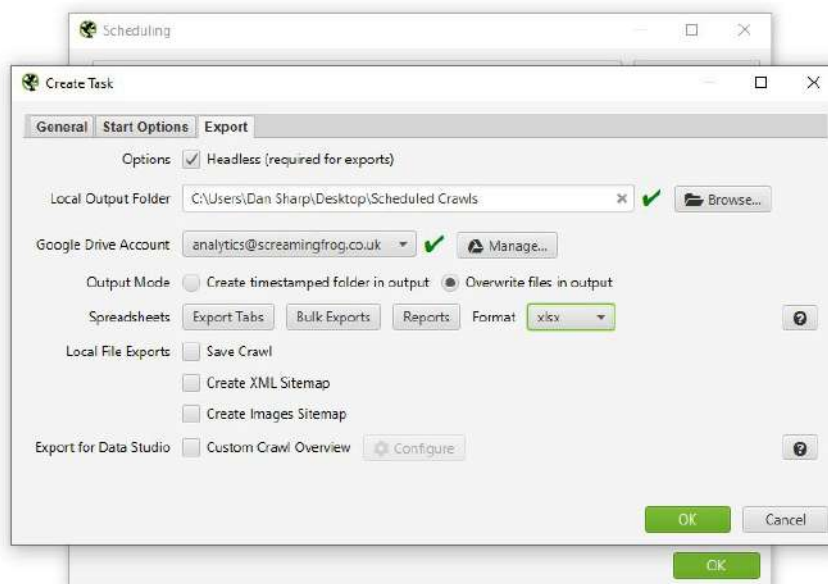
Nella seconda tab “Start Options” puoi scegliere la modalità di scansione tra “spider” e “list”, inserire il “root domain” o sottodominio e definire quale configurazione applicare al crawl, scegliendo tra i vari “Crawl Profiles” (se esistono).

In questa scheda puoi inoltre integrare attraverso API la scansione con i dati di Google Analytics, Search Console, Moz, Majestic e Ahrefs per una visione maggiormente granulare del progetto.

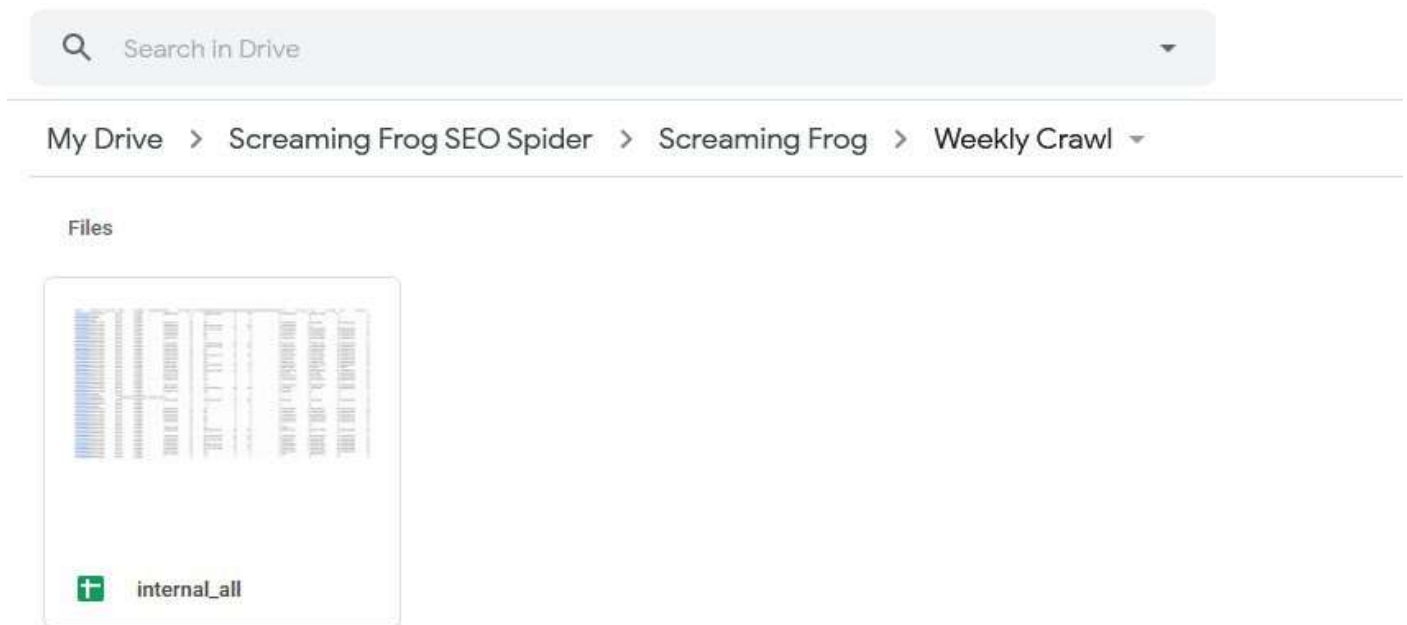


3. Impostare il metodo di salvataggio del Seo Audit

Attraverso la scheda “Export” sei in grado di definire le tabs, i “Bulk Exports” e i rapporti che vuoi esportare. Puoi scegliere di scaricare solamente una specifica scheda (es. “Internal”, “meta description” o “headings”) o combinare più tabs e rapporti selezionandoli dalla finestra a disposizione. I dati saranno serviti di default in formato “.csv” e salvati in locale o in formato “gsheet”. Scegliendo questa seconda opzione troverai i documenti esportati nella cartella “Screaming Frog SEO Spider” di Google Drive (dovrai aver collegato il tuo account Google).



Di seguito un esempio di un documento esportato su Google Drive.

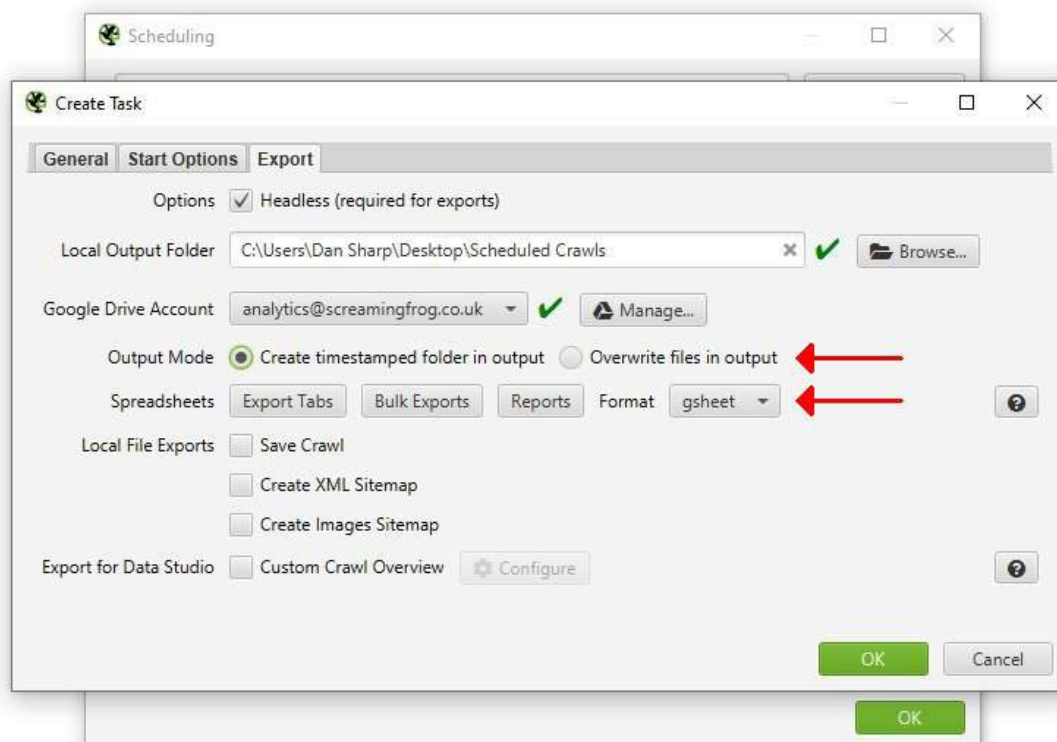


Un'ulteriore opzione molto comoda dello "Scheduling" periodico è quella di poter scegliere se, al termine di ogni crawl programmato, sovrascrivere in automatico il documento originario così da conservare solamente l'ultima scansione, oppure creare un "timestamped" che aggiunge un nuovo documento nella cartella e mantiene traccia dello storico dei crawl.

Nota Bene: utilizzando la modalità di archiviazione "Database Storage" anche la scansione programmata è sempre a tua disposizione per essere riutilizzata in un secondo momento.

File > Crawls

La funzione "Scheduling" lavora in modalità "Headless" (senza interfaccia) per evitare interazioni dell'utente non volute. Nel caso in cui ci siano più istanze di crawl programmate e sovrapposte Screaming Frog le esegue nello stesso momento. Questa condizione richiede maggiori risorse al sistema, i tempi potrebbero allungarsi e il computer potrebbe rallentare.



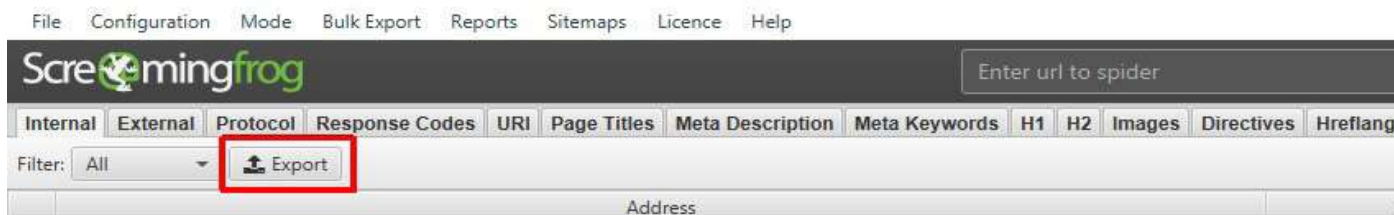
ESPORTAZIONE DEI DATI DI UN CRAWL

Il Seo Spider offre essenzialmente tre metodi per esportare le metriche ed indicatori ottenuti nella scansione:

1. Utilizzare la funzione “Export” presente sulle singole schede

La prima opzione a tua disposizione ti consente di esportare i dati popolati direttamente dalle singole tab cliccando sul pulsante “Export”

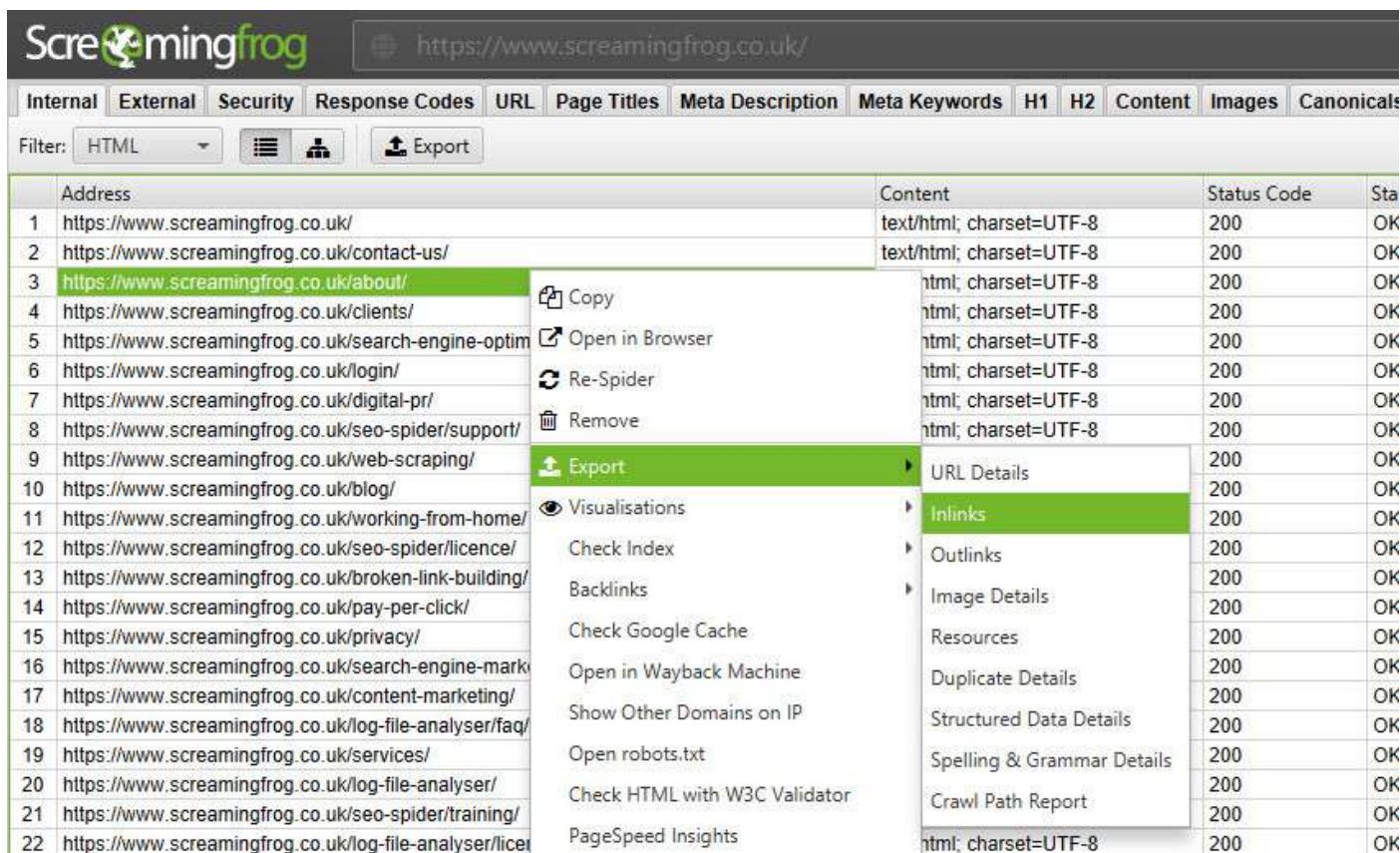
Questo metodo permette di esportare le metriche di ogni specifica Tab singolarmente. Ricorda però che se hai applicato un filtro alla scheda e successivamente scelto l'esportazione il documento scaricato sarà parziale e conterrà solamente i dati riferiti al filtro applicato!



2. Esportazione dei dati presenti nella finestra inferiore

Per esportare i dati della finestra inferiore di Screaming Frog è sufficiente selezionare nella parte superiore l'URL di riferimento (possibile selezionare anche più URL contemporaneamente), cliccare con il tasto destro del mouse, scegliere “Export” e optare per il report più adeguato.

Alternativamente dopo aver selezionato gli URL nella scheda in alto puoi utilizzare il tasto “Export” che si trova vicino ai filtri.



3. Esportazione in blocco con il "Bulk Export"

La terza opzione prevede l'esportazione tramite "Bulk Export" che ti permette di ottenere tutti i dati divisi per tipologia di elemento. Sarà quindi possibile ottenere tutte le istanze di un link tramite l'opzione "All inlinks", scaricare gli status code, l'alt text delle immagini e molto altro ancora.

Attraverso questo metodo sei in grado di ottenere i seguenti report:

- **Links > All Inlinks:** rappresenta tutti i link collegati ad ogni singolo URL che Screaming Frog ha scansionato durante la scansione del sito web; questo report considera tutti i collegamenti ipertestuali delle pagine, delle immagini, gli hreflang, eventuali next/prev, canonical etc.

ESEMPIO DELLA "TAB INLINKS":

La pagina www.dominio.it/servizi riceve link dalle pagine "home", "contatti" e "dove siamo". Nella tab "Inlinks" trovi i collegamenti delle tre pagine nella colonna "From" con destinazione unica verso la pagina "/servizi" (colonna "To"). L'esportazione "All Inlinks" comprende tutti questi riferimenti (nell'esempio 3) per ogni pagina scansionata.

- **Outlinks > All Outlinks:** considera tutti i collegamenti verso risorse esterne scoperte dallo Spider. Corrisponde ai risultati della Tab "Response code" con filtro "All".

ESEMPIO DELLA "TAB OUTLINKS":

La pagina www.dominio.it/servizi include due link verso le pagine "chi-siamo" e "home". Nella tab "Outlinks" trovi i 2 collegamenti nella colonna "To" con sorgente unica rappresentata dalla pagina "/servizi" (colonna "From"). L'esportazione "All Outlinks" comprende tutti i collegamenti esterni (nell'esempio 2) per ogni pagina scansionata.

- **Links > All Anchor Text:** definisce tutti i collegamenti <a href> scoperti (non comprende i links delle immagini, css, hreflang etc.) e presenti nella tab Response Code con filtro "All".
- **Links > External Links:** presenta tutti i links con destinazione esterna al sottodominio scansionato.
- **Web > Screenshots:** esporta tutti gli "screenshot" che sono presenti nella tab "Rendered Page" (finestra in basso) se scegli un crawl in modalità Javascript Rendering Mode (di default la scansione è in modalità "Text Only").

Config > Spider > Rendering > Javascript

Web > All Page Source: rappresenta l'html statico o il rendering dell'html delle pagine scansionate. Opzione disponibile solamente nella versione "Javascript Rendering Mode"

Web > HTTP Headers: i dati esportati corrispondono allo status code dell'HTTP headers. Per usufruire di questa funzione è necessario selezionare l'opzione "http headers" dal menu di configurazione.

Config > Spider > Extraction

Web > All Cookies: comprende tutti gli URL e rispettivi cookie emessi in un crawl. Per ottenere ed estrarre i 'Cookies' devi abilitare la relativa funzione.

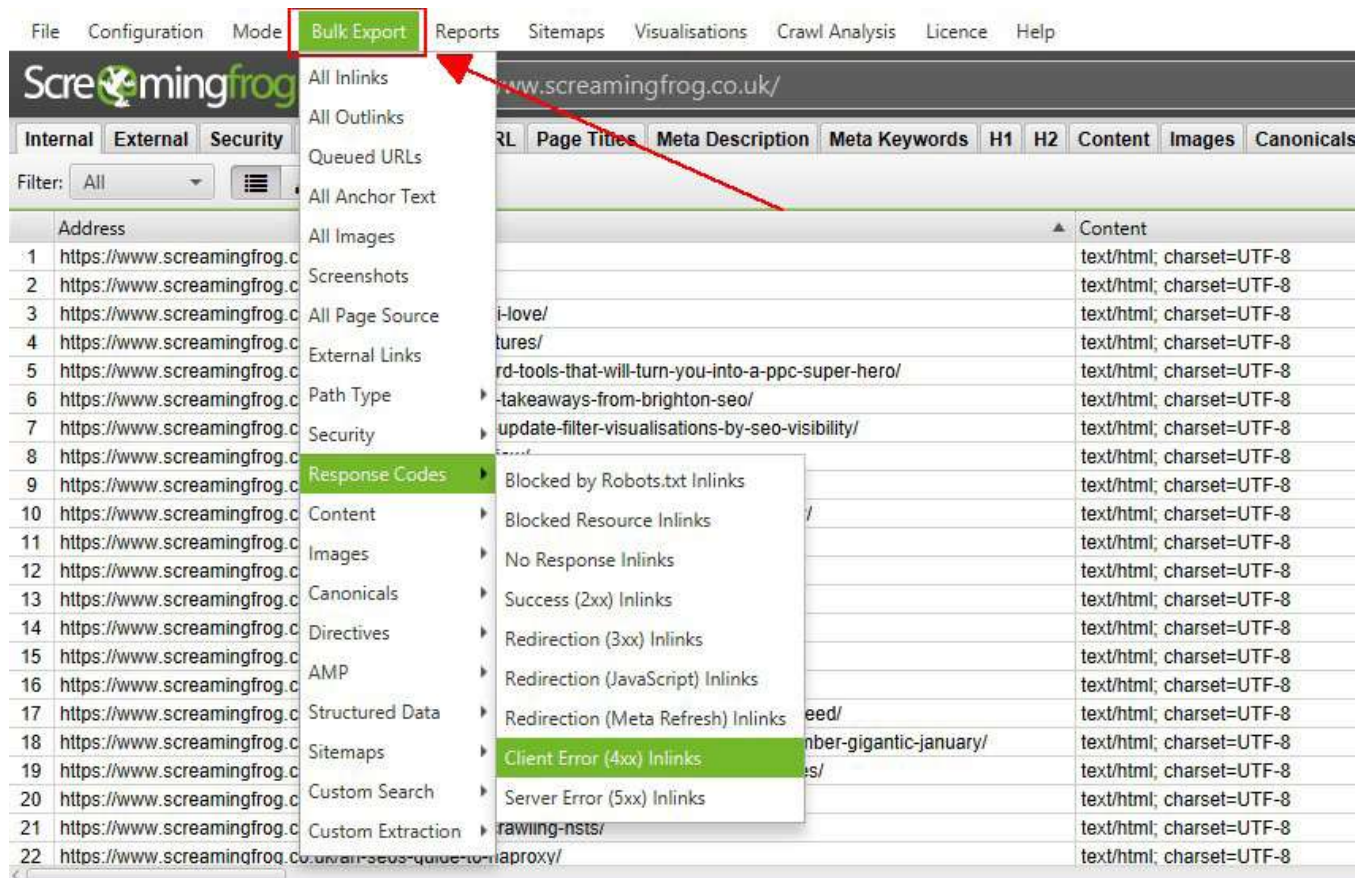
Config > Spider > Extraction > Cookies

Per una diagnosi completa dei cookie ti consiglio di abilitare anche la modalità di rendering JavaScript per includere anche quelli che vengono definiti solamente dopo l'esecuzione Javascript o tag immagine pixel. Con questo report ottieni la tipologia di cookie, il nome, la sua durata (Expiration Time) ed altre informazioni utili per gestire al meglio la privacy policy e cookie declaration.

- **Path Type:** questo report ti permette di esportare specifici links in base al tipo di percorso e può includere: absolut links, protocol-relative, root-relative e path-relative.
- **Response Code:** con questa esportazione è possibile ottenere in modo isolato i diversi status code degli URL scansionati dallo Spider (1xx,200,3xx,4xx,5xx etc.).
- **Content:** questo report è molto utile per migliorare la parte semantica del tuo sito web e controllare se esistono eventuali contenuti duplicati o pagine troppo simili fra di loro assecondando le sempre più evolute esigenze dei Motori di Ricerca.
- **Images:** l'ottimizzazione delle immagini è un'attività primaria per il posizionamento organico di un sito web. I report a disposizione ti permettono di scaricare in un solo click tutte le immagini scansionate dal crawler, capire se queste hanno un peso ottimizzato, controllare lo stato degli Alt Tag, in quale Url e posizione della pagina sono state scovate dallo Spider e decidere se esportare un solo file completo o isolare i parametri per te più congeniali (es. solo le immagini a cui manca l'Alt Tag) per alleggerire il file di export.
- **Canonicals:** attraverso questi documenti puoi decidere se comprendere tutti i links e i loro riferimenti presenti nella tab Canonical con filtro "All" o ad esempio isolare quelli a cui manca "self referencing Inlinks" o estrarre solamente le pagine "canonicalizzate".
- **Directives:** scegliendo questi report sei in grado di ottenere tutti i dati della Tab "Directives", isolare eventuali pagine che presentano il tag "No Index" o noODP etc.
- **Amp:** il report presenta tutti i links presenti nella tab "Amp". Molto utile per isolare eventuali pagine con "Amp html" ma status code diverso da 200.
- **Structured Data:** utili report per analizzare le pagine che presentano dati strutturati e per verificare quelle che presentano errori di convalida. I dati possono essere testati anche da <https://developers.google.com/search/docs/advanced/structured-data>.
- **Sitemaps:** report molto utile per scoprire eventuali pagine inserite nella sitemap ma con status code diverso da 200, pagine canonicalizzate o "No index". Ricorda che per popolare il filtro devi prima utilizzare la funzione "Crawl Analysis".
- **Custom Search/Extraction:** in questi report sono presenti tutti i dati della tab "Custom Search/Extraction".

Configuration > Custom > Search/Extraction

Questi rapporti "Custom" si riveleranno molto utili per estrapolare ad esempio links specifici che matchano una certa condizione o per isolare delle informazioni specifiche da un sito (es. gli autori di un blog, il numero di commenti ottenuti, pagine a cui manca Google Analytics).



CUSTOM EXTRACTION

Screaming Frog è un Seo Tool molto versatile che non si limita a restituire dati e metriche predefinite ma consente anche delle analisi avanzate personalizzabili. Una di queste è la “Custom Extraction”, una funzionalità molto potente che ti permette di collezionare tutti i dati dall'html di ogni pagina web (modalità text-only) o dati renderizzati con la modalità di scansione “Javascript Rendering Mode”.

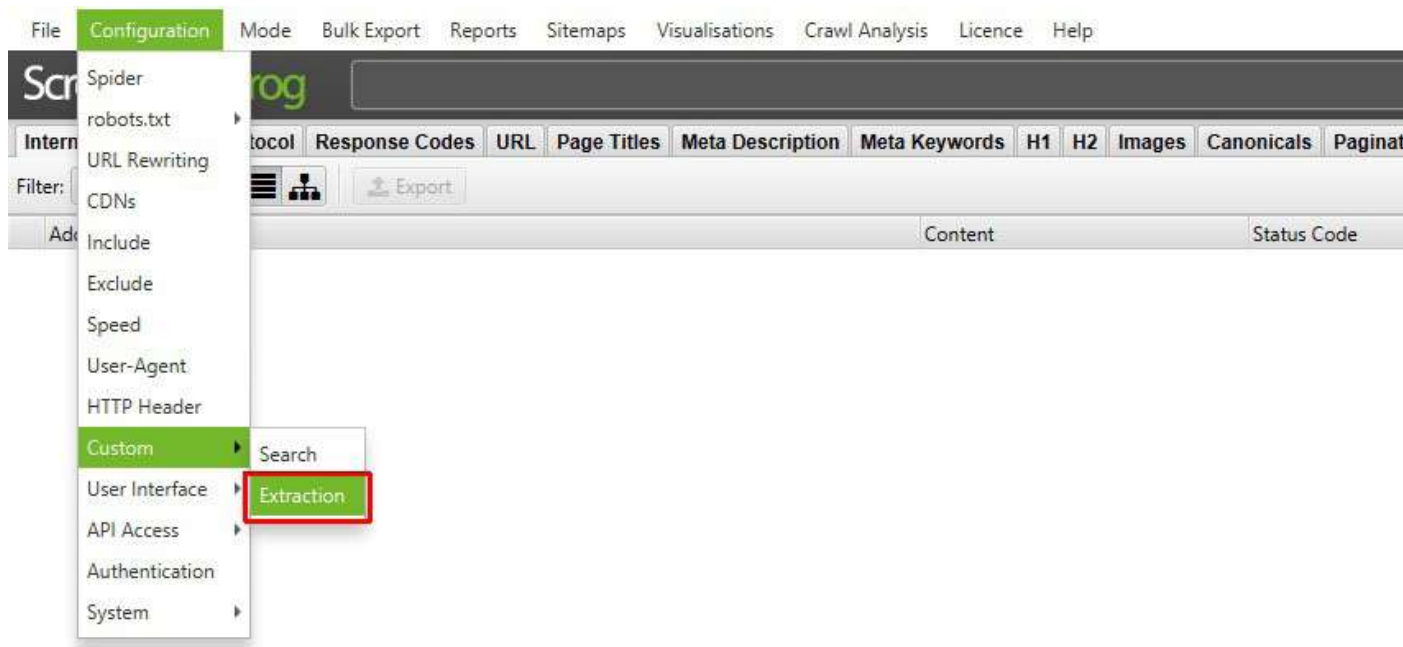
Il Seo Spider prevede 3 modalità di estrazione dei dati con la “Custom Extraction”:

1. **XPath:** XPath selectors;
2. **CSS Path;**
3. **Regex:** per estrazione di dati più avanzati (approfondimento: <https://docs.oracle.com/javase/8/docs/api/java/util/regex/Pattern.html>)

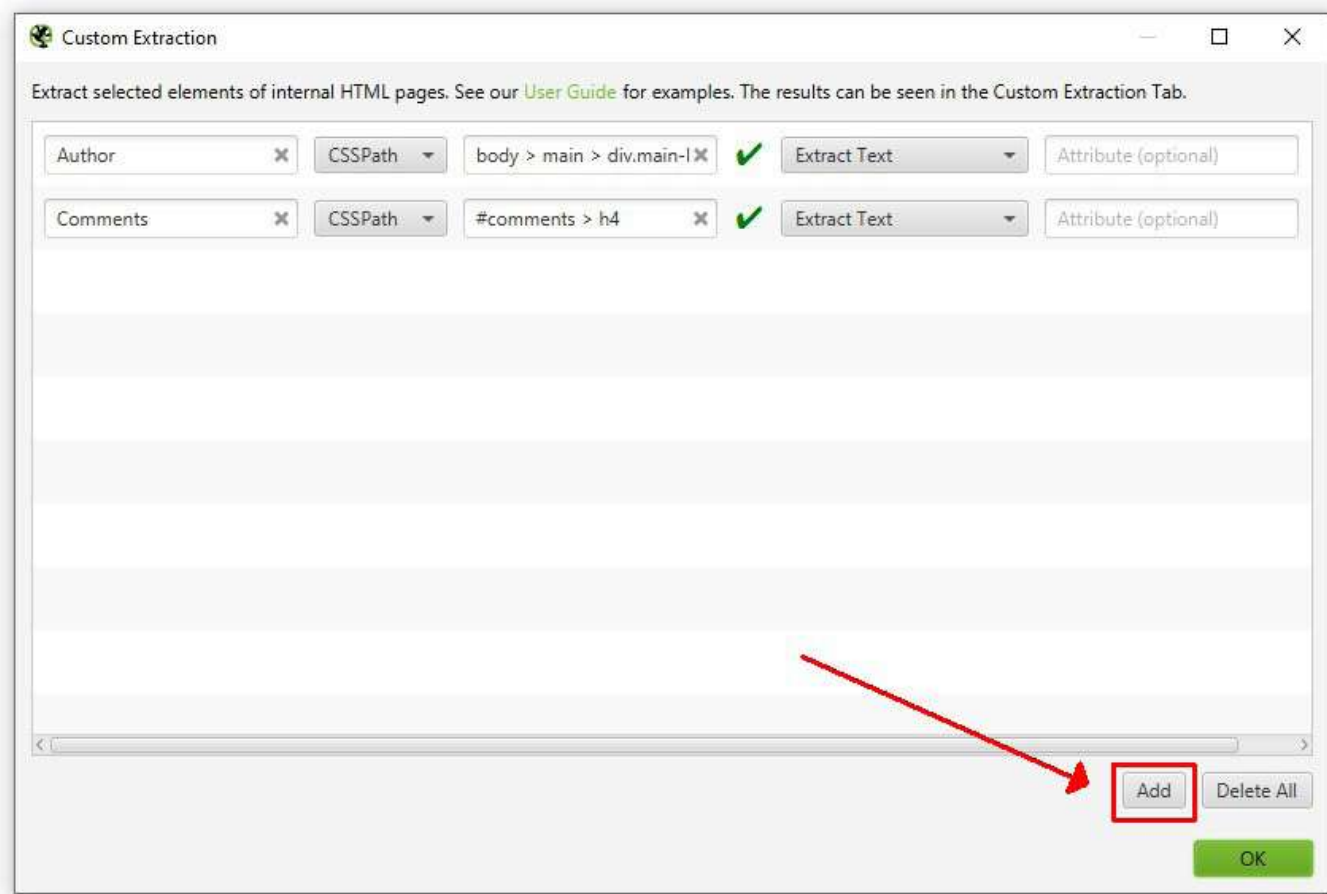
Se decidi di approfittare delle prime due soluzioni “XPath” e “CSS Path” puoi scegliere quali elementi estrapolare tramite:

- **Extract Html elements:** permette di collezionare dati dagli elementi selezionati e dal loro contenuto interno (“Inner Html Content”).
- **Extract Inner HTML:** permette di estrapolare il contenuto interno di un determinato elemento selezionato. Nel caso in cui l'elemento Html contenga altri elementi Html, anche i sottoelementi saranno a tua disposizione.
- **Extract Text:** colleziona il contenuto testuale dell'elemento selezionato e il contenuto testuale dei suoi sottoelementi.
- **Function Value:** restituisce il risultato della funzione fornita. Se ad esempio utilizzi una funzione come “count(//h1)” ottieni il numero di <h1> presenti nella pagina.

Search Extraction: Config > Custom > Extraction



Una volta selezionata la “Custom Extraction” ti basta cliccare su “Add” e inserire le tue istruzioni per l'estrazione dei dati.



I dati ottenuti sono disponibili nella tab “Custom Extraction” e nella scheda “Internal” in una colonna dedicata.

File Configuration Mode Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help

Screamingfrog <https://www.screamingfrog.co.uk/blog/>

Filter:	Internal	External	Protocol	Response Codes	URL	Page Titles	Meta Description	Meta Keywords	H1	H2	Custom Extraction	Images	Canonical
All													
	4	https://www.screamingfrog.co.uk/seo-spider-7-0/	200	OK	screamingfrog	85 Comments							
	5	https://www.screamingfrog.co.uk/seo-spider-companion-tools/	200	OK	Richard Lawther	7 Comments							
	6	https://www.screamingfrog.co.uk/log-file-analyser-2-0/	200	OK	screamingfrog	7 Comments							
	7	https://www.screamingfrog.co.uk/find-killer-content/	200	OK	Daniel Cartland	7 Comments							
	8	https://www.screamingfrog.co.uk/seo-spider-10/	200	OK	screamingfrog	72 Comments							
	9	https://www.screamingfrog.co.uk/the-beginners-guide-to-seo-competitor...	200	OK	Freddy Barnes	6 Comments							
	10	https://www.screamingfrog.co.uk/learn-to-crawl/	200	OK	Oliver Brett	6 Comments							
	11	https://www.screamingfrog.co.uk/seo-spider-9-0/	200	OK	screamingfrog	62 Comments							
	12	https://www.screamingfrog.co.uk/seo-spider-12/	200	OK	screamingfrog	60 Comments							
	13	https://www.screamingfrog.co.uk/chasing-for-a-link/	200	OK	Mark Porter	5 Comments							
	14	https://www.screamingfrog.co.uk/brightonseo-crawling-clinic/	200	OK		5 Comments							
	15	https://www.screamingfrog.co.uk/seo-spider-6-0/	200	OK	screamingfrog	57 Comments							

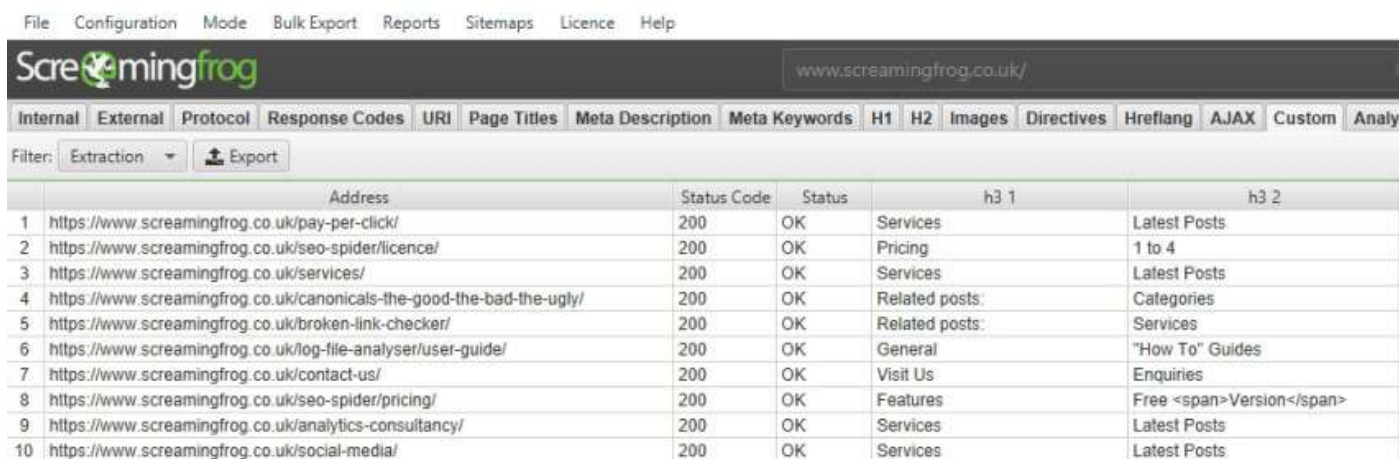
Ora che sai impostare la “Custom Extraction” e hai capito le potenzialità vediamo alcuni esempi di applicazione che credo ti possano essere utili per le tue scorribande nella seo tecnica.

EXTRACTION CON XPATH

Di seguito alcuni esempi di estrapolazioni utilizzando l'XPath.

1. **Headings:** di default il Seo Spider colleziona solamente i titoli principali della pagina (H1 e H2) ma per un'analisi più specifica e completa potresti avere la necessità di reperire informazioni specifiche anche sugli altri "headings" presenti in pagina.

Estrazione di tutti gli heading H3: XPath: //h3



	Address	Status Code	Status	h3 1	h3 2
1	https://www.screamingfrog.co.uk/pay-per-click/	200	OK	Services	Latest Posts
2	https://www.screamingfrog.co.uk/seo-spider/licence/	200	OK	Pricing	1 to 4
3	https://www.screamingfrog.co.uk/services/	200	OK	Services	Latest Posts
4	https://www.screamingfrog.co.uk/canonicals-the-good-the-bad-the-ugly/	200	OK	Related posts:	Categories
5	https://www.screamingfrog.co.uk/broken-link-checker/	200	OK	Related posts:	Services
6	https://www.screamingfrog.co.uk/log-file-analyser/user-guide/	200	OK	General	"How To" Guides
7	https://www.screamingfrog.co.uk/contact-us/	200	OK	Visit Us	Enquiries
8	https://www.screamingfrog.co.uk/seo-spider/pricing/	200	OK	Features	Free Version
9	https://www.screamingfrog.co.uk/analytics-consultancy/	200	OK	Services	Latest Posts
10	https://www.screamingfrog.co.uk/social-media/	200	OK	Services	Latest Posts

Estrazione del primo heading H3:

XPath: /descendant::h3[1]

Estrazione dei primi 5 heading H3:

XPath: /descendant::h3[position() >= 0 and position() <= 5]

Se vuoi contare il numero di H3 presenti nella pagina:

XPath: count(//H3)

Se vuoi conoscere la lunghezza di una stringa puoi sempre usare XPath ma impostando l'opzione "Function Value".

XPath: string-length (//h3)

2. Estrapolazione delle singole versioni degli "Hreflang"

XPath: //*[@hreflang]/@hreflang

File Configuration Mode Bulk Export Reports Sitemaps Licence Help

Screamingfrog

www.moo.com

Results

Internal External Protocol Response Codes URI Page Titles Meta Description Meta Keywords H1 H2 Images Directives Hreflang AJAX Custom Analytics Search

Filter: Extraction Export

	Address	Status Code	Status	hreflang 1	hreflang 2
1	https://www.moo.com/uk/	200	OK	<link href="/us/" hreflang="en-US" rel="alternate">	<link href="/es-us/" hreflang="es...
2	https://www.moo.com/au/	200	OK	<link href="/us/" hreflang="en-US" rel="alternate">	<link href="/es-us/" hreflang="es...
3	https://www.moo.com/uk/design-templates/minicards/	200	OK	<link href="https://www.moo.com/au/design-templates/m...	<link href="https://www.moo.co...
4	https://www.moo.com/es/	200	OK	<link href="/us/" hreflang="en-US" rel="alternate">	<link href="/es-us/" hreflang="es...
5	https://www.moo.com/uk/about/company-information.html	200	OK	<link href="/us/about/company-information.html" hreflan...	<link href="/es-us/about/compan...
6	https://www.moo.com/uk/sitemap/	200	OK	<link href="https://www.moo.com/au/sitemap/" hreflang=...	<link href="https://www.moo.co...
7	https://www.moo.com/uk/products/square-business-cards.html	200	OK	<link href="/us/products/square-business-cards.html" hre...	<link href="/es-us/products/squa...
8	https://www.moo.com/uk/products/greeting-cards.html	200	OK	<link href="/us/products/greeting-cards.html" hreflang="e...	<link href="/es-us/products/greet...
9	https://www.moo.com/uk/products/nfc/business-cards-plus.html	200	OK	<link href="/us/products/nfc/business-cards-plus.html" hr...	<link href="/es-us/products/nfc/b...
10	https://www.moo.com/uk/accessories.html	200	OK	<link href="/us/accessories.html" hreflang="en-US" rel="...	<link href="/es-us/accessories.ht...

3. Scoprire e collezionare le diverse tipologie ("Types") nei **dati strutturati**:

```
XPath: //*[@itemtype]/@itemtype
```

Nota Bene: per la convalida dei dati strutturati non è necessario utilizzare una "Custom Extraction" ma puoi visualizzare i dati nella tab "Structured Data".

4. Collezionare i **Social Media Tag, Open Graph Tags e Twitter cards**:

XPath:

```
//meta[starts-with(@property, 'og:title')]/@content
//meta[starts-with(@property, 'og:description')]/@content
//meta[starts-with(@property, 'og:type')]/@content
//meta[starts-with(@property, 'og:site_name')]/@content
//meta[starts-with(@property, 'og:image')]/@content
//meta[starts-with(@property, 'og:url')]/@content
//meta[starts-with(@property, 'fb:page_id')]/@content
//meta[starts-with(@property, 'fb:admins')]/@content
```

```
//meta[starts-with(@property, 'twitter:title')]/@content
//meta[starts-with(@property, 'twitter:description')]/@content
//meta[starts-with(@property, 'twitter:account_id')]/@content
//meta[starts-with(@property, 'twitter:card')]/@content
//meta[starts-with(@property, 'twitter:image:src')]/@content
//meta[starts-with(@property, 'twitter:creator')]/@content
```

5. Estrarre gli **indirizzi email** e/o i numeri di telefono da un sito web

```
XPath: //a[starts-with(@href, 'mailto')]
```

```
XPath: //a[starts-with(@href, 'tel')]
```

6. Estrarre particolari frames dal sito web come Google Tag Manager, Video Youtube.

Collezionare tutti gli IFrame:

```
XPath: //iframe/@src
```

Restituisce gli IFrame dei video Youtube:

```
XPath: //iframe[contains(@src, 'www.youtube.com/embed/')]
```

Collezionare lo script di Google Tag Manager

```
XPath: //iframe[not(contains(@src, 'https://www.googletagmanager.com/'))]/@src
```

7. Estrazione di contenuti da **specifici Divs o Span** fornendo la classe (da sostituire al posto di "example")

```
XPath: //div[@class="example"]  
//span[@class="example"]
```

Collezionare contenuti concatenati in un'unica estrapolazione

```
XPath: //div[contains(@class,'main-blog--posts_single-inner--text--inner')]/h3//a[@class="comments-link"]
```

Questo esempio colleziona i titoli e numero di commenti dei blog post (per funzionare dovrai adeguare le classi specifiche del tuo sito web).



	Address	Status Code	Status	Extractor 1 1	Extractor 1 2	Extractor 1 3	Extractor 1 4
1	https://www.screamingfrog.co.uk/blog/	200	OK	How to Scrape Google Search Features Using XPath	6 Comments	The Do's and Don'ts of Chasing for a Link	3 Comments
2	https://www.screamingfrog.co.uk/blog/page/4/	200	OK	How To Fetch & Render (Almost) Any Site	33 Comments	Screaming Frog Log File Analyser Update – Version 2.0	7 Comments
3	https://www.screamingfrog.co.uk/blog/page/10/	200	OK	Google AdWords: New Changes to Ad Rotation	5 Comments	15 Top Quotes and Takeaways from Brighton SEO	2 Comments
4	https://www.screamingfrog.co.uk/blog/page/5/	200	OK	How To Analyse Your Ideas To Find Killer Content	7 Comments	How Accurate Are Website Traffic Estimators?	31 Comments
5	https://www.screamingfrog.co.uk/blog/page/3/	200	OK	Win Screaming Frog T-Shirts	190 Comments	Content Idestion: Tools & Techniques	6 Comments
6	https://www.screamingfrog.co.uk/blog/page/13/	200	OK	What SEM's Are Saying About The Screaming Frog ...	22 Comments	After Black Friday & Mega Monday Comes 'Dramatic Deca...	1 Comment
7	https://www.screamingfrog.co.uk/blog/page/2/	200	OK	Screaming Frog Wins Best SEO Campaign at the UK ...	2 Comments	The BrightonSEO Crawling Clinic	6 Comments

EXTRACTION CON REGEX

Le regole Regex sono uno strumento molto potente per collezionare dati con Screaming Frog; vediamo alcuni esempi applicativi da poter utilizzare immediatamente nel tuo prossimo Seo Audit:

1. Estrazione dell'ID di Google Analytics e di Google Tag Manager:

```
["](UA-.*?)["]
```

```
["](GTM-.*?)["]
```

Custom Extraction

Extract selected elements of internal HTML pages. See our [User Guide](#) for examples. The results can be seen in the Custom Tab in the Extraction filter.

GA UA ID

Regex

```
["](UA-.*?)["]
```

2. Estrazione dei **Structured Data**.

Per estrapolare i dati strutturati abbiamo visto che è possibile utilizzare anche la modalità XPath ma se questi hanno una formattazione in JSON-LD è consigliabile utilizzare la sintassi RegEX:

```
"product": "(.*)"  
"ratingValue": "(.*)"  
"reviewCount": "(.*)"
```

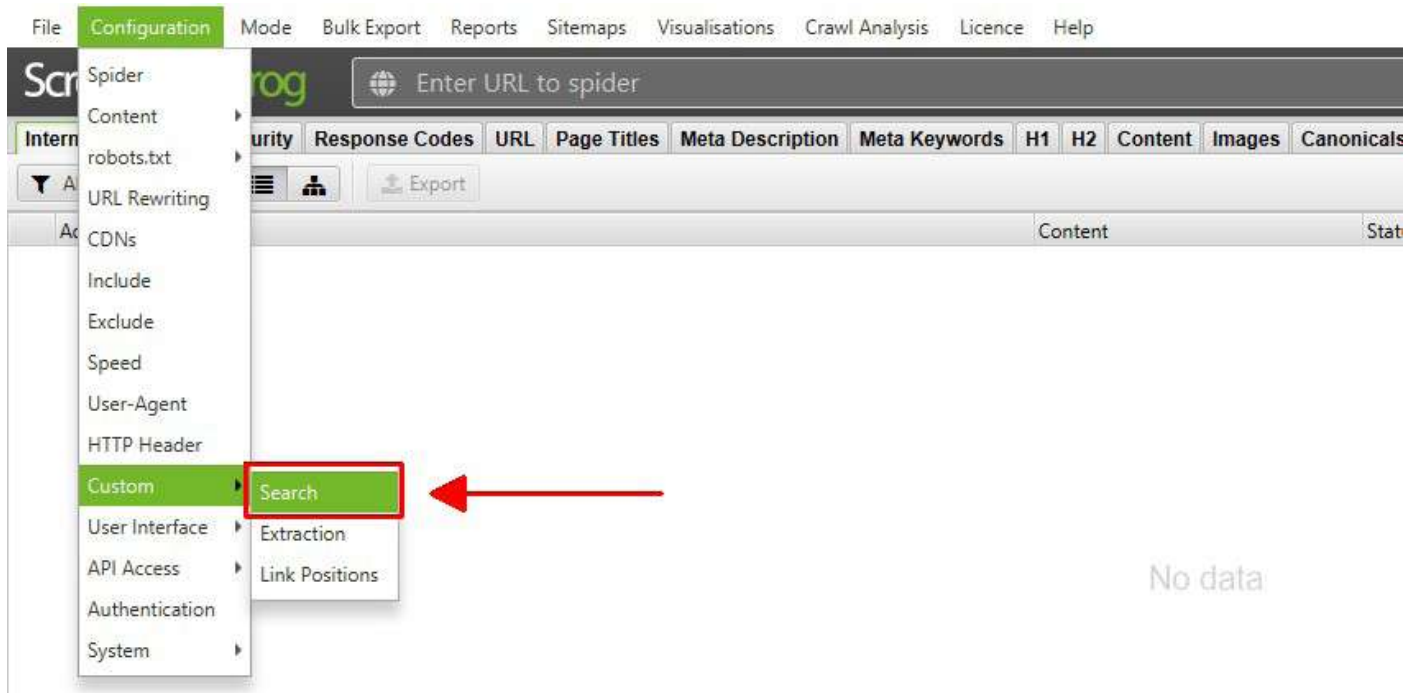
Estrapolazione di ogni dato dallo script JSON-LD:

```
<script type="application/ld+json">(.*?)</script>  
Estrazione di indirizzi Email: [a-zA-Z0-9-_.]+@[a-zA-Z0-9-_.]+
```

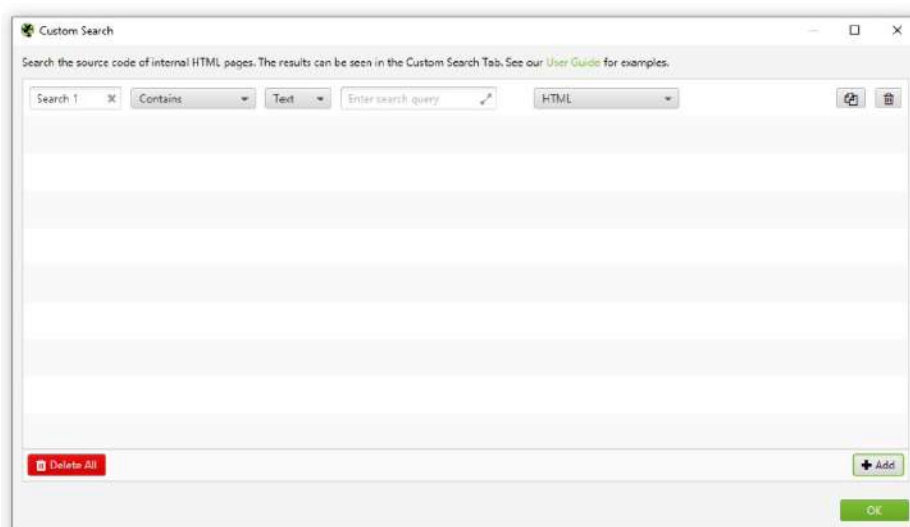
CUSTOM SEARCH

Il Seo Spider ti permette di verificare la presenza di qualsiasi elemento contenuto nel codice sorgente delle pagine di un sito web attraverso la funzionalità “Custom Search”. Per default il sistema ricerca l’HTML raw mentre passando alla modalità di scansione “Javascript Rendered” ti permette di estrapolare anche il dato renderizzato. Il tool consente di configurare fino a 100 ricerche personalizzate che ti restituiscono URL che contengono o non contengono l’input ricercato.

Config > Custom > Search

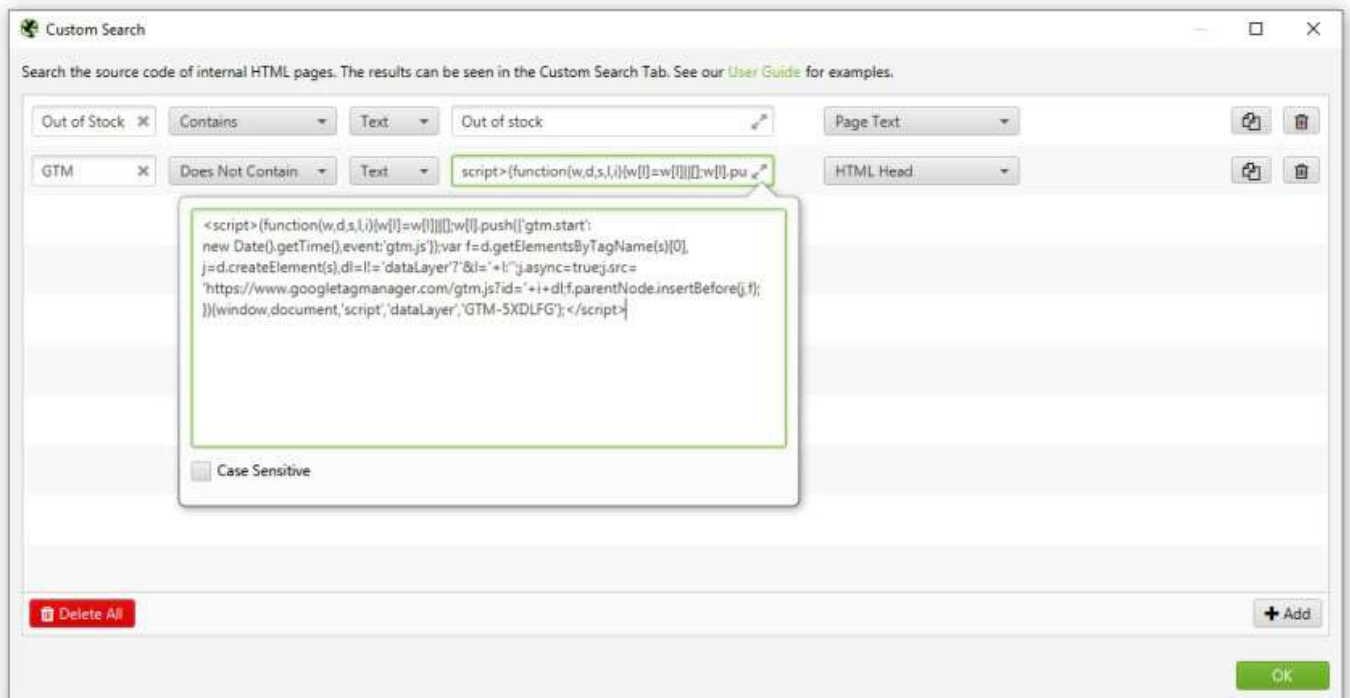


Per utilizzare la funzionalità “Custom Search” è sufficiente cliccare sul tasto “Add” della schermata, nominare il filtro, scegliere l’opzione “Contains” (contiene) o “Does Not Contains” (non contiene), scegliere la tipologia di ricerca tra testo o RegEX, inserire l’input da ricercare e definire dove effettuare la ricerca (html, testo della pagina, un particolare elemento, XPath etc.).



Nell'esempio seguente abbiamo ipotizzato una ricerca personalizzata con due obiettivi:

1. l'identificazione dei prodotti non disponibili di un ecommerce ricercando la dicitura "Out of Stock" presente nelle pagine di dettaglio (Contains).
2. la definizione delle pagine in cui manca lo script di Google Tag Manager (Does Not Contains).



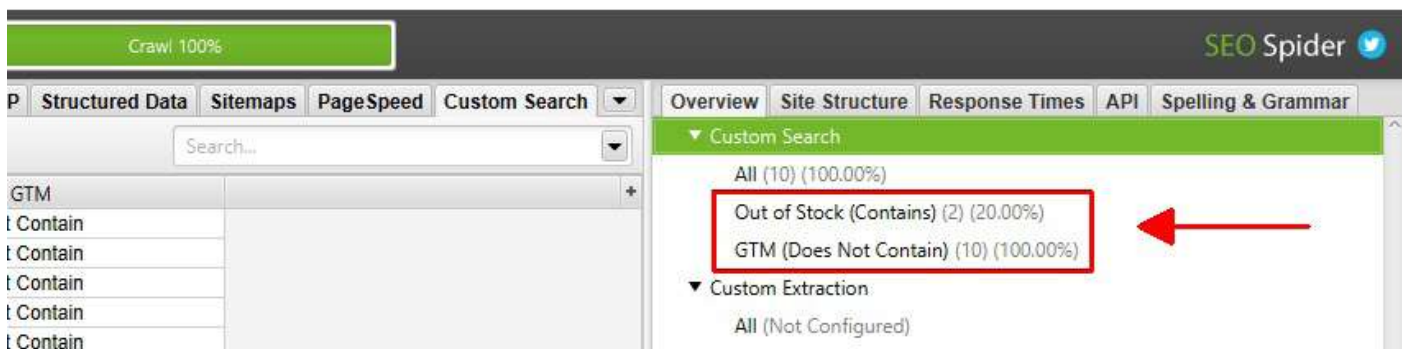
Per collezionare i risultati è sufficiente dare "ok" ed eseguire il crawl.

Una volta processata la scansione i dati saranno popolati in tempo reale e divisi in colonne nella tab "Custom Search".

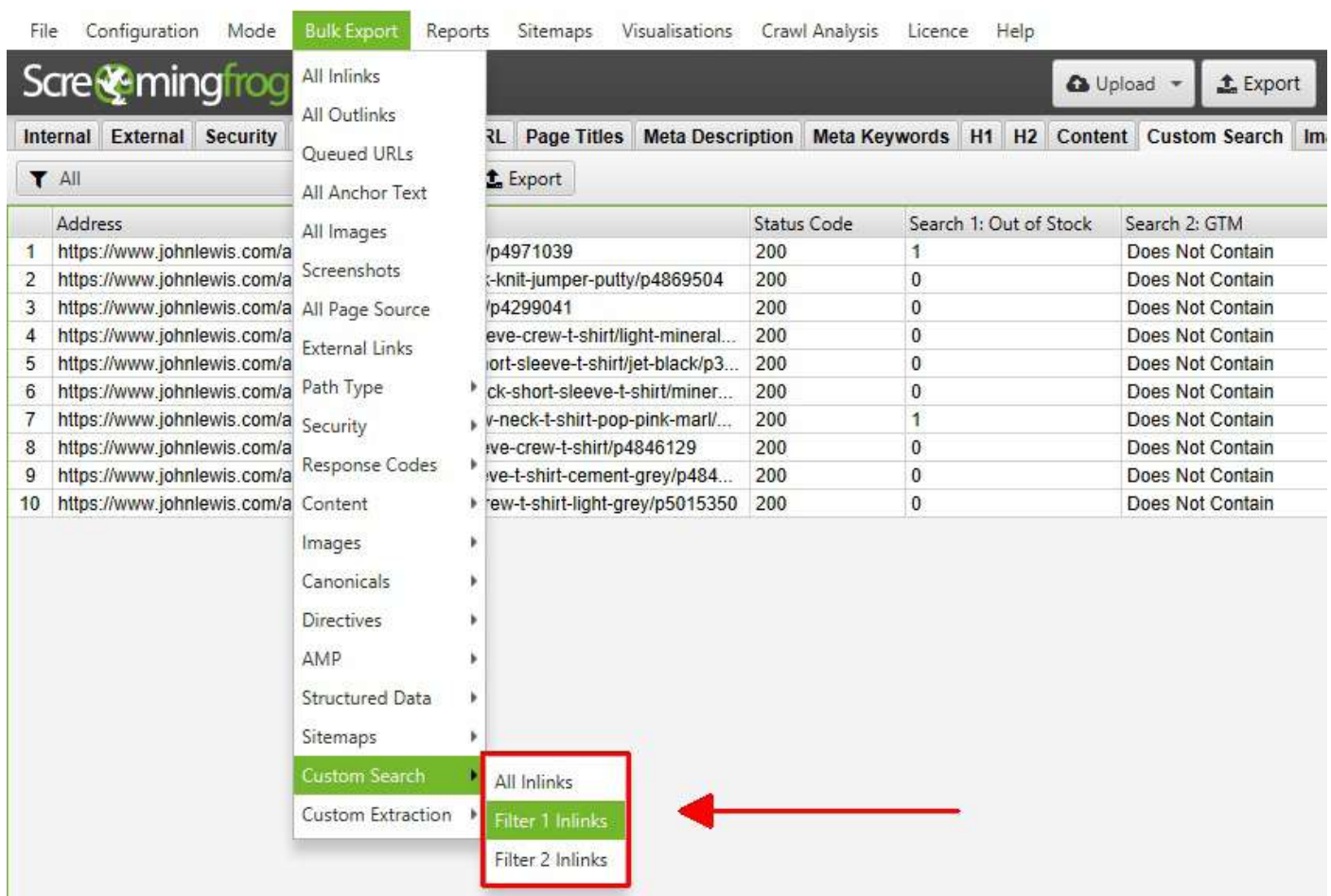
Volendo puoi isolare una o l'altra "Custom Search" utilizzando il filtro a discesa che trovi a sinistra della finestra superiore.

File Configuration Mode Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help									
Screamingfrog									
Upload Export Start Clear									
External	Security	Response Codes	URL	Page Titles	Meta Description	Meta Keywords	H1	H2	Custom Search
All Export									
All			Status Code		Search 1: Out of Stock		Search 2: GTM		
✓ All			200		0		Does Not Contain		
Out of Stock (Contains)			200		0		Does Not Contain		
GTM (Does Not Contain)			200		0		Does Not Contain		
4	https://www.johnlewis.com/allsa...		200		1		Does Not Contain		
5	https://www.johnlewis.com/allsa...		200		0		Does Not Contain		
6	https://www.johnlewis.com/allsa...		200		0		Does Not Contain		
7	https://www.johnlewis.com/allsa...		200		0		Does Not Contain		
8	https://www.johnlewis.com/allsa...		200		0		Does Not Contain		
9	https://www.johnlewis.com/allsa...		200		0		Does Not Contain		
10	https://www.johnlewis.com/allsa...		200		1		Does Not Contain		

Per comodità gli stessi risultati li puoi anche consultare dal pannello della sidebar laterale sotto la voce "Custom Search".

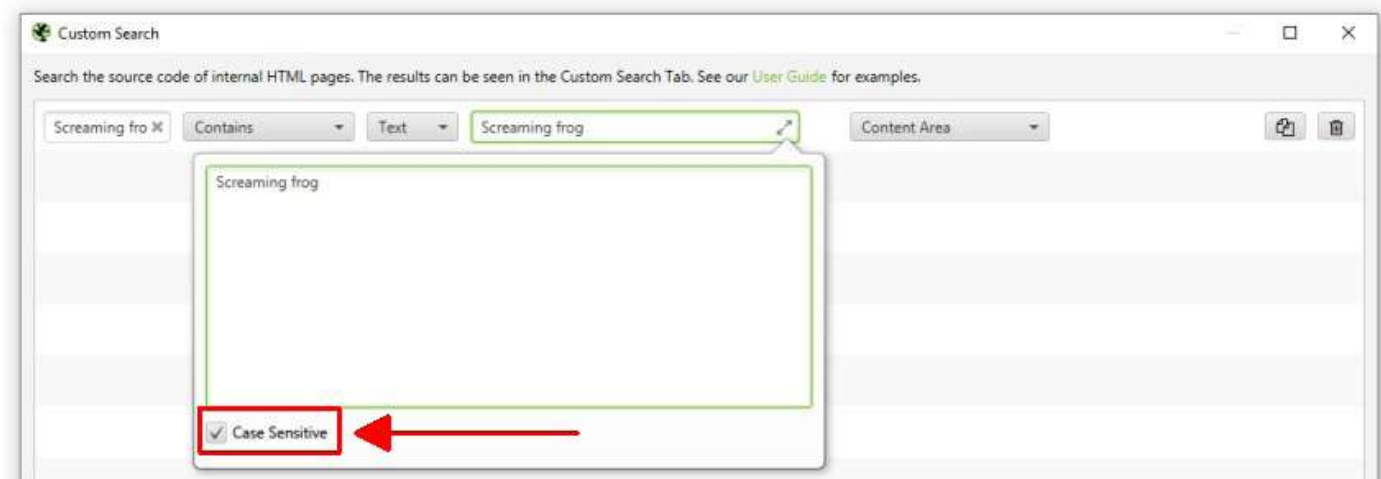


Per completare l'analisi puoi esportare i risultati cliccando direttamente sul tasto "Export" della scheda o usare la funzione "Bulk Export" per ottenere tutti gli Inlinks della tua ricerca personalizzata.



Ora che hai compreso la versatilità della "Custom Search" ti propongo alcune best practice molto utili nei Seo Audit quotidiani.

Un primo utilizzo molto interessante in modalità "Text Only" è quello di scovare eventuali errori di misspelling, acronimi etc. considerando l'aspetto "case sensitive" delle parole (le RegEX sono già di default "case sensitive") abilitando la relativa opzione.



Un'altra applicazione della "Custom Search" in modalità RegEX è la ricerca di specifiche query (a corrispondenza esatta). Questa ricerca ti permette di scoprire eventuali problemi di cannibalizzazione con parole chiave strategiche presenti in diverse landing page che potrebbero penalizzare il Search Intent su Google.

RegEX: \bquery\b

La ricerca precedente può essere anche ampliata a più parole utilizzando la stessa "Custom Search" e dividendo le singole RegEX con la "|" (Simbolo Pipe). Questa ricerca di permette di valutare la densità delle query nel testo e di evitare eventuali penalizzazioni di overstuffing.

RegEX: \bquery1\b|\bquery2\b|\bquery3\b

Nell'esempio seguente sono state ricercate le query: "search engine optimization", "natural" e "free".

Address	Content	Status Code	Status	Search 1: Multiple Words
1 https://www.screamingfrog.co.uk/search-engine-optimisation/	text/html; charset=UTF-8	200	OK	5
2 https://www.screamingfrog.co.uk/	text/html; charset=UTF-8	200	OK	2
3 https://www.screamingfrog.co.uk/about/	text/html; charset=UTF-8	200	OK	1
4 https://www.screamingfrog.co.uk/digital-pr/	text/html; charset=UTF-8	200	OK	1
5 https://www.screamingfrog.co.uk/contact-us/	text/html; charset=UTF-8	200	OK	0
6 https://www.screamingfrog.co.uk/login/	text/html; charset=UTF-8	200	OK	0
7 https://www.screamingfrog.co.uk/clients/	text/html; charset=UTF-8	200	OK	0
8 https://www.screamingfrog.co.uk/seo-spider/support/	text/html; charset=UTF-8	200	OK	0
9 https://www.screamingfrog.co.uk/web-scraping/	text/html; charset=UTF-8	200	OK	0
10 https://www.screamingfrog.co.uk/blog/	text/html; charset=UTF-8	200	OK	0
11 https://www.screamingfrog.co.uk/seo-spider/licence/	text/html; charset=UTF-8	200	OK	0

Il risultato della scansione presenta per la prima pagina 5 istanze (3 volte "search engine optimization", 1 volta a testa "natural" e "free").

LE OPZIONI DELLA FUNZIONE "CUSTOM SEARCH"

Quando utilizzi la funzione "Custom Search" hai a disposizione 7 opzioni che ti permettono di limitare l'area di ricerca:

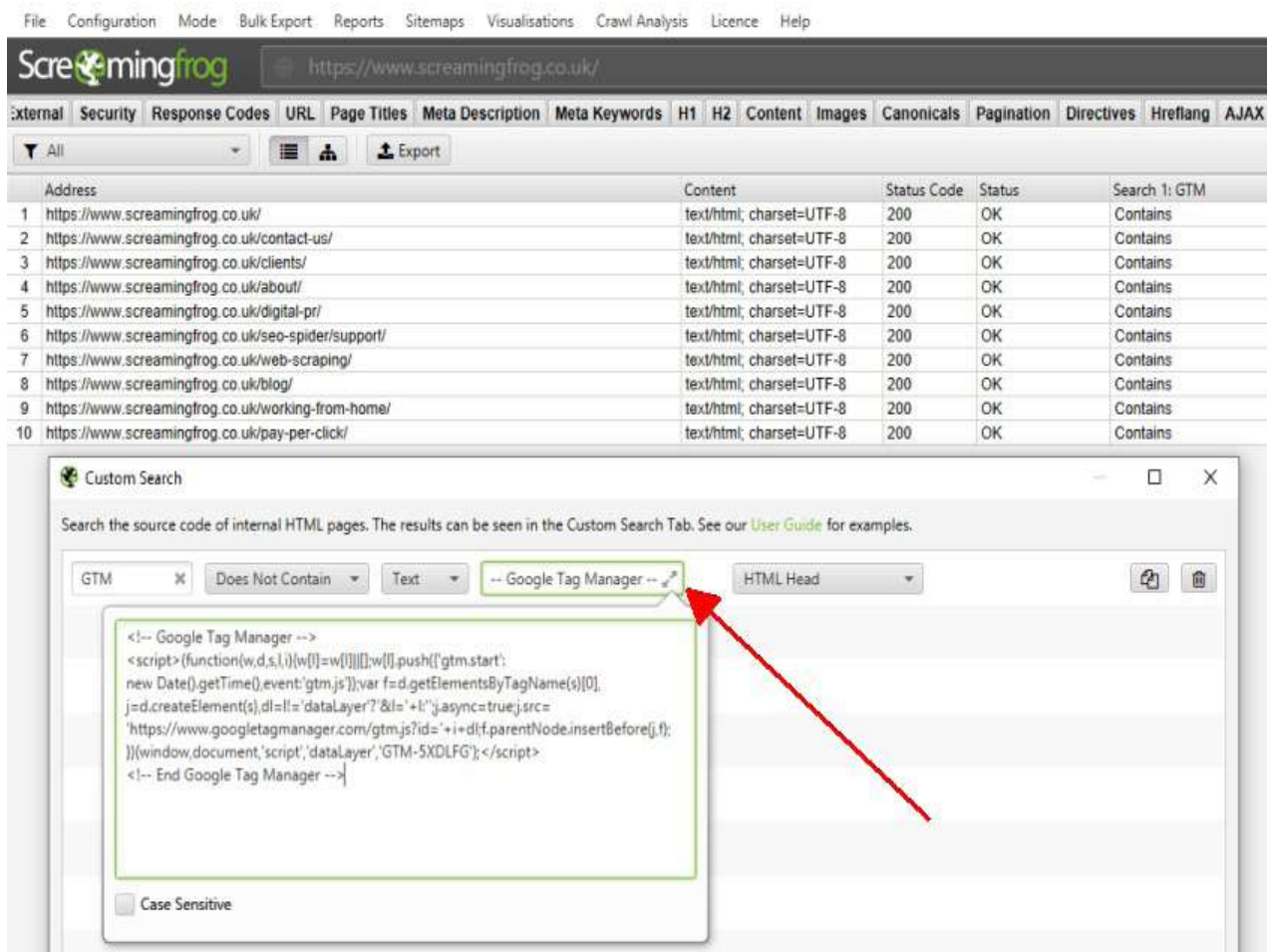
- **HTML:** comprende l'intero codice HTML.
- **Page Text:** considera solamente il testo della pagina web ma vengono esclusi tutti gli elementi HTML.
- **Page Text No Anchors:** comprende il testo delle pagine web, escludendo qualsiasi HTML o qualsiasi testo contenuto all'interno dei tag HTML di ancoraggio (noto anche come A Elements). Questo può essere utile quando si cercano parole che sono anche incluse nel testo dei link all'interno dei menu.
- **HTML Head:** la ricerca sarà isolata all'html dell'header.
- **HTML Body:** la ricerca esamina gli elementi di testo e html del tag html Body.
- **XPath:** ti permette di specificare un XPath per definire la posizione nell'HTML dove deve essere eseguita la ricerca. Per esempio, se vuoi limitare l'indagine solamente al testo contenuto nei titoli h3, puoi fornire //h3.
- **Content Area:** la ricerca è effettuata di default all'interno dell'elemento html Body.

ESEMPIO

Ad esempio potremmo ricercare se il nome Screaming Frog è stato sempre scritto correttamente o presenta dei "Misspelling" negli articoli del blog. Nello specifico abbiamo interrogato il Seo Spider per verificare se il nome aziendale del Seo Tool sia scritto in modo corretto o presenti la lettera minuscola ad inizio della seconda parola: "Screaming frog".



Nota Bene: La schermata di impostazione delle ricerche personalizzate presenta un campo di input "singular-line" ma, cliccando sulle due freccette del campo, hai la possibilità di espanderlo per inserire anche più righe di codice o intere frasi.



IL FILE ROBOTS.TXT

Se ti occupi di Seo sai benissimo l'importanza del file robots.txt per gestire il comportamento dei Motori di Ricerca. Anche Screaming Frog ne conosce l'importanza e rispetta le direttive "Allow/Disallow" comportandosi esattamente come i Bot.

Vediamo insieme alcune situazioni tipiche in cui devi gestire questo file per svolgere un Seo Audit:

- **Fase di staging:** il sito presenta un file robots.txt completamente chiuso per i Bot per non rischiare di essere indicizzato ma hai la necessità di fare una scansione per controllarlo prima della pubblicazione online.
- **Fase Post messa online:** il sito è appena stato rilasciato e devi controllare che il file Robots.txt sia stato settato correttamente e le direttive di Allow/Disallow corrispondano alle strategie Seo pianificate.

SOLUZIONI

Come detto nella fase di "Staging" il file Robots.txt può risultare completamente bloccato dalla direttiva Disallow che non permette al Seo Spider o ai BOT dei Motori di Ricerca di fare un crawl del sito web.

```
User-agent: *
Disallow: /
```

Le soluzioni per ovviare a questo problema sono essenzialmente 3:

1. **Ignore file Robots.txt:** attraverso questa opzione Screaming Frog non considera il file e scansiona il sito web nonostante la direttiva. Per default lo Spider Seo è impostato su "Respect Robots.txt".

Config > robots.txt > Ignore robots.txt

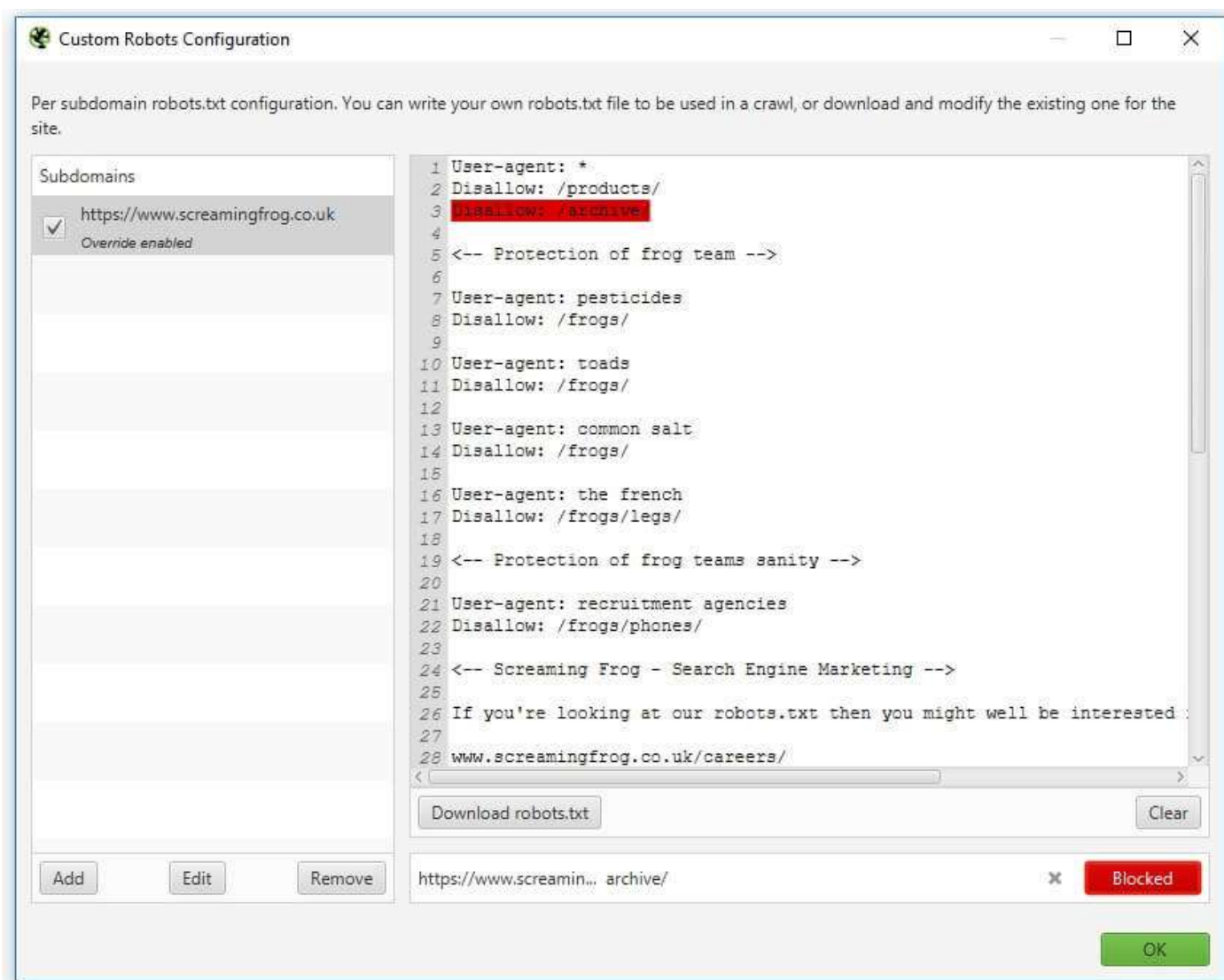
In alternativa puoi scegliere l'opzione "Ignore file Robots.txt, but report status" che ugualmente ignora le direttive del robots.txt ma lascia traccia delle risorse che sarebbero state bloccate dal file.

2. Modifica "Live" del file Robots.txt: è possibile chiedere agli sviluppatori di modificare il file del sito in staging permettendo solamente a Screaming Frog Spider di poter scansionare il sito web.

```
User-agent: *  
Disallow: /  
User-agent: Screaming Frog SEO Spider  
Allow: /
```

3. Utilizza la funzione "Custom robots.txt": ti permette di scaricare il file, modificarlo e testarlo. Molto intuitivo l'editor previsto da Screaming Frog in cui modificare le direttive sulla console per sovrascrivere il file robots.txt e permettere il crawl. Questa attività non impatta sulla versione "Live" dello staging.

Config > robots.txt > Custom



Nota Bene: per controllare preliminarmente le direttive inserite nel file Robots.txt è sufficiente avviare il crawl scegliendo uno dei diversi "User-Agent" a disposizione e controllare quali sono le risorse bloccate nella scheda "Response Code" impostando come filtro "Blocked by Robots.txt". (ricorda di controllare che sia stata selezionata la voce "respect file robots.txt" dalle configurazioni).

File Configuration Mode Bulk Export Reports Sitemaps Licence Help

Screamingfrog

Upload Export

Internal External Protocol Response Codes URI Page Titles Meta Description Meta Keywords H1 H2 Images Directives Hreflang

Filter: Blocked by Robots.txt Export

	Address	Status	Matched Robots.txt Line
1	https://www.screamingfrog.co.uk/archive/this-is-blocked/	Blocked by Robots.txt	Line:3 https://www.screamingfrog.co.uk/archive/
2	https://www.screamingfrog.co.uk/archive/not-a-real-url/	Blocked by Robots.txt	Line:3 https://www.screamingfrog.co.uk/archive/
3	https://www.screamingfrog.co.uk/archive/archive/	Blocked by Robots.txt	Line:3 https://www.screamingfrog.co.uk/archive/

L'opzione "Custom Robots.txt" è ideale per testare nuove direttive prima di cambiare il file robots.txt nella versione "Live" in caso di ottimizzazioni post messa online.

Vediamo ora alcune possibili configurazioni del file robots.txt che possono essere utili per definire i limiti di scansione dei BOT.

Il File Robots.txt blocca completamente tutti i Bot

```
User-agent: *
Disallow: /
```

Blocco di una specifica cartella a tutti gli user-agent

```
User-agent: *
Disallow: /folder/
```

Blocco dei Bot verso uno specifico URL

```
User-agent: *
Disallow: /a-specific-url.html
```

Blocco completo per Googlebot

```
User-agent: Googlebot
Disallow: /
```

Direttive multiple per Googlebot

```
User-agent: Googlebot
Disallow: /
Allow: /crawl-this/
```

Se hai impostato delle direttive in conflitto (cioè un "Allow" e un "Disallow" che si sovrappongono per lo stesso percorso di file), Google seguirà quella meno restrittiva tranne che la regola "Disallow" non sia più specifica.

Un'ultima best practice comprende l'utilizzo delle "Wildcard" nel file robots.txt (permesso sia da Google che da Bing).

```
User-agent: *
Disallow: /*?
```

Con questa direttiva puoi bloccare l'accesso a tutti gli Urls che presentano dei parametri dopo il simbolo "?"

Di seguito alcuni esempi delle linee guida di Google rispetto all'ordine di precedenza delle direttive nel file robots.txt

ESEMPIO URL	DIRETTIVE ROBOTS.TXT
http://example.com/page	allow: /p disallow: / Regola applicabile: allow: /p, perché è più specifica.
http://example.com/folder/page	allow: /folder disallow: /folder Regola applicabile: allow: /folder perché, in caso di regole corrispondenti, Google utilizza quella meno restrittiva.
http://example.com/page.htm	allow: /page disallow: /*.htm Regola applicabile: disallow: /*.htm, perché corrisponde a più caratteri nell'URL, quindi è più specifica.
http://example.com/page.php5	allow: /page disallow: /*.ph Regola applicabile: allow: /page perché, in caso di regole corrispondenti, Google utilizza quella meno restrittiva.
http://example.com/	allow: /\$ disallow: / Regola applicabile: allow: /\$, perché è più specifica.
http://example.com/page.htm	allow: /\$ disallow: / Regola applicabile: disallow: /, perché la regola allow si applica solo all'URL principale.

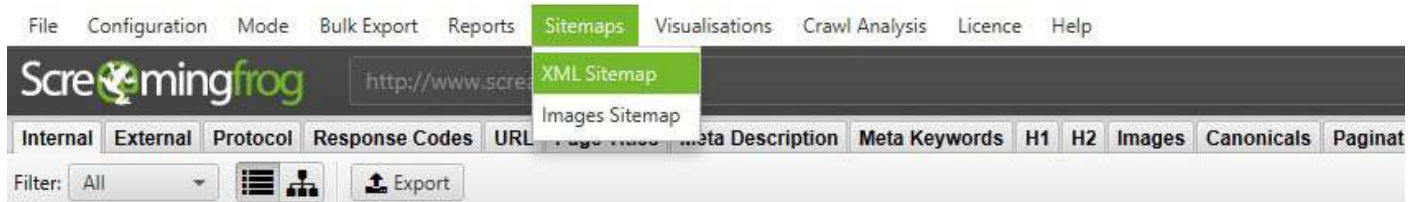
Fonte: https://developers.google.com/search/docs/advanced/robots/robots_txt

SITEMAP

Screaming Frog è un ottimo strumento per generare con semplicità sia Sitemap XML che Sitemap Images.

Rispetto ai tool che offrono lo stesso servizio online il Seo Spider permette di scegliere gli URL che compongono la Sitemap in modo estremamente granulare evitando l'inserimento di pagine con status code diverso da 200, pagine canonicalizzate, con tag "no index" etc. che consumano inutilmente Crawling Budget.

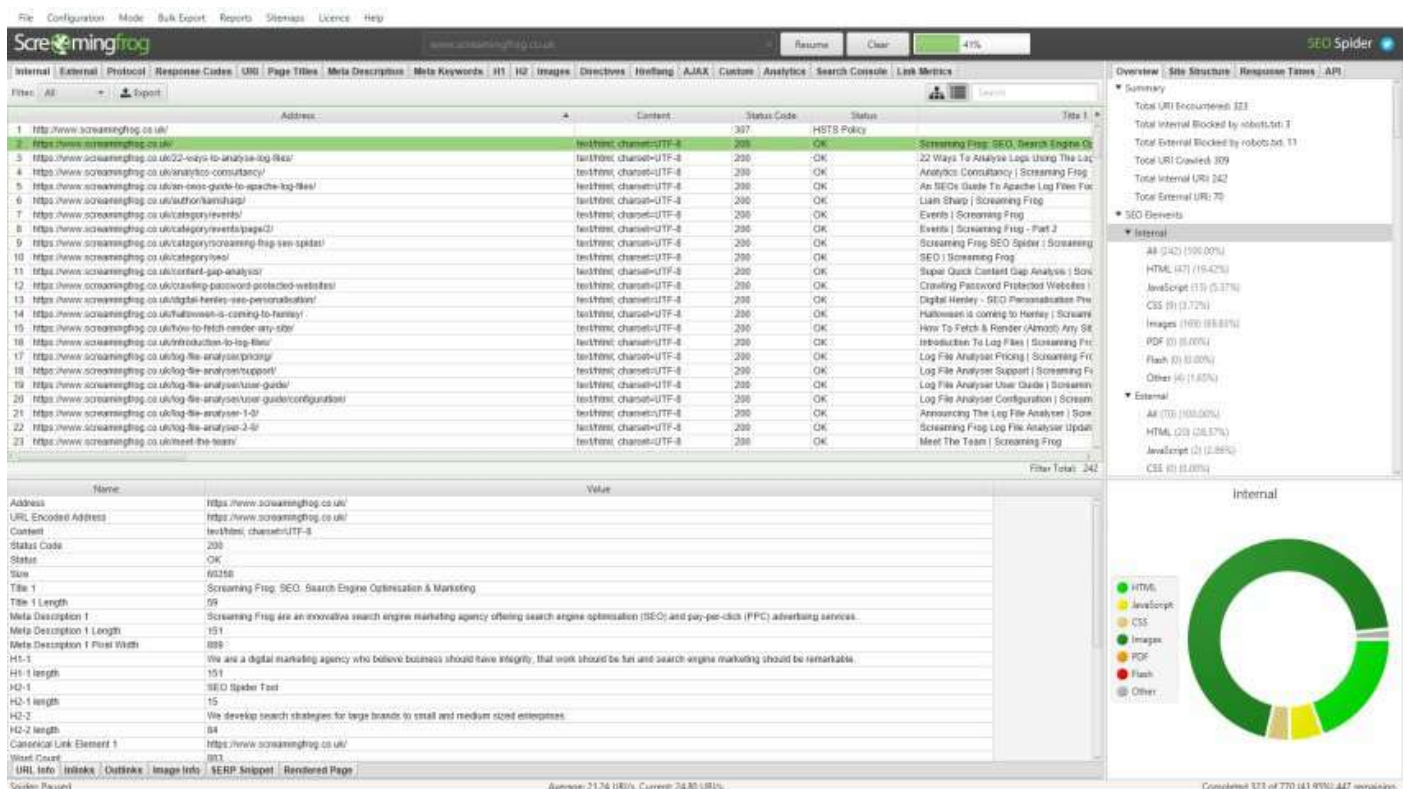
Nel caso il numero degli URL superino le direttive di Google, il Seo Spider genera in automatico una "Sitemap Index" in base agli standard di Sitemap.org.



HOW TO:

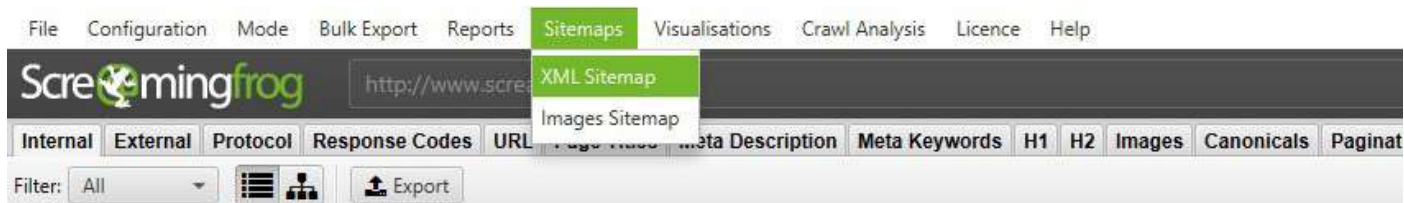
GENERARE UNA SITEMAP

1. Per creare la tua Sitemap è sufficiente inserire il dominio da scansione e avviare il crawl.



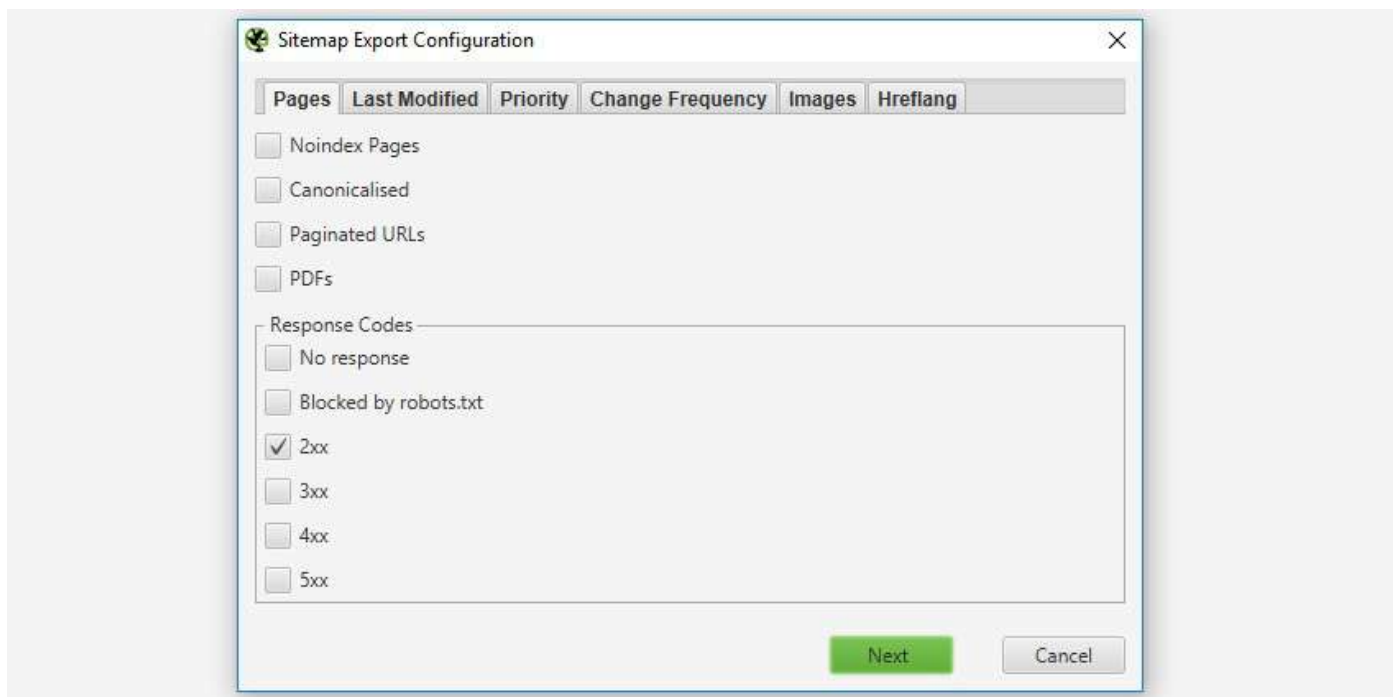
2. Una volta completata la scansione puoi cliccare sulla voce del menu "Sitemap" e scegliere tra due opzioni:

- Sitemap XML.
- Images Sitemap.



3. All'apertura del "Sitemap Export Configuration" puoi decidere quali attributi devono caratterizzare gli URL per essere aggiunti alla Sitemap XML.

Considera che per default vengono inserite solo pagine html con status code 200 mentre sono automaticamente escluse le risorse "Canonicalizzate", le paginazioni, le pagine bloccate dal robots.txt e pdf o che rispondono con "Status Code" 5xx, 4xx o con reindirizzamenti.



Nella maggior parte dei casi puoi lasciare inalterata la tab "Pages" con le configurazioni di default mentre è buona norma gestire la scheda "Priority" gestendo il range compreso tra 0.1 con bassa priorità di scansione e 1.0 ad alta priorità.

Sitemap Export Configuration

Pages | Last Modified | **Priority** | Change Frequency | Images | Hreflang

☒ Include <priority> tag

Crawl Depth 0: 1

Crawl Depth 1: 0.9

Crawl Depth 2: 0.8

Crawl Depth 3: 0.7

Crawl Depth 4: 0.6

Crawl Depth 5+: 0.5

Next Cancel

Puoi anche definire la “Change Frequency”, attributo opzionale che ‘suggerisce’ la frequenza con cui è probabile che la pagina cambi (imprescindibile nel caso tu gestisca un sito di news).

Il Seo Spider calcola la “Frequency Change” in base all’ultima modifica dell’Header HTTP o al “livello” (profondità) degli URL. L’opzione ‘Calculate from Last Modified Header’ imposterà “daily” se la pagina è stata modificata nelle ultime 24 ore altrimenti sarà impostata su ‘monthly’.

Puoi integrare la Sitemap anche con le immagini e gli attributi Hreflang in modo facoltativo. Se l’opzione “include images” è spuntata, allora tutte le immagini che popolano la scheda “Internal” (e tab ‘Images’) saranno incluse per default.

Come mostrato nello screenshot qui sotto, puoi anche scegliere di includere solo le immagini che risiedono su una CDN.

Sitemap Export Configuration

Pages | Last Modified | Priority | Change Frequency | **Images** | Hreflang

☒ Include Images

☐ Include Noindex Images

☒ Include only relevant Images with up to: 10 inlinks

Regex list of CDNs hosting images to be included

.*cdn.screamingfrog.co.uk.*

Next Cancel

Per ottimizzare al meglio quali immagini inserire nella Sitemap ti consiglio di selezionare l'opzione "Include only relevant images with up X inlinks" che ti permette di non considerare i loghi, icone e in generale immagini di bassa importanza che mediamente ricevono link sitewide. Per gestire questa opportunità ti basta consultare la tab "Images", scoprire il numero di referenze di ogni singola immagine e decidere quali escludere in base al numero di collegamenti.

File Configuration Mode Bulk Export Reports Sitemaps Licence Help

Screamingfrog

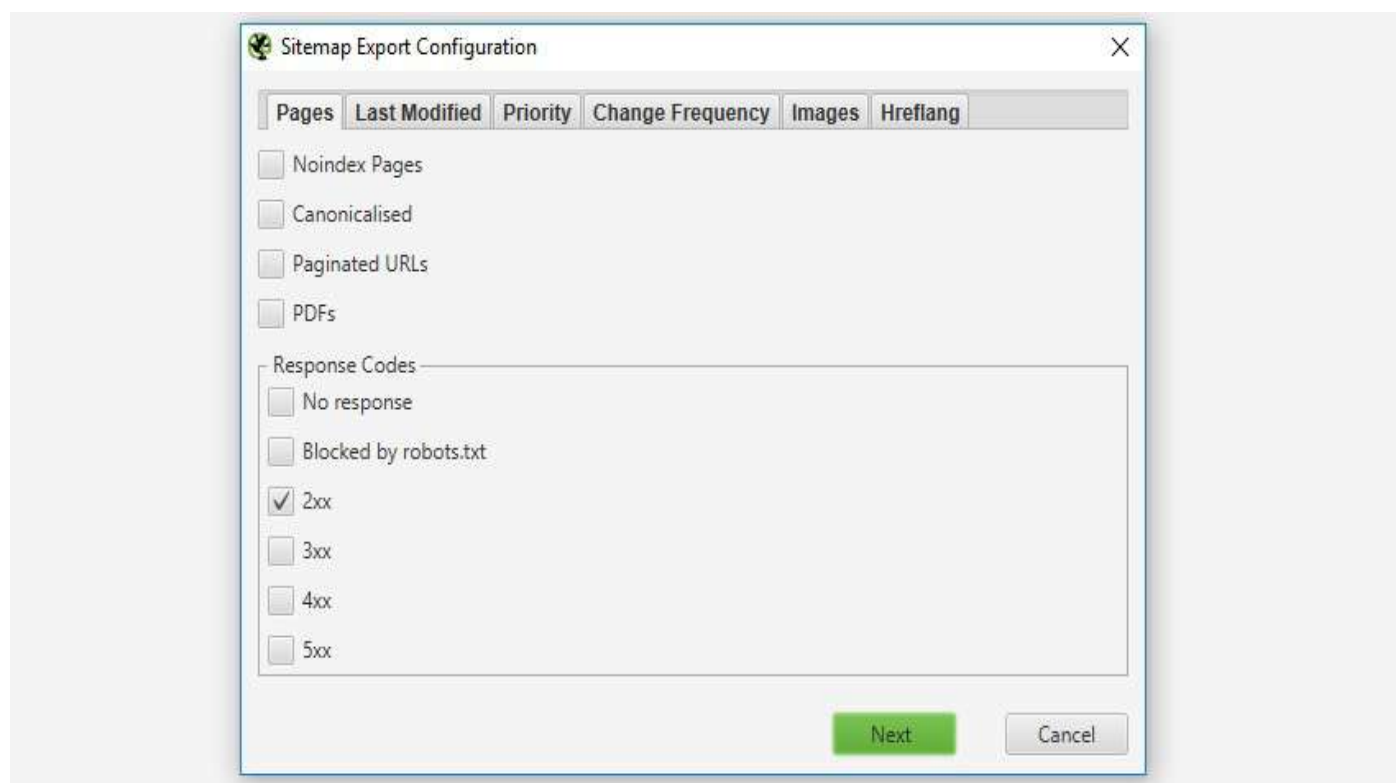
www.screamingfrog.co.uk

Internal External Protocol Response Codes URI Page Titles Meta Description Meta Keywords H1 H2 Images Directives Hreflang

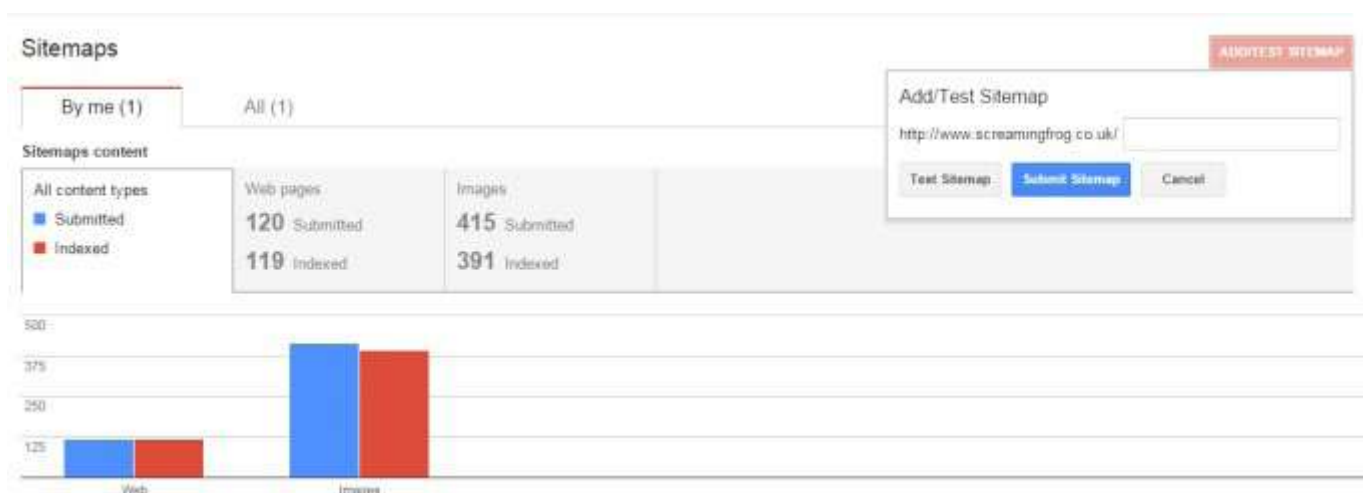
Filter: All Export

	Address	Content	Size	IMG Inlinks▼
1	https://secure.gravatar.com/avatar/3a6a739ea17db3d307dc910a689abbcd?s=100&d=mm&r=g	image/png	10351	407
2	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/social/twitter.png	image/png	2154	379
3	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/social/google.png	image/png	2556	379
4	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/social/facebook.png	image/png	1937	379
5	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/social/rss.png	image/png	1871	378
6	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/social/linkedin.png	image/png	2280	378
7	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/hamburger.png	image/png	234	216
8	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/logos/logo.png	image/png	4694	216
9	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/frogbg2.png	image/png	1617	216
10	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/footer-fly.png	image/png	1071	216
11	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/frogbg1.png	image/png	2303	216
12	https://www.screamingfrog.co.uk/wp-content/uploads/2012/04/dan-sharp.jpg	image/jpeg	30747	51
13	https://www.screamingfrog.co.uk/wp-content/uploads/2012/04/matt-hopson.jpg	image/jpeg	35179	23
14	https://www.screamingfrog.co.uk/wp-content/uploads/2015/12/morphcostumes.png	image/png	1459	18
15	https://www.screamingfrog.co.uk/wp-content/uploads/2015/12/namecheap2.png	image/png	1048	17

4. Completato il processo di configurazione, clicca "Next" ed esporta la tua nuova sitemap.



5. Il Seo Spider ha concluso il suo lavoro, ora tocca a te fare conoscere ai Motori di Ricerca la tua Sitemap XML inviandola tramite Search Console (Google), Bing Webmaster Tool o altro ed aggiungendo il riferimento della Sitemap nel file robots.txt.

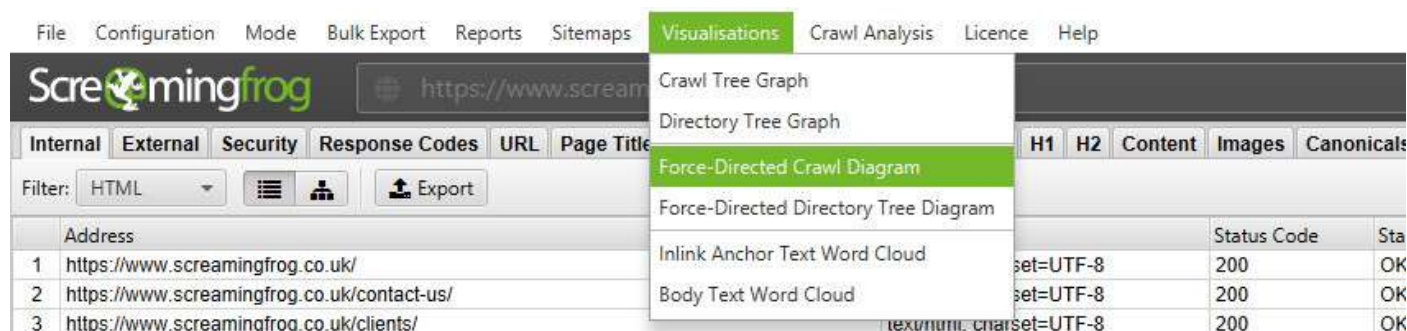


VISUALISATION

Screaming Frog offre tre tipi di visualizzazione:

1. crawl visualisations.
2. Directory tree visualisations.
3. Inlinks Anchor Text e Body Text Word Clouds.

In sintesi la "Crawl Visualisation" ti permette di comprendere al meglio i collegamenti interni, le "Directory Tree Visualisations" l'architettura e organizzazione del sito web mentre la "Cloud Words" ti consente di visualizzare quanto una parola è stata usata come anchor text o nel testo per una visione semantica delle pagine Html.

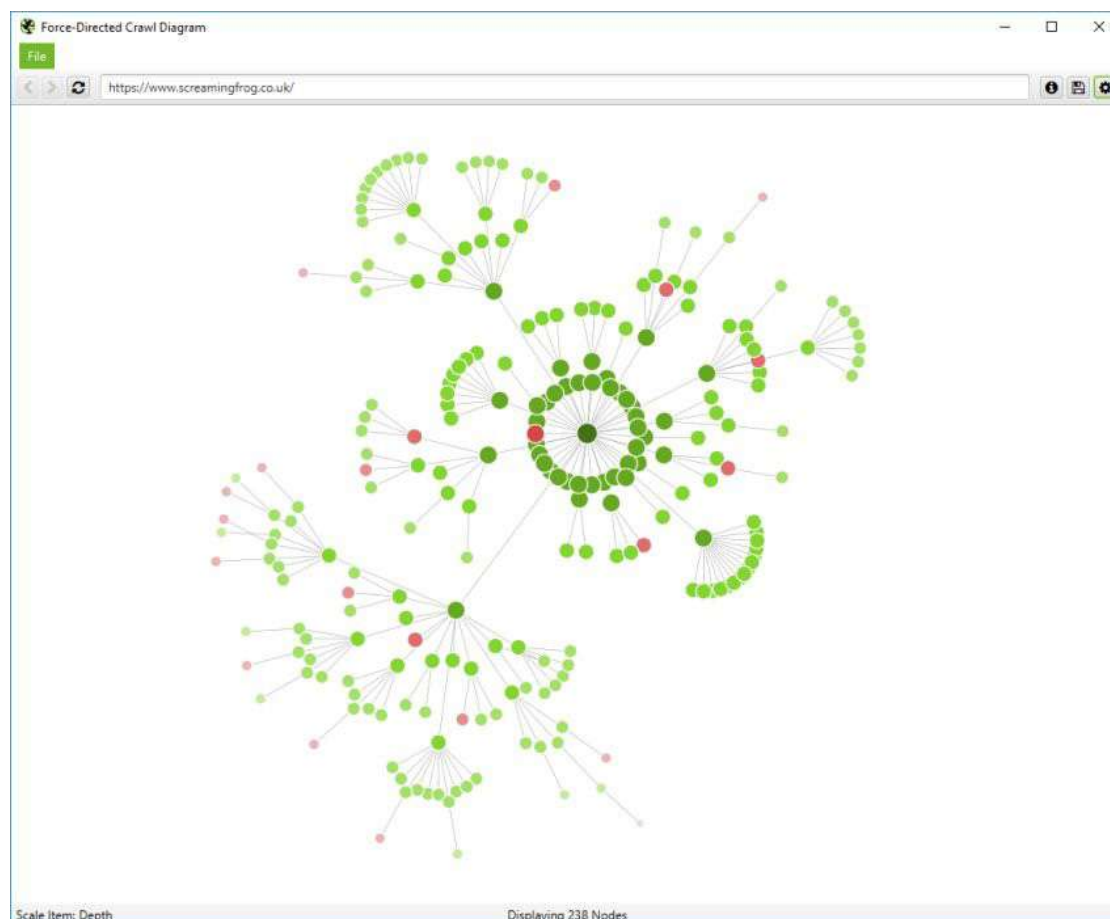


CRAWL VISUALIZATION

Le "Crawl visualisations" includono due opzioni:

1. Force-Directed Crawl Diagram.
2. Crawl Tree Graph.

Queste visualizzazioni sono utili per l'analisi del linking interno e per comprendere i percorsi di navigazione del Seo Spider nella scansione del sito web. Nella rappresentazione grafica del tool nel caso in cui ci siano più percorsi brevi verso una pagina, cioè un URL è collegato da due pagine alla stessa profondità di scansione, Screaming Frog considererà il primo URL scansionato (spesso il primo nella fonte).



Il grafico è interattivo e visualizza i diversi stati dei nodi presenti con colori e intensità differenti. Per comprendere al meglio la rappresentazione grafica ti invito a cliccare sul tasto "i" in alto a destra della schermata dove troverai una legenda.

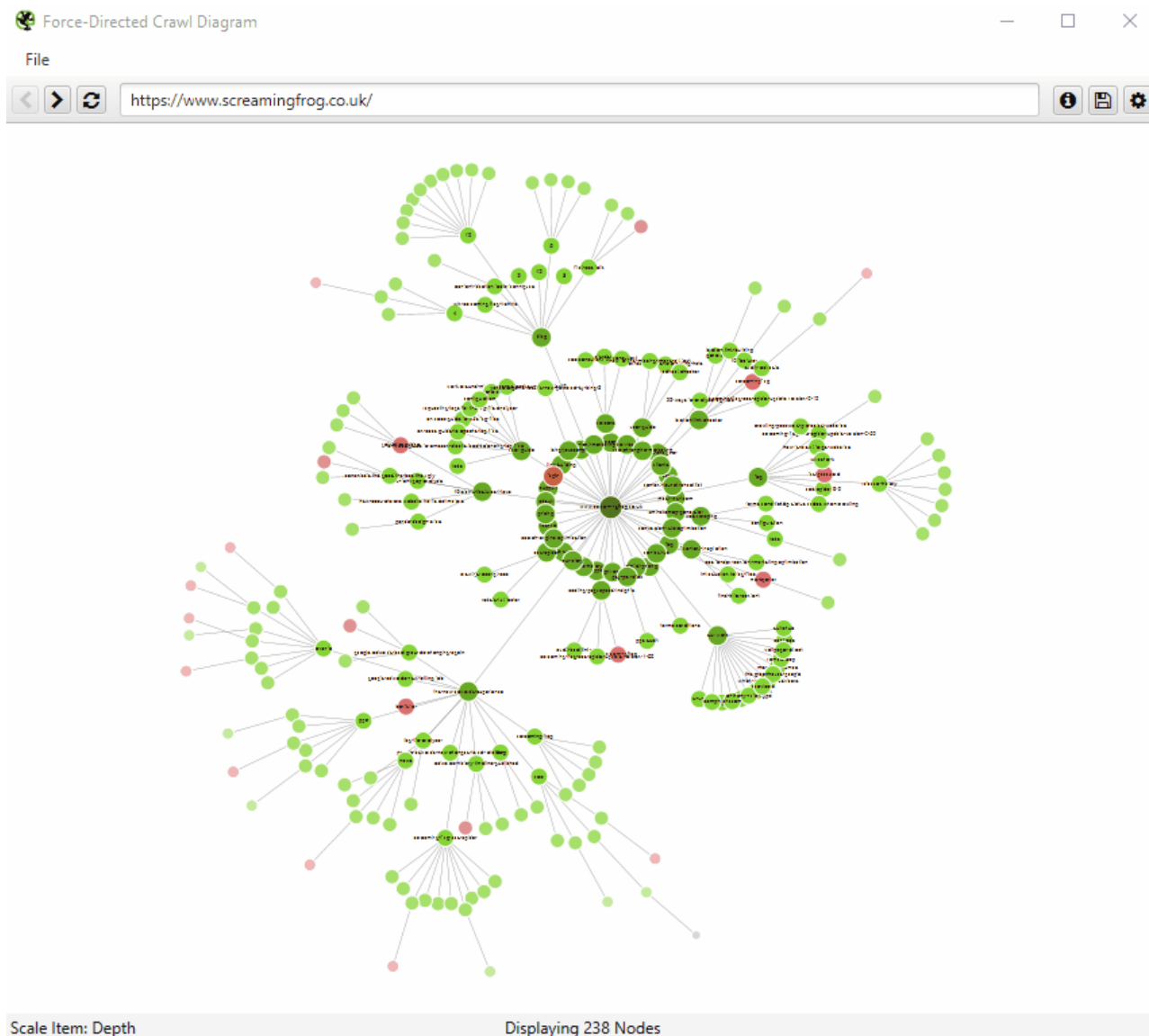
Nella configurazione di default le pagine indicizzabili sono rappresentate dai nodi verdi mentre il cerchio più scuro e più grande al centro è l'URL di partenza (es. la homepage). Man mano che le pagine sono più distanti dal nodo principale in termini di profondità di scansione, sono rappresentate con nodi più piccoli e di colore più chiaro, come fosse una heatmap.

Il rosso pastello evidenzia gli URL che non sono indicizzabili, il che rende abbastanza facile individuare le aree problematiche di un sito web. Ci possono essere delle ragioni valide per mantenere delle pagine non indicizzabili, ma visualizzare la loro proporzione e dove sono, può esserti utile per identificare rapidamente le aree di interesse ed indagare se rappresentano un errore.

La visualizzazione mostra fino a 10k URL alla volta, in quanto questa modalità necessita di molta memoria.

Una delle funzionalità che prediligo di questa visualizzazione è la possibilità di mettere un "focus" su alcuni nodi (tasto destro del mouse) espandendo l'interesse verso specifiche aree e mostrare più URL di quella sezione (fino ad altri 10k URL alla volta).

Per la navigazione fra i nodi puoi usare il mouse o più semplicemente il browser come navigatore, digitando direttamente l'indirizzo della risorsa.

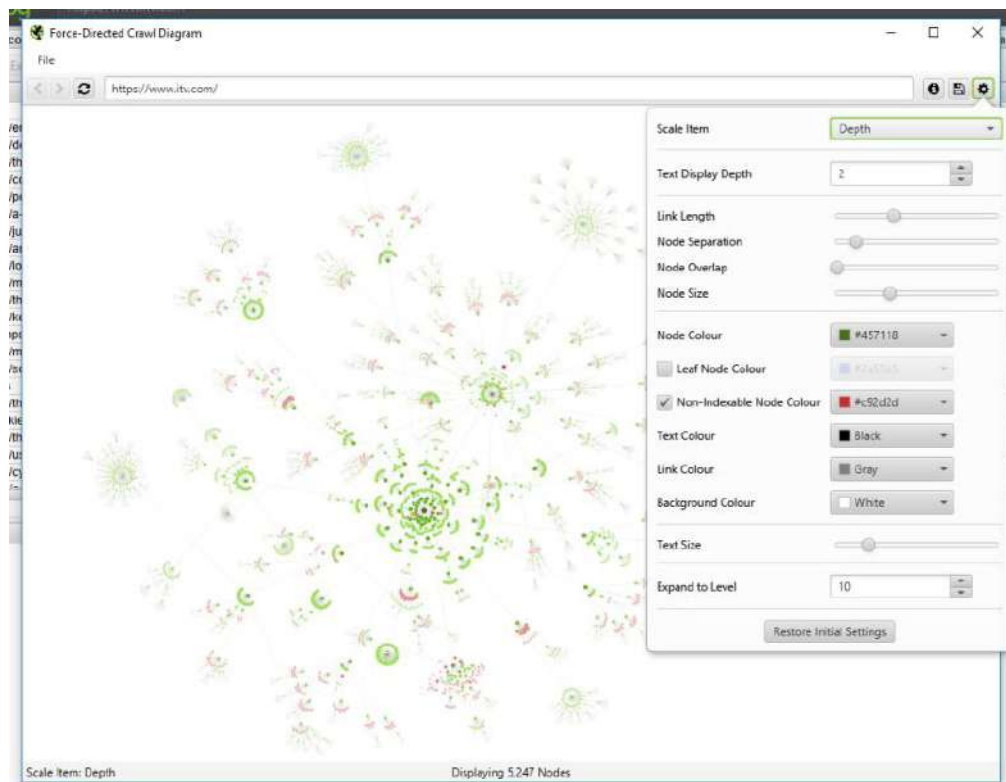


Quando una visualizzazione ha raggiunto il limite di 10k URL, troverai un nodo di colore grigio che identifica che i figli sono stati troncati (a causa dei limiti di dimensione).

Sarà quindi possibile, come già accennato, “esplorare” il nodo e focalizzare l’interesse su quella specifica area espandendola.

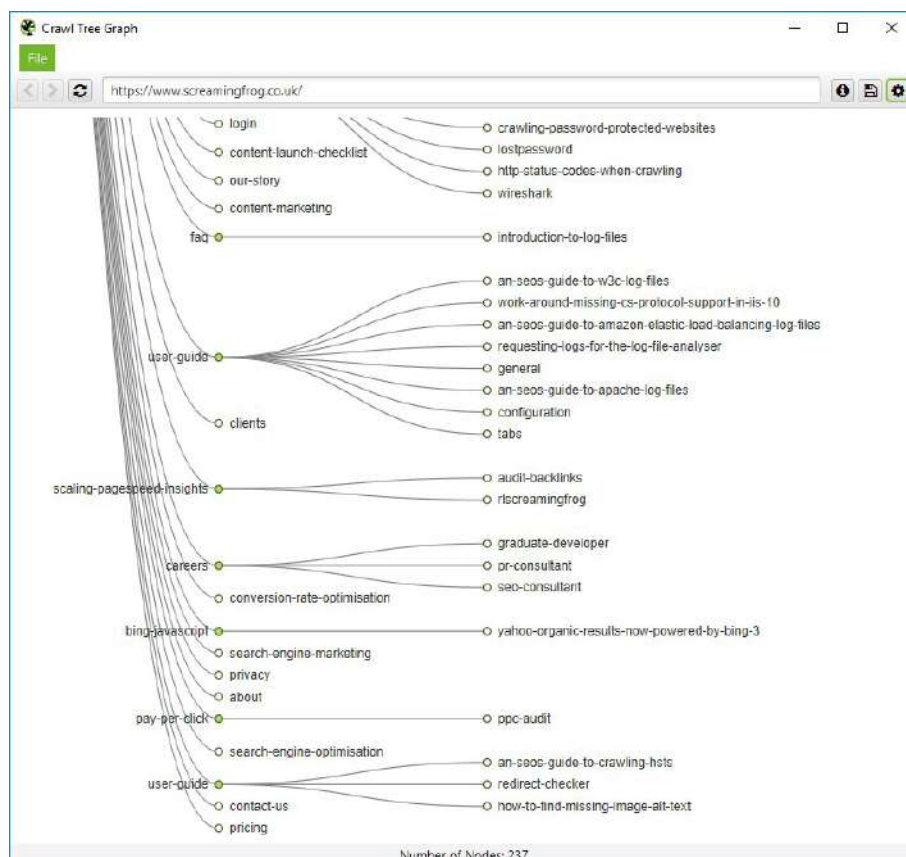
Il Seo Spider offre anche l’opportunità di personalizzare queste visualizzazioni e puoi scegliere a tuo piacimento le colorazioni più congeniali, le dimensioni dei nodi e dei testi etc.

Per attivare questa funzione è sufficiente cliccare sul simbolo dell’ingranaggio in alto a destra della visualizzazione e utilizzare l’editor dedicato.



Inoltre modificando l'opzione "Scale Item" (opzione "Force-Directed Crawl Diagram") la rappresentazione dei nodi sarà scalabile in base ad altre metriche come gli inlink unici, il conteggio delle parole, le sessioni GA, i clic GSC, il Link Score, la Page Authority di Moz e altro ancora. La dimensione e il colore dei nodi saranno scalati in base a queste metriche, il che ti aiuta a visualizzare molti aspetti diversi oltre al linking interno, come ad esempio le sezioni che potrebbero avere un contenuto troppo ridotto (Low Content).

Attraverso questi diagrammi è molto semplice individuare e risolvere problemi di "Orphan Page" e l'uso congiunto di "Force-Directed Crawl Diagram" e "Crawl Tree Graph visualisations" ti permette una visione intuitiva e veloce di eventuali criticità anche in fase di pre analisi.



Entrambe le rappresentazioni sono configurabili per una vista “left to right” (o viceversa) oppure “Top to Bottom” (o viceversa) e ti permettono di applicare il “Focus” sui singoli nodi.

La versione “Top to Bottom” la prediligo quando voglio valutare la divisione in “Silos” del sito web e potenzialmente quali sezioni possono essere ottimizzate nell’architettura del progetto in termini di “Search Intent”.

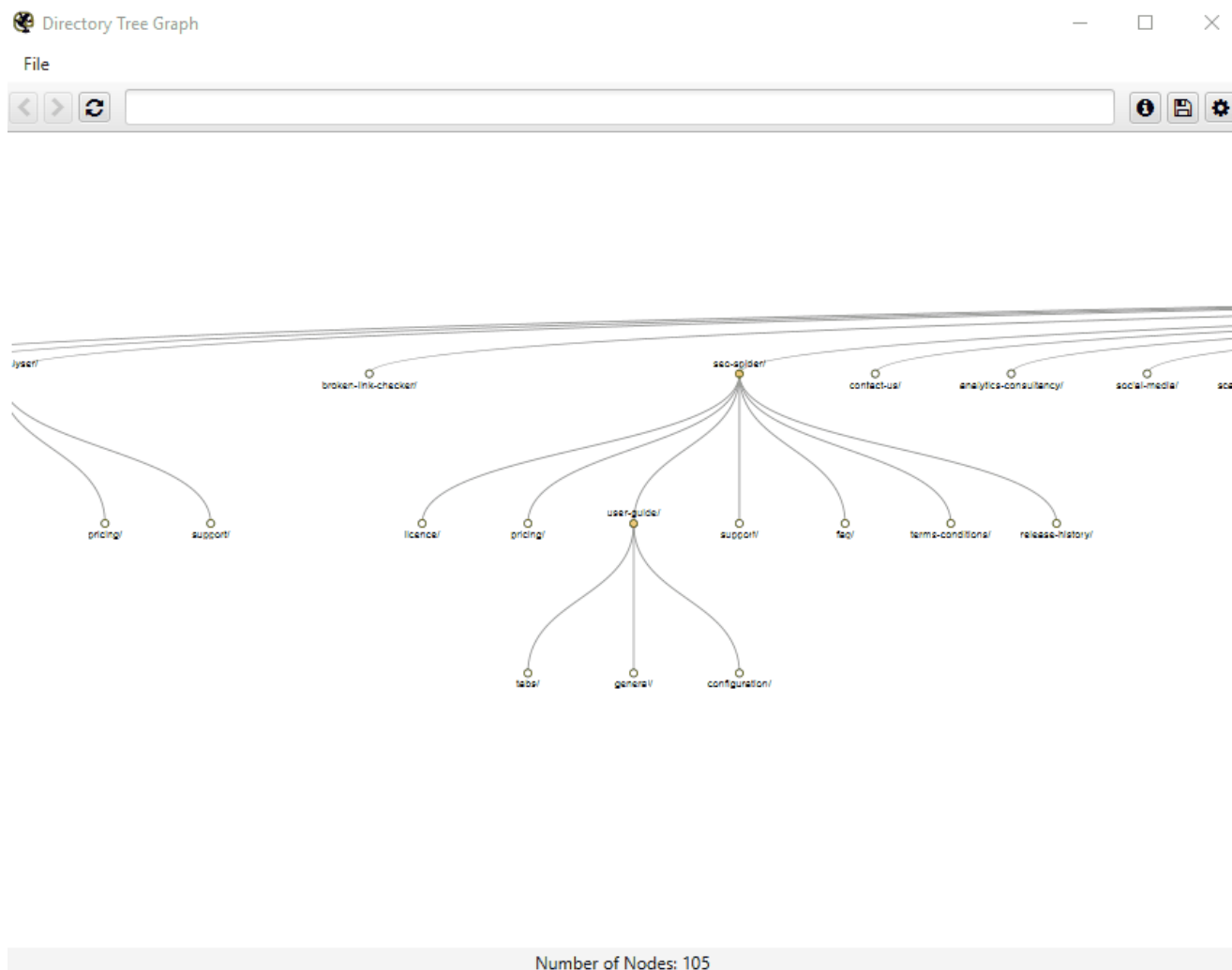
VISUALIZZAZIONE A DIRECTORY

La seconda modalità di visualizzazione che il Seo Spider propone tiene conto delle “directory” del sito web e include il “Force-Directed Directory Tree Diagram” e il “Directory Tree Graph”.

La visualizzazione “Directory Tree” ti può aiutare a comprendere l’architettura degli URL a livello di sito e il modo in cui è organizzato in base al raggruppamento per modelli. Questo ti può essere molto utile, poiché molti raggruppamenti spesso condividono gli stessi template di pagina e problemi SEO identici su cui concentrare il focus.



Con le aggregazioni per modelli puoi visualizzare ad esempio come le pagine “non indicizzabili (nodi rossi)” sono organizzate insieme, poiché hanno lo stesso modello, mentre nel diagramma di crawl sono distribuite ovunque. Questa vista ti permette a colpo d’occhio di prendere confidenza con la struttura del sito e relative criticità in modo già segmentato. La visualizzazione sulle “Directory”, come la precedente “crawl visualisation”, prevede anche la versione ad albero (“Directory Tree Graph”) con le stesse funzionalità di “Force-Directed Directory Tree Diagram”.



Utilizzando la visualizzazione a directory è sempre importante ricordare che i nodi rappresentati non sempre rappresentano degli URL. Possono semplicemente rappresentare un percorso, che non esiste però come pagina.

ESEMPIO:

Un esempio lo puoi vedere eseguendo un crawl del sito di Screaming Frog nel quale esiste una sotto-cartella “screamingfrog.co.uk/**author**/name”, ma non un corrispettivo URL “/author/” che risponde con status code 404.



Come si nota dall'immagine, selezionando il nodo "/Author/" le informazioni del popup sono limitate al solo percorso; se fosse stata una pagina ci sarebbero molte più informazioni come la profondità di scansione, se risponde come "index" o "noindex", dettagli sugli Inlinks e Outlinks etc.



VISUALIZZAZIONE INLINKS TEXT & WORD CLOUDS

Questa visualizzazione ti mostra una “nuvola di parole” per ogni singolo Url del sito web. Per attivare questa funzionalità puoi seguire principalmente due strade:

1. Visualisation > Inlinks Anchor Text Word Clouds/Body Text Word Clouds.
2. Selezionare uno specifico link dalla schermata “Internal” con il tasto destro, scegliere Visualisations e “Inlinks Anchor Text Word Clouds/Body Text Word Clouds”.

The screenshot shows the Screaming Frog SEO Spider interface. The top menu bar includes File, Configuration, Mode, Bulk Export, Reports, Sitemaps, Visualisations, Crawl Analysis, Licence, and Help. The main toolbar has tabs for Internal, External, Protocol, Response Codes, URI, Page Titles, Meta Description, Meta Keywords, H1, H2, Images, Canonicals, and Paginati. Below the toolbar is a filter dropdown set to 'All' and an 'Export' button. The main table lists crawled URLs with columns for Address, Content, and Status Code. A right-click context menu is open over the URL 'https://www.screamingfrog.co.uk/about/'. The menu includes options like Copy, Open in Browser, Re-Spider, Remove, Export, Visualisations, Check Index, Backlinks, Check Google Cache, Open in Wayback Machine, Show Other Domains on IP, and Open robots.txt. The 'Visualisations' option is highlighted, and a submenu is displayed with options: Crawl Tree Graph, Directory Tree Graph, Force-Directed Crawl Diagram, Force-Directed Directory Tree Diagram, Inlink Anchor Text Word Cloud (highlighted), and Body Text Word Cloud.

DIFFERENZE TRA “INLINK ANCHOR TEXT WORD CLOUD” E “BODY TEXT WORD CLOUD”

La “Inlink Anchor Text Word Cloud” include tutti gli Anchor Text interni verso un dato URL o immagine e l’alt tag delle immagini.

Il “Body Text Word Cloud” include tutto il testo all’interno del corpo HTML di una pagina. Per utilizzare questa visualizzazione devi abilitare “Store HTML”.

Config > Spider > Extraction

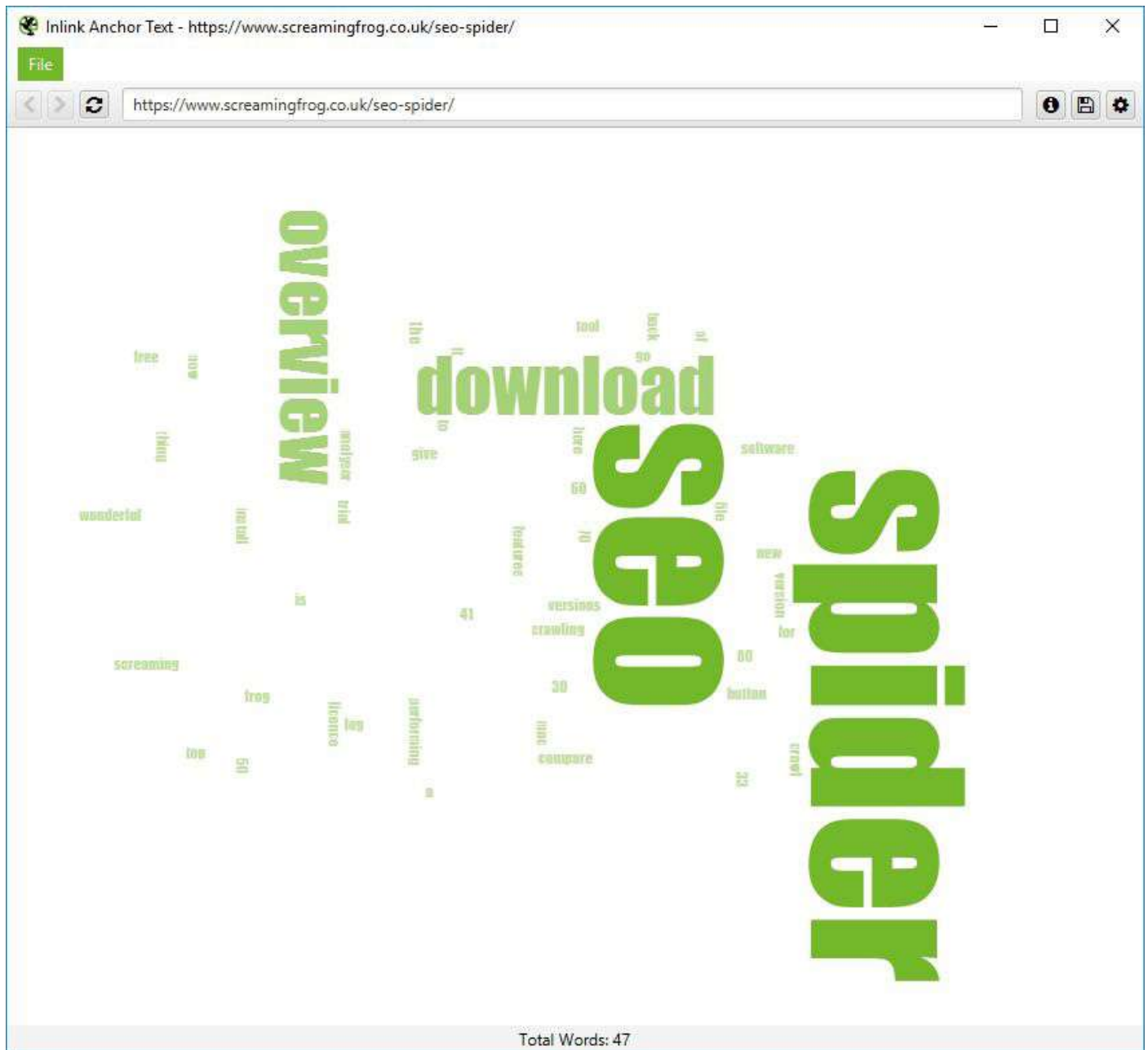
In questo modo sei in grado di salvare l’html statico nel tuo disco e vederlo nella tab (finestra in basso) “View Source” > “Original Html” prima di una eventuale esecuzione dello script Javascript.

Funzione molto comoda per confrontare la versione prima e dopo la renderizzazione e scovare eventuali aree irraggiungibili per lo Spider dei Motori di Ricerca.

Questa modalità di visualizzazione ti permette di avere una prima visione sulla densità di certe query in ottica semantica o se ci sia la presenza di un uso inconsapevole di Anchor Text di default come “Clicca qui”, “scopri” o altro che si ripetono in tutto il sito e devono essere ottimizzati.

Tutte le visualizzazioni possono essere salvate ed esportate in formato SVG o HTML.

Di seguito un esempio di visualizzazione “Inlinks Anchor Text”.

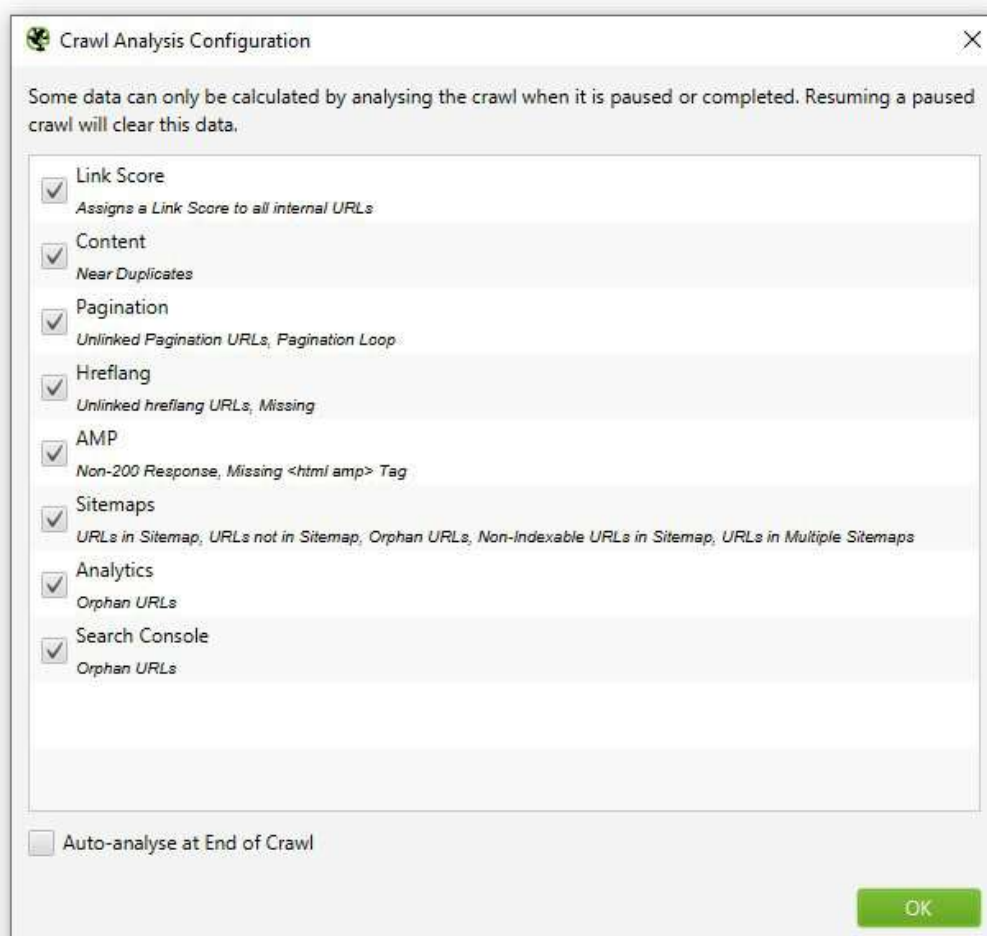


CRAWL ANALYSIS

Il SEO Spider analizza e riporta i dati in tempo reale con metriche, schede e filtri popolati durante il processo. Tuttavia, ci sono alcuni elementi che richiedono un successivo calcolo alla fine della scansione o quando il crawl è stato interrotto.

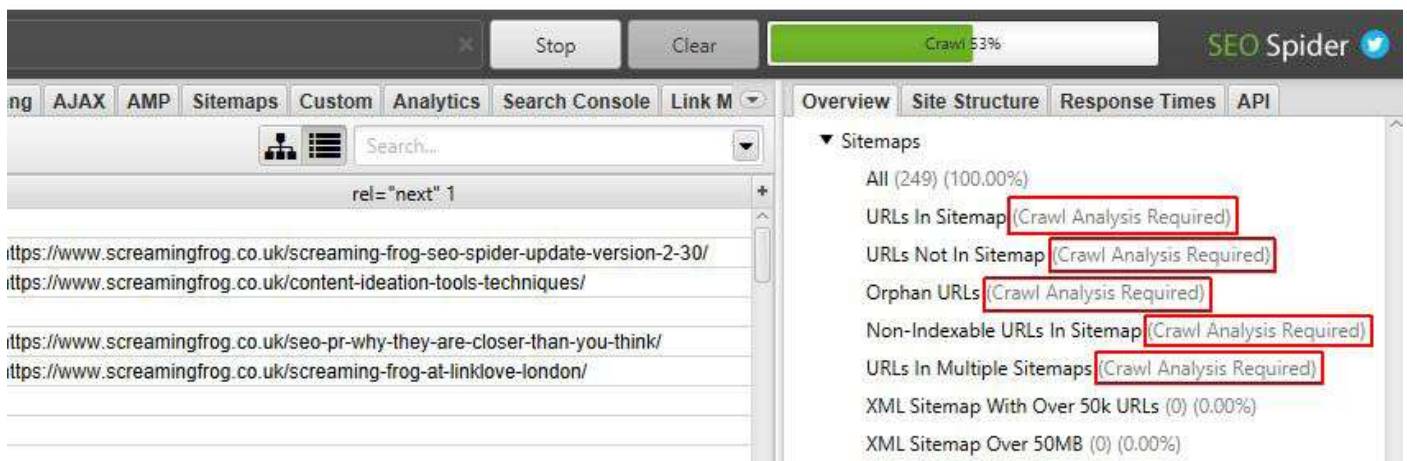
Ad esempio alcune metriche come il “Link Score” e la “Sitemap” richiedono “La Crawl Analysis” rispettivamente per calcolare il potenziale ranking degli URLs in base ai collegamenti interni e per definire quali Urls sono stati scansionati ma non sono presenti nella mappa del sito.

Crawl Analysis > Configure



Nella sidebar del Seo Spider puoi vedere quali filtri richiedono un’analisi successiva al crawl controllando la dicitura “Crawl Analysis Required”.

Nella configurazione della “Crawl Analysis” ti consiglio di abilitare la voce “Auto-analyse at End of Crawl” per automatizzare questa operazione e usufruire di questi dati anche in contesti di scansioni programmate (“scheduling”).



Per utilizzare la “Crawl Analysis” in modo manuale ti basta cliccare sul menu corrispondente e premere start al completamento della scansione.

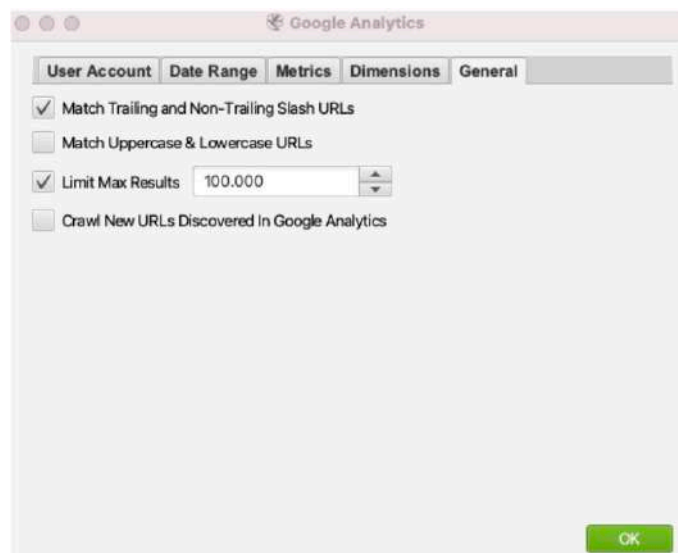


Nota Bene: i dati relativi alle pagine orfane (Orphan Pages) scoperte tramite Google Analytics e Search Console saranno a tua disposizione solamente se hai connesso il Seo Spider alle rispettive API e selezionato l’opzione “Crawl New Urls Discovered in Google Analytics/Google Search Console” nella scheda generale.

Config > API Access > Google Analytics/Search Console >General .

Altrimenti le “Orphan Pages” saranno visibili solamente nei Report

Report > “Orphan Pages”.



Nel caso in cui, dopo aver eseguito la “Crawl Analysis”, non trovi i dati relativi alla Sitemap prova a controllare le configurazioni e verifica di aver abilitato la voce “Crawl Linked XML” e definito manualmente l’indirizzo della sitemap o lasciato allo Spider l’onere di scoprire l’indirizzo mediante la consultazione del robots.txt.

Config > Spider > Crawl Linked XML

Nella seconda opzione verifica la presenza del riferimento alla sitemap nel file robots.txt; se risulta mancante puoi modificarlo direttamente con la funzione “Custom Robots”.

CAPITOLO 3: REPORTS

Screaming Frog ha investito molto negli ultimi anni nell'aggiornare il Spider Tool e contestualmente sono cresciuti anche i "Report" a disposizione. Il tool ti permette di sbizzarrirti nelle esportazioni dei dati garantendo tutta una serie di documenti ad hoc per analizzare la maggior parte delle criticità in modo specifico.

La lista dei documenti sono a tua disposizione nel menu principale e puoi salvarli in locale con formato ".csv", ".xls", ".xlsx" o in "gsheet" sfruttando Google Drive. Nell'ultimo caso li trovi nella cartella "Screaming Frog SEO Spider" di Drive.

CRAWL OVERVIEW REPORT

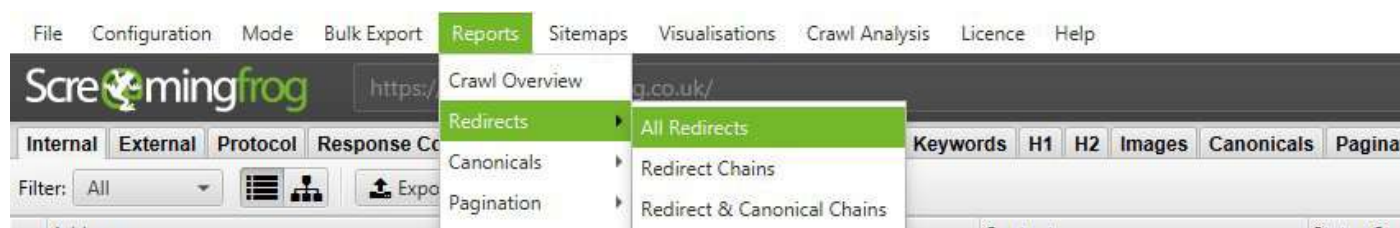
Questo rapporto ti fornisce i dati che trovi nella sidebar laterale del Seo Spider. Il documento rappresenta un riepilogo sintetico della scansione molto utile da utilizzare in prima istanza per una supervisione dall'alto di ciascuna delle schede e dei rispettivi filtri.

REDIRECTS

Questi documenti mettono il focus sui reindirizzamenti scoperti durante la scansione identificando l'origine del collegamento.

Al momento trovi tre diversi report a tua disposizione:

1. **All Redirects:** mostra tutti i singoli reindirizzamenti e segnala la presenza di eventuali catene di link.
2. **Redirect Chains:** comprende gli URL che presentano 2 o più reindirizzamenti consecutivi creando una pericolosa "Redirect Chain".
3. **Redirect and Canonical Chains:** identifica gli URL con almeno 2 reindirizzamenti e catene di URL "canonical".



I documenti "Redirect Chains" e "Redirect and Canonical Chains" sono molto utili per mappare le catene di "redirect", "canonical" e situazioni di "Loop" permettendoti di identificare il numero di salti lungo il percorso e quali siano gli URL sorgente.

In modalità Spider (Mode > Spider) questi rapporti ti mostrano tutti i redirect scoperti da un singolo hop in su.

Il documento è diviso in colonne e comprende:

1. il "numero di reindirizzamenti".
2. La tipologia di catena identificata:
 - I. HTTP Redirect,
 - II. JavaScript Redirect,
 - III. Canonical.
 - IV. etc.
3. Se esistono cicli di reindirizzamenti (Loop).

Se il documento non presenta record, significa che lo Spider non ha trovato loop o catene di redirect che devono essere ottimizzate.

I rapporti "Redirects", "Redirect Chains" e "Redirect & Canonical Chains" li puoi anche scaricare nel caso di scansione in modalità elenco (Mode > List). In questo caso il documento mostrerà una riga per ogni URL fornito nella lista.

Spuntando le opzioni “Always Follow Redirects - Segui sempre i Redirect” e “Always Follow Canonicals - Segui sempre i canonici” il Seo Spider continuerà a scansionare i redirect e i canonici in modalità elenco ignorando la profondità di scansione. In sintesi, seguirà tutti i passaggi fino alla destinazione finale.

Nota Bene: i documenti relativi ai reindirizzamenti ti saranno particolarmente utili in fase di migrazione di un sito web per controllare la corretta redistribuzione del Ranking acquisito e non sprecare inutilmente Crawling Budget.

REPORTS CANONICAL

I documenti “Canonical Chains” e “Non-Indexable Canonicals” evidenziano errori e problemi con gli elementi di link canonici o implementazione dei canonici HTTP.

1. Il rapporto “**Canonical Chains**” evidenzia qualsiasi URL che ha più di 2 link canonici in una catena.

ESEMPIO:

- I. L’URL A è “canonicalizzato” e presenta un “canonical URL” verso l’URL B
- II. L’URL B è anch’esso “canonicalizzato” verso l’URL C.

2. Il rapporto “**Non-Indexable Canonicals**” evidenzia errori e problemi con i link canonici. In particolare, questo rapporto mostra tutti i link canonici che hanno come destinazione URL bloccati da robots.txt, presentano status code 3XX, 4XX o 5XX o qualsiasi cosa diversa da una risposta 200 ‘OK’.

Questo rapporto fornisce anche dati su tutti gli URL che sono stati scoperti dal Seo Spider tramite il link “canonical” ma non presentano collegamenti (internal link) dal sito.

Questo dato sarà a disposizione nella colonna “unlinked” con la dicitura “true”.

HOW TO:

COME FARE UNA SCANSIONE DEI CANONICAL

L’elemento “Canonical” (rel=”canonical”) è un attributo utilizzato per specificare ai Bot quale sia la versione preferita di un contenuto presente in più pagine (con slug differenti) per evitare contenuti duplicati.

Questo tag è molto importante in tutti i siti web ma diviene imprescindibile negli e-commerce, vediamo il motivo.

ESEMPIO:

Immagina di avere un e-commerce che vende abbigliamento sportivo delle migliori marche e la pagina di dettaglio della “scarpa da running” possa essere raggiunta sia dalla pagina brand che dalla categoria “running”.

Ipotizziamo due scenari:

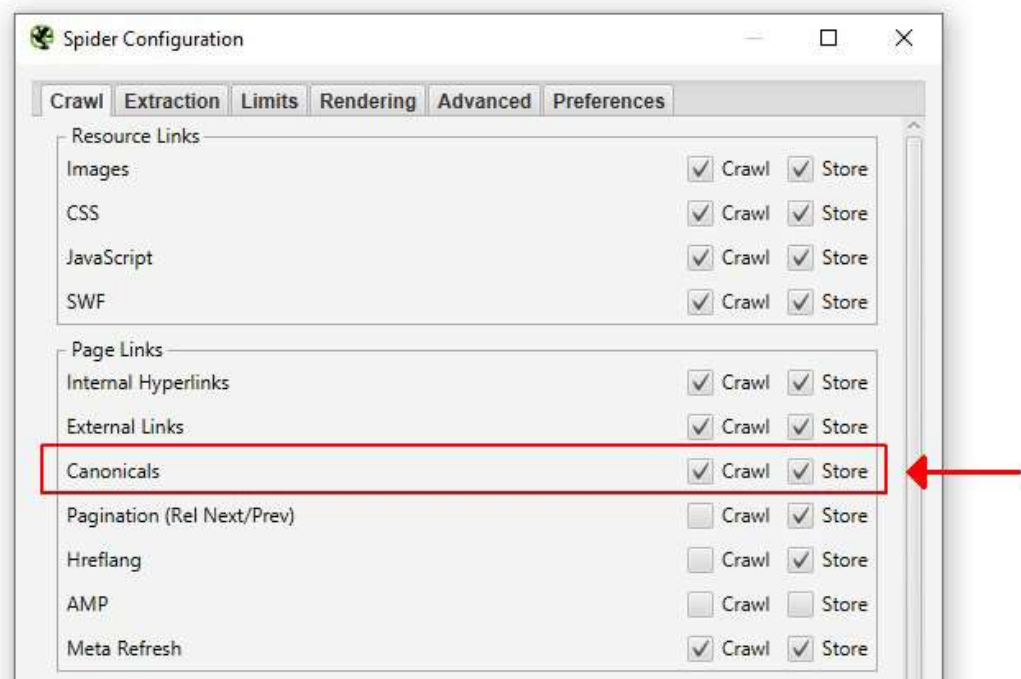
1. il navigatore clicca sulla “scarpa sportiva” dalla pagina “brand”.
Pagina di dettaglio prodotto dalla pagina ‘brand’ Asics: **dominio.com/asics/gel-venture-8**
2. il navigatore clicca sul prodotto della categoria “Running”.
Pagina di dettaglio prodotto dalla pagina di categoria ‘Running’: **dominio.com/running/gel-venture-8**.

In questo caso la pagina di dettaglio si porta nello slug anche i riferimenti di categoria e il sito presenta lo stesso contenuto con due percorsi differenti. L’uso del tag canonical ti aiuta a definire quale delle due dovrà essere considerata dallo Spider nella definizione del ranking.

Il Seo Spider scansiona gli elementi “rel=canonical” che si trovano nell'HTML e negli header HTTP e consente di avere un rapporto preciso sulla loro configurazione e sugli errori più comuni.

Per la diagnosi dei Canonical è sufficiente abilitare la scansione e la memorizzazione dei dati dal menu configurazioni ed eseguire il crawl.

Config > Spider > Crawl



I dati sono a tua disposizione nella scheda “canonicals” della finestra superiore e possono essere filtrati in base ai seguenti attributi:

- **Contains Canonicals:** la pagina ha un URL canonico impostato (tramite elemento di collegamento, intestazione HTTP o entrambi). Questo potrebbe essere un URL canonico autoreferenziale, dove l'URL della pagina è lo stesso dell'URL canonico, o potrebbe essere “canonicalised” (canonizzato), dove l'URL canonico è diverso dall'URL della pagina.
- **Self Referencing:** c'è la presenza di un link canonico che corrisponde allo stesso URL della pagina come autoreferenziazione che ne determina lo status di contenuto primario (vedi esempio sopra dell'e-commerce e scarpa sportiva); quindi se esiste un altro “Url Path” con lo stesso contenuto dovrebbe prevedere un tag “Canonicalised”.
- **Canonicalised:** la pagina presenta un URL canonico diverso da se stesso. L'URL in questione è “canonicalizzato” in un'altra posizione. Questo significa che i Motori di Ricerca vengono istruiti a non indicizzare la pagina, e il Ranking viene consolidato alla versione canonica.
- **Missing:** non c'è un URL canonico presente né come elemento di collegamento, né tramite intestazione HTTP. Se una pagina non indica un URL canonico, Google identifica quella che ritiene essere la migliore versione o URL. Questo può portare ad una imprevedibilità del ranking, e quindi generalmente tutti gli URL dovrebbero specificare una versione canonica (self-referencing).
- **Multiple:** ci sono più canonici impostati per un singolo URL (o più elementi di collegamento, intestazione HTTP, o entrambi combinati). Questo può portare all'imprevedibilità, poiché ci dovrebbe essere solo un singolo URL canonico impostato da una singola implementazione (elemento di collegamento o intestazione HTTP) per ogni pagina.
- **Non-Indexable Canonical:** l'URL canonico è verso una pagina non indicizzabile. Questo include i canonici che sono bloccati da robots.txt, nessuna risposta, redirect (3XX), errore del client (4XX), errore del server (5XX) o 'noindex'.

Le versioni canoniche degli URL dovrebbero essere sempre indicizzabili, pagine con risposta '200'. Pertanto, i canonici che vanno a pagine non indicizzabili dovrebbero essere corretti alle versioni indicizzabili ('indexable').

Filter	Address	Occurrences	Indexability	Indexability ..	Canonical Link Element 1
✓ All					
1 Contains Canonical	e.co.uk/	1	Indexable		https://www.theightingsuperstore.co.uk/
2 Self Referencing	e.co.uk/interior-lighting/kitchen-lights-and-undersheif-lighting	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/kitchen-lights-a
3 Canonicalised	e.co.uk/interior-lighting/floor-lamps/modern-floor-lamps	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/floor-lamps/mod
4 Missing	e.co.uk/outdoor-lighting/garden-wall-lights/white-outdoor-w...	1	Indexable		https://www.theightingsuperstore.co.uk/outdoor-lighting/garden-wall-lig
5 Multiple	e.co.uk/light-bulbs	1	Indexable		https://www.theightingsuperstore.co.uk/light-bulbs
6 Non-Indexable Canonical	e.co.uk/showroom	1	Indexable		https://www.theightingsuperstore.co.uk/showroom
7	e.co.uk/blog/post/the-most-ecofriendly-offices-in-the-world	1	Indexable		https://www.theightingsuperstore.co.uk/blog/post/the-most-ecofriendly...
8	e.co.uk/outdoor-lighting/outdoor-post-lights/specially-finish...	1	Indexable		https://www.theightingsuperstore.co.uk/outdoor-lighting/outdoor-post-li
9	e.co.uk/outdoor-lighting/spot-lights	1	Indexable		https://www.theightingsuperstore.co.uk/outdoor-lighting/spot-lights
10	https://www.theightingsuperstore.co.uk/interior-lighting/ceiling-lights/stairwell-staircase-pe...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/ceiling-lights/sta
11	https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lighting/bathroom-wall-li...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lightin...
12	https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lighting/bathroom-mirror...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lightin...
13	https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lighting/bathroom-mirror...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lightin...
14	https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lighting/bathroom-mirror...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lightin...
15	https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lighting/bathroom-mirror...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lightin...
16	https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lighting/bathroom-mirror...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lightin...
17	https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lighting/bathroom-mirror...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lightin...
18	https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lighting/bathroom-mirror...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lightin...
19	https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lighting/bathroom-mirror...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lightin...
20	https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lighting/bathroom-mirror...	1	Indexable		https://www.theightingsuperstore.co.uk/interior-lighting/bathroom-lightin...

La colonna "Occurences" conta il numero di elementi rel="canonical" che sono stati scoperti per ogni singolo URL.

Dalla sidebar laterale puoi verificare in tempo reale tutti gli URL che hanno delle criticità nella gestione dei canonical. Nell'immagine qui sotto, puoi vedere un URL che è "canonicalizzato" e 1 URL che presenta un "Non-Indexable Canonical".

Overview	Site Structure	Response Times	API
Canonicals All (4227) (100.00%) Contains Canonical (4227) (100.00%) Self Referencing (4226) (99.98%) Canonicalised (1) (0.02%) Missing (0) (0.00%) Multiple (0) (0.00%) Non-Indexable Canonical (1) (0.02%)			

PAGINATION REPORTS

Il Seo Spider prevede due report dedicati alla paginazione:

1. "Non-200 Pagination URLs".
2. "Unlinked Pagination URLs".

Questi due documenti sono molto utili per evidenziare errori e criticità con gli attributi "rel=prev" e "rel=next" delle paginazioni.

Consultando "Non-200 Pagination URLs" sei in grado di identificare ogni URL con gli attributi "rel=prev" e "rel=next" che richiamano risorse con uno status code diverso da 200.

Il filtro "Unlinked Pagination URLs" ti aiuterà a definire tutti gli url che sono stati scoperti dallo Spider Seo grazie agli attributi "next/prev" ma che non presentano alcun collegamento (link) interno al sito web.

Al momento della guida questi attributi non vengono più considerati da Google ma rimangono ancora utilizzati dagli altri Motori di Ricerca.



HREFLANG REPORTS

Il tag Hreflang permette a Google di conoscere le diverse versioni linguistiche di una pagina web e di restituire la versione più appropriata al navigatore in base al suo browser.

Screaming Frog dedica 7 documenti da poter scaricare ognuno con un focus diverso sull'attributo:

- **All Hreflang URLs:** questo rapporto è il più completo e colleziona tutti gli URL & Hreflang URL includendo le indicazioni per regione e lingua scoperti durante la fase di crawl.
- **Non-200 Hreflang URLs:** attraverso questo documento sei in grado di scoprire quali sono le pagine che comprendono collegamenti Hreflang verso pagine con status code diverso da 200 (no response, blocked by robots.txt, 3XX, 4XX or 5XX responses).
- **Unlinked Hreflang URLs:** documento che rileva tutti gli URL scoperti tramite collegamento "Hreflang" ma che non presentano nemmeno un Hyperlink nel sito (es. Orphan Pages).
- **Missing Confirmation Links:** attraverso questo rapporto sei in grado di individuare quali siano gli URL a cui manca il "Confirmation Links". Ad esempio la pagina A presenta un hreflang verso la pagina B e informa il Bot del Motore di Ricerca che A e B sono due versioni linguistiche dello stesso contenuto ma B non ha un hreflang verso la pagina A creando incongruenza.

- **Inconsistent Language Confirmation Links:** questo report mostra le pagine di conferma che usano codici di lingua incoerenti e diversi per lo stesso contenuto.
- **Non Canonical Confirmation Links:** questo documento mostra gli URLs che presentano l'hreflang con destinazione verso elementi "non canonical". Se decidi di "canonicalizzare" un elemento stai informando lo Spider che quello specifico contenuto ha una versione primaria "Canonical" quindi nel tuo "hreflang" avresti dovuto inserire quella versione.
- **Noindex Confirmation Links:** rapporto ideale per scovare URLs che presentano il tag Hreflang verso risorse con attributo "No index". Anche in questo caso inviando il Bot verso risorse "Non indicizzabili" si crea un enorme incongruenza tra la scelta di inibire un contenuto e quella di invitarlo a servire proprio quell'elemento come versione linguistica principale.

HOW TO:

COME FARE UN AUDIT CORRETTO DELL'ATTRIBUTO HREFLANG

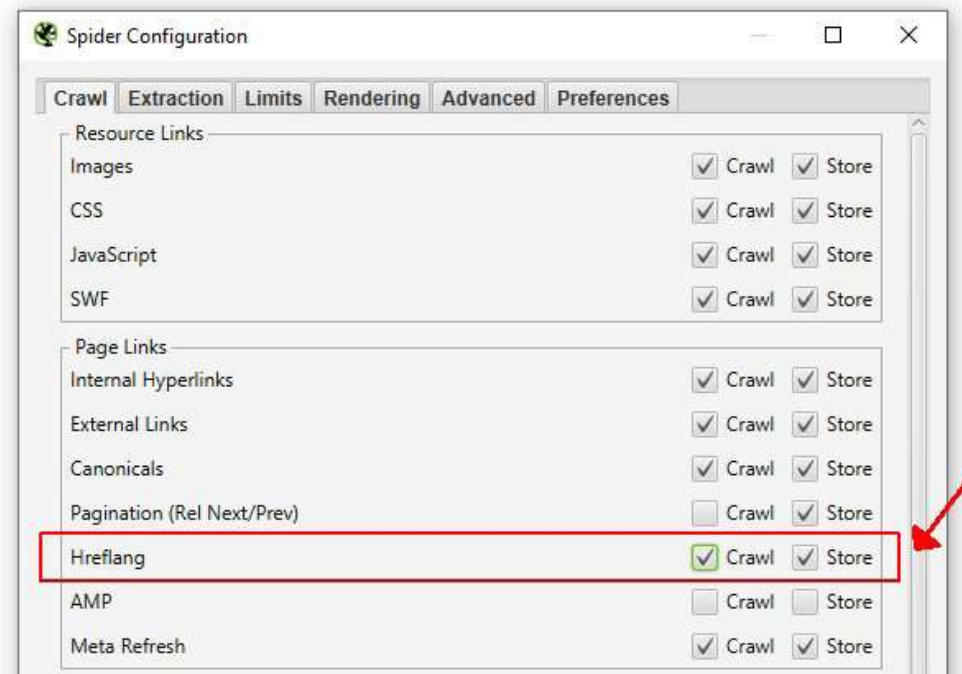
Il Seo Spider scansiona le annotazioni hreflang rel="alternate" nel codice HTML, via HTTP Header o nelle XML Sitemaps e riporta eventuali criticità ed errori nella loro implementazione.

Vediamo insieme il processo per ottenere un Seo Audit del tag Hreflang:

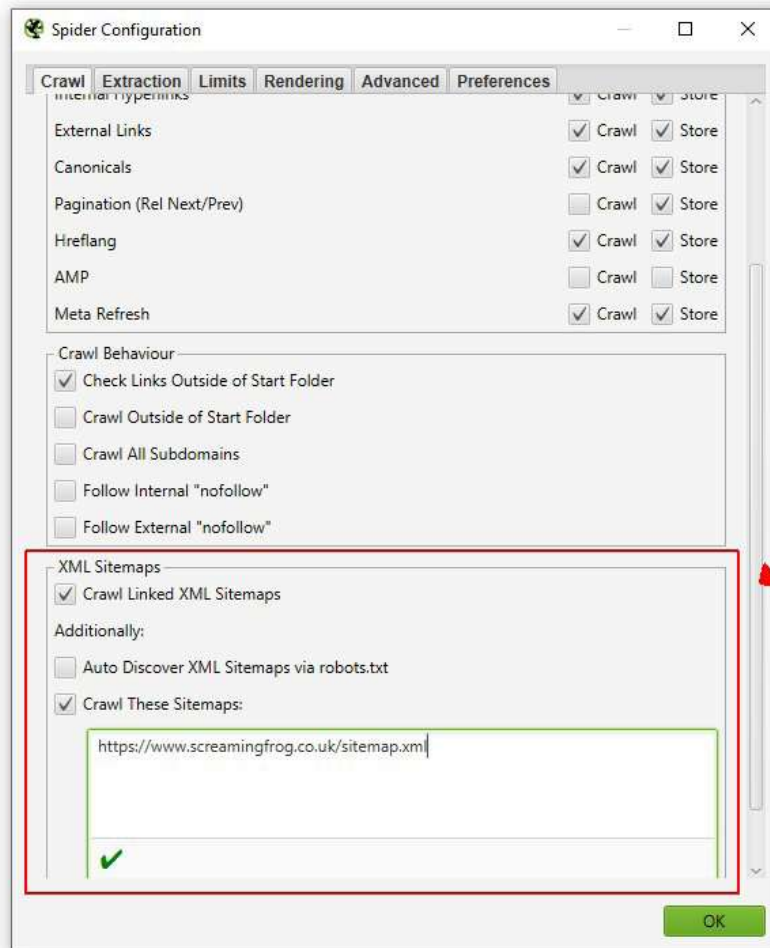
1. Abilitare la scansione e memorizzazione del tag

Per collezionare i dati relativi all'attributo Hreflang è sufficiente abilitare la relativa opzione nelle configurazioni del Seo Spider.

Config > Spider > Crawl



Nel caso volessi scansionare e memorizzare il rel="alternate" della Sitemap XML ti basta abilitare anche la voce "Crawl Linked XML Sitemap" e impostare manualmente il suo indirizzo o lasciarla scoprire dal Seo Spider tramite file Robots.txt.

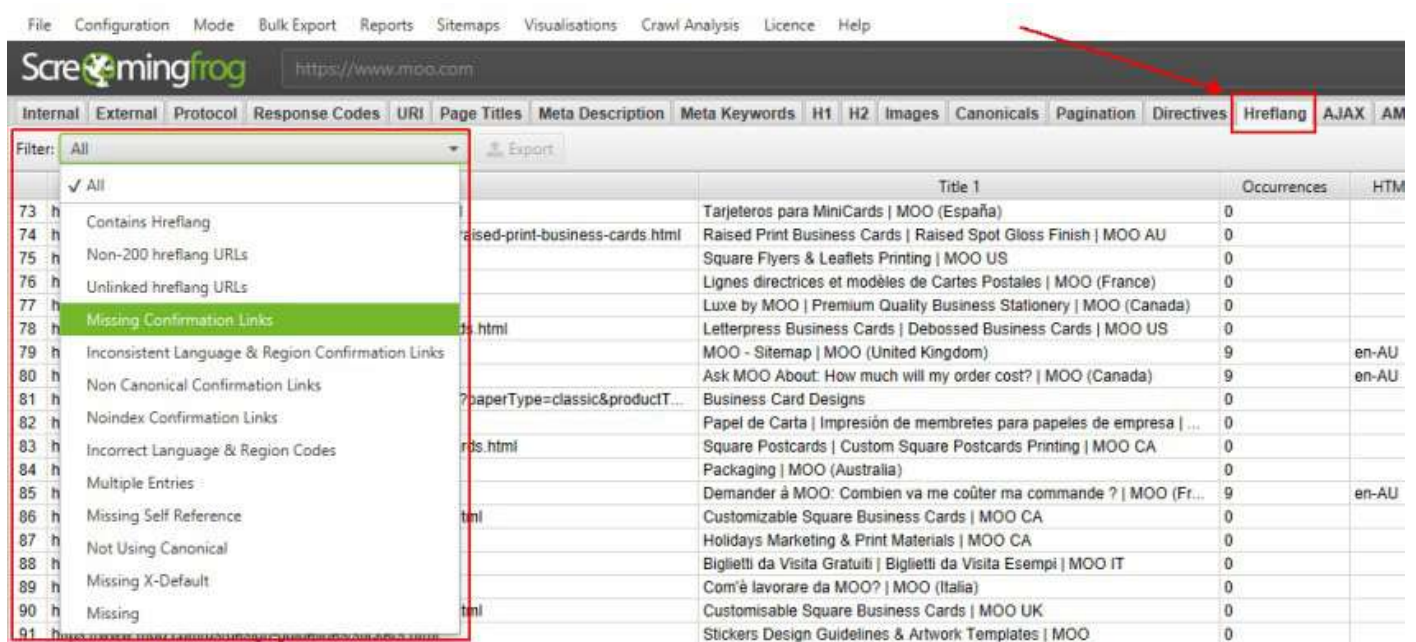


2. Consultazione dei dati nella Tab Hreflang

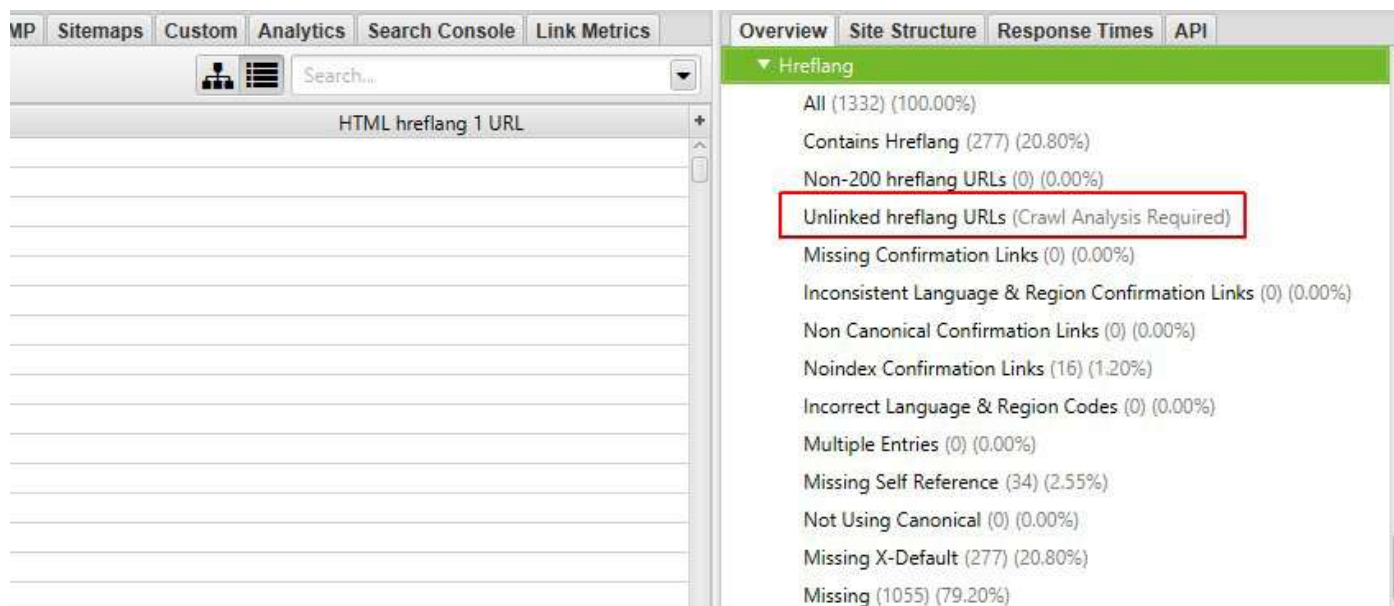
Dopo aver eseguito la scansione trovi i dati memorizzati nella scheda Hreflang con le relative occorrenze per ogni URL.

In questa tab hai a disposizione 13 preziosi filtri che ti permettono di valutare ogni aspetto della hreflang e le eventuali criticità rilevate.

Tutti i filtri sono a tua disposizione al termine della scansione tranne "Unlinked hreflang URLs" che per essere popolato necessita della "Crawl Analysis".

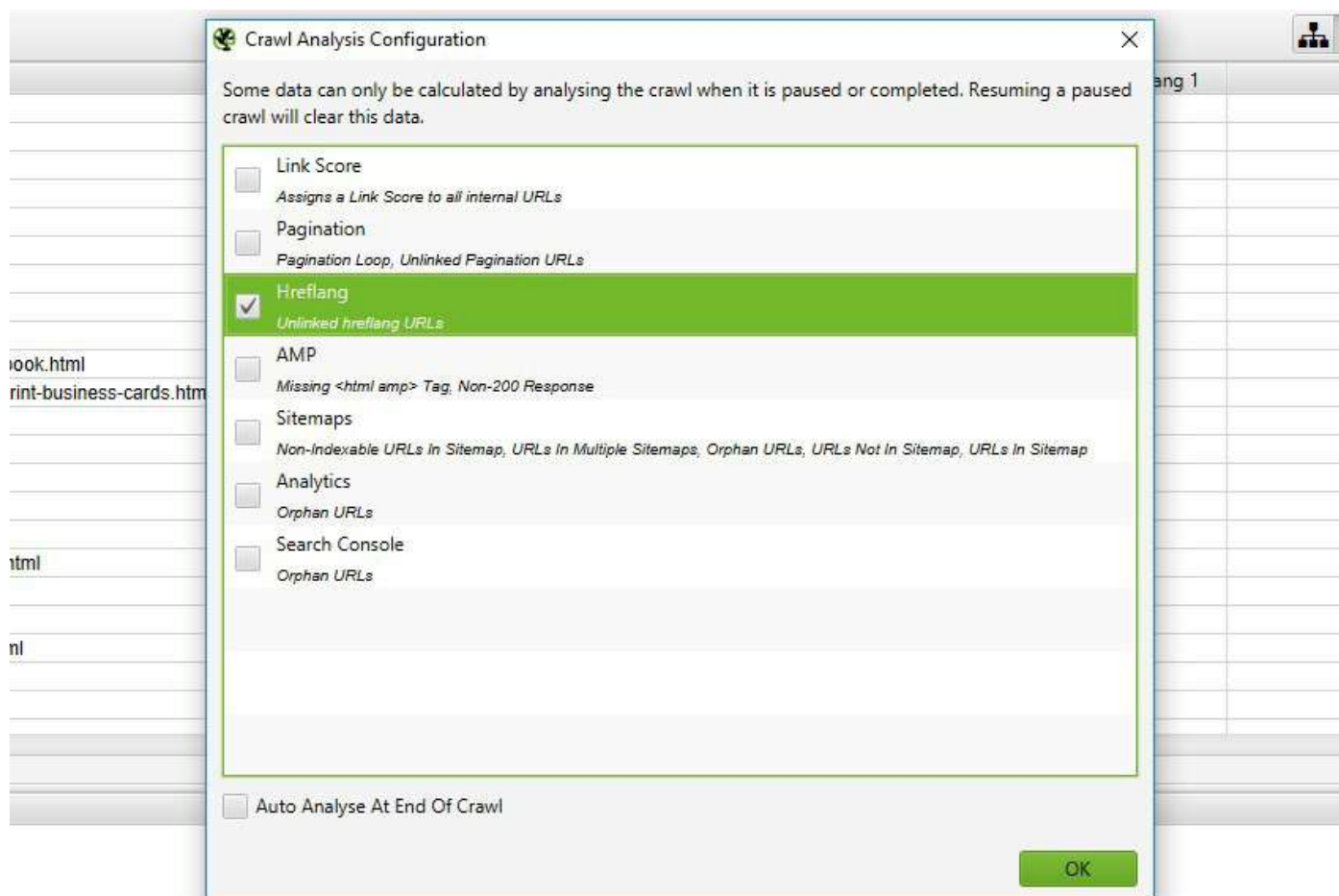


I risultati della scansione sono a disposizione anche nella sidebar laterale con le stesse modalità sopra citate.



3. Avviare la Crawl Analysis

Per completare l'analisi con il filtro mancante al termine della scansione è sufficiente configurare la "Crawl Analysis" dal menu abilitando "Hreflang - Unlinked hreflang URLs". Questo step può essere automatizzato abilitando nella configurazione "Auto Analyse At End Of Crawl".



Una volta completata la scansione con la “Crawl Analysis” tutti i dati sono a tua disposizione. Analizziamo ora nel dettaglio le caratteristiche dei filtri:

1. Contains Hreflang: comprende tutti gli URL che presentano annotazioni hreflang rel=”alternate” da qualsiasi implementazione, sia che si tratti di elementi di collegamento, intestazione HTTP o Sitemap XML.

2. Non-200 Hreflang URLs: si tratta di URL contenuti nelle annotazioni hreflang rel=”alternate” che non hanno un codice di risposta 200. Appartengono a questo segmento tutti gli URL: bloccati da robots.txt, con risposta “no response”, 3XX (reindirizzamenti), 4XX (errori del client) o 5XX (errori del server).

Come abbiamo già trattato precedentemente gli URL hreflang devono essere scansionabili e indicizzabili e quindi gli URL che non rispecchiano queste caratteristiche vengono trattati come errori e ignorati dai Motori di Ricerca.

Gli URL hreflang non-200 li puoi consultare dal pannello “Dettagli URL” della finestra inferiore con uno stato di conferma “non-200”.

Possono essere esportati in blocco tramite l’esportazione del documento “Non-200 Hreflang URLs”.

Report > Hreflang > Non-200 Hreflang URLs.

3. Unlinked Hreflang URLs: questo filtro ti permette di individuare quelle pagine che sono state scoperte solamente attraverso annotazioni hreflang e non attraverso il link interno del sito web.

Ricorda bene : le annotazioni hreflang non passano PageRank come un tag di ancoraggio tradizionale, quindi questo potrebbe rappresentare un campanello di allarme nell’architettura degli “internal link”.

PROGETTO CON DIVERSI TLD

Caso Particolare: stai gestendo un progetto basato su più domini con TLD differenziati per nazionalità e, nello specifico esiste una versione per la lingua italiana “www.dominio.it”, una per la versione inglese “www.dominio.com” e una quella tedesca “www.dominio.de”.

In questo scenario per gestire la scansione degli Hreflang devi considerare la funzione CDNs.

Config > CDN

Per scoprire esattamente quali URL hreflang su queste pagine non sono collegati, usa l’esportazione “Unlinked Hreflang URLs”.

Rapporti > Hreflang > Unlinked Hreflang URLs

4. Missing Return Links: filtro che colleziona URL a cui manca il link di ritorno (o ‘return tags’ in Google Search Console), dalle loro pagine alternative. L’attributo Hreflang è reciproco, quindi tutte le versioni alternative devono confermare la relazione.

ESEMPIO:

Quando ad esempio la pagina X si collega alla pagina Y usando hreflang per definirla come pagina alternativa, la pagina Y deve avere un link di ritorno di conferma. Nessun link di ritorno significa che le annotazioni hreflang possono essere ignorate o non interpretate correttamente dal BOT. Gli URL di ritorno mancanti possono essere visti nel riquadro inferiore della finestra “URL Info” con uno stato di conferma “mancante”.

Possono essere esportati in blocco tramite l'esportazione "Missing return Links - Link di ritorno mancanti"

Rapporti > Hreflang > Missing return Links

5. Inconsistent Language & Region Return Links: questo filtro include gli URL con link di ritorno per lingua e regione incoerenti. Questo avviene quando un link di ritorno ha una lingua o un valore regionale diverso da quello a cui fa riferimento l'URL stesso. Questa metrica può essere vista nella finestra inferiore "Url Details - Dettagli URL" con uno stato di conferma "Inconsistent - Incoerente". I dati possono essere esportati in blocco tramite l'esportazione "Inconsistent Language and Region Return Links"

Reports > Hreflang > Inconsistent Language and Region Return Links.

6. Non Canonical Return Links: URL con link di ritorno hreflang non canonici. Quindi queste sono essenzialmente pagine che si collegano ad altre pagine tramite il set hreflang ma non sono le versioni canoniche, cioè hanno un canonical a un URL diverso. Hreflang dovrebbe includere solo le versioni canoniche degli URL. Gli URL di ritorno non canonici possono essere visti nella finestra inferiore "Url Details - Dettagli URL" con uno stato di conferma "Non canonical". Possono essere esportati in blocco tramite l'esportazione del report "Non Canonical Return Links".

Reports > Hreflang > Non Canonical Return Links

7. Noindex Return Links: riporta i collegamenti che presentano un meta tag "noindex". Quindi queste sono essenzialmente pagine che si collegano ad altre pagine nel loro hreflang, che però presentano un attributo "noindex". Quindi il loro link di ritorno non è valido, poiché è noindex. Tutte le pagine all'interno di un set hreflang dovrebbero essere indicizzabili, e quindi qualsiasi URL di ritorno con "noindex" può portare i Bot a ignorare l'attributo hreflang. Gli URL dei link di ritorno non indicizzabili possono essere visti nel riquadro inferiore della finestra "URL Details- Dettagli URL" con uno stato di conferma "noindex". Possono essere esportati in blocco tramite l'esportazione del rapporto "Noindex Return Links"

Rapporti > Hreflang > Noindex Return Links.

8. Codici lingua e regione errati: questo filtro verifica che la lingua (in formato ISO 639-1) e i valori del codice regionale opzionale (in formato ISO 3166-1 Alpha 2) siano validi. I valori hreflang non supportati possono essere visualizzati nella finestra inferiore "Dettagli URL" con uno stato "invalid - non valido".

9. Multiple Entries: URL che presentano annotazioni multiple per una lingua o un codice regionale. Per esempio, se la pagina X si collega alla pagina Y e Z usando la stessa annotazione di valore hreflang "en" creerà incongruenza per il Bot. Questo filtro raccoglie anche implementazioni multiple, per esempio, se le annotazioni hreflang sono state scoperte come elementi di collegamento combinate con intestazione HTTP.

10. Missing Self Reference: URL che mancano della propria annotazione hreflang di auto-riferimento rel="alternate". Nonostante Google abbia definito "Opzionale" questa indicazione la sua impostazione rimane utile per gli altri Motori di Ricerca. In sintesi ogni URL dovrebbe prevedere l'attributo hreflang a se stesso.

11. Missing X-Default: URL mancanti di un attributo X-Default nell'impostazione hreflang. Questo è opzionale e non è necessariamente un errore o un problema ma è consigliato come suggerimento per i Bot.

12. Missing: URL che mancano completamente dell'attributo hreflang. L'assenza non è necessariamente un errore se il sito web è tradotto solamente in una lingua e ha un ambito territoriale. Attraverso l'utilizzo dei filtri e la Tab "URL Info", nella finestra inferiore, hai a disposizione tutti i dettagli di ogni singolo URL per una visione più analitica dello stato di salute degli Hreflang.

Name	Value
HTML hreflang 1	en-AU
HTML hreflang 1 URL	https://www.moo.com/au/reseller/
HTML hreflang 1 Language Code Valid	Valid
HTML hreflang 1 Confirmation Status	noindex
HTML hreflang 2	en-CA
HTML hreflang 2 URL	https://www.moo.com/ca/reseller/
HTML hreflang 2 Language Code Valid	Valid
HTML hreflang 2 Confirmation Status	noindex
HTML hreflang 3	de-DE
HTML hreflang 3 URL	https://www.moo.com/de/reseller/
HTML hreflang 3 Language Code Valid	Valid
HTML hreflang 3 Confirmation Status	noindex
HTML hreflang 4	es-ES
HTML hreflang 4 URL	https://www.moo.com/es/reseller/
HTML hreflang 4 Language Code Valid	Valid
HTML hreflang 4 Confirmation Status	noindex
HTML hreflang 5	es-US
HTML hreflang 5 URL	https://www.moo.com/es-us/reseller/
HTML hreflang 5 Language Code Valid	Valid
HTML hreflang 5 Confirmation Status	noindex

[URL Info](#)
[Inlinks](#)
[Outlinks](#)
[Image Info](#)
[Resources](#)
[SERP Snippet](#)
[Rendered Page](#)
[View Source](#)

Ogni singolo filtro è a tua disposizione per l'esportazione tramite il pulsante "Export" o mediante la voce di menu Report.

Report > Hreflang > Scelta del report specifico

ORPHAN PAGES REPORT

Le pagine orfane rappresentano gli URL che non ricevono alcun collegamento interno da altre pagine del sito web e non sono raggiungibili dal crawler perché isolate.

Il rapporto “Orphan Pages” fornisce un elenco di URL raccolti tramite la connessione tra il Seo Spider e le API di Google Analytics, Google Search Console (Search Analytics API) o scovati tramite Sitemap XML che non trovano alcun riscontro con gli URL scoperti durante la scansione.

In sintesi ti vengono serviti gli URL non raggiungibili dal Seo Spider per la mancanza di link da seguire o che presentano l'attributo “no follow”.

Ma se non hanno collegamenti perché esistono nei database di Google Analytics o Search Console? Probabilmente sono stati indicizzati con la versione precedente del sito o da altre fonti come XML Site-maps, oppure esistono dei referral che le richiamano.

Per collezionare dati sulle Orphan Pages è sufficiente configurare le API mediante le configurazioni del Seo Spider.

Config > API Access > Scegliere le API da connettere > Connect to new Account (o connessione verso account configurati in precedenza)

Alla conclusione della scansione è indispensabile utilizzare la “Crawl Analysis” per popolare con i dati la scheda relativa.

Le fonti a disposizione per il controllo delle pagine orfane sono:

1. **GA:** l'URL è stato scoperto tramite l'API di Google Analytics.
2. **GSC:** l'URL è stato scoperto in Google Search Console, dalla Search Analytics API.
3. **Sitemap:** l'URL è stato scoperto tramite la Sitemap XML.
4. **GA & GSC & Sitemap:** l'URL è stato scoperto in Google Analytics, Google Search Console & XML Sitemap.

Best practice: per ottimizzare la tua analisi Seo con le API di Google Analytics ti consiglio di scegliere come dimensione “Landing Page” e come segmento “Traffico organico” così da ottenere dati maggiormente rilevanti ed evitare pagine di login, pagina di carrello e-commerce etc.

HOW TO:

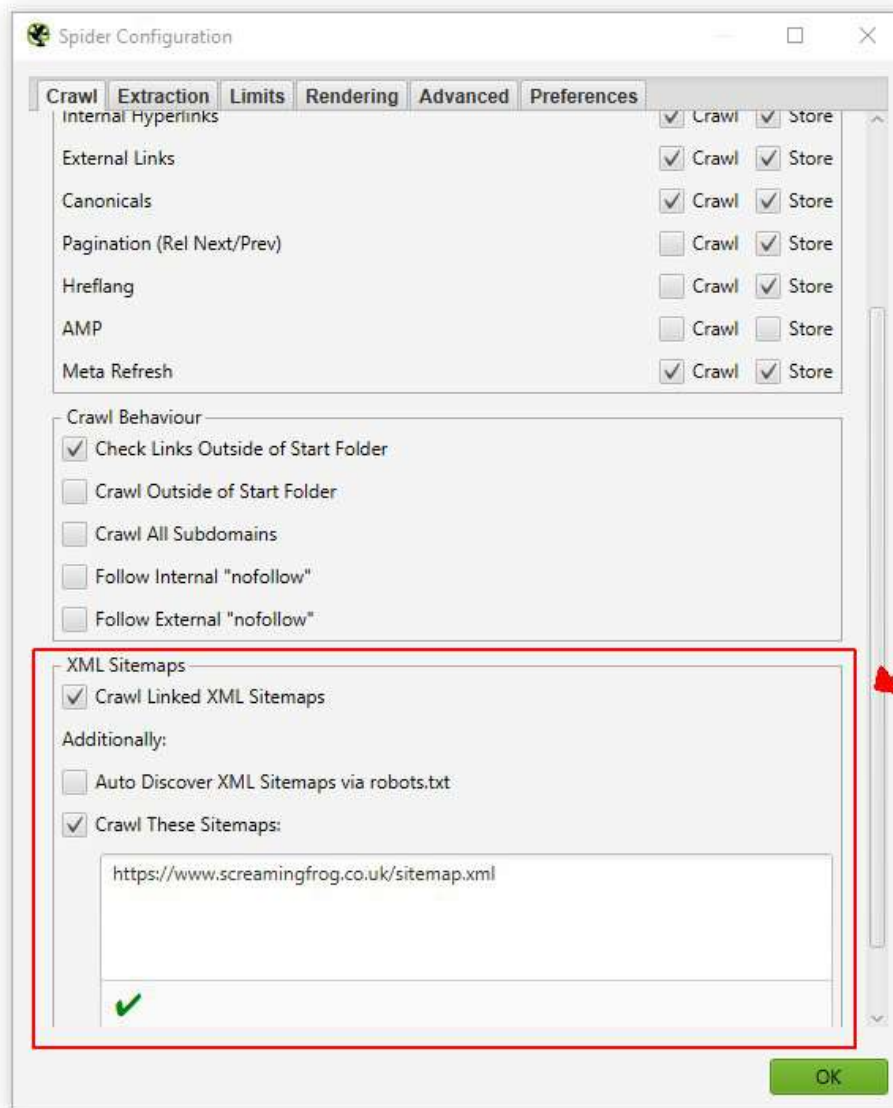
COME SCANSIONARE LE “ORPHAN PAGE”

Compresa l'importanza di individuare le pagine Orfane vediamo la procedura da seguire con il Seo Spider:

1. Impostazione della scansione della Sitemap XML

Preliminarmente devi abilitare l'opzione “Crawl Linked XML Sitemaps” per scansionare la Sitemap del tuo sito web.

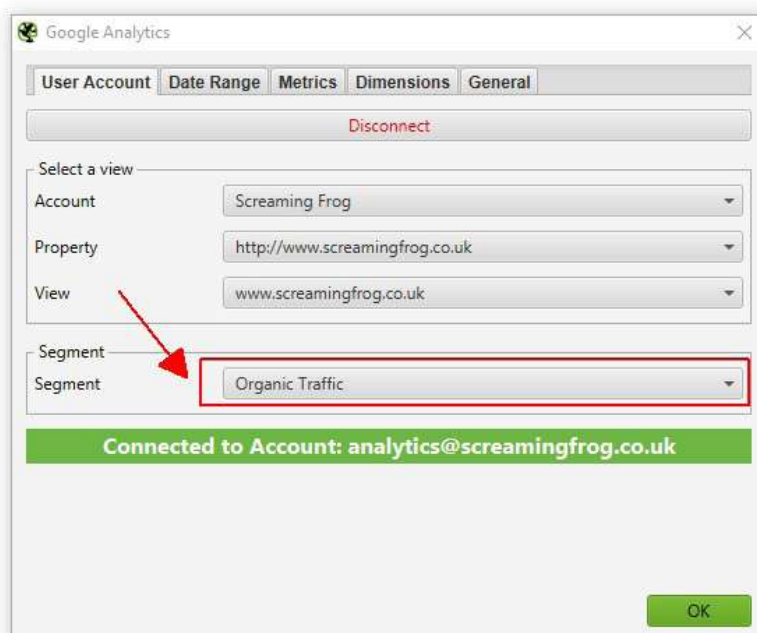
Config > Spider > Crawl > Crawl Linked XML Sitemaps



2. Connessione alle API di Google Analytics

Configura la connessione alle API di Google Analytics, per trovare le Orphan Pages dalla ricerca organica di uno specifico Account, Proprietà o Vista.

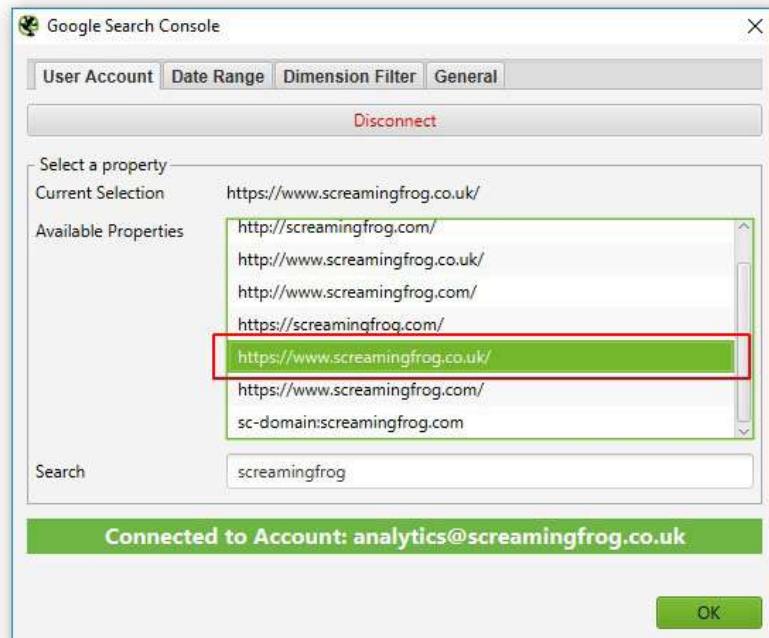
Ricordati di scegliere il segmento "Traffico organico", impostare e personalizzare l'intervallo di date da analizzare (Data Range) che di default è di un mese e definire le metriche e le dimensioni di tuo interesse.



Ricordati di abilitare l'opzione "Crawl New URLs Discovered in Google Analytics". Se questa opzione è disabilitata, gli URL scoperti tramite Google Analytics sono disponibili solo nel rapporto "Pagine orfane". Non saranno aggiunti alla coda di scansione, visualizzabili all'interno dell'interfaccia utente e appariranno sotto le rispettive schede e filtri.

3. Connessione alle API di Search Console

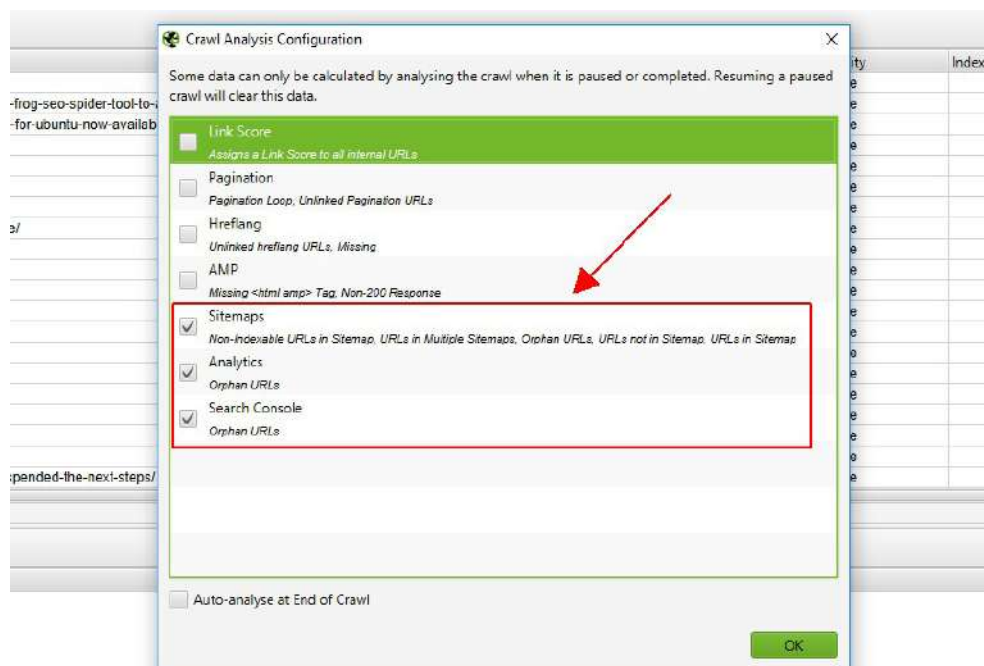
Potrai inoltre connetterti all'API di Google Search Console per trovare tutte le pagine che nonostante non abbiano alcun collegamento interno al sito web stanno ricevendo Impressions e Click in un certo intervallo di tempo.



Anche in questo caso è possibile modificare la data di riferimento su cui collezionare i dati. Allo stesso modo di Google Analytics, se questa opzione non è attivata, i nuovi URL scoperti tramite Google Search Console saranno disponibili solo nel rapporto "Orphan Pages". Non saranno aggiunti alla coda di scansione, visualizzabili all'interno dell'interfaccia utente e appariranno sotto le rispettive schede e filtri.

4. Avvia la scansione e popola i dati con la Crawl Analysis

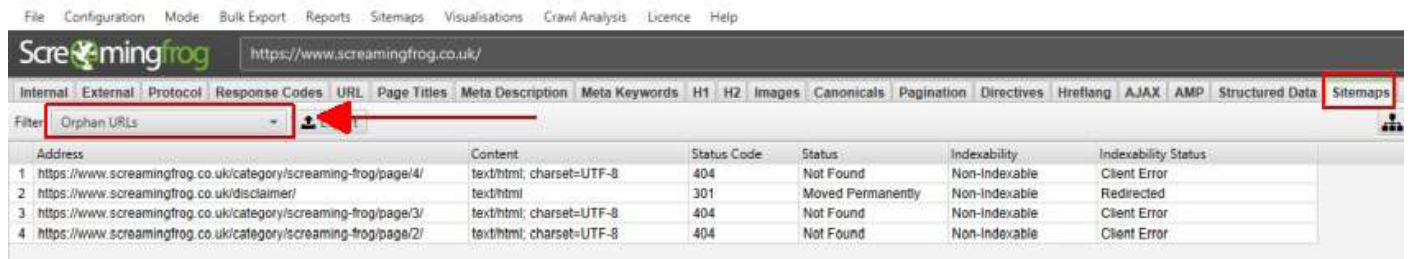
Conclusa la fase di configurazione non ti resta che eseguire la scansione e attivare la "Crawl Analysis" (se non l'hai già automatizzata precedentemente dalle configurazioni).



Completate queste attività sarai in grado di sfogliare ogni scheda e il rispettivo filtro "Orphan URLs" per visualizzare tutte le pagine orfane scoperte.

Mettiamo in pratica quanto abbiamo appena descritto scansionando il sito di Screaming Frog. Lo andiamo ad impostare per la scansione della Sitemap XML e lo connettiamo alle API di Google Search Console.

SITEMAP XML



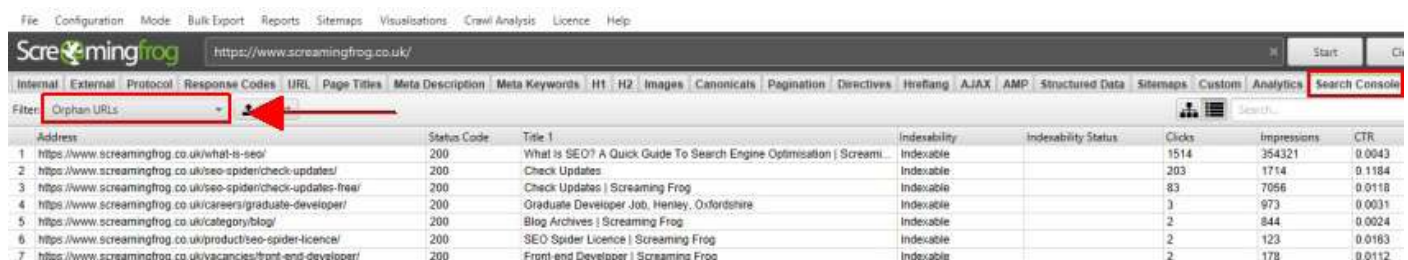
Address	Content	Status Code	Status	Indexability	Indexability Status
1 https://www.screamingfrog.co.uk/category/screaming-frog/page/4/	text/html, charset=UTF-8	404	Not Found	Non-Indexable	Client Error
2 https://www.screamingfrog.co.uk/disclaimer/	text/html	301	Moved Permanently	Non-Indexable	Redirected
3 https://www.screamingfrog.co.uk/category/screaming-frog/page/3/	text/html, charset=UTF-8	404	Not Found	Non-Indexable	Client Error
4 https://www.screamingfrog.co.uk/category/screaming-frog/page/2/	text/html, charset=UTF-8	404	Not Found	Non-Indexable	Client Error

Dall'esempio qui sopra puoi notare come il sito di Screaming Frog presenti alcune pagine orfane (Tab Sitemap) che sono state scoperte dal crawl della Sitemap XML.

In questo caso le pagine presentano uno status code 404 e 301 e potrebbero rappresentare delle vecchie pagine che non sono state rimosse dalla Sitemap XML dopo la pubblicazione del nuovo portale.

API SEARCH CONSOLE

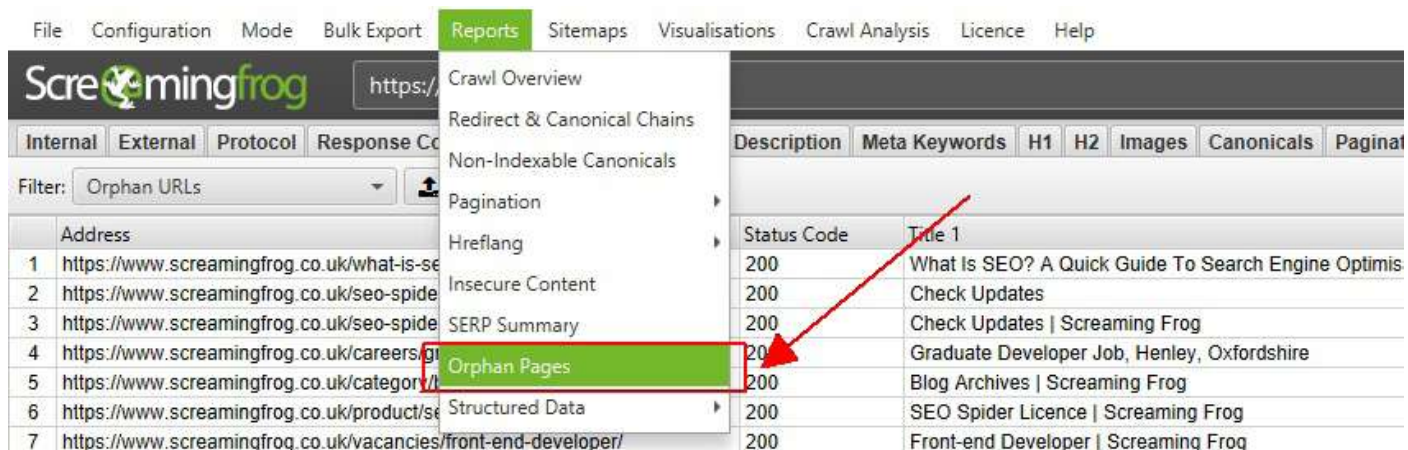
Connettendo il sito di Screaming Frog alle API di Search Console (figura in basso) puoi notare che, a differenza del caso precedente (analisi con sitemap.xml), le pagine orfane sono ancora presenti nel sito web (status code 200) e ricevono Impressions da Google nonostante non presentano alcun link interno.



Address	Status Code	Title 1	Indexability	Indexability Status	Clicks	Impressions	CTR
1 https://www.screamingfrog.co.uk/what-is-seo/	200	What is SEO? A Quick Guide To Search Engine Optimisation Screami	Indexable		1514	354321	0.0043
2 https://www.screamingfrog.co.uk/seo-spidercheck-updates/	200	Check Updates	Indexable		203	1714	0.1184
3 https://www.screamingfrog.co.uk/seo-spidercheck-updates-frear/	200	Check Updates Screaming Frog	Indexable		83	7056	0.0118
4 https://www.screamingfrog.co.uk/careers/graduate-developer/	200	Graduate Developer Job, Henley, Oxfordshire	Indexable		3	973	0.0031
5 https://www.screamingfrog.co.uk/category/blog/	200	Blog Archives Screaming Frog	Indexable		2	844	0.0024
6 https://www.screamingfrog.co.uk/product/seo-spider-licence/	200	SEO Spider Licence Screaming Frog	Indexable		2	123	0.0163
7 https://www.screamingfrog.co.uk/vacancies/front-end-developer/	200	Front-end Developer Screaming Frog	Indexable		2	178	0.0112

Ricorda Bene: entrambi i connettori di Google permettono di esportare singolarmente i risultati semplicemente con il pulsante "Export" presente nella finestra superiore del Seo Spider.

Se preferisci avere una panoramica di tutte le Orphan Pages scoperte da Google Analytics, Search Console e Sitemap.xml puoi utilizzare il report "Orphan pages", documento che trovi nel menu principale.



Description	Meta Keywords	H1	H2	Images	Canonicals	Pagination
Status Code	Title 1					
200	What Is SEO? A Quick Guide To Search Engine Optimis					
200	Check Updates					
200	Check Updates Screaming Frog					
200	Graduate Developer Job, Henley, Oxfordshire					
200	Blog Archives Screaming Frog					
200	SEO Spider Licence Screaming Frog					
200	Front-end Developer Screaming Frog					

SCOPRIRE ORPHAN PAGES ATTRAVERSO IL CRAWL DEPTH

L'ultima strategia possibile per scovare la presenza di Orphan pages (dopo la configurazione della Site-map XML, API Google Analytics e Search Console API) è quella di utilizzare la scheda 'Internal' la quale include ogni URL trovato durante il crawl e "Crawl Analysis".

Le pagine non trovate attraverso la scansione dei link interni avranno sempre un "Crawl depth" vuoto.

File Configuration Mode Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help

Screamingfrog <https://www.screamingfrog.co.uk/>

Internal External Response Codes URL Page Titles Meta Description Meta Keywords H1 H2 Images Canonicals Pagination Direct

Filter: HTML    Export

	Address	Status Code	Status	Indexability	Indexability Status	Crawl Depth ▼
1	https://www.screamingfrog.co.uk/wp-content/themes/screami...	200	OK	Indexable		
2	https://www.screamingfrog.co.uk/vacancies/ppc-consultant/	200	OK	Indexable		
3	https://www.screamingfrog.co.uk/vacancies/front-end-develo...	200	OK	Indexable		
4	https://www.screamingfrog.co.uk/vacancies/digital-designer/	200	OK	Indexable		
5	https://www.screamingfrog.co.uk/team/valia-kalfa/	200	OK	Non-Indexable	noindex	
6	https://www.screamingfrog.co.uk/rocketfrog/	200	OK	Indexable		
7	https://www.screamingfrog.co.uk/log-file-analyser/check-upd...	200	OK	Indexable		
8	https://www.screamingfrog.co.uk/faqs/can-i-get-a-quote-in-a-...	200	OK	Non-Indexable	noindex	
9	https://www.screamingfrog.co.uk/client/wallpaper-direct/	200	OK	Non-Indexable	noindex	
10	https://www.screamingfrog.co.uk/?attachment_id=1862&utm...	301	Moved Permanently	Non-Indexable	Redirected	
11	https://www.screamingfrog.co.uk/client/e-j-churchill/	200	OK	Non-Indexable	noindex	
12	https://www.screamingfrog.co.uk/client/barclays/	200	OK	Non-Indexable	noindex	
13	https://www.screamingfrog.co.uk/careers/freelance-copywrit...	200	OK	Indexable		
14	https://www.screamingfrog.co.uk/wp-login.php	200	OK	Indexable		
15	https://www.screamingfrog.co.uk/vacancies/seo-manager/	200	OK	Indexable		

REPORT INSECURE CONTENT

Questo rapporto è molto utile perché ti segnala gli URL che presentano contenuto ritenuto insicuro e ti permette di evitare gli avvertimenti sulla sicurezza dei Browser che potrebbero allontanare il navigatore.

Il documento “insecure Content” ti presenta tutti gli URL sicuri (HTTPS) che hanno elementi insicuri con protocollo HTTP come i link interni, immagini, JS, CSS, SWF o immagini esterne su un CDN, profili sociali ecc.

Il report diviene molto utile soprattutto in fase di migrazione per avere una visione completa sulle potenziali criticità.



Ecco un esempio del file in cui le destinazioni delle immagini hanno protocollo “http” e non vengono considerate attendibili dallo Spider Tool.

Insecure Content					
Type	Source	Destination	Alt Text	Status Code	Status
IMG	https://example.com/top-cities-revealed/	http://example.com/wp-content/uploads/2014/09/small-latest.png	Small latest (1)	200	OK
IMG	https://example.com/how-to-buy-a-house-guide/	http://example.com/wp-content/themes/rep/img/how-to-buy-a-house.jpg	Buy a house	200	OK
IMG	https://example.com/first-time-buyer-guide/	http://example.com/wp-content/themes/rep/img/first-time-buyer.jpg	first time buyer	200	OK
IMG	https://example.com/careers/join-our-team/	http://example.com/wp-content/uploads/2013/06/join.jpg	join us	200	OK
IMG	https://example.com/careers/sales-manager/	http://example.com/wp-content/uploads/2012/10/sales-manager.jpg	sales manager	200	OK
IMG	https://example.com/careers/office-manager/	http://example.com/wp-content/uploads/2013/01/office-manager.jpg	office manager	200	OK

REPORT SUI DATI STRUTTURATI

Come ben sai i dati strutturati rivestono un ruolo informativo per il Motore di Ricerca e se gestiti correttamente, garantiscono degli enormi benefici in termini di visibilità grazie alle diverse features del Motore di Ricerca.

Screaming Frog ti mette a disposizione 4 reports per valutare ed ottimizzare eventuali criticità di convalida dei "Structured Data":

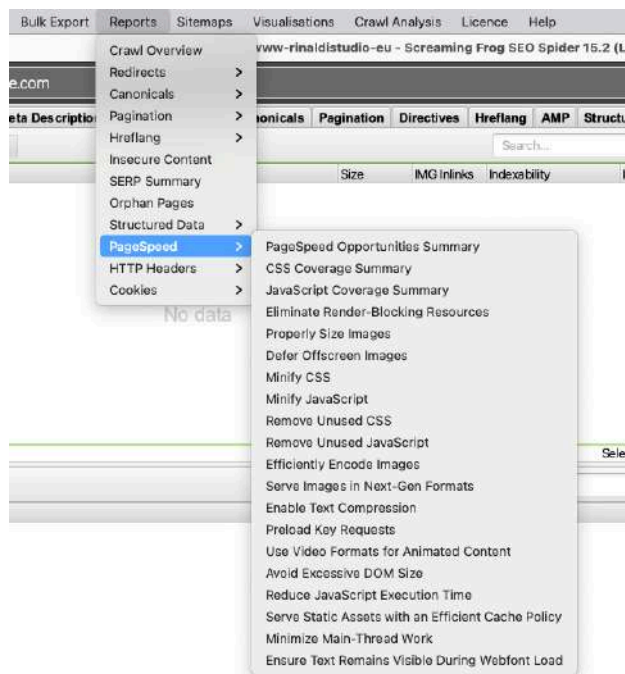
- 1. Validation Errors & Warnings Summary:** aggrega i dati strutturati in base agli errori di convalida unici e agli avvertimenti scoperti mostrando il numero di URL interessati da ogni criticità, con degli URL campionari che identificano il problema specifico.
- 2. Validation Errors & Warnings:** visualizza tutti gli errori di convalida dei dati strutturati in modo granulare per ogni URL includendo il nome della proprietà (Organization etc), il formato (JSON-LD etc), la gravità del problema (errore o avviso), il tipo di convalida (Google Product etc) e il messaggio relativo del problema (esempio: /review property is required).
- 3. Google Rich Results Features Summary:** aggrega i dati per "Google Rich Results Features" rilevati in un crawl mostrando il numero di URL che presentano ogni singola caratteristica.
- 4. Google Rich Results Features:** mappa in modo capillare ogni URL in base alle singole caratteristiche disponibili e mostra quali sono state rilevate per ogni URL.

Structured Data Validation Errors & Warnings Summary Report						
Name	Format	Validation Type	Issue Severity	Number of URL's Affected	Issue Message	Sample URL Affected
ContactPoint	JSON-LD	Google Corporate Contact	Warning	16180	'http://schema.org/contactOption' property is recommended for 'ContactPoint'	https://www.example.com/
BreadcrumbList	JSON-LD	Google Breadcrumb	Error	16069	One of name or item.name must be provided in 'BreadcrumbList'	https://www.example.com/
Listitem	JSON-LD	Google Breadcrumb	Error	16075	'http://schema.org/position' property is required for 'Listitem'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288	'http://schema.org/review' property is recommended for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288	'http://schema.org/offers' property is recommended for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288	'http://schema.org/aggregateRating' property is recommended for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288	'http://schema.org/sku' property is recommended for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Error	10288	'http://schema.org/image' property is required for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288	'http://schema.org/description' property is recommended for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288	'http://schema.org/gtin13' or 'http://schema.org/gtin14' or 'http://schema.org/mpn' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945	'http://schema.org/review' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945	'http://schema.org/offers' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945	'http://schema.org/brand' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Carousel	Error	3945	'http://schema.org/url' or 'http://schema.org/item' property is required for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945	'http://schema.org/aggregateRating' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Error	3945	'http://schema.org/name' property is required for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945	'http://schema.org/sku' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Error	3945	'http://schema.org/image' property is required for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945	'http://schema.org/description' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Carousel	Error	3945	'http://schema.org/position' property is required for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945	'http://schema.org/gtin8' or 'http://schema.org/gtin13' or 'http://schema.org/gtin14' or 'http://schema.org/mpn' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Carousel	Error	3945	'http://schema.org/itemListElement' is required to be of type 'http://schema.org/ListItem' in 'ItemList' (currently=Product)	https://www.example.com/
Article	JSON-LD	Google Article	Warning	1306	'http://schema.org/dateModified' property is recommended for 'Article'	https://www.example.com/
Article	JSON-LD	Google Article	Warning	1306	'http://schema.org/headline' property is recommended for 'Article'	https://www.example.com/
Article	JSON-LD	Google Article	Error	1305	'http://schema.org/image' is required to be 'http://schema.org/ImageObject' or 'http://schema.org/URL' in 'Article'	https://www.example.com/
Article	Microdata	Google Article	Warning	118	'http://schema.org/dateModified' property is recommended for 'Article'	https://www.example.com/
Article	Microdata	Google Article	Warning	118	'http://schema.org/headline' property is recommended for 'Article'	https://www.example.com/
Article	Microdata	Google Article	Warning	118	'http://schema.org/datePublished' property is recommended for 'Article'	https://www.example.com/
Article	Microdata	Google Article	Warning	118	'http://schema.org/image' property is recommended for 'Article'	https://www.example.com/
ItemList	Microdata	Google Carousel	Error	16	'http://schema.org/itemListElement' property is required for 'ItemList'	https://www.example.com/

REPORT PAGESPEED REPORTS

Questo report si basa sulle API di Page Speed Insight (PSI) e fornisce tutti i principali indicatori sul tempo di caricamento delle pagine web permettendo una diagnosi molto precisa su eventuali blocchi o cause che ostacolano le performance.

Config > API Access > Pagespeed Insights



I documenti a disposizione sono davvero molti, al momento ci concentriamo sui 3 report agglomerati, gli altri filtri sono richiamati in diverse aree della guida.

- 1. PageSpeed Opportunities Summary:** riassume tutte le opportunità uniche scoperte nel sito, il numero di URL interessati e il risparmio potenziale medio e totale in termini di dimensioni e millisecondi per definire la priorità di intervento.
- 2. CSS Coverage Summary:** determina un potenziale risparmio di tempo se viene rimosso il codice inutilizzato di ogni CSS che viene caricato nel sito.
- 3. Riepilogo copertura JavaScript:** evidenzia quanto di ogni file JS è inutilizzato durante la scansione e il potenziale risparmio eliminando il codice inutilizzato.

REPORT HTTP HEADER SUMMARY

Questo documento restituisce una vista aggregata di tutte le intestazioni di risposta HTTP scoperte durante una scansione. Mostra ogni intestazione unica di risposta HTTP e il numero di URL unici che hanno risposto con l'intestazione.

Per popolare i dati con le "intestazioni HTTP" devi abilitare la relativa funzione.

Config > Spider > Estrazione

I dettagli più specifici sugli URL e intestazioni li puoi trovare nella scheda "HTTP Headers" della finestra inferiore e tramite l'esportazione del documento "All HTTP Headers".

Bulk Export > Web > All HTTP Headers

In alternativa, puoi interrogare le intestazioni HTTP direttamente nella scheda "Internal", dove vengono aggiunte in colonne uniche separate.

COOKIE SUMMARY

Il rapporto "Cookie Summary" mostra una vista aggregata dei cookie unici scoperti durante una scansione, considerando il loro nome, dominio di riferimento, la scadenza, i valori di sicurezza e l'HttpOnly. Viene visualizzato anche il numero di URL su cui ogni cookie unico è stato emesso.

Per utilizzare questo rapporto è sufficiente abilitare l'estrazione dei "cookie".

Config > Spider > Extraction

Non dimenticare di abilitare anche la modalità di rendering JavaScript per una visione accurata dei cookie che vengono caricati sulla pagina utilizzando JavaScript o tag di immagini pixel.

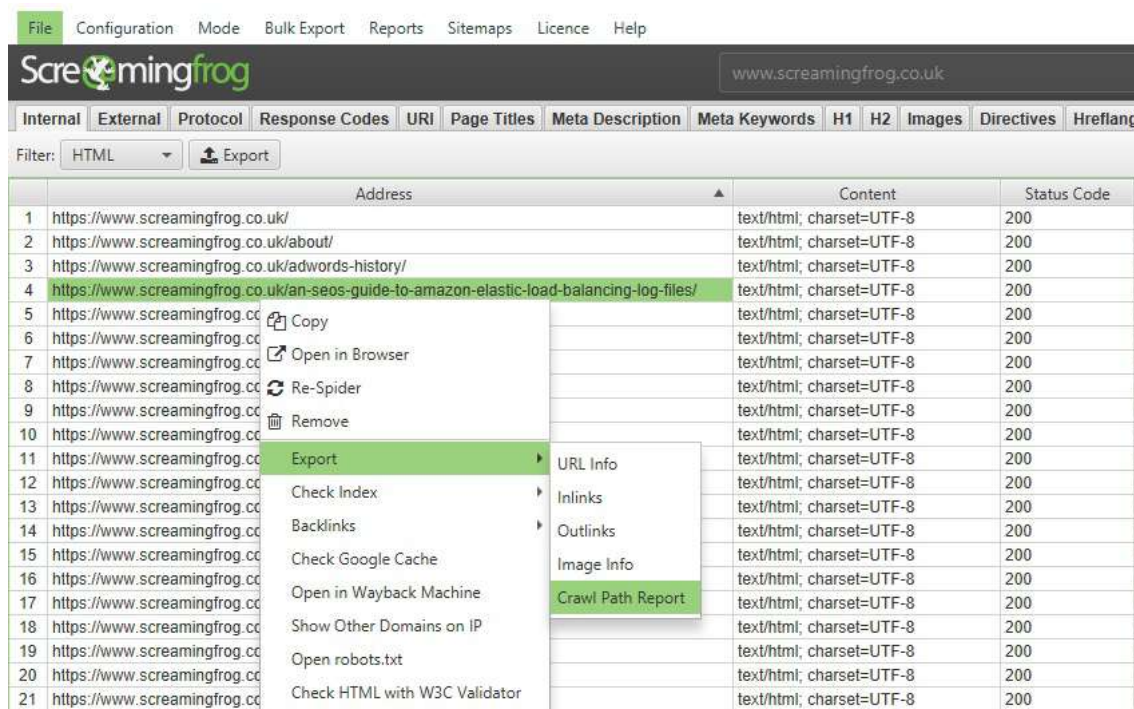
Questo rapporto aggregato è estremamente utile per essere conforme con il GDPR sulla privacy. Puoi consultare ulteriori dettagli sulla scheda "Cookies" della finestra inferiore e tramite l'esportazione "All Cookies".

Bulk Export > Web > All Cookies

CRAWL PATH REPORT

Il rapporto “Crawl Path Report” non lo trovi nel menu principale a differenza dei precedenti.

Il documento è disponibile cliccando con il tasto destro del mouse su un URL della finestra superiore del Seo Spider e selezionando l'opzione “esporta”.



Questo rapporto mostra il percorso più breve che il SEO Spider ha seguito durante il crawl per scoprire l'URL.

Questa informazione ti può essere molto utile per le pagine con un “Crawl Depth” molto profondo e ti evita di dover visualizzare gli “inlink” di molti URL per scoprire la sorgente originale; una sua applicazione potrebbe rientrare nella diagnostica degli URL infiniti causati da plugin o calendari periodici.

Il rapporto lo devi consultare dal basso verso l'alto. Il primo URL in fondo alla colonna “source” è il primo URL scansionato (con un livello '0').

La colonna “destination” invece ti mostra quali URL sono stati scansionati successivamente, e questi costituiscono i conseguenti URL “sorgente” per il livello successivo (1) e così via, verso l'alto.

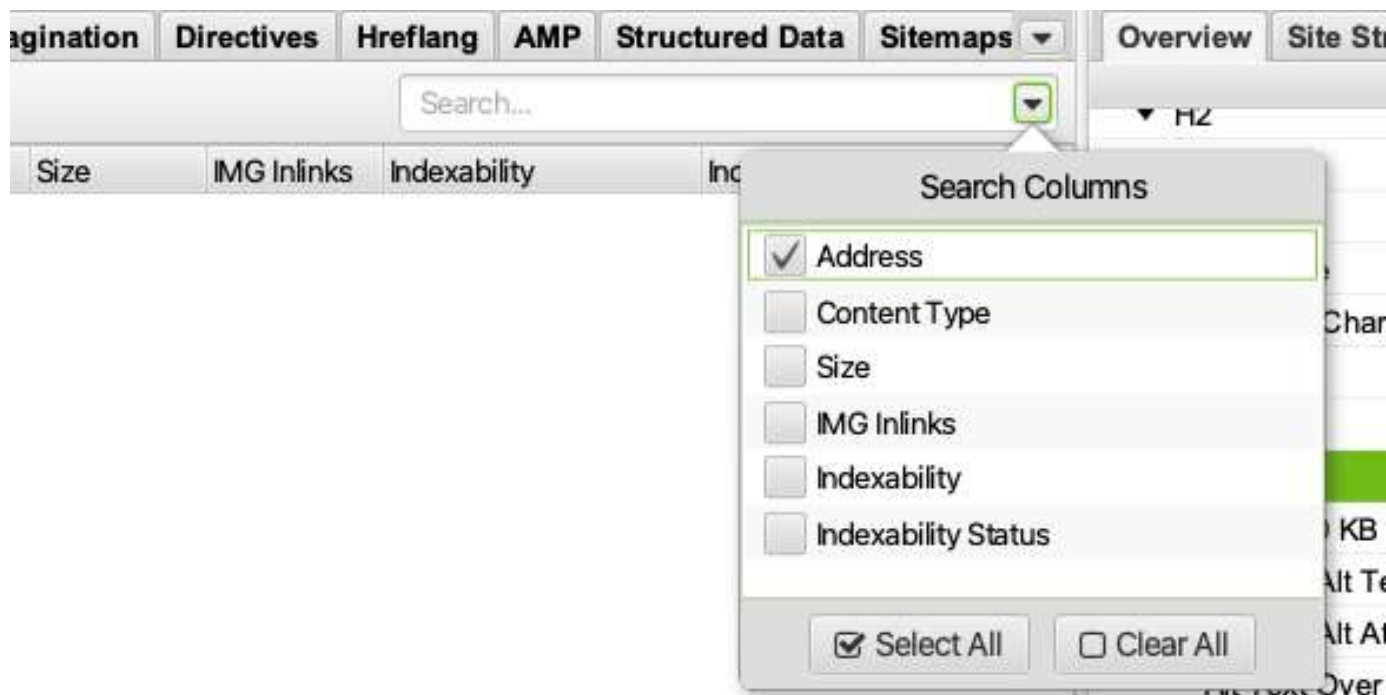
Il primo URL della colonna “destination”, all'inizio del rapporto, sarà l'URL del rapporto del percorso di scansione.

SEARCH FUNCTION

Screaming Frog prevede una barra di ricerca interna che puoi utilizzare per filtrare i dati del crawl in base a tuoi specifici focus.

La casella di ricerca (Search Box) la trovi in alto a destra dell'interfaccia e si basa sull'utilizzo di espressioni RegEX (espressione regolare).

Per impostazione predefinita, il SEO Spider cerca solo sulla colonna URL, ma anche questa funzionalità la puoi personalizzare cliccando sul menu a tendina a destra della Search Box.



Dopo aver definito l'input di ricerca, è sufficiente premere invio e ti verranno restituite solo le righe che matchano con la tua richiesta.

Se vuoi rendere questa ricerca insensibile alle maiuscole e alle minuscole, puoi farlo usando la seguente espressione regolare:

```
(?i)KEYWORD(?-i)
```

Questa espressione regolare corrisponderà a: keyword, Keyword e KEYWORD.

Per escludere una certa parola o lettera puoi usare un “look-ahead” o “look-behind” negativo.

```
^(?!.*?parola chiave).*
```

Se hai bisogno di escludere solo un carattere speciale, come un punto interrogativo (?), puoi fare l'escape con un backslash.

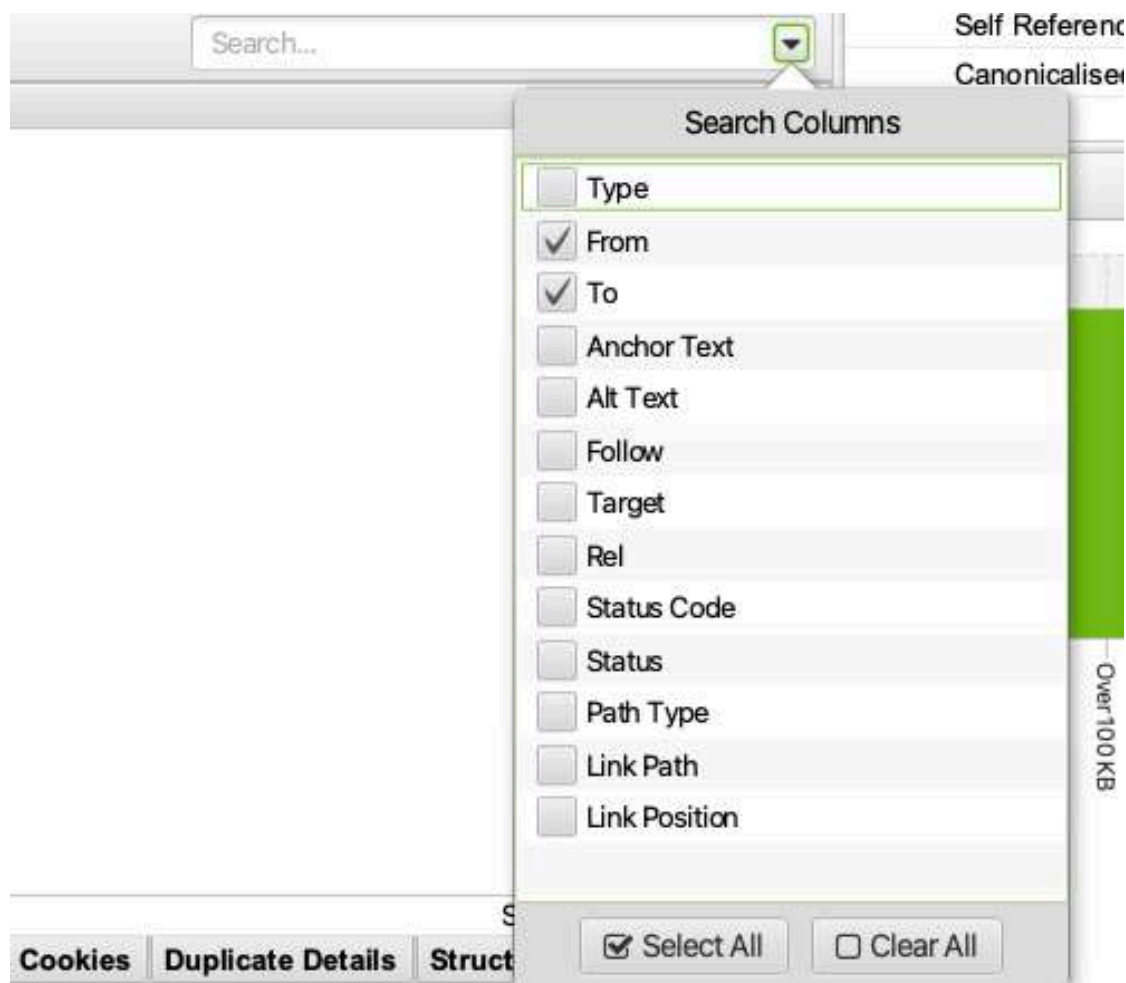
```
^(?!.*?\?).*
```

Per cercare parole duplicate o ripetute:

```
(parola chiave)(.*\1+)
```

Un'ulteriore casella di ricerca con le stesse caratteristiche e funzionalità la puoi trovare nelle schede della finestra inferiore.

Molto utile la funzione "Search" nella tab "View Source" per ricercare e filtrare particolari elementi presenti negli html collezionati.



Ci sono molte applicazioni utili di questa funzionalità, vediamo alcuni esempi:

- ricerca di una query all'interno degli headings per comprendere il search intent della pagina durante lo scraping di un concorrente;
- la ricerca di una query nei "meta titles" per vedere se ci sono duplicazioni che porterebbero a problemi di cannibalizzazione
- il confronto fra le query rilevate mediante la Search Console e le parole chiavi presenti nell'h1.

SCHEDA INTERNAL

Sulla console del Seo Spider la prima scheda a tua disposizione è “Internal” e combina tutti gli URL che sono presenti sullo stesso sottodominio della pagina iniziale di scansione. Questa scheda comprende tutte le metriche del Seo Spider, fatta eccezione per i dati collezionati dalle schede “external”, “hreflang” e “dati strutturati” (data structured).

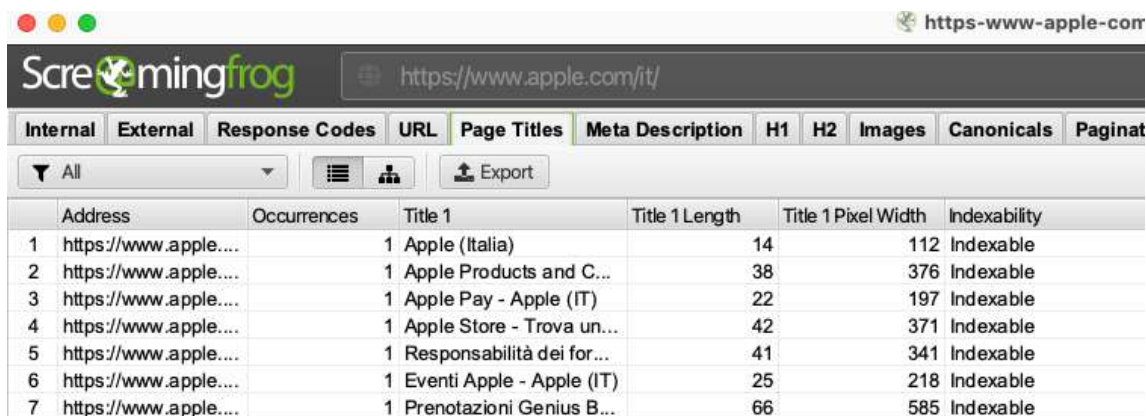
Eventuali sottodomini possono essere resi come Interni (e non considerati altrimenti come referral esterni) se utilizzi la configurazione “crawl all subdomains”, la modalità “list” o la funzione CDNs.

In questa tab puoi consultare i seguenti dati:

- **Address:** identifica l'indirizzo (URL) della risorsa scansionata (html, Javascript, Css, Immagini, Pdf, Flash, altro).
- **Content:** in corrispondenza dello specifico “URL” esaminato restituisce la tipologia dell'elemento:
 - Html;
 - Javascript;
 - Css;
 - Immagini (Images);
 - Pdf;
 - Flash;
 - Altro.
- **Status Code:** identifica la risposta Http del server:
 - code 200;
 - code 301;
 - code 302;
 - code 404;
 - code 500;
- **Status:** identifica la risposta Http dell'header quando uno specifico URL viene richiamato:
 - code 200: OK;
 - code 301: Moved Permanently (redirect definitivo);
 - code 302: Moved Temporarily (redirect temporaneo);
 - code 404: Not Found (Non Trovato);
 - code 500: Server Error (Problema legato al Server),
 - Connection Timeout
- **Indexability:** indica se un URL è “Indexable” (indicizzabile da parte del Motore di Ricerca) o “Non-Indexable” (non indicizzabile).
- **Indexability Status:** specifica il motivo per il quale un URL non può essere indicizzato. Ad esempio perché “Canonicalised” verso un'altra risorsa, status code diverso da 200 etc, presenta “Connection Timeout” nello status e quindi viene considerato “No Response” etc.
- **Title 1:** identifica il “Meta Title” (Page Title) che viene utilizzato da Google nella sua Serp di risultati. Il Seo Spider colleziona il testo presente nel codice html:

```
<title> Il titolo della pagina </title>
```

- **Title 1 Length:** lunghezza del “Title 1” in termini di numero di caratteri. E' consigliato un titolo compreso tra i 30 e 55 caratteri per non incorrere in troncature dello snippet in Serp.



	Address	Occurrences	Title 1	Title 1 Length	Title 1 Pixel Width	Indexability
1	https://www.apple...	1	Apple (Italia)	14	112	Indexable
2	https://www.apple...	1	Apple Products and C...	38	376	Indexable
3	https://www.apple...	1	Apple Pay - Apple (IT)	22	197	Indexable
4	https://www.apple...	1	Apple Store - Trova un...	42	371	Indexable
5	https://www.apple...	1	Responsabilità dei for...	41	341	Indexable
6	https://www.apple...	1	Eventi Apple - Apple (IT)	25	218	Indexable
7	https://www.apple...	1	Prenotazioni Genius B...	66	585	Indexable

- **Title 1 Pixel Width:** larghezza del "Title 1" in termini di pixel. Corrisponde come funzionalità al precedente ma ti consiglio di ragionare sempre in pixel perché la diversità dello spazio occupato dai caratteri potrebbe ingannarti.

ESEMPIO:

Se utilizzi una lettera "M" oppure una "i" con il primo indicatore otterrai sempre 1 ma i pixel utilizzati dalle due lettere saranno diversi con un'incidenza maggiore della "M" maiuscola in termini di spazio nello Snippet.

- **Meta Description 1:** identifica la "Description" utilizzata nello snippet di Google.
- **Meta Description Length:** numero di caratteri utilizzati per la "Meta Description". E' consigliato un testo compreso tra 70 e 155 caratteri.
- **Meta Description Pixel Width:** larghezza in pixel della Meta Description. Si consiglia di rimanere nel range 400 - 1011 pixels.

```
<meta name="description" content="Le funzioni base di Screaming Frog." />
```

- **Meta Keyword 1:** parole chiavi utilizzate per il singolo URL. L'inserimento delle Meta Keyword è opzionale, Google ha dichiarato di non considerare nella valutazione dei contenuti e relativo ranking.
- **Meta Keywords Length:** numero di caratteri utilizzati per le meta keyword.
- **H1- 1:** identifica il primo heading "H1" dell'URL (titolo principale della pagina).
- **H1 Length-1:** lunghezza in caratteri del primo heading "H1".
- **H2-1:** identifica il primo heading "H2" (sottotitolo).
- **H2-Length-1:** identifica la lunghezza in caratteri del primo heading "H2".

Codice Html per gli headings H1 e H2

```
<H1>Titolo principale della pagina</H1>
```

```
<H2>Sottotitolo</H2>
```

Il Seo Spider raccoglie solamente i primi due heading H2 che incontra nel codice sorgente. Per estrarre i dati relativi agli altri headings (h3,h4 etc) ti consiglio di utilizzare la funzione "Custom Extraction".

- **Meta Robots 1:** identifica le direttive del meta robots (es. max-image-preview:large)
- **X-Robots-Tag 1 – X-Robots-tag HTTP:** direttive dell'header per l'URL (es. noindex).
- **Meta Refresh 1:** ricerca se nel codice di uno specifico URL c'è la stringa "refresh" dei dati.

```
<meta http-equiv="refresh" content="5">
```

- **Canonical Link Element:** identifica i dati dell'elemento di collegamento "canonical".

```
<link rel="canonical" href="https://www.nostrodominio.com/servizi" />
```

- **rel="next" 1:** il Seo Spider raccoglie questi elementi di collegamento HTML progettati per indicare la relazione tra gli URL in una serie paginata, in particolare indica la relazione con la pagina successiva. Questo elemento e il successivo sono spesso presenti nei blog, forum e e-commerce. Nonostante Google abbia dichiarato di essere in grado di scansionare completamente un sito web senza queste indicazioni ritengo debbano essere ancora considerate per gli altri Bot. Il "next" indica la relazione con la pagina successiva.
- **rel="prev" 1:** il Seo Spider raccoglie questi elementi di collegamento HTML progettati per indicare la relazione tra gli URL in una serie paginata. Il "prev" indica la relazione con la pagina precedente.
- **Size:** questa dimensione viene definita in kilobyte. Il valore è impostato dall'intestazione Content-Length se fornito, se non è disponibile, è impostato a zero. Per le pagine HTML il valore considera la dimensione dell'HTML (non compresso) in KB. All'esportazione, la dimensione viene servita in byte Equivalenza → Kilobyte = 1.024 byte
- **Word Count:** questo indice identifica la somma di tutte le "parole" scansionate all'interno del tag body, escluso il markup HTML. Il conteggio si basa sull'area del contenuto escludendo gli elementi nav e footer. Questa metrica la puoi anche personalizzare definendo un'area di analisi diversa. Per una diagnosi più granulare puoi includere o escludere elementi HTML, classi e ID. Ricorda che, nonostante il Seo Spider sia molto preciso ci possono essere delle divergenze fra i valori riportati rispetto ad un calcolo manuale. Queste incongruenze sono dovute al fatto che il parser esegue alcune correzioni quando si imbatte sull'HTML non valido. Da non sottovalutare inoltre eventuali impostazioni personali di rendering che potrebbero influenzare anche quale HTML venga esaminato. Screaming Frog conteggia una parola prendendo il testo e dividendolo per spazi senza alcuna considerazione in base alla visibilità del contenuto (come il testo all'interno di un div impostato come nascosto).

Config > Content > Area

- **Text Ratio:** per restituire questo valore il Seo Spider considera il testo come numero di caratteri non-HTML contenuti nel tag "<body>...</body>" di una pagina diviso per il totale dei caratteri totali di cui è composta la pagina html (testo + elementi html) e presenta il dato in forma percentuale.
- **Crawl Depth:** rappresenta la profondità di un URL rispetto alla pagina iniziale di scansione (normalmente l'homepage) intesa come numero di click necessari per raggiungerla. Ricordati che anche i reindirizzamenti sono contati come un livello di profondità.
- **Link Score:** una metrica compresa tra 0-100, che calcola il valore relativo di una pagina in base ai suoi link interni, simile al PageRank di Google. Per ottenere i dati relativi è necessario utilizzare la "Crawl Analysis" alla fine della scansione.
- **InLinks:** identifica il numero di collegamenti interni ricevuti da un URL da parte di altri URL dello stesso sottodominio.
- **Unique InLinks:** corrisponde al numero di Internal Links "unici" con destinazione l'URL in analisi.

ESEMPIO:

Se 'pagina A' si collega a 'pagina B' 3 volte, ci sarà un solo "Unique InLinks" (e 3 InLinks).

- **% del totale:** rappresenta la percentuale di inlink interni unici (considera solo le pagine Html con status code 200) all'URL.
- **Outlinks:** numero di collegamenti interni verso altri URL dello stesso sottodominio.

- **Unique External Outlinks:** numero di link unici presenti in elementi di ancoraggio da un dato URL ad un altro sottodominio.

ESEMPIO:

Se la 'pagina A' si collega alla "pagina B" su un sottodominio diverso 3 volte, questo verrebbe contato come 3 outlink esterni e 1 outlink esterno unico alla "pagina B".

- **External Outlinks:** numero di link dell'Url esaminato verso sottodomini esterni (diversi di quello scansionato).
- **Closest Similarity Match:** attraverso questa funzionalità puoi vedere l'indice di similarità fra più pagine ed evitare situazioni di duplicazione. Utilizzando le soglie di default del Seo Spider le pagine sono identificate come "quasi duplicati" se presentano una corrispondenza pari o superiore al 90%. Anche questa funzionalità la puoi personalizzare definendo delle soglie ad hoc.
Per popolare questa colonna dovrai abilitare la "Crawl Analysis" al termine della scansione. Solo gli URL con contenuto superiore alla soglia di somiglianza selezionata conterranno dati, gli altri rimarranno vuoti. Quindi, per impostazione predefinita, questa colonna conterrà solo dati per gli URL con il 90% o più di similarità.

Config > Content > Duplicates > Enable Near Duplicates > Near Duplicate Similarity Threshold

- **Hash:** il Seo Spider considera il valore "hash" attraverso algoritmo MD5 dell'URL. Nel caso di due valori identici i contenuti sono considerati duplicati e possono essere penalizzati dal Motore di Ricerca. Questo indice, a differenza della scheda "Closest Similarity Match", è un controllo dei contenuti duplicati esatti.

Nota bene: se due valori di hash corrispondono, le pagine sono esattamente uguali nel contenuto. Se c'è la differenza anche di un solo carattere, avranno valori di hash unici e non saranno rilevati come contenuti duplicati esatti.

URL > Duplicate

- **No. Near Duplicates:** identifica il numero di URL quasi duplicati scoperti durante la scansione che soddisfano o superano la soglia impostata nella "Near Duplicate Similarity Threshold" (di default è del 90%). Questa impostazione può essere anche modificata.

Config > Content > Duplicates

Per popolare questa colonna devi aver precedentemente selezionato l'opzione "Abilita i quasi duplicati" e avviare la "Crawl Analysis".

Config > Content > Duplicates

- **Spelling Errors:** il numero totale di errori ortografici scoperti per un URL. Affinché questa colonna sia popolata, deve essere selezionato "Enable Spell Check" (Abilita controllo ortografico).

Config > Content > Spelling & Grammar

- **Grammar Errors:** il numero totale di errori di grammatica scoperti per ogni singolo URL. Affinché questa colonna sia popolata è necessario selezionare "Abilita controllo grammaticale". Nelle impostazioni è possibile definire la lingua di controllo, definire le regole grammaticali e il dizionario di riferimento.

Per consultare i dati relativi agli errori grammaticali e di “Spelling” è sufficiente consultare le colonne dedicate nella scheda “Internal” selezionare gli URL che presentano gli errori e vederli in dettaglio nella scheda “Spelling & Grammar Details” della finestra in basso del Seo Spider. Ottima funzionalità per il copywriter nello scovare eventuali errori di battitura o utilizzo improprio di regole grammaticali. Oltre alla lista di errori sono presenti suggerimenti per la correzione e la sezione dove sono stati riscontrati con relativa anteprima. Per la lingua italiana è ancora poco affidabile nei suggerimenti ma molto utile in termini di errori di battitura.

Config > Contenuto > Ortografia e grammatica

- **Response Time:** tempo impiegato in secondi per scaricare l'URL. Viene calcolato in base al tempo necessario per emettere una richiesta HTTP e ottenere la risposta HTTP completa dal server. Questo parametro non include il tempo per scaricare risorse aggiuntive con “modalità di rendering JavaScript”.
- **Last Modified:** indice letto dall'intestazione “Last-Modified” nella risposta HTTP del server. Se il server non lo fornisce, il valore sarà vuoto.
- **Redirect Url:** se l'URL presenta un reindirizzamento questa colonna includerà la destinazione dell'URL.
- **Redirect Type:** lo Spider Seo identifica:
 - **HTTP Redirect:** attivato da un'intestazione HTTP.
 - **HSTS Policy:** reindirizzamento da intestazione HSTS.
 - **JavaScript Redirect:** reindirizzamento innescato dall'esecuzione di JavaScript (può essere presente solo quando usi la modalità “rendering JavaScript”).
 - **MetaRefresh Redirect:** innescato da un tag meta refresh nell'HTML.
- **HTTP Version:** questa colonna mostra la versione HTTP con cui il crawl è stato effettuato. Il Seo Spider attualmente effettua la scansione solo utilizzando HTTP/1.1, ma questa colonna è stata rilasciata in preparazione del supporto di HTTP/2 in un futuro aggiornamento.
- **URL Encoded Address:** questa colonna identifica l'URL effettivamente richiesto dal SEO Spider. Tutti i caratteri non ASCII sono codificati con le percentuali.

Tutti questi dati raccolti nella Tab “Internal” possono essere filtrati (Filter) in base ai seguenti parametri:

- **Html:** pagine Html.
- **Javascript:** ogni elemento Javascript.
- **CSS:** ogni foglio di stile Css.
- **Immagini.**
- **Pdf.**
- **Flash:** ogni file con estensione .swf.
- **Other:** ogni altra tipologia di file come ad esempio docs.
- **Unknown:** comprende ogni URL che presenta un tipo di contenuto sconosciuto. Tra gli URL che troverai con questo filtro saranno compresi quelli bloccati da robots.txt, perché il loro tipo di file è sconosciuto per il Seo Spider non avendoli scansati.

EXTERNAL TAB

La scheda "External" include tutti i dati relativi agli URL esterni. Vengono classificati come "esterni" gli URL dei sottodomini diversi da quello della pagina iniziale della scansione.

La Tab include i seguenti valori divisi per colonne:

- **Address:** l'indirizzo dell'URL esterno.
- **Content Type:** la tipologia del contenuto scoperto nell'URL.
- **Status Code:** restituisce il codice di risposta HTTP (200, 3xx, 4xx, 5xx).
- **Status:** rappresenta la risposta dell'intestazione HTTP ("Ok" per le risposte 200 del server, "Internal Server Error" nel caso di problemi con il server, "Forbidden" nel caso di status 403, etc.).
- **Crawl Depth** (Profondità di scansione): numeri di click necessari per raggiungere un URL dalla homepage o dalla pagina iniziale di scansione.
- **Inlinks:** numero di link trovati durante il crawl che puntano ad un sottodominio esterno. Se selezioni un URL dalla finestra superiore puoi vedere i dettagli (Es. la sorgente del link nella tab InLinks >From) dei collegamenti nella finestra inferiore del Seo Spider.

I filtri a tua disposizione sono i medesimi della Tab Internal.

SECURITY TAB

La scheda "Security" ti permette di consultare i dati relativi alla sicurezza degli URL interni in un crawl e include i seguenti dati:

- **Address:** l'URL sottoposto a scansione.
- **Content:** rappresenta la tipologia di contenuto dell'URL.
- **Status Code:** il codice di risposta HTTP.
- **Status:** la risposta dell'intestazione HTTP.
- **Indexability:** identifica se l'URL è indicizzabile o non indicizzabile.
- **Indexability Status:** presenta il motivo per cui un URL non è indicizzabile, ad esempio potrebbe essere "canonicalizzato" verso un altro URL o presentare il tag "no index".
- **Canonical Link Element 1/2 etc:** identifica tutte le istanze trovate dallo Spider relative ai dati canonici dell'URL.
- **Meta Robots 1/2 etc:** presenta tutte le istanze dei meta robots trovati nell'URL.
- **X-Robots-Tag 1/2 etc - Dati X-Robots-tag.**

Per la scheda dedicata alla sicurezza sono a tua disposizione diversi filtri che ti permettono una visione molto dettagliata di eventuali criticità:

- **HTTP URLs:** questo filtro ti permette di isolare tutti gli URL che presentano come protocollo (HTTP) o che contengano contenuti misti nelle quali si caricano risorse insicure. Il protocollo "Https" è fondamentale per la sicurezza e al contempo vitale per garantire una navigazione sicura degli utenti. Chrome e altri browser utilizzano messaggi come "Contenuto Non Sicuro" vicino al sottodominio per disincentivare la navigazione verso pagine con questo protocollo. Se devi controllare quali URL presentano "contenuti misti" ti basta consultare la scheda "inlink" della finestra inferiore del Seo Spider. Sei in grado anche di esportare questi risultati con il report "HTTP URLs Inlinks".

- **URL HTTPS:** comprende tutti gli URL che presentano un protocollo attendibile come l'HTTPS.
- **Mixed Content:** attraverso il filtro sarai in grado di consultare tutte le pagine HTML con protocollo HTTPS ma che presentano risorse come immagini, JavaScript o CSS con protocollo HTTP che non vengono considerate sicure. Il contenuto misto indebolisce l'HTTPS rendendo le pagine più vulnerabili e compromissibili.
Le risorse HTTP le puoi visualizzare per ogni URL nella scheda "outlinks" ed esportare insieme alle pagine sorgente con il report "Mixed Content".

- **Form URL Insecure:** il filtro visualizza le pagine HTML con un Form di contatto che presenta come attributo di azione un collegamento insicuro (HTTP). Tutti gli URL contenuti nei moduli di un sito web dovrebbero essere criptati e quindi devono essere HTTPS. L'URL del modulo HTTP può essere visualizzato cliccando sull'URL sorgente nella finestra superiore e poi sulla scheda della schermata inferiore "Dettagli URL". Questi possono essere esportati insieme alle pagine in cui si trovano con il report "Form URL Insecure".

- **Form on HTTP URL:** identifica le pagine con protocollo HTTP che presentano un form e potrebbero essere bloccate dal browser. Tutti i dati inseriti nel modulo, compresi nomi utente e password potrebbero essere intercettati e creare un danno al navigatore. Il modulo può essere visualizzato facendo clic sull'URL nella finestra superiore e poi sulla scheda della finestra inferiore "Dettagli URL", che ti mostra i dettagli del modulo sull'URL HTTP.
- **Unsafe Cross-Origin Links:** vengono visualizzate tutte le pagine che si collegano a siti esterni usando l'attributo `target="_blank"` (per aprire una pagina in una nuova scheda), senza usare `rel="noopener"` (o `rel="noreferrer"`) allo stesso tempo. Il solo utilizzo di `target="_blank"` espone le pagine a problemi di sicurezza e di prestazioni. Idealmente `rel="noopener"` dovrebbe essere usato su tutti i link che contengono l'attributo `target="_blank"` per evitare il "Tabnabbing Inverso".

Tabnabbing inverso: situazione in cui la pagina aperta mediante l'attributo `target="_blank"` può ottenere il controllo della pagina di origine mediante l'utilizzo di `window.opener.location.assign()` per sostituire la scheda in background con documenti malevoli. Attraverso il `rel="noopener"` l'utente malintenzionato non sarà in grado di accedere all'oggetto della finestra tramite `window.opener`.

L'utilizzo dell'attributo `"noopener"` non influisce a livello Seo ma preserva la sicurezza e le performance evitando che i processi della pagina di destinazione siano eseguiti nello stesso thread della pagina di origine. Alcuni dubbi rimangono sull'utilizzo dell'attributo `"noreferrer"` che rimuove i dati di riferimento del click a livello di browser e, non permette al proprietario del sito collegato e aperto tramite il click, di conoscerne la provenienza. Nel caso di Google Analytics l'attributo `"noreferrer"`, modifica la "sorgente di traffico" da `"referral"` a `"diretta"` alterando le statistiche. Partendo dal presupposto che l'inserimento di un link esterno sia voluto non riteniamo l'attributo imprescindibile.

I link esterni che contengono l'attributo `target="_blank"` possono essere visualizzati nella scheda "outlinks" e nella colonna "target". Possono essere esportati insieme alle pagine sorgente dal report "Unsafe Cross-Origin Links"

Protocol-Relative Resource Links: questo filtro ti mostra tutte le pagine che caricano risorse come immagini, JavaScript e CSS usando link relativi senza specificare il protocollo (ad esempio `///screamingfrog.co.uk`).

Questa situazione è molto comune ed adottata dagli sviluppatori per risparmiare tempo lasciando al browser di specificare il protocollo. Tuttavia questa abitudine può esporre alcuni siti ad essere compromessi e potrebbe delineare problemi di prestazione.

I link delle risorse che sono relativi al protocollo possono essere visualizzati per ogni URL cliccando sulla scheda "outlinks" e visualizzando la colonna "Path Type" per "Protocol Relative". Possono essere esportati insieme alle pagine sorgente dal rapporto Protocol-Relative Resource Links.

Bulk Export > Security > Protocol-Relative Resource Links

- **Missing HSTS Header:** identifica tutti gli URL che mancano dell'intestazione di risposta HSTS. La risposta HTTP Strict-Transport-Security (HSTS) indica ai browser che si dovrebbe accedere solo tramite HTTPS, piuttosto che HTTP. Se un sito web accetta una connessione a HTTP, prima di essere reindirizzato a HTTPS, i visitatori inizialmente comunicheranno ancora su HTTP. L'intestazione HSTS indica al browser di non caricare mai su HTTP e di convertire automaticamente tutte le richieste in HTTPS. Il Seo Spider stesso seguirà le istruzioni dell'intestazione HSTS, ma riporterà qualsiasi link incontrato su URL HTTP con un codice di stato 307 e lo stato della "HSTS Policy".
- **Missing Content-Security-Policy-Header:** identifica qualsiasi URL che manca dell'header di risposta Content-Security-Policy. Questa intestazione permette ad un sito web di controllare quali risorse vengono caricate per una pagina. Questa politica può aiutare a proteggere contro gli attacchi cross-site scripting (XSS) che sfruttano la fiducia del browser sul contenuto ricevuto dal server. Il Seo Spider controlla solo l'esistenza dell'intestazione, e non interroga le direttive trovate all'interno per determinare se sono ben impostate per il sito web. Questo dovrebbe essere eseguito manualmente.
- **Bad Content Type:** visualizza qualsiasi URL in cui il tipo di contenuto effettivo non corrisponde al tipo di contenuto impostato nell'intestazione dell'Header. Identifica anche qualsiasi tipo MIME utilizzato ma non valido.

RESPONSE CODE TAB

La scheda dedicata al “Response Tab” mostra lo stato HTTP e gli Status Code degli URL interni ed esterni scoperti durante la scansione. Ecco nel dettaglio tutte le colonne a tua disposizione:

- **Address:** L'URL sottoposto a scansione.
- **Content:** identifica la tipologia di contenuto presente nell'URL (text/html, image/png, image/jpg, etc).
- **Status Code:** visualizza il codice di risposta HTTP (200, 3xx, 4xx, 5xx, no response).
- **Status:** restituisce la risposta dell'intestazione HTTP (ok se risponde 200, Not Found, Moved Permanently, etc.).
- **Indexability:** definisce se l'URL è indicizzabile o non indicizzabile.
- **Indexability Status:** il motivo per cui un URL non è indicizzabile. Per esempio, se è canonicalizzato ad un altro URL o “noindex” o “Redirected”.
- **Inlinks:** corrisponde al numero di inlinks interni presenti nella pagina. Gli “inlink interni” sono link che puntano a un dato URL da risorse dello stesso sottodominio scansionato.
- **Response Time:** identifica il tempo necessario in secondi per scaricare l'URL.
- **Redirect URL:** nel caso in cui l'indirizzo URL faccia un reindirizzamento, questa colonna include la destinazione dell'URL.
- **Redirect Type:** la colonna può contenere
 - HTTP Redirect: reindirizzamento innescato da un'intestazione HTTP.
 - HSTS Policy: reindirizzamento causato da una precedente intestazione HSTS.
 - JavaScript Redirect: innescato dall'esecuzione di JavaScript. Lo puoi ottenere solo se hai scelto la modalità di scansione rendering JavaScript.
 - Meta Refresh Redirect: innescato da un tag meta refresh nell'HTML della pagina.

Nella scheda “Response Code” hai a disposizione i seguenti filtri:

1. **Blocked by Robots.txt:** il filtro definisce tutti gli URL bloccati dal robots.txt del sito che non permettono allo Spider di scansionare quelle risorse e, di conseguenza di non poter indicizzarle sui Motori di Ricerca.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/frogs/legs/		0	Blocked by Robots.txt

Ogni risposta “0” nella colonna Status Code identifica la mancanza di una risposta HTTP del server, la colonna Status ne identifica il motivo. Nell'esempio sopra lo user-agent dello Spider è stato bloccato dal robots.txt attraverso la direttiva “disallow” ed essendo bloccato non può valutare la risposta HTTP.

2. **Blocked Resource:** tutte le risorse che sono bloccate dal robots.txt di un sito web come le immagini, il JavaScript e i CSS. Questo filtro si popola solo se hai abilitato il rendering JavaScript. Ricorda che le risorse bloccate appaiono sotto “Blocked by Robots.txt” se non è stata abilitata la modalità Javascript dalle configurazioni.

3. No Response: visualizza dati relativi agli URL che non inviano una risposta alla richiesta HTTP del Seo Spider. Nel caso più comune corrisponde ad un URL digitato in modo scorretto, timeout di connessione, connessione rifiutata o errore di connessione. Con lo Status “No Response” il SEO Spider ha problemi a creare connessioni o a ricevere risposte. Il problema potrebbe essere il Proxy.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/no-response/		0	No Response

Lo Status “DNS lookup failed” identifica che il sito non è stato trovato per nulla per un errore di immissione del dominio o per una mancanza di connessione alla rete.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.zone/		0	DNS lookup failed

L'avviso “Connection Timeout” identifica che Screaming Frog ha richiesto l'HTTP response dal server ma non ha ricevuto risposta entro 20 secondi. Per ovviare al problema è possibile aumentare il “Response Timeout” dalle configurazioni o diminuire la velocità dello spider (Lowering Speed) per non caricare il Server e dargli più tempo per la risposta.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/timeout/		0	Connection Timeout

Una ‘Connection Refused’ viene restituita quando il tentativo di connessione del Seo Spider è stato rifiutato in qualche punto tra la macchina locale e il sito web. I problemi più comuni sono relativi alla presenza di firewall e dal tentativo di bloccare un user agent sconosciuto. Per risolvere questa condizione ti consiglio di cambiare lo “user-agent” da “Screaming Frog Seo Spider” a “Chrome”.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/refused/		0	Connection Refused

4. Success (2XX): l'URL richiesto è stato ricevuto, compreso, accettato ed elaborato con successo. Idealmente tutti gli URL incontrati in un crawl dovrebbero restituire un codice di stato “200” con uno stato ‘OK’, che è perfetto per il crawling e l'indicizzazione dei contenuti.

5. Redirection (3XX): è stato riscontrato un reindirizzamento come 301 o 302. Idealmente ogni link interno dovrebbe collegarsi ad una risposta con status code 200 ed evitare collegamenti che facciano fare dei passi in più allo Spider in un'ottica di salvaguardare il Crawling Budget.

6. Redirection (JavaScript): è stato incontrato un reindirizzamento lato JavaScript.

7. Redirect (Meta Refresh): è stato riscontrato un reindirizzamento mediante “Meta Refresh”.

8. Client Error (4xx): il filtro visualizza le risorse per le quali si è verificato un problema con la richiesta. Queste possono includere risposte come 400 bad request, 403 Forbidden, 404 Page Not Found, 410 Removed, 429 Too Many Requests e altro. I Client Errors come i 404 dovrebbero essere ottimizzati per non sprecare il Crawling Budget. Per farlo puoi modificare la sorgente del collegamento, rimuovere questi collegamenti con il file htaccess modificando lo status code a 410 o reindirizzarli dove appropriato.

9. Errore del server (5XX): il server non è riuscito a soddisfare una richiesta apparentemente valida. Questo può includere risposte comuni come “500 Internal Server Errors” e “503 Server Unavailable”. Una percentuale alta di questo errore potrebbe indicare un server sotto stress o una configurazione errata che richiede ulteriori indagini.

TAB URL

La scheda presenta i dettagli relativi agli URL scoperti durante il crawl e i rispettivi problemi.

- **Address:** identifica l'URL sottoposto a scansione.
- **Content Type:** restituisce la tipologia di contenuto dell'URL.
- **Codice di stato:** il codice di risposta HTTP.
- **Status Code:** la risposta dell'intestazione HTTP.
- **Indexability:** se l'URL è indicizzabile o non indicizzabile.
- **Indexability Status:** il motivo per cui un URL non è indicizzabile.
- **Hash:** valore seriale di ogni pagina per identificazione di eventuali contenuti duplicati.
- **Length:** la lunghezza dell'URL calcolato in base ai caratteri.
- **Canonical 1:** i dati canonici dell'elemento di collegamento.
- **URL Encoded Address:** l'URL effettivamente richiesto dal Seo Spider. Tutti i caratteri non-ASCII vengono codificati in percentuale, in base all'RFC 3986.

Per analizzare i dati collezionati puoi utilizzare uno dei 9 filtri messi a disposizione:

- **Non ASCII Characters:** vengono mostrati gli URL che contengono attributi che non sono inclusi nel set di caratteri ASCII. Gli URL devono essere convertiti in un formato ASCII valido, codificando i link all'URL con caratteri sicuri (composti da % seguito da due cifre esadecimali). Oggi i browser e i Motori di Ricerca sono ampiamente in grado di trasformare ed adeguare accuratamente gli URL in modo automatico.
- **Underscores:** gli URL presentano il simbolo del "trattino basso" ("_") al loro interno. A livello Seo non sempre questi caratteri vengono interpretati dal Motore di Ricerca come separatori di parole e potrebbero penalizzare l'indicizzazione. Consiglio l'utilizzo del simbolo "-" fra le parole per identificare lo spazio fra le query da indicizzare.
- **Uppercase:** mostrano gli URL che hanno dei caratteri maiuscoli al loro interno. Gli URL sono sensibili alle maiuscole e alle minuscole, quindi, come best practice, gli URL dovrebbero essere generalmente minuscoli, per evitare qualsiasi potenziale confusione e duplicazione.
- **Multiply Slashes:** l'URL presenta più slash nel "path" (esempio: screamingfrog.co.uk/seo//).
- **Repetitive Path:** l'URL ha un percorso ripetuto nella stringa (esempio: screamingfrog.co.uk/seo/seo/, o screamingfrog.co.uk/services/seo/technical/seo/). In alcuni casi questo può essere legittimo e logico per la struttura del sito web, ma potrebbe essere ridondante e visto come "overstuffing" dal Motore di Ricerca. Questo filtro ti può anche aiutare ad identificare problemi con collegamenti relativi non corretti, che potrebbero causare URL infiniti.
- **Contains a Space:** vengono visualizzati gli URL che hanno uno spazio al loro interno. Questi sono considerati non sicuri e potrebbero causare la rottura del link quando lo si condivide. I trattini dovrebbero essere preferiti agli spazi come separatori.
- **Internal Search:** l'URL potrebbe essere parte della funzione di ricerca interna del sito. Google e gli altri Motori di Ricerca raccomandano di bloccare le pagine di ricerca interna dalla scansione e i collegamenti ad essa dovrebbero includere un attributo di link "nofollow".
- **Parameters:** l'URL include parametri come '?' o '&'. Questo non è un problema per Google, ma si raccomanda di limitare il numero di parametri in un URL che possono risultare complicati per gli utenti, e potrebbero venir considerati a basso valore aggiunto.

- **Over 115 Characters:** l'URL ha una lunghezza superiore a 115 caratteri di lunghezza. Questo non è necessariamente un problema Seo in senso stretto, ma di user experience, in base alle ricerche del settore è stato dimostrato che gli utenti preferiscono stringhe di URL più brevi e concise.

TAB PAGE TITLES

I Meta Titles sono considerati un fattore di "Ranking" per il Motore di Ricerca e rappresentano il primo punto di contatto tra l'utente e il tuo sito web. L'ottimizzazione dei "Meta Titles" permette di migliorare il posizionamento nei risultati dei Motori di Ricerca e di aumentare il CTR (Click Through Rate) degli snippet in Serp.

Attraverso la scheda "Titles" sei in grado di visualizzare tutti gli elementi da monitorare nell'ottimizzazione di questo tag.

```
<title>Questo è un titolo di pagina</title>
```

La Tab include le seguenti colonne:

- **Address:** considera i diversi URL scansionati.
- **Occurrences:** Il numero di "meta titles" trovati sulla pagina per un massimo di 2 entità.
- **Title 1:** il contenuto testuale del tag Meta Title.
- **Title 1/2 length:** il numero di caratteri del Meta Titles.
- **Indexability:** se l'URL è indicizzabile o non indicizzabile.
- **Indexability Status:** definisce il motivo per cui un URL viene considerato non indicizzabile.

Questi elementi possono essere filtrati secondo le seguenti caratteristiche:

- **Missing:** visualizza tutte le pagine che non presentano un Meta Title e ti permette di scoprire quali siano gli URLs da ottimizzare. Questo tag partecipa alla formazione del ranking del sito web ed aiuta a migliorare il CTR e relativo traffico. La sua ottimizzazione ti permette di evitare che il Motore di Ricerca scelga autonomamente il titolo delle pagine con alta percentuale di storpiature.
- **Duplicate:** mostra tutte le pagine che hanno titoli di pagina duplicati. Vista la sua funzionalità di ranking è molto importante che ogni pagina abbia un Meta Titles distinto che permetta agli utenti di comprendere immediatamente il Search Intent direttamente dallo Snippet ed eviti di creare confusione alle scorribande del Motore di Ricerca.
- **Over 60 characters:** include tutte le pagine che presentano titoli di pagina più lunghi di 60 caratteri. I caratteri oltre questo limite potrebbero essere troncati nei risultati di ricerca di Google e avere meno peso nel punteggio assegnato in termini di ranking. Il consiglio personale è quello di non superare i 55 caratteri. Il filtro per default è impostato su 60 ma può essere personalizzato cambiando la soglia massima da "Preferences".

Config > Spider > Preferences

- **Below 30 characters:** identifica tutte le pagine che hanno titoli di pagina sotto i 30 caratteri di lunghezza. Questo non è necessariamente un problema a livello Seo, ma un Meta Title troppo stringato potrebbe essere poco incisivo e non comunicare bene il Search Intent.
- **Over X Pixel:** la lunghezza dello snippet di Google è in realtà basata su limiti di pixel, piuttosto che su un numero complessivo di caratteri. Attraverso questo filtro sarà possibile personalizzare la soglia massima in base alle direttive del Motore di Ricerca e monitorare le pagine che hanno titoli di pagina di lunghezza superiore a X pixel. Per default Screaming frog imposta come limite 562 pixels.

- **Below X Pixels:** tutte le pagine che hanno titoli di pagina di lunghezza inferiore a X pixel. Questo non rappresenta un elemento critico, ma avendo più spazio a disposizione sarebbe ottimale comunicare in modo più esaustivo i tuoi USP.
- **Same as h1:** vengono mostrati tutti i Meta Title che corrispondono esattamente all'H1 della pagina. Non deve essere considerato un vero e proprio problema ma diversificando il titolo e l'H1 ci sono maggiori opportunità di utilizzare parole chiave alternative, sinonimi o frasi chiave correlate che meglio strutturano il keyword set della pagina e definiscono il search intent.
- **Multiple:** identifica tutte le pagine che presentano più di un Meta Titles. I titoli di pagina multipli sono spesso causati da plugin o moduli in conflitto nel CMS ed andrebbero ottimizzati per non creare confusione nelle valutazioni del Motore di Ricerca.

TAB META DESCRIPTION

La Meta Description non corrisponde ad un fattore di ranking ma rappresenta il punto di incontro con un potenziale visitatore; una “description” scritta bene e persuasiva permette, insieme al “page title” di aumentare il numero di click e di conseguenza la conoscenza del tuo sito web.

```
<meta name="description" content="This is a meta description."/>
```

- **Address:** visualizza l'url scansionato.
- **Occurrences:** visualizza il numero di “meta descriptions” presenti nella pagina.
- **Meta Description 1-2:** il contenuto testuale della meta description.
- **Meta Description 1/2 length:** mostra la lunghezza della Meta Description in base al numero di caratteri. Il consiglio è quello di scrivere un testo compreso fra i 70 e 155 caratteri.
- **Indexability:** identifica se l'URL è indicizzabile o non indicizzabile.
- **Indexability Status:** definisce il motivo per cui un URL viene considerato non indicizzabile.

Anche con questa scheda hai a disposizione diversi filtri interessanti:

- **Missing:** qualsiasi pagina che ha una meta descrizione mancante, il contenuto è vuoto o presenta uno spazio bianco. Le pagine inserite in questo elenco rappresentano delle opportunità sprecate nel comunicare i benefici del tuo prodotto o servizio e influenzare i tassi di clic per gli URL più importanti.
- **Duplicate:** tutte le pagine che hanno meta description duplicate. È molto importante curare questo tag per comunicare al meglio lo scopo di ogni pagina. Se sono duplicate o irrilevanti, saranno ignorate dai Motori di Ricerca.
- **Over 155 characters:** tutte le pagine che hanno meta description che superano i 155 caratteri di lunghezza. I caratteri oltre questo limite potrebbero essere troncati nei risultati di ricerca di Google.
- **Below 70 characters:** tutte le pagine che hanno meta descrizioni sotto i 70 caratteri di lunghezza.
- **Over X Pixels:** questo filtro mostra tutte le pagine che hanno descrizioni di lunghezza superiore a X pixel e che potrebbero essere troncate nei risultati di ricerca di Google.

- **Below X Pixels:** mostra tutte le pagine che hanno meta descrizioni di lunghezza inferiore a X pixel.
- **Multiple:** tutte le pagine che hanno meta descrizioni multiple e che manderebbero in confusione il Bot.

TAB META KEYWORDS

Le meta keywords sono ampiamente ignorate dai Motori di Ricerca e Google non le considera nella definizione del ranking nei risultati di ricerca. Perciò ti consiglio di ignorarle completamente a meno che non ti rivolgi a Motori di Ricerca alternativi.

Il tag meta keywords dovrebbe essere collocato nella testa del documento e si presenta così nell'HTML:

```
<meta name="keywords" content="seo, agenzia seo, servizi seo"/>
```

TAB HEADING H1

I tag da <h1> a <h6> sono utilizzati per definire la struttura del contenuto delle pagine web in base ad una scala di importanza decrescente.

Le intestazioni dovrebbero essere ordinate per dimensione e importanza e aiutano gli utenti e i Motori di Ricerca a capire il contenuto della pagina e le sezioni creando una sorta di gerarchia del contenuto. Gli heading e in particolare l'<h1> dovrebbero descrivere l'intento di ricerca della pagina rivestendo un ruolo chiave come segnale di ranking on-page.

La Tab h1 mostra i dati relativi all'intestazione HTML <h1> e le sue potenziali criticità.

```
<h1>Questo è un h1</h1>
```

Per impostazione predefinita, il SEO Spider estrae e riporta solo i primi due <h1> scoperti su una pagina. Se desideri estrarre tutti gli h1, ti consiglio di utilizzare l'estrazione personalizzata (Custom Extraction).

- **Address:** l'URL scansionato.
- **Occurrences:** numero di heading <h1> trovati nella pagina (il seo Spider estrae i primi due).
- **h1-1-2:** mostra il contenuto testuale del <h1>.
- **h1-length-1-2:** definisce la lunghezza dell'heading in base al numero di caratteri.
- **Indexability:** se l'URL è indicizzabile o non indicizzabile.
- **Indexability Status:** Il motivo per cui un URL viene considerato non indicizzabile.

Il crawl dei titoli principali comprende i seguenti filtri:

- **Missing:** identifica le pagine che non hanno un <h1>, il contenuto è vuoto o presenta uno spazio bianco (es. <h1> </h1>).
- **Duplicate:** tutte le pagine che hanno <h1> duplicati. Essendo un segnale forte per il Motore di Ricerca è fondamentale avere degli headings unici e distinti, semanticamente ottimizzati che rappresentino bene il Search Intent della pagina e rendano uniche le diverse proposte di contenuto.
- **Over 70 characters:** tutte le pagine che hanno gli <h1> con più di 70 caratteri di lunghezza. Questo non è strettamente un problema, poiché non c'è un limite di caratteri per le intestazioni ma tuttavia, dovrebbero essere concisi e descrittivi per gli utenti e i Motori di Ricerca.

- **Multiple:** tutte le pagine che hanno più <h1>. Si consiglia di usare il rango dei titoli (h1-h6) per trasmettere la struttura del documento. Lo standard classico HTML4 definisce che ci dovrebbe essere solo un singolo <h1> per pagina, e questo è ancora generalmente raccomandato per gli utenti e il Seo.

TAB HEADING H2

Come ampiamente descritto per il titolo principale anche l'heading H2 rappresenta un segnale di ranking che aiuta a definire il Search Intent della pagina e la sua struttura semantica. Valgono le stesse regole descritte precedentemente e il Seo Spider prevede gli stessi filtri e limitazioni.

Per default vengono estratti solo i primi due h2 presenti in pagina; nel caso volessi controllare un numero più elevato ti consiglio di usare la funzione "Custom Extraction"

```
<h2>Questo è un h2</h2>
```

TAB CONTENT

La scheda "Content" mostra i dati relativi al contenuto degli URL scoperti durante la scansione in termini di numero di parole, contenuti duplicati e quasi duplicati, ed errori di ortografia e di grammatica.

- **Address:** l'indirizzo dell'URL.
- **Word Count:** questo indice identifica la somma di tutte le "parole" scansionate all'interno del tag body, escluso il markup HTML.
Il conteggio si basa sull'area di contenuto definito dal Seo Spider ma puoi anche personalizzarlo da "Config > Content > Area".
Per impostazione predefinita, gli elementi <nav> e <footer> sono esclusi.
Per un'analisi più analitica puoi decidere di includere o escludere elementi HTML, classi e ID. Ci potrebbero essere delle divergenze fra i valori riportati da Screaming Frog e un calcolo manuale delle stesse. Queste incongruenze sono dovute al fatto che il parser esegue alcune correzioni quando si imbatte su HTML non valido. Da non sottovalutare eventuali impostazioni personali di rendering che potrebbero influenzare anche quale HTML viene considerato. Screaming Frog conteggia una parola prendendo il testo e dividendolo per spazi senza alcuna considerazione in base alla visibilità del contenuto (come il testo all'interno di un div impostato come nascosto).
- **Closest Similarity Match:** attraverso questa funzionalità puoi vedere l'indice di similarità fra più pagine ed evitare situazioni di duplicazione. Utilizzando le soglie di default il Seo Spider identifica come contenuti "quasi duplicati" se presentano una corrispondenza del 90% o superiore. La soglia la puoi personalizzare da:

```
Config > Content > Duplicates > Enable Near Duplicates > Near Duplicate Similarity Threshold (definizione di una nuova soglia).
```

Per popolare questa colonna devi utilizzare la "Crawl Analysis". Solo gli URL con contenuto superiore alla soglia di somiglianza selezionata conterranno dati, gli altri rimarranno vuoti. In sintesi, per impostazione predefinita, questa colonna conterrà solo dati per gli URL con il 90% o più di similarità.

- **No. Near Duplicates:** identifica il numero di URL quasi duplicati scoperti in un crawl che soddisfano o superano la "Near Duplicate Similarity Threshold", che è una corrispondenza del 90% per impostazione predefinita. Questa impostazione la puoi regolare da "Config > Content > Duplicates". Per popolare questa colonna devi abilitare la funzione "Abilita i quasi duplicati" ed eseguire la "Crawl Analysis".
- **Errori di lingua totali:** il numero totale di errori di ortografia e grammatica scoperti per un URL. Affinché questa colonna sia popolata, è necessario selezionare "Enable Spell Check - Abilita controllo ortografico" e/o "Enable Grammar Check - Abilita controllo grammaticale"

- **Spelling Errors:** il numero totale di errori ortografici scoperti per un URL. Affinché questa colonna sia popolata, devi abilitare “Enable Spell Check” (Abilita controllo ortografico).
- **Grammar Errors:** identifica il numero totale di errori di grammatica scoperti per un URL. Affinché questa colonna sia popolata è necessario selezionare “Enable Grammar Check - Abilita controllo grammaticale”. Nelle impostazioni puoi definire la lingua di controllo, definire le regole grammaticali e il dizionario di riferimento.
Per consultare gli errori grammaticali e di “Spelling” nella scheda inferiore del Seo Spider è sufficiente selezionare gli URL presenti nella colonna dedicata della scheda “Internal” (finestra superiore). Ottima funzionalità per il copywriter per scovare eventuali errori di battitura o utilizzo improprio di regole grammaticali. Oltre alla lista di errori sono presenti suggerimenti per la correzione e la sezione del sito web dove sono stati riscontrati con relativa anteprima. Per la lingua italiana è ancora poco affidabile nei suggerimenti ma molto utile in termini di errori di battitura grazie al dizionario.
- **Language:** mostra la lingua prescelta per i controlli ortografici e grammaticali. Questo indice è basato sull'attributo della lingua impostato nell'HTML. La lingua può anche essere impostata da

- **Hash:** considera il valore della “hash” attraverso algoritmo MD5 dell'URL. Questo indice è un controllo dei contenuti duplicati esatti a differenza della scheda “Closest Similarity Match” in cui si esplicita una soglia di similarità. Nota bene che se due valori di hash corrispondono, le pagine sono esattamente uguali nel contenuto. Se c'è la differenza anche di un solo carattere, presentano valori di hash unici e non saranno rilevati come contenuti duplicati.
I duplicati esatti possono essere visti sotto “URL > Duplicate”.
- **Indexability:** se l'URL è indicizzabile o non indicizzabile.
- **Indexability Status:** il motivo per cui un URL non è indicizzabile. Potrebbe essere canonicalizzato verso un altro URL o presentare il tag “noindex”.

A disposizione i seguenti filtri:

- **Exact Duplicates:** questo filtro mostra tutte le pagine che presentano lo stesso “hash” e che potrebbero venir considerate duplicate dal Motore di Ricerca peggiorando il PageRank. Questo filtro ti permette di isolare le pagine duplicate ed ottimizzarle impostando correttamente la versione canonica e quella “Canonicalizzata”.
- **Near Duplicates:** attraverso questo filtro sei in grado di visualizzare tutte le pagine simili (non identiche!) in base ad una soglia di similarità che per default è al 90% e può essere personalizzata dalle impostazioni.

La colonna “Closest Similarity Match” mostra la più alta percentuale di somiglianza con un'altra pagina. La colonna “No. Near Duplicates” mostra il numero di pagine che sono simili alla pagina in base alla soglia di somiglianza. L'algoritmo viene eseguito sul testo della pagina, piuttosto che sull'intero HTML come i duplicati esatti.

Il contenuto utilizzato per questa analisi può essere configurato sotto “Config > Content > Area”. Ricorda che le pagine possono avere una somiglianza del 100%, ma essere considerate come “quasi duplicato” per degli arrotondamenti, quindi il 99,5% o superiore sarà visualizzato come 100%. Per popolare questa colonna devi abilitare “Enable Near Duplicates” ed eseguire la “Crawl Analysis”.

- **Low Content Pages:** questo filtro evidenzia tutte le pagine con un numero di parole inferiore a 200. Il conteggio delle parole è basato sulle impostazioni dell'area di contenuto usata nell'analisi che può essere configurata da "Config > Content > Area". Non esistono dichiarazioni ufficiali da parte del Motore di Ricerca rispetto al numero minimo di parole per far sì che un contenuto sia considerato valido ma, considerando la citazione "Content is the King", il contenuto descrittivo è uno degli elementi più importanti in un'ottica di posizionamento organico. Un'ottima analisi che ti consiglio è quella di confrontare le performance della Search Console in termini di Impressions e il filtro "Low Content Pages" per vedere se le pagine con poco contenuto vengono penalizzate e il contenuto debba essere approfondito. Ovviamente queste considerazioni dipendono dal sito web. Gli e-commerce potrebbero non necessitare di testi molto lunghi ed essere performanti anche con contenuti più ridotti.

La soglia per definire le pagine a basso contenuto la puoi regolare tramite "Low Content Word Count".

Config > Spider > Preferences > Low Content Word Count.

- **Spelling Errors:** questo filtro contiene tutte le pagine HTML con errori di ortografia.
- **Grammar Errors:** contiene tutte le pagine HTML con errori grammaticali.

TAB IMAGES

Le immagini sono una delle risorse che maggiormente popolano i siti web ed e-commerce per valorizzare e enfatizzare prodotti o servizi. Questi elementi sono un fattore a cui il Motore di Ricerca concede parecchia attenzione e nel caso di una gestione non ottimizzata possono essere un punto molto dolente per delle soft penalizzazioni in Serp.

Fra i fattori da ottimizzare ti consiglio di porre attenzione a:

- **Peso delle immagini:** il peso delle immagini potrebbe influire sul tempo di caricamento della pagina web e potrebbe inficiare anche il Crawling Budget. Essendo, per loro natura, uno degli elementi più grandi (intesi come "spazio occupato nella pagina") della risorsa html se non configurate in modo corretto potrebbero essere penalizzate in termini di LCP (Core Web Vitals). Lo spider Seo considera i Kilobytes come unità di misura.
- **Nome delle Immagini:** un nome file appropriato indirizza il Bot verso la corretta indicizzazione a database e inserimento nella serp dedicata alle "images" anche per query correlate.
- **Gestione dell'Alt Tag:** questo attributo svolge una funzione molto importante perché permette allo Spider di comprendere immediatamente il contenuto dell'immagine. Questo elemento, assieme al nome file e posizione nel testo consente una corretta indicizzazione nella "Serp images" e salvaguarda il sito web da potenziali "soft penalization". Gli attributi Alt dovrebbero specificare un testo alternativo pertinente e descrittivo sullo scopo di un'immagine e appaiono nel sorgente dell'HTML come l'esempio seguente:

```

```

La scheda dedicata alle immagini mostra i dati relativi a tutte le immagini scoperte durante una scansione. Questo include sia le immagini interne che quelle esterne, scoperte tramite i tag o i tag . Gli attributi alt delle immagini (spesso indicati erroneamente come "alt tag") possono essere visualizzati facendo clic sull'URL di un'immagine e poi sulla scheda "Image Details - Dettagli immagine" della finestra inferiore.

Le immagini decorative dovrebbero fornire un testo alt nullo (alt="") e non mancante in modo che possano essere ignorate dalle tecnologie assistive, come gli screen reader.

```

```

Le colonne a disposizione sono relative a :

- **Address:** l'URL scansionato.
- **Content:** la tipologia di contenuto dell'immagine (jpeg, gif, png ecc.).
- **Size:** dimensione dell'immagine in kilobyte. Nel caso di esportazione dei dati relativi alle immagini la dimensione utilizza come unità di misura il byte perciò è necessario dividere per 1.024 per convertire in kilobyte.
- **Indexability:** se l'URL dell'immagine è indicizzabile o non indicizzabile.
- **Indexability Status:** il motivo per cui un URL immagine non è indicizzabile, ad esempio, se fosse canonicalizzato verso un altro URL o presenti il tag "noindex".

Molto utili i 4 filtri a disposizione:

- **Over 100kb:** mostra le immagini che superano i 100kb. La velocità della pagina è estremamente importante per gli utenti e per il SEO e spesso le grandi risorse come le immagini sono uno dei problemi più comuni nel rallentamento delle pagine web. Questo filtro agisce semplicemente come regola generale per aiutare a identificare le immagini che sono abbastanza grandi nelle dimensioni del file e che possono richiedere più tempo per essere caricate. Queste dovrebbero essere ottimizzate, insieme alle opportunità identificate nella scheda PageSpeed che utilizza l'API PSI e Lighthouse per verificarne la velocità.

Questo può aiutare a identificare le immagini che non sono state ottimizzate nelle dimensioni, che si caricano fuori dallo schermo, che non sono ottimizzate ecc. Anche questo filtro lo puoi personalizzare da "Config > Spider > Preferences" per definire una soglia diversa dai 100 Kb impostati di default.

- **Missing Alt Text:** immagini che presentano nel codice html un attributo alt, ma mancano del testo. Per visualizzare lo status clicca sull'indirizzo (URL) dell'immagine e poi sulla scheda "Image Details". Le immagini dovrebbero avere un testo alternativo descrittivo ottimizzato sia per una valorizzazione semantica e descrittiva dell'immagine stessa sia per aiutare gli screen readers utilizzati da non vedenti ed ipovedenti.
- **Missing Alt Attribute:** immagini che mancano completamente di un attributo alt.
- **Alt Text Over 100 Characters:** comprende le immagini che presentano un "Alt Text" che supera i 100 caratteri. Ti consiglio di evitare l'utilizzo di parole chiavi in serie in questo attributo e di preferire testi semanticamente corretti rispetto al contesto in cui l'immagine viene inserita.

HOW TO:

COME SCANSIONARE GLI ALT TEXT

L'alt text (testo alternativo) viene usato all'interno dell'HTML di un sito web per descrivere testualmente un'immagine.

Questo tag è fondamentale per l'accessibilità web (per non vedenti e ipovedenti) e per servire un testo alternativo all'immagine se per caso la stessa non può essere caricata.

E' molto usato dai Motori di Ricerca per comprendere il contenuto delle immagini nonostante la sua funzione si sia affievolita vista l'interpretazione autonoma di Google mediante l'intelligenza artificiale.

Vediamo assieme quali siano i principali step per trovare immagini a cui manca questo attributo.

1. Eseguito il crawl di un sottodominio è sufficiente cliccare sulla scheda "Images" nella quale saranno presenti tutti gli URL relativi alle immagini, le rispettive dimensioni (Size) e il numero di link IMG di origine.

File Configuration Mode Bulk Export Reports Sitemaps Licence Help

Screamingfrog

www.screamingfrog.co.uk

Internal External Protocol Response Codes URI Page Titles Meta Description Meta Keywords H1 H2 Images Directives Hreflang

Filter: All Export

	Address	Content	Size	IMG Inlinks
1	https://www.screamingfrog.co.uk/wp-content/uploads/2017/09/banner.png	image/png	94639	4
2	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/hamburger.png	image/png	234	221
3	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/green-small/social.png	image/png	1114	1
4	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/green-small/analytics.png	image/png	1008	1
5	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/green-small/linkbuilding...	image/png	1429	1
6	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/logos/logo.png	image/png	4694	221
7	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	1526	4
8	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/green-small/contentmark...	image/png	1262	1
9	https://www.screamingfrog.co.uk/wp-content/uploads/2017/09/crawl-sites-geo-ip-redirection.jpg	image/jpeg	51364	5
10	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/story/lily-frog.png	image/png	5943	1

2. Cliccando sulla scheda "Images Details", nella finestra inferiore del SEO Spider, vedrai tutte le informazioni sulle pagine di origine che fanno riferimento all'immagine (From - l'URL 'da') e l'alt text di ogni elemento dell'immagine scoperta tra cui l'anteprima scorrendo sul lato destro della schermata.

The screenshot displays the 'Images Details' tab in the Screamingfrog SEO Spider. The top table lists various image URLs, their content types, sizes, and inlink counts. The bottom section shows a detailed view of a specific image (https://www.screamingfrog.co.uk/wp-content/uploads/2014/03/rand.jpg) with a list of URLs that link to it. The 'Alt Text' column shows that most links have an empty alt attribute, except for one which has the value 'rand'.

Nell'esempio qui sopra, puoi vedere come quattro pagine del sito web di Screaming Frog (elencati nella colonna 'From') presentano la stessa immagine (https://www.screamingfrog.co.uk/wp-content/uploads/2014/03/rand.jpg).

L'incongruenza rilevata è che solamente una di queste quattro pagine html presenta un Alt Text (nell'immagine 'rand'), le altre sono vuote, cioè mancano di testo alt.

Le immagini possono avere molti URL diversi che fanno riferimento a loro, e quindi, molti testi alt diversi da poter ottimizzare. Nonostante l'immagine sia presente in più pagine è plausibile pensare che il Search Intent delle stesse sia diverso e quindi ti consiglio, in fase di ottimizzazione, di prevedere diverse versioni di Alt Text per contestualizzare al meglio il contenuto.

3. Utilizza i filtri "Missing Alt Text" o "Missing Alt Attribute".

Il filtro "Missing Alt Text" include tutte le immagini che hanno un attributo alt, ma si presenta vuoto o include solo uno spazio bianco.

```

```

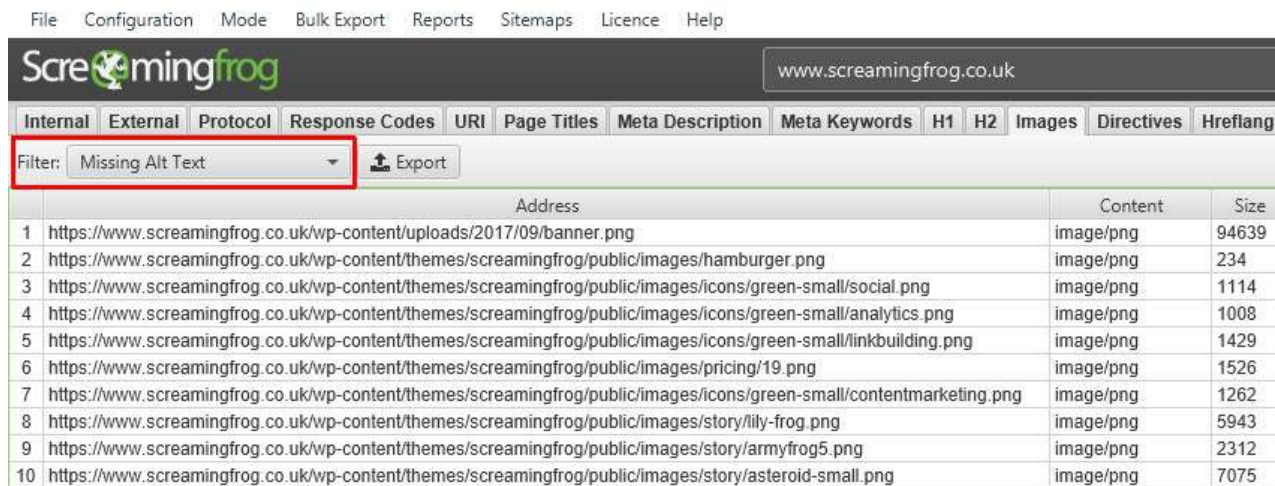
Il filtro 'Missing Alt Attribute' include le immagini che non hanno un attributo alt.

```

```

Screaming Frog prevede questi due filtri separati per identificare le immagini decorative che come best practice dovrebbero prevedere un testo alt vuoto, piuttosto che non presente per non causare problemi agli screen reader. Una buona norma potrebbe essere quella di fornire le immagini decorative come immagini di sfondo CSS.

Approfondimento: [Alt Tag Immagini - W3C](#)



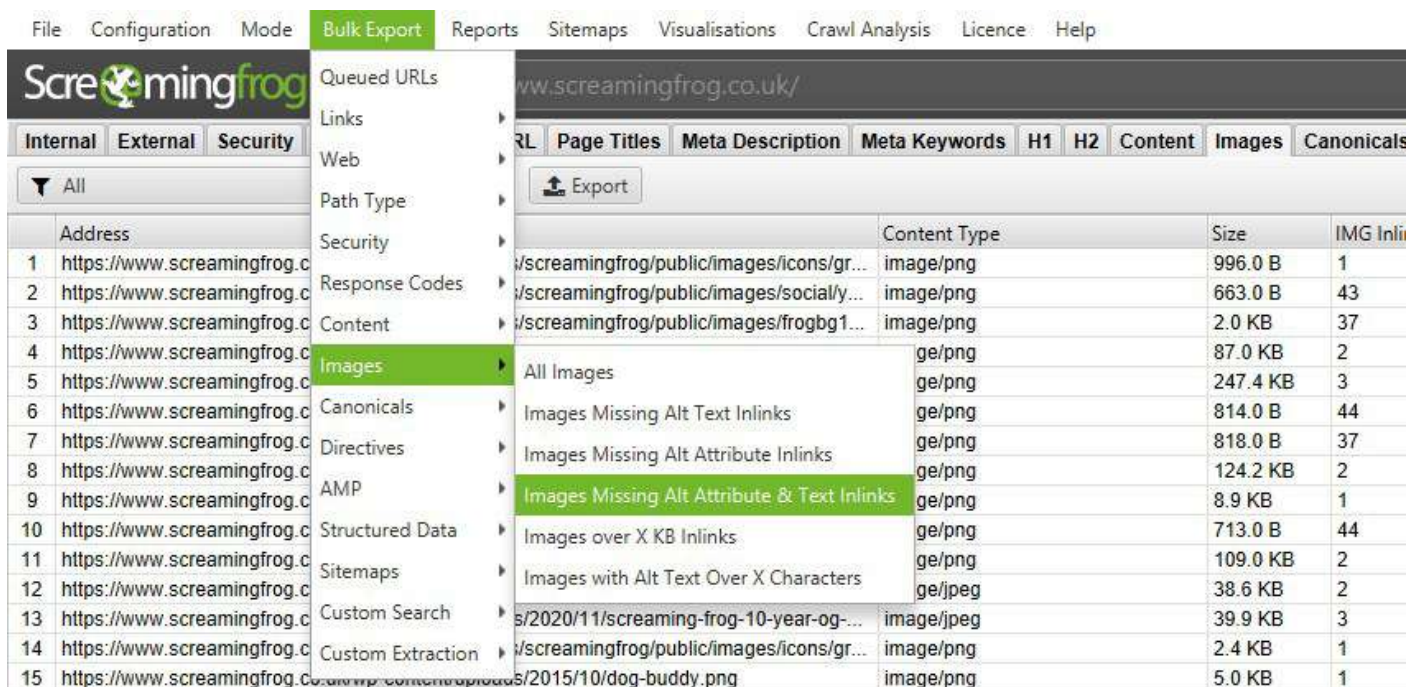
The screenshot shows the Screaming Frog SEO Spider interface. The 'Filter' dropdown is set to 'Missing Alt Text'. The 'Images' tab is selected, displaying a list of 10 images with their addresses, content types, and sizes.

	Address	Content	Size
1	https://www.screamingfrog.co.uk/wp-content/uploads/2017/09/banner.png	image/png	94639
2	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/hamburger.png	image/png	234
3	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/green-small/social.png	image/png	1114
4	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/green-small/analytics.png	image/png	1008
5	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/green-small/linkbuilding.png	image/png	1429
6	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	1526
7	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/green-small/contentmarketing.png	image/png	1262
8	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/story/lily-frog.png	image/png	5943
9	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/story/armyfrog5.png	image/png	2312
10	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/story/asteroid-small.png	image/png	7075

4. Collezionato i dati sei in grado di esportarli per singolo filtro direttamente dalla Tab o in modo massivo con la funzione "Bulk Export"

Bulk Export > Images > Images Alt Attribute & Text Inlinks

Attraverso questi documenti puoi visualizzare anche i dati delle pagine di riferimento "source", il percorso delle immagini e l'alt Text per una visione più accurata di eventuali criticità.



The screenshot shows the 'Bulk Export' menu in the Screaming Frog SEO Spider interface. The 'Images' option is selected, and the sub-menu 'Images Missing Alt Attribute & Text Inlinks' is highlighted. The main table displays the results of this export, showing columns for Address, Content Type, Size, and IMG Inlinks.

	Address	Content Type	Size	IMG Inlinks
1	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/green-small/social.png	image/png	996.0 B	1
2	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/social/y...	image/png	663.0 B	43
3	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/frogbg1...	image/png	2.0 KB	37
4	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/icons/green-small/analytics.png	image/png	87.0 KB	2
5	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/hamburger.png	image/png	247.4 KB	3
6	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	814.0 B	44
7	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	818.0 B	37
8	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	124.2 KB	2
9	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	8.9 KB	1
10	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	713.0 B	44
11	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	109.0 KB	2
12	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	38.6 KB	2
13	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/jpeg	39.9 KB	3
14	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	2.4 KB	1
15	https://www.screamingfrog.co.uk/wp-content/themes/screamingfrog/public/images/pricing/19.png	image/png	5.0 KB	1

In fase di ottimizzazione delle immagini potrebbe essere molto utile esportare tutte le risorse (comprese quelle che mancano degli Alt Text) in un'unica soluzione.

Bulk Export > Images > All Images

The screenshot shows the Screaming Frog SEO Spider interface. The 'Bulk Export' menu is open, and the 'Images' > 'All Images' path is selected. The main table displays the following data:

URL	Page Titles	Meta Description	Meta Keywords	H1	H2	Images	Content	Canonicals	
Export									
							Content Type	Size	IMG Inli
1	https://www.screamingfrog.c	s/2020/10/data-blogheader.png	image/png					247.4 KB	3
2	https://www.screamingfrog.c	s/2020/08/CWV-Header.jpg	image/jpeg					38.6 KB	4
3	https://www.screamingfrog.c	s/2016/02/which-testimonial-logo.jpg	image/jpeg					5.9 KB	1
4	https://www.screamingfrog.c		ge/png					10.1 KB	1
5	https://www.screamingfrog.c		ge/jpeg					11.0 KB	1
6	https://www.screamingfrog.c		ge/jpeg					4.9 KB	1
7	https://www.screamingfrog.c		ge/svg+xml					583.0 B	3
8	https://www.screamingfrog.c		ge/jpeg					10.9 KB	1
9	https://www.screamingfrog.c		ge/jpeg					8.3 KB	1
10	https://www.screamingfrog.c		ge/png					124.2 KB	2
11	https://www.screamingfrog.c		ge/png					86.6 KB	2
12	https://www.screamingfrog.c		ge/png					3.0 KB	2
13	https://www.screamingfrog.c	s/2016/07/Racevisors-logo.png	image/png					13.3 KB	2
14	https://www.screamingfrog.c	s/2015/10/anthony-nolan.jpg	image/jpeg					6.7 KB	1
15	https://www.screamingfrog.c	s/2016/04/red-candy.jpg	image/jpeg					8.9 KB	2

Potrebbe capitare che alcune immagini non vengano scoperte dal Seo Spider. Nella maggior parte dei casi il motivo è legato al file robots.txt che le blocca o sono caricate usando JavaScript o CSS (come CSS "background-image").

Nel primo caso puoi customizzare il file robots (config>>file robots>>custom), nel secondo è sufficiente abilitare l'opzione "Rendering JavaScript" e riavviare il crawl.

TAB CANONICAL

L'elemento `rel="canonical"` aiuta a specificare un'unica versione preferita di una pagina quando è disponibile tramite più URL. È un suggerimento ai Motori di Ricerca per evitare contenuti duplicati, consolidando l'indicizzazione e le proprietà dei link in un unico URL da utilizzare nel ranking.

L'elemento di collegamento canonico dovrebbe essere collocato nella testa del documento e assomiglia a questo in HTML:

```
<link rel="canonical" href="https://www.example.com/" >
```

Puoi anche usare le intestazioni HTTP `rel="canonical"`, che hanno questo codice:

```
Link: <http://www.example.com>; rel="canonical"
```

La Tab "canonicals" mostra gli elementi canonici dei link e i canonici HTTP scoperti durante una scansione fornendo in colonne i seguenti dati:

- **Address:** l'URL scansionato.
- **Occurences:** il numero di link canonici trovati (sia tramite l'elemento link che HTTP).
- **Indexability:** mostra se l'URL è indicizzabile o non indicizzabile.
- **Indexability Status:** la ragione perché un URL non è indicizzabile. Per esempio, se è canonizzato ad un altro URL.
- **Canonical Link Element 1/2 etc:** tutte le istanze relative ai link canonici.
- **HTTP Canonical 1/2 etc:** tutte le istanze canoniche emesse tramite HTTP.
- **Meta Robots 1/2 etc:** informazioni sul Meta robots trovato sull'URL.
- **X-Robots-Tag 1/2 etc - Dati X-Robots-tag:** il Seo Spider trova tutte le istanze se ce ne sono più di una.
- **rel="next" e rel="prev":** il Seo tool raccoglie questi elementi di collegamento HTML progettati per indicare la relazione tra gli URL in una serie paginata.

Questa tab include i seguenti filtri:

- **Contains Canonical:** identifica se nelle pagine sono presenti degli URL canonici (sia tramite elemento di collegamento, intestazione HTTP o entrambi). Il risultato potrebbe corrispondere ad un URL canonico auto referenziato dove l'URL della pagina è lo stesso dell'URL canonico (Self Referencing), o potrebbe essere "canonicalizzato", dove l'URL canonico è diverso dall'URL della pagina.
- **Self Referencing:** l'URL ha un canonico che è lo stesso URL della pagina scansionata (quindi, è auto referenziato). Idealmente solo le versioni canoniche degli URL dovrebbero essere collegate internamente, e ogni URL dovrebbe prevedere un canonico auto referenziato per aiutare ad evitare qualsiasi potenziale problema di contenuto duplicato che possa verificarsi.
- **Canonicalised:** la pagina contiene un URL canonico diverso da se stesso. L'URL è "canonicalizzato" verso un'altra pagina HTML. Questo significa che i Motori di Ricerca vengono istruiti a non indicizzare la pagina, e il Ranking dovrebbero essere assegnato all'URL canonico di destinazione. In un mondo perfetto, un sito web non avrebbe bisogno di canonicalizzare alcun URL in quanto solo le versioni canoniche dovrebbero essere collegate, ma spesso sono richieste a causa di varie circostanze al di fuori del controllo, e per prevenire contenuti duplicati.
- **Missing:** non esistono dei link "canonical" né come elemento di collegamento, né tramite intestazione HTTP. In questo caso, Google identifica quello che pensa sia la migliore versione o URL componendo imprevedibilità del ranking.

- **Multiple:** siamo in presenza di più canonici impostati per uno stesso URL (sia elementi di link multipli, intestazione HTTP, o entrambi combinati). Questo può portare all'imprevedibilità, poiché ci dovrebbe essere solo un singolo URL canonico impostato da una singola implementazione (elemento di collegamento o intestazione HTTP) per ogni pagina.
- **Non-Indexable Canonical:** l'URL canonico è una pagina non indicizzabile, cioè non risponde con status code 200. Questo filtro include i canonici bloccati da robots.txt, nessuna risposta, redirect (3XX), errore del client (4XX), errore del server (5XX) o 'noindex'.

TAB PAGINATION

La scheda della paginazione include informazioni sugli elementi di collegamento HTML `rel="next"` e `rel="prev"` scoperti in una scansione, che sono usati per indicare la relazione tra gli URL dei componenti in una serie paginata.

Mentre Google ha annunciato il 21 marzo 2019 che non utilizza `rel="next"` e `rel="prev"` nell'indicizzazione da molto tempo, altri Motori di Ricerca come Bing (che alimenta anche Yahoo), lo usano ancora come suggerimento per la scoperta e la comprensione dell'architettura del sito web.

Gli attributi di paginazione dovrebbero essere messi nella testa del documento e si presentano così in HTML:

```
<link rel="prev" href="https://www.example.com/seo/">
<link rel="next" href="https://www.example.com/seo/page/2/">
```

- **Address:** l'URL scansionato.
- **Occurrences:** Il numero assoluto di canonici trovati in un Url (sia tramite elemento di collegamento che HTTP).
- **Indexability:** mostra se l'URL è indicizzabile o non indicizzabile.
- **Indexability Status:** la ragione perché un URL non è indicizzabile.
- **rel="next":** il Seo Spider raccoglie questi elementi di collegamento HTML progettati per indicare la relazione tra gli URL in una serie paginata. Il collegamento è riferito all'URL della pagina successiva.
- **rel="prev":** lo Spider raccoglie questi elementi di collegamento HTML progettati per indicare la relazione tra gli URL in una serie paginata. Il collegamento è riferito all'URL della pagina precedente.
- **Canonical Link Element 1/2 etc:** indicazione dell'URL canonical.
- **HTTP Canonical 1/2 etc:** indicazioni sul canonical emesso via HTTP.
- **Meta Robots 1/2 etc:** Meta robots trovati sull'URL.
- **X-Robots-Tag 1/2 etc - Dati X-Robots-tag:** il Seo Spider trova tutte le istanze se ce ne sono più di una.

I filtri a disposizione sono:

- **Contains Pagination:** l'URL scansionato presenta nel codice html un attributo `rel="next"` e/o `rel="prev"`, indicando che è parte di una serie paginata.
- **First Page:** l'URL ha solo un attributo `rel="next"`, indicando che è la prima pagina di una serie paginata. È facile e utile scorrere questi URL e assicurarsi che siano accuratamente implementati nella pagina madre della serie.
- **Paginated 2+ Pages:** l'URL ha un `rel="prev"` su di esso, indicando che non è la prima pagina, ma una pagina paginata in una serie.
- **Pagination URL Not In Anchor Tag:** un URL contenuto in uno o entrambi gli attributi `rel="next"` e `rel="prev"` della pagina, non si trova come collegamento ipertestuale in un anchor text html e non permette il clic da parte del navigatore impedendo la navigazione della serie.

- **Non-200 Pagination URL:** gli URL contenuti negli attributi rel="next" e rel="prev" non rispondono con un codice di stato 200 'OK'. Questo può includere URL bloccati da robots.txt, nessuna risposta, 3XX (reindirizzamenti), 4XX (errori client) o 5XX (errori server). Gli URL di paginazione devono essere scansionabili e indicizzabili e quindi gli URL non-200 sono trattati come errori e ignorati dai Motori di Ricerca. Gli URL di paginazione non-200 possono essere esportati in blocco tramite l'esportazione del report "Non-200 Pagination URLs".

Report > Pagination > Non-200 Pagination URLs

- **Unlinked Pagination URL:** l'URL contenuto negli attributi rel="next" e rel="prev" non presenta alcun collegamento ulteriore a quello dell'attributo (nessun link da altre risorse del sito web). Gli attributi di paginazione non possono passare PageRank come un elemento di ancoraggio tradizionale, quindi questo potrebbe essere un segno di un problema con i link interni, o gli URL contenuti nell'attributo di paginazione. Questi URL critici possono essere esportati in blocco tramite l'esportazione del report "Unlinked Pagination URLs".

Report > Pagination > Unlinked Pagination URLs

- **Non indexable:** l'URL di paginazione non è indicizzabile. A livello Seo gli url utilizzati nella paginazione dovrebbero essere tutti indicizzabili a meno che non si sia prevista una pagina di servizio "View All" o, nel caso ci siano parametri extra sugli URL di paginazione che richiedono la canonicalizzazione. Un errore molto comune è l'utilizzo della canonicalizzazione delle pagine paginate 2+ alla prima pagina di una serie. Un altro errore comune è l'uso di "noindex" sulle paginazioni, che inibisce Google di indicizzare le pagine e di seguire gli outlink.
- **Multiple Pagination URLs:** ci sono più attributi rel="next" e rel="prev" sulla pagina che potrebbero mandare in confusione lo spider del Motore di Ricerca.
- **Pagination Loop:** questo filtro mostra gli URL che hanno attributi rel="next" e rel="prev" che creano dei loop non permettendo la continuazione della navigazione nella serie.

LOOP ESEMPIO:

1. La pagina 1: dominio.com/page/1 → presenta il "rel=next" a se stessa
2. La pagina 2 dominio.com/page/2 presenta il "rel=next" → dominio.com/page/3 e il rel=prev → dominio.com/page/2
3. La pagina 3 dominio.com/page/3 presenta il "rel=next" → dominio.com/page/2

- **Sequence Error:** il filtro visualizza gli URL che hanno un errore nella sequenza degli elementi di collegamento HTML rel="next" e rel="prev". Questo controllo assicura che gli URL contenuti all'interno degli elementi di collegamento HTML rel="next" e rel="prev" siano reciproci e confermino la loro relazione nella serie.

COME FARE UNA SCANSIONE DELLA PAGINAZIONE

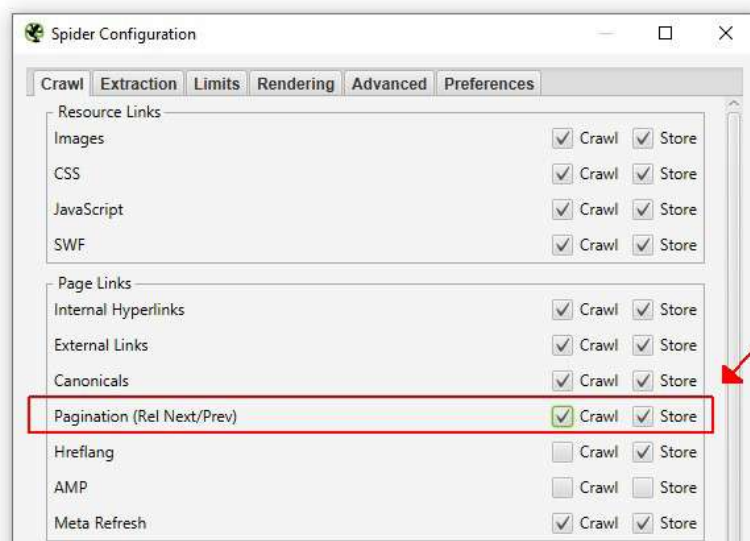
Gli attributi di paginazione `rel="next"` e `rel="prev"` sono usati per indicare ai Motori di Ricerca la relazione tra gli URL in una serie paginata. Aiutano a consolidare le proprietà di indicizzazione nella sequenza e indirizzano gli utenti all'URL più rilevante all'interno della serie.

Le paginazioni sono molto diffuse negli ecommerce e nei blog o thread di forum. Nonostante questi elementi non vengano considerati da Google per l'indicizzazione (i link nel `rel=prev/next` non passano ranking) rimangono ugualmente importanti per altri Motori di Ricerca come Bing e Yahoo per la "crawlerabilità" del sito e per la definizione della struttura.

Vediamo assieme quali siano gli step da seguire per collezionare eventuali problemi legati a questa tematica.

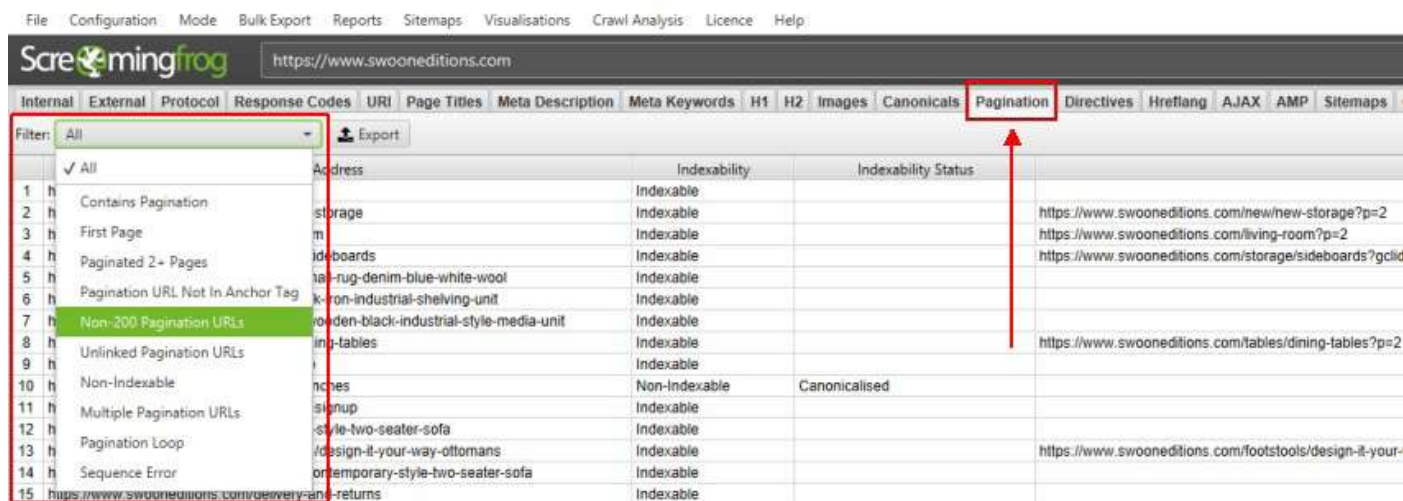
1. Definizione delle opzioni di paginazione

Per prima cosa devi abilitare l'opzione dedicata alle paginazioni per consentire a Screaming Frog di fare il crawl, memorizzare i dati collezionati e permettere la creazione dei report. Completata la configurazione puoi eseguire la scansione.



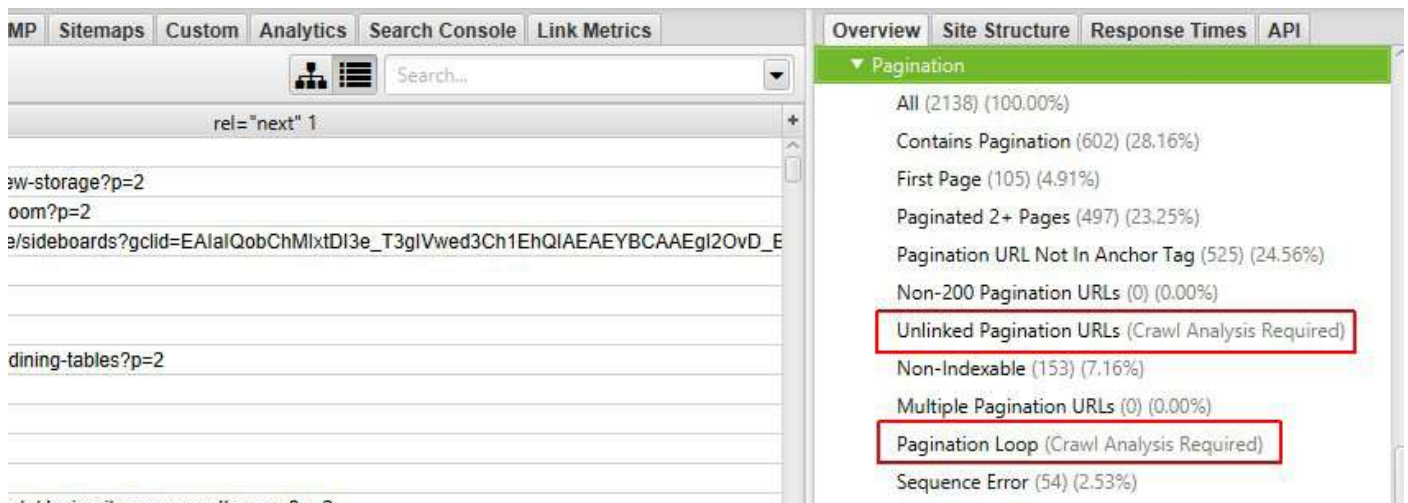
2. Consultazione dei dati della Paginazione

Completata la scansione puoi visualizzare i dati nella scheda "Pagination" e utilizzare uno dei filtri a disposizione per valutare l'implementazione degli attributi identificando eventuali criticità.

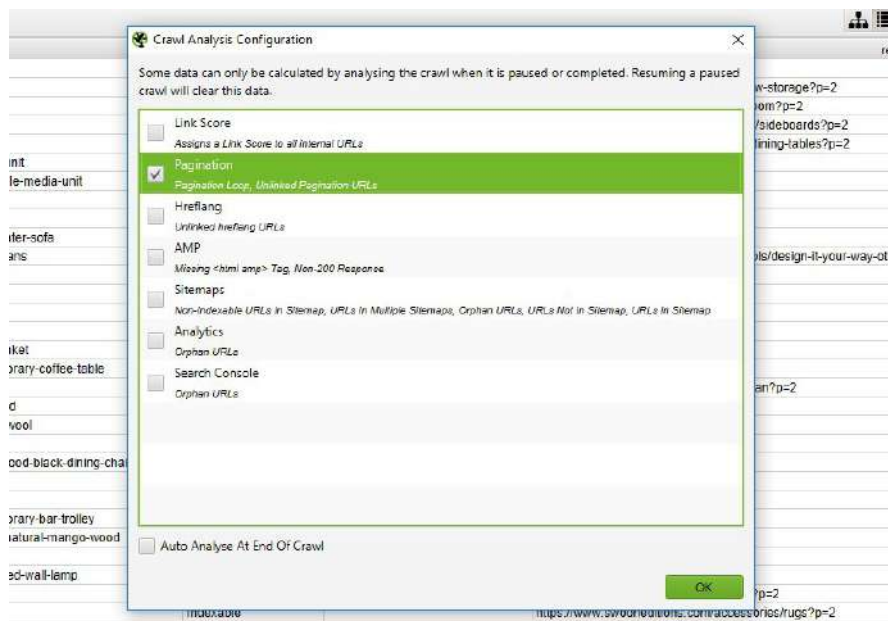


3. Esegui la Crawl Analysis

Tutti i filtri sono popolati durante la scansione tranne “Unlinked Pagination URLs” e “Pagination Loop” che necessitano della “Crawl Analysis”.



Non ti resta che abilitare la voce “Pagination” nel popup dedicato ed avviare il calcolo del Seo Spider.



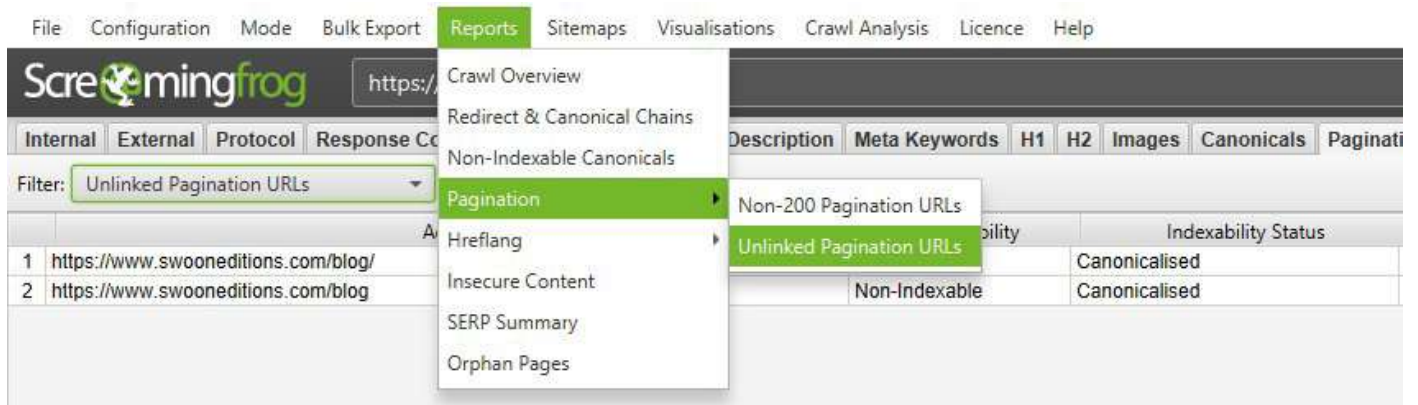
4. Esportazione dei dati della Paginazione

Scelti i filtri più adatti al tuo Audit Seo non ti resta che esportare i dati.

In questo caso puoi scegliere se utilizzare la funzione “Export” presente nella tab che ti fornisce i dati parziali del filtro applicato o scegliere l’esportazione aggregata con tutti i dettagli relativi alle pagine sorgente che contengono errori o problemi di paginazione.

Reports > Pagination

Ad esempio, l’esportazione “Reports > Pagination > Unlinked Pagination URLs” include i dettagli delle pagine sorgente che contengono gli attributi rel=”next” e rel=”prev” verso determinati URL che però non ricevono alcun collegamento supplementare da altre risorse del sito web.



TAB DIRECTIVES

La scheda delle direttive mostra i dati relativi al tag meta robots e il tag X-Robots nell'intestazione HTTP. Queste direttive robots possono controllare come il tuo contenuto e gli URL vengano visualizzati nei risultati dei Motori di Ricerca, come Google.

Per comprendere al meglio l'importanza di queste direttive considera due scenari possibili in cui si potrebbe imbattere il Bot del Motore di Ricerca:

1. Il file robots.txt è mancante o ha autorizzato il crawl di un URL, i Motori di Ricerca considerano quella risorsa scansionabile e archiviabile nell'indice a meno che non ci siano delle direttive più specifiche definite attraverso i Meta Robots.
2. il robots.txt ha in "disallow" l'url e non permette la sua scansione. In questo caso i meta robots non saranno rilevati e le loro istruzioni non saranno eseguite.

Il meta robots tag dovrebbe essere posto nell'head di una pagina HTML; per esempio un meta tag "noindex" apparirà così:

```
<meta name="robots" content="noindex">
```

La stessa direttiva può essere emessa nell'intestazione HTTP usando il X-Robots-Tag, che si presenta così:

```
X-Robots-Tag: noindex
```

Le istruzioni dei Meta Robots possono essere anche combinate seguendo questa sintassi standard che prevede più direttive separate dalla virgola:

```
<meta name="robots" content="noindex,nofollow">
```

La tab "Directives" si presenta con tre colonne, la prima che riguarda l'indirizzo dell'URL (Address) e le restanti dedicate rispettivamente al Meta Robots e X-Robots-Tag per le quali il Seo Spider colleziona tutte le istanze scoperte.

I filtri a tua disposizione sono i seguenti:

- **Index:** identifica le pagine che presentano l'attributo "index"; non è necessaria la sua presenza perché i Motori di Ricerca indicizzano gli URL anche senza questa direttiva.
- **Noindex:** questo attributo indica ai Motori di Ricerca di non indicizzare le pagine. Questa direttiva non blocca la scansione dello Spider! Può essere impostata a livello globale o specifica per ogni Motore di Ricerca.

Esempio di direttiva noindex applicata a tutti i MdR:

```
<meta name="robots" content="noindex">
```

```
<meta name="google" content="noindex">
```

- **Follow:** attraverso la direttiva "follow" si istruisce lo Spider a seguire tutti i link della pagina durante il crawling. Se non presente i Motori di Ricerca seguiranno i link di default.
- **Nofollow:** consiste in un "suggerimento" che dice ai Motori di Ricerca di non seguire alcun link sulla pagina per la scansione. Questa direttiva è comunemente applicata nei siti di staging assieme a "noindex" per evitare le scansioni dei Motori di Ricerca prima della pubblicazione. Per poter eseguire un crawl che presenta questa direttiva puoi abilitare l'opzione "Follow Internal Nofollow" dal menu "Config > Spider". Ricorda bene: con questa direttiva i link della pagina non passano PageRank.
- **None:** il meta tag "none" non corrisponde alla mancanza di direttive, anzi è molto restrittivo e combina il "noindex" e il "nofollow".
- **NoArchive:** questo tag robots indica a Google di non mostrare la copia cache della pagina nei risultati di ricerca. Di default la copia cache di una risorsa è disponibile in Chrome utilizzando il comando "cache:www.dominio.xxx".
- **NoSnippet:** questo meta robots inibisce il Motore di Ricerca (al momento solo a Google) di mostrare la "Meta Description" e il link alla versione cache nei suoi risultati.
- **Max-Snippet:** questo valore permette di limitare la lunghezza dello snippet di testo per questa pagina a [numero] di caratteri in Google. I valori speciali includono - 0 per nessun snippet, o -1 per permettere qualsiasi lunghezza di snippet.
- **Max-Image-Preview:** questo valore imposta la dimensione massima dell'anteprima di un'immagine di una pagina Html in Google. I valori dell'indice possono essere:
 - "none": non deve essere mostrata alcuna anteprima dell'immagine;
 - "standard": è possibile mostrare un'anteprima predefinita dell'immagine;
 - "large": è possibile mostrare un'anteprima dell'immagine che sia larga al massimo come l'area visibile.

```
Es: <meta name="robots" content="max-image-preview:standard">
```

- **Max-Video-Preview:** questa direttiva indica a Google di usare un massimo di [numero] secondi come snippet per i video di una pagina nei risultati di ricerca. Puoi anche specificare 0 per permettere solo un'immagine statica, o -1 per permettere qualsiasi lunghezza di anteprima.

```
<meta name="robots" content="max-video-preview:-1">
```

- **NoODP:** questo meta tag, ormai in disuso, indicava a Google di non usare l'Open Directory Project (DMOZ) per i suoi snippet. Può tranquillamente essere rimosso.
- **NoYDIR:** questo è un vecchio meta tag che indicava a Google di non usare Yahoo Directory per i suoi snippet. Questo meta robots può essere rimosso.
- **NoImageIndex:** il meta robots con questa direttiva dice a Google di non indicizzare le immagini di una pagina html. Può capitare che le immagini presenti in uno specifico Url con istruzione "NoimageIndex" siano linkate anche da altre pagine, in questo caso potrebbero comunque venire indicizzate. Per evitare questa condizione una soluzione è quella di impostare X-Robots-Tag nell'HTTP Header del file immagine.
- **NoTranslate:** questo valore dice a Google di non proporre la traduzione della pagina nei risultati di ricerca. Se non presente, il Motore di Ricerca potrebbe mostrare un link accanto al risultato in Serp con la possibilità di vedere la pagina tradotta.
- **Unavailable_After:** questa istruzione ti permette di specificare l'ora e la data esatta in cui vuoi che Google smetta di mostrare la pagina nei risultati di ricerca. Istruzione molto utile per pagine legate ad eventi, fiere, promozioni a tempo etc.

```
<meta name="robots" content="unavailable_after: 2020-09-21">
```

- **Refresh:** questa direttiva reindirizza l'utente ad un nuovo URL dopo un certo periodo di tempo.

TAB HREFLANG

In un mercato sempre più globale cresce la necessità di internazionalizzare il proprio business con siti web che presentano contenuti tradotti che soddisfino le diverse esigenze del pubblico. Google utilizza gli attributi "rel="alternate" hreflang="x" per definire quali siano i contenuti più appropriati da servire nei risultati di ricerca in base alla geolocalizzazione del visitatore.

Per questa ragione l'hreflang riveste un ruolo chiave nel posizionamento organico internazionale garantendo al Motore di Ricerca di comprendere al meglio la relazione esistente tra pagine con versioni linguistiche differenti e per evitare a livello Seo di ridurre potenziali contenuti considerati duplicati. L'attributo hreflang può essere inserito nel tag <head>... </head>, nell'intestazione HTTP oppure nella sitemap xml e prevede nella sua sintassi l'indicazione della pagina in cui è inserita (autoreferenziazione) e tutte le lingue alternative in cui viene servito il contenuto.

```
<link rel="alternate" href="https://dominio.com/en" hreflang="en" />
<link rel="alternate" href="https://dominio.com/de" hreflang="de" />
<link rel="alternate" href="https://dominio.com/en" hreflang="x-default" />
```

E' inoltre possibile definire per la stessa lingua anche la zona geografica di un URL alternativo:

```
<link rel="alternate" hreflang="en-gb" href="http://en-gb.example.com/page.html" />
<link rel="alternate" hreflang="en-us" href="http://en-us.example.com/page.html" />
<link rel="alternate" hreflang="en" href="http://en.example.com/page.html" />
<link rel="alternate" hreflang="de" href="http://de.example.com/page.html" />
<link rel="alternate" hreflang="x-default" href="http://en.example.com/page.html" />
```

Nel caso appena trattato la pagina prevede una versione per l'inglese britannico, quello americano, una versione tedesca e una versione in inglese che viene servita da Google nel caso in cui il visitatore non possa essere soddisfatto linguisticamente dalle versioni geo-localizzate (Es. potrebbe essere il caso di un australiano, o di un italiano che visita il sito web).

Per collezionare i dati relativi all'internazionalizzazione di un sito web puoi far affidamento su Screaming Frog e, nello specifico sulla scheda dedicata che include i dettagli sulle annotazioni hreflang scansionate dal SEO Spider, fornite da elementi di link HTML, HTTP Header o XML Sitemap.

Per collezionare questi preziosi dati devi abilitare le opzioni "Store Hreflang" e "Crawl Hreflang" (Config > Spider). Per estrarre le annotazioni hreflang dalla Sitemap XML, devi abilitare anche l'opzione "Crawl Linked XML Sitemaps".

Oltre all'indirizzo dell'URL e il titolo della pagina nella scheda trovi importanti informazioni relative a:

- **Occurrences:** il numero di "rel=alternate" scoperte durante il crawl.
- **HTML hreflang 1-2:** la lingua dell'attributo hreflang e il codice della regione del collegamento HTML.
- **HTML hreflang 1/2 URL etc:** l'URL utilizzato come Hreflang.
- **HTTP hreflang 1/2 etc:** la lingua dell'hreflang e il codice della regione dell'intestazione HTTP.
- **HTTP hreflang 1/2 URL etc:** l'URL hreflang dall'intestazione HTTP.
- **Sitemap hreflang 1/2 etc:** la lingua nell'hreflang e il codice regionale scoperto dalla Sitemap XML. Attenzione, questo dato è a tua disposizione solo se esegui il crawling della Sitemap XML in modalità elenco.
- **Sitemap hreflang 1/2 URL etc:** l'URL hreflang della Sitemap XML. Nota bene: questo dato viene inserito solo quando si esegue la scansione della Sitemap XML in modalità elenco.

La Tab prevede i seguenti filtri:

- **Contiene Hreflang:** questo filtro considera tutti gli URL che hanno annotazioni hreflang rel="alternate" da qualsiasi implementazione, sia che si tratti di elementi di collegamento, intestazione HTTP o Sitemap XML.
- **Non-200 Hreflang URLs:** visualizza tutti gli url che sono stati utilizzati nelle annotazioni hreflang rel="alternate" ma che non rispondono con lo "Status Code" 200. Vengono inclusi ad esempio gli URL bloccati da robots.txt, nessuna risposta, 3XX (redirect), 4XX (errori client) o 5XX (errori server). Come puoi comprendere, vista l'importanza di questo elemento è vitale che gli URL hreflang siano scansionabili e indicizzabili e quindi gli URL che non lo permettono devono essere ottimizzati per non essere ignorati da parte dei Motori di Ricerca. Gli URL hreflang non-200 possono anche essere esaminati nel riquadro inferiore della finestra "URL Info" controllando lo "Status Code".

Report > Hreflang > Non-200 Hreflang URLs

- **Unlinked Hreflang URLs:** il filtro identifica le pagine che contengono uno o più URL hreflang che sono scopribili solo attraverso le annotazioni "rel=alternate hreflang" di link hreflang rel="alternate". Considerando che le annotazioni hreflang non passano PageRank come un tag di ancoraggio tradizionale, questa condizione potrebbe essere un avvertimento importante in ottica di ottimizzazione dei collegamenti interni, o un errore di battitura o strategia delle annotazione hreflang. Per scoprire esattamente quali URL hreflang su queste pagine sono prive di collegamento, puoi utilizzare l'esportazione del rapporto "Unlinked Hreflang URLs".

Report > Hreflang > "Unlinked Hreflang URLs"

- **Missing Return Links:** comprende tutti gli URL scoperti nelle annotazioni "hreflang" a cui manca un link di ritorno (o 'return tags' in Google Search Console) ad essi, dalle loro pagine alternative. L'Hreflang è reciproco, quindi tutte le versioni alternative devono confermare la relazione. Quando la pagina A si collega alla pagina B usando l'annotazione hreflang per specificarla come pagina alternativa, la pagina B deve prevedere un link di ritorno. Mancando un link di ritorno il Motore di Ricerca potrebbe ignorare completamente l'annotazione o mal interpretarla. Gli URL di ritorno mancanti possono essere visti nella tab della finestra inferiore "URL Info" con uno stato di conferma "missing" e li puoi esportare in blocco tramite l'esportazione "Missing return Links".

Reports > Hreflang > Missing return Links

- **Inconsistent Language & Region Return Links:** questo filtro include gli URL con link di ritorno in lingua e regione incoerenti. Questo avviene quando un link di ritorno ha una lingua o un valore regionale diverso da quello a cui fa riferimento l'URL stesso. Gli URL di ritorno in lingua incoerente possono essere visti nella finestra inferiore "URL Info" con uno stato di conferma "Inconsistent".

Reports > Hreflang > Inconsistent Language Return Links

- **Non Canonical Return Links:** URL con link di ritorno hreflang non canonici. Hreflang dovrebbe includere solo le versioni canoniche degli URL. Quindi questo filtro raccoglie i link di ritorno che vanno a URL che non sono le versioni canoniche (canonicalised). Questo filtro è più interessato ai link di ritorno dalle annotazioni hreflang. Quindi immagina che l'URL A sia canonicalizzato verso l'URL B e un'annotazione hreflang verso l'URL C, sull'URL C l'hreflang di ritorno dovrebbe essere all'URL canonico (URL B) piuttosto che all'URL A. Gli URL di ritorno non canonici possono essere visti nella finestra inferiore "URL Details" con uno stato di conferma "Not canonical" e possono essere esportati in blocco tramite l'esportazione "Non Canonical Return Links".

Reports > Hreflang > Non Canonical Return Links

- **Noindex Return Links:** comprende i link inseriti nel “rel=alternate hreflang” che presentano un meta tag “noindex”. Gli URL dei link di ritorno noindex possono essere analizzati nel riquadro inferiore della finestra “URL Info” con uno stato di conferma “noindex”. Possono essere esportati in blocco tramite l’esportazione “Noindex Return Links”.

Report > Hreflang > Noindex Return Links

- **Incorrect Language & Region Codes:** verifica semplicemente che la lingua (in formato ISO 639-1) e i valori del codice regionale opzionale (in formato ISO 3166-1 Alpha 2) siano validi. I valori hreflang non supportati possono essere visualizzati nella finestra inferiore “URL Info” con uno stato “non valido”.
- **Multiple Entries:** il filtro identifica gli URL con voci multiple per una lingua o un codice regionale. Per esempio, se la pagina X si collega alla pagina Y e Z usando la stessa annotazione di valore hreflang “en”.
- **Missing Self Reference:** colleziona gli URL a cui manca l'autoreferenziazione nel rel=”alternate”.
- **Not Using Canonical:** filtro che evidenzia la condizione per cui l'URL A è canonicalizzato verso l'URL B, ma l'annotazione hreflang continua ad auto-referenziarsi all'URL A. Hreflang dovrebbe includere solo le versioni canoniche degli URL.
- **Missing X-Default:** colleziona gli URL mancanti di un attributo X-Default hreflang.
- **Missing:** insieme degli URL ai quali mancano gli attributi hreflang.

TAB AMP

AMP è l'acronimo di Accelerated Mobile Pages e corrisponde ad una tecnologia (Web Component Framework) nata dalla collaborazione fra Google e Twitter per consentire di velocizzare il caricamento delle pagine web sui dispositivi mobili. Questo progetto, assieme all'orientamento “Mobile First Indexing” mostra quanto il Motore di Ricerca sia indirizzato verso il mercato “Mobile”.

AMP permette di ridurre in modo drastico i tempi di caricamento rispetto alle pagine tradizionali per un'esperienza utente completamente rivoluzionata che può trasformarsi nella riduzione dell'abbandono della pagina, maggiori conversioni, leads etc. oltre a migliorare le performance che sono uno dei più significativi segnali di ranking.

Al momento AMP include principalmente siti di news ma si crede che il bacino di applicazione nel tempo possa aumentare.

Le pagine AMP sono alternative alle pagine tradizionali ma seguono delle regole sintattiche e di markup specifiche che le differiscono dalle pagine canoniche.

Per comunicare al Motore di Ricerca l'esistenza di queste pagine AMP si utilizzano dei link tag di corrispondenza fra la pagina canonica e la versione AMP:

```
<link rel="amphtml" href="https://www.dominio.com/amp/page.html">
```

Tag Link inserito nella pagina canonica che identifica la versione AMP

```
<link rel="canonical" href="https://www.dominio.com/page.html">
```

Tag Link inserito nella pagina AMP che identifica la versione canonica

In questo modo quando Google scansiona la pagina canonica (nell'esempio: <https://www.dominio.com/page.html>) scopre anche la versione AMP (<https://www.dominio.com/amp/page.html>) e, se la considera valida la inserisce nella sua CDN.

Molto importante ricordare che la versione semplificata non impatta negativamente sul Crawling Budget del sito web dato che Google ne dedica uno per l'AMP.

La Tab AMP include tutte le “Accelerated Mobile Pages (AMP)” scoperte durante il crawl. Queste sono identificate tramite il tag HTML AMP e gli inlink rel=”amphtml”.

Per popolare questa tab devi abilitare entrambe le opzioni “Store” e “Crawl” AMP (“Config > Spider”).

- **Address:** identifica l'URL scansionato.
- **Occurrences:** il numero di tag link canonici scoperti (sia tramite l'elemento link che HTTP).
- **Indexability:** identifica se l'URL è indicizzabile o non indicizzabile.
- **Indexability Status:** identifica il motivo per cui un URL non può essere indicizzato.
- **Title 1:** corrisponde al primo titolo della pagina.
- **Title 1 Length:** visualizza la lunghezza del titolo in base al numero di caratteri.
- **Title 1 Pixel Width:** visualizza la larghezza del titolo della pagina misurato in pixel.
- **h1 - 1:** corrisponde al primo heading H1 della pagina.
- **h1 - Len:** visualizza il numero di caratteri della heading H1.
- **Size:** definisce la dimensione in byte della pagina AMP.
- **Word Count:** determina il numero complessivo di parole all'interno del tag body escludendo il markup HTML.
- **Text Ratio:** corrisponde al rapporto fra il numero di caratteri del testo ("non-HTML") trovati nel tag HTML body di una pagina e il numero totale di caratteri di cui è composta la pagina HTML visualizzando il dato in percentuale.
- **Crawl Depth:** identifica la distanza della pagina AMP rispetto alla pagina iniziale di scansione in base al numero di click. Nel Seo Spider i reindirizzamenti sono contati come un livello supplementare.
- **Response Time:** visualizza il tempo necessario per scaricare l'Url. L'unità di misura è il secondo.

La Tab AMP include due tipologie di filtri: la prima tipologia relativa a controlli puramente Seo, la seconda per specifiche peculiari delle versioni AMP.

- **Non-200 Response:** identifica quali URL in versione AMP non rispondono con un codice di stato 200 "OK". Le potenziali cause potrebbero essere: bloccati da robots.txt, nessuna risposta, reindirizzamenti, errori del client e del server.
- **Missing Non-AMP Return Link:** identifica gli URL che hanno una versione AMP ma non prevedono il Tag Link rel="amphtml" di ritorno all'URL AMP.
- **Missing Canonical to Non-AMP:** visualizza gli URL AMP a cui manca il Tag Link verso URL canonical o che hanno un Tag Link verso un'altra risorsa AMP.
- **Non-Indexable Canonical:** vengono mostrati gli URL AMP che hanno come canonical URL non indicizzabili. Normalmente i link canonici dovrebbero essere indicizzabili.
- **Non Indexable:** identifica gli "Url AMP" che non sono indicizzabili perché correttamente canonicalizzati verso la versione desktop (canonical).
- **Indexable:** visualizza gli "URL AMP" che sono indicizzabili. Gli URL AMP che hanno un equivalente desktop dovrebbero sempre essere non indicizzabili perché canonicalizzati verso l'equivalente desktop. Solamente gli URL AMP che non hanno una versione equivalente desktop dovrebbero essere indicizzabili.

Vediamo un esempio di una pagina base AMP:

```
<!doctype html>
<html amp lang="en">
  <head>
    <meta charset="utf-8">
    <script async src="https://cdn.ampproject.org/v0.js"></script>
    <title>Hello, AMPs</title>
    <link rel="canonical" href="https://amp.dev/documentation/guides-and-tutorials/start/create/basic_markup/">
    <meta name="viewport" content="width=device-width,minimum-scale=1,initial-scale=1">

    <style amp-boilerplate>body{-webkit-animation:-amp-start 8s steps(1,end) 0s 1 normal both;-moz-animation:-amp-start 8s
    steps(1,end) 0s 1 normal both;-ms-animation:-amp-start 8s steps(1,end) 0s 1 normal both;animation:-amp-start 8s steps(1,end) 0s
    1 normal both}@-webkit-keyframes -amp-start{from{visibility:hidden}to{visibility:visible}}@-moz-keyframes -amp-start{from{visi-
    bility:hidden}to{visibility:visible}}@-ms-keyframes -amp-start{from{visibility:hidden}to{visibility:visible}}@-o-keyframes -amp-star-
    t{from{visibility:hidden}to{visibility:visible}}@keyframes -amp-start{from{visibility:hidden}to{visibility:visible}}</style><noscrip-
    t><style amp-boilerplate>body{-webkit-animation:none;-moz-animation:none;-ms-animation:none;animation:none}</style></
    noscript>

  </head>
  <body>
    <h1 id="hello">Hello AMPHTML World!</h1>
  </body>
</html>
```

- **Missing HTML AMP Tag:** le pagine AMP HTML devono contenere un tag HTML o HTML AMP di primo livello.

```
<html amp lang="en">
```

- **Missing/Invalid HTML Doctype HTML Tag:** le pagine AMP HTML devono iniziare con il doctype, doctype HTML tipico di Html5 e attraverso questo filtro ti vengono segnalate le pagine che non lo prevedono o lo hanno impostato in modo errato.

```
<!doctype html>
```

- **Missing Head Tag:** le pagine AMP HTML devono contenere il tag head che invece è considerato opzionale in HTML.

```
<head>...</head>
```

- **Missing Tag Body:** i documenti AMP HTML devono contenere il tag <body>. Questo filtro visualizza le pagine in cui il tag è mancante.
- **Missing Canonical:** gli URL AMP devono contenere un tag canonical all'interno dell'Head che punta alla versione HTML regolare (non AMP) del documento HTML AMP, o a se stesso se non esiste tale versione HTML.
- **Missing/Invalid Meta Charset Tag:** le pagine AMP HTML devono contenere un tag meta charset="utf-8" subito dopo il tag <head> che serve per codificare la pagina.
- **Missing/Invalid Meta Viewport Tag:** le pagine AMP HTML devono contenere il meta name="viewport" content="width=device-width,minimum-scale=1" tag all'interno del loro tag head. Si raccomanda anche di includere initial-scale=1 per ottenere una viewport responsive.
- **Missing/Invalid AMP Script:** il Seo Spider controlla che le pagine AMP HTML contengano un tag script async src="https://cdn.ampproject.org/v0.js" all'interno del loro tag head.
- **Missing/Invalid AMP Boilerplate:** identifica le pagine AMP HTML che non contengono il codice boilerplate AMP nel loro tag head oppure lo stesso è invalido. Il codice AMP-boilerplate <style amp-boilerplate>...</style> nasconde il contenuto della pagina impostando l'opacità della sezione body a 0. Dopo aver caricato l'AMP il sistema calcola il layout della pagina e il sistema AMP imposta l'opacità della sezione body a 1 permettendo di vedere il contenuto.
- **Contains Disallowed HTML:** identifica ogni URL AMP con HTML non consentito per AMP.

TAB STRUCTURED DATA

I Motori di Ricerca e in particolare Google si aggiornano continuamente per riuscire a comprendere al meglio i contenuti di una pagina web. I dati strutturati servono proprio a questo perché semplificano la scansione e interpretazione fornendo indizi più precisi al Motore di Ricerca.

Sono elementi molto preziosi che aiutano il ranking della pagina e permettono di abilitare funzionalità e miglioramenti nei risultati di ricerca. I dati strutturati vengono codificati con uno specifico markup e seguono il vocabolario di Schema.org anche se il riferimento principale rimane il “Google Search Central” che identifica quali siano i dati strutturati obbligatori, facoltativi e opzionali nei diversi casi specifici. Schema.org prevede molti più attributi che però non sono considerati obbligatori in ottica di Ricerca Google e quindi possono essere tralasciati durante un’ottimizzazione di un sito web.

La “Ricerca Google” supporta 3 formati di Structured data:

1. JSON-LD (consigliato);
2. Microdati;
3. RDFa.

I dati strutturati possono venire testati e convalidati attraverso il “Test dei risultati multimediali” e monitorato il loro stato di salute anche con Search Console nella sezione dedicata.

Anche Screaming Frog è un utile alleato nel monitoraggio di questo markup e la scheda “Structured Data” include tutti i principali dettagli segnalando i relativi problemi di convalida scoperti durante il crawl.

Per collezionare i dati e rispettivi filtri devi abilitare le opzioni di configurazione “JSON-LD”, “Microdata”, “RDFa”, “Schema.org Validation” e “Google Rich Result Feature Validation”.

Config > Spider > Extraction

Le colonne a tua disposizione nella tab “Structured Data”:

- **Address:** l’URL scansionato.
- **Errors:** visualizza il numero complessivo di errori di convalida.
- **Warnings:** corrisponde al numero complessivo di avvertimenti scoperti durante il crawl.
- **Total Types:** identifica il numero totale di tipologie di dati strutturati scoperti per l’URL.
- **Unique Types:** mostra quante tipologie uniche di “structured data” sono stati scoperti per l’URL.
- **Type 1:** primo dato strutturato scoperto.
- **Type 2 etc:** secondo dato strutturato scoperto.

I filtri a tua disposizione sono i seguenti:

- **Contains Structured Data:** elenca tutti gli URL che contengono dati strutturati.
- **Missing Structured Data:** mostra gli URL che non contengono dati strutturati.
- **Validation Errors:** vengono visualizzati tutti gli URL che presentano degli errori di convalida in base al vocabolario di Schema.org, di Google rich result o di entrambi, a seconda della tua configurazione. Il Seo Spider classifica i problemi di Schema.org sempre come errori, piuttosto che come avvisi. La convalida di Google rich result feature ti mostra gli errori per le proprietà richieste mancanti o problemi con l’implementazione delle proprietà richieste.
- **Validation Warnings:** mostra tutti gli URL che presentano degli avvisi (“warnings”) di convalida rispetto ai “Google Rich Results”. Il Seo Spider popola questo filtro considerando solamente i problemi di convalida dei “dati strutturati facoltativi”.

- **Parse Errors:** il filtro presenta tutti gli URL che contengono dati strutturati che non sono stati analizzati correttamente. Questo è spesso dovuto a un mark-up errato. Se stai usando il formato preferito di Google, JSON-LD, allora il [JSON-LD Playground](#) è uno strumento eccellente per aiutare il debug degli errori di analisi.
- **Microdata URLs:** corrisponde agli URL che contengono dati strutturati in formato microdati.
- **Microdata URLs | JSON-LD URLs | RDFa URLs:** vengono mostrati rispettivamente gli URL che contengono dati strutturati in formato JSON-LD, RDFa o microdata.

TAB SITEMAP

La scheda raccoglie tutti i dati più significativi estrapolati dalla scansione di una Sitemap XML.

Per ottenere i dati devi abilitare "Crawl Linked XML Sitemaps" in "Configurazione > Spider", avviare la scansione ed eseguire una "Crawl Analysis" successiva.

Le colonne a disposizione nella tab includono informazioni generali sugli URL e sulla loro indicizzabilità. Molto più interessanti e specifici i filtri dedicati:

- **URLs in Sitemap:** vengono visualizzati gli URL che sono presenti in una Sitemap XML. A livello Seo questi elementi dovrebbero contenere solamente versioni indicizzabili e canoniche evitando pagine con status code diverso da 200, bloccate da robots.txt o canonicalizzate.
- **URLs Not In Sitemap:** identifica tutti gli URL che sono stati trovati durante la scansione ma non appartengono alla Sitemap XML. In questo caso consiglio sempre un controllo per constatare se questi URL sono stati tralasciati in modo voluto o sono delle proprie mancanze ed andrebbero aggiunti. Questo filtro non considera gli URL non indicizzabili, assumendoli come scelte consapevoli, e quindi non li segnala.
- **Orphan URLs:** visualizza gli URL che sono presenti nella Sitemap XML, ma non sono stati scoperti durante la scansione evidenziando un potenziale problema di link interno.
- **Non. Indexable URLs in Sitemap:** URL che sono in una Sitemap XML, ma per qualche ragione non sono indicizzabili. In questo caso le ottimizzazioni possono essere essenzialmente due: la prima rimuovere dalla Sitemap le pagine o, se lo "status" "non-indexable" rappresenta un errore andare a sistemarlo.
- **URLs in Multiple Sitemaps:** URL che sono in più di una Sitemap XML. Potrebbe essere il caso di creare una sitemap index.
- **XML Sitemap with Over 50k URLs:** identifica le XML Sitemap che hanno più di 50k URLs in contraddizione con i limiti imposti dal Motore di Ricerca. La soluzione è quella di creare una "sitemap index" da inviare tramite Search Console a Google.
- **XML Sitemap with Over 50mb:** identifica ogni Sitemap XML che presenti una dimensione del file superiore a 50 Mb.

TAB PAGE SPEED

Oggigiorno le performance di un sito web sono vitali per il posizionamento organico nella Serp di Google, questa tendenza viene ancora di più avvalorata dai nuovi Core Web Vitals entrati tra i fattori di Ranking da maggio 2021.

La scheda "PageSpeed" del Seo Spider ti aiuta a collezionare i dati da due fonti:

- **PageSpeed Insights (PSI)** che utilizza Lighthouse per verificare la velocità delle pagine web mediante "dati di laboratorio";
- **Chrome User Experience Report (CrUX, o "dati sul campo")** che ti permette di ottenere dati reali di navigazione (ultimi 28 giorni).

Per popolare i dati della tab “PageSpeed” è necessario connettere le API di PSI.

Configuration > API > PageSpeed Insights

Dalla scheda Pagespeed hai a disposizione diverse metriche generiche:

- Total Size Savings.
- Total Time Savings.
- Total Requests.
- Total Page Size.
- HTML Size.
- HTML Count.
- Image Size.
- Image Count.
- CSS Size.
- CSS Count.
- JavaScript Size.
- JavaScript Count.
- Font Size.
- Font Count.
- Media Size.
- Media Count.
- Other Size.
- Other Count.
- Third Party Size.
- Third Party Count.

Ulteriori metriche più specifiche a disposizione sono divise fra “CrUX Metrics” che corrispondono ai dati sul campo di Chrome e “Lighthouse Metrix” con dati di laboratorio rilasciati da PageSpeed Insight.

CRUX METRICS

- CrUX Performance.
- CrUX First Contentful Paint Time (sec).
- CrUX First Contentful Paint Category.
- CrUX First Input Delay Time (sec).
- CrUX First Input Delay Category.
- CrUX Origin Performance.
- CrUX Origin First Contentful Paint Time (sec).
- CrUX Origin First Contentful Paint Category.
- CrUX Origin First Input Delay Time (sec).
- CrUX Origin First Input Delay Category.

LIGHTHOUSE METRICS:

- Performance Score.
- Time to First Byte (ms).
- First Contentful Paint Time (sec).
- Speed Index Time (sec).
- Time to Interactive (sec).
- First Contentful Paint Score.
- First Meaningful Paint Time (sec).
- First Meaningful Paint Score.
- Speed Index Score.
- Estimated Input Latency (ms).
- Estimated Input Latency Score.
- First CPU Idle (sec).
- First CPU Idle Score.
- Time to Interactive Score.

La Tab, oltre alle metriche comprende i seguenti filtri:

- **Eliminate Render-Blocking Resources:** sono visualizzate tutte le pagine con risorse che stanno bloccando il “First Paint” della pagina, insieme al potenziale risparmio.
- **Properly Size Images:** il filtro evidenzia tutte le pagine che contengono immagini non correttamente dimensionate ed evidenzia il potenziale risparmio di dati se venissero corrette.

▲ Properly size images 9.77 s ^

Serve images that are appropriately-sized to save cellular data and improve load time. [Learn more.](#)

☐ Show 3rd-party resources (0)

URL	Size	Potential Savings
 /flower_photo.jpg (correct-dimensions.glitch.me)	1,953 KB	1,918 KB
 /flower_logo.png (correct-dimensions.glitch.me)	71 KB	69 KB

Idealmente la tua pagina web non dovrebbe servire immagini più grandi della versione resa sullo schermo dell'utente. La strategia più consona è quella di servire “responsive images” generando più versioni e specificando quali usare nel tuo Html o CSS mediante le media queries, le dimensioni della viewport etc.

Altre soluzioni possono essere l'uso di CDN immagini o utilizzare formati di immagini vettoriali SVG per scalare a qualsiasi dimensione ad esempio le icone.

- **Defer Offscreen Images:** identifica tutte le pagine che contengono immagini nascoste o fuori dallo schermo assieme al potenziale risparmio. Per ovviare al problema puoi considerare il “Lazy-loading” di queste immagini dopo le risorse critiche per abbassare l'indice “Time to Interactive”.
- **Minify CSS:** tutte le pagine che presentano file CSS non minificati, insieme al risparmio potenziale se ottimizzati.

■ Minify CSS 0.15 s ^

Minifying CSS files can reduce network payload sizes. [Learn more.](#)

☐ Show 3rd-party resources (0)

URL	Size	Potential Savings
/styles/bootstrap.css (lh-perf-failures.glitch.me)	188 KB	35 KB

Minificare i CSS permette di migliorare il tempo di caricamento delle pagine perché normalmente i CSS utilizzati sono più grandi di quanto serve. I file CSS possono contenere caratteri non necessari, come ad esempio i commenti, gli spazi bianchi e l'indentazione.

Al momento della messa online, questi caratteri possono essere tranquillamente rimossi, per ridurre le dimensioni del file senza influenzare il modo in cui il browser elabora gli stili. Questa tecnica è chiamata minificazione.

```
/* Header background should match brand colors. */
h1 {
  background-color: #000000;
}
h2 {
  background-color: #000000;
}
```

Questo codice può venir minificato in una sola riga che permette al browser di utilizzare un numero di bytes inferiore.

```
h1,h2 { background-color: #000000; }
```

- **Minify JavaScript:** mostra tutte le pagine con file JavaScript non minificati, insieme al potenziale risparmio. Minificare i file JavaScript può ridurre le dimensioni del payload e il tempo di analisi dello script.
La minificazione è il processo di rimozione degli spazi bianchi e di qualsiasi codice che non è necessario per creare un file di codice più piccolo ma perfettamente valido. Ad esempio Terser è un popolare strumento di compressione JavaScript. Webpack v4 include di default un plugin per questa libreria per creare file di build minificati.
- **Remove Unused CSS:** identifica le pagine con CSS inutilizzati, insieme al potenziale risparmio in bytes. Di solito gli sviluppatori utilizzano <link> per aggiungere i fogli di stile di una pagina:

```
<!doctype html>
<html>
<head>
<link href="main.css" rel="stylesheet">
```

Il file main.css che il browser scarica è definito foglio di stile esterno e viene memorizzato separatamente dall'HTML che lo include nel codice. In questo modo ogni qual volta un browser incontra questa condizione dovrà scaricare, analizzare ed elaborare tutti i fogli di stile esterni che incontra prima di poter servire qualsiasi contenuto sullo schermo dell'utente.

- **Efficiently Encode Images:** il filtro visualizza le pagine con immagini non ottimizzate, insieme al potenziale risparmio. Lighthouse raccoglie tutte le immagini JPEG o BMP sulla pagina, imposta il livello di compressione di ogni immagine a 85, e poi confronta la versione originale con quella compressa. Se il risparmio potenziale è di 4 KiB o maggiore, Lighthouse segnala l'immagine come ottimizzabile.
- **Serve Images in Next-Gen Formats:** evidenzia le pagine con immagini servite in vecchi formati come jpeg e png. Il consiglio è quello di utilizzare formati come AVIF e WebP che hanno caratteristiche di compressione e qualità superiori.
- **Enable Text Compression:** tutte le pagine con risorse basate sul testo che non sono state compresse, insieme al potenziale risparmio.
- **Preconnect to Required Origin :** il filtro identifica tutte le pagine che presentano delle connessioni a domini diversi dal tuo ma non stanno ancora dando priorità alle richieste di fetch con link rel=preconnect, insieme al potenziale risparmio. Il Browser prima di richiedere una risorsa ad un server necessita di stabilire una connessione con lo stesso per cercare il nome dominio e definirlo in un indirizzo IP, per impostare una connessione al server e per crittografare la connessione. In tutti questi steps il browser invia dati e riceve risposte dal server, questo processo si definisce "Round Trip" e, in base alle condizioni della rete richiede del tempo che influisce inesorabilmente sulle performance.

- Per ottimizzare questo aspetto puoi avvalerti del “preconnect” delle risorse esterne (connessioni verso altri domini) che istruisce il browser su quali risorse connettersi in prima istanza perché considerate prioritarie.

```
<head>
<link rel=preconnect href="https://dominio.com">
</head>
```

Questi suggerimenti sono applicabili anche attraverso l’HTTP Header

```
Link: <https://dominio.com; rel=preconnect
```

ESEMPIO PRECONNECT

Un esempio di applicazione del “preconnect” potrebbe essere per Google Tag Manager o Google Analytics

```
Link: <https://www.google-analytics.com>; rel=preconnect; crossorigin; probability=1.0;
```

```
Link: <https://www.googletagmanager.com>; rel=preconnect; crossorigin; probability=1.0;
```

- **Reduce Server Response Times (TTFB):** la velocità di caricamento è essenziale per una user-experience adeguata e gli utenti sono molto sensibili su questo aspetto. Un sito che impiega molti secondi per caricarsi ha delle percentuali di abbandono direttamente proporzionali. Una delle cause più comuni è legata alle risposte lente dei server nel restituire il contenuto della pagina al browser. Questo filtro identifica tutte le pagine in cui il browser ha dovuto aspettare più di 600 ms per la risposta del server nella richiesta del documento principale.
- **Avoid Multiple Redirects:** visualizza le pagine che hanno risorse che reindirizzano, e il potenziale risparmio usando l’URL diretto. I reindirizzamenti rallentano la velocità di caricamento della pagina. Quando un browser richiede una risorsa che è stata reindirizzata, il server di solito restituisce una risposta HTTP come questa:

```
HTTP/1.1 301 Moved Permanently
Location: /path/to/new/location
```

e il browser è costretto a fare un’altra richiesta HTTP per recuperare la risorsa nella nuova posizione. Questa considerazione non vuole mettere al bando l’utilizzo dei 3xx ma farti evitare le catene di reindirizzamenti o reindirizzamenti multipli nella stessa pagina. Da evitare sempre i redirect nelle risorse necessarie al “Critical Rendering Path”.

- **Preload Key Requests:** identifica le pagine html che presentano elementi che potrebbero essere gestiti con il “preload”. Quando viene richiesta una pagina il browser richiede al server la pagina html analizzando il contenuto e inviando delle richieste distinte per ogni risorsa di riferimento inserita nel documento. Per migliorare le performance è consigliato accelerare questo processo richiedendo le risorse critiche in anticipo. Un esempio di applicazione potrebbe essere la richiesta di un font in un file CSS dato che il browser lo scoprirebbe solamente dopo la lettura del foglio di stile. Utilizzando la funzione “Preload” andrai a precaricare e recuperare una risorsa che il browser considererebbe solamente in una fase successiva ad esempio nelle critical chain request. Ricorda che il browser memorizza nella cache la risorsa in “preload” senza eseguire script ma avendoli già considerati li servirà immediatamente una volta richiesti. Questa funzionalità è molto indicata per dipendenze come JS, Font e CSS.

- Questa funzionalità è una direttiva per il browser che non blocca il caricamento del documento html. Per implementare la direttiva “Preload” è sufficiente aggiungere il tag <link> nella sezione <head> dell’html o utilizzare l’intestazione HTTP Link.

Sezione Head

```
<link rel="stylesheet" href="styles/main.css">
<link rel="preload" href="main.js" as="script">
```

HTTP Header

```
Link: </css/style.css>; rel="preload"; as="style"
```

- **Use Video Format for Animated Images:** identifica le pagine con GIF animate, insieme al potenziale risparmio se convertite in video.
- **Avoid Excessive DOM size:** identifica tutte le pagine che presentano una dimensione del DOM che supera i 1.500 nodi totali raccomandati. Considerare il DOM è di vitale importanza perché potrebbe influire negativamente sulle performance e sul rendimento in diversi modi:
 - spesso il DOM include molti nodi non visibili all’utente nel caricamento della pagina per la prima volta; questa condizione aumenta inutilmente il volume di dati da servire agli utenti e rallenta il tempo di caricamento (prestazioni di carico).
 - durante la navigazione ci sono numerose interazioni con la pagina web da parte degli utenti e degli script che portano il browser a riorganizzare la posizione e lo stile dei nodi del DOM creando rallentamenti per il rendering (prestazioni di runtime).
 - Se il JavaScript usa selettori di query generali come document.querySelectorAll(‘li’), potrebbe inconsapevolmente memorizzare riferimenti a un numero molto grande di nodi, il che può sovraccaricare le capacità di memoria dei dispositivi degli utenti (prestazioni di memoria).
- **Reduce Javascript Execution Time:** filtro che segnala le pagine con un tempo di esecuzione JavaScript medio o lento. Quando una pagina html impiega molto tempo nell’esecuzione degli script Javascript rallenta le prestazioni:
 - aumentando il costo della rete: utilizzando più byte aumenta il tempo di download.
 - Il Javascript viene analizzato e compilato nel thread principale che non rimane a disposizione per l’interazione dell’utente.
 - Occupando il thread principale Javascript si ritarda anche il TTI (Time to Interactive) penalizzando la Ux della pagina e il Core Web Vitals dedicato.
 - Se il Javascript contiene molti riferimenti potrebbe consumare inesorabilmente molta memoria penalizzando la consultazione della pagina da parte dell’utente.

Per migliorare le performance di esecuzione dei Javascript si consiglia in linea generale di “minificare” e comprimere il codice, rimuovere quello inutilizzato e memorizzarlo in cache.

- **Serve Static Assets With An Efficient Cache Policy:** restituisce tutte le pagine con risorse che non sono “cachate”, insieme al potenziale risparmio. Quando un browser richiede una risorsa, il server che fornisce la risorsa può dire al browser per quanto tempo dovrebbe memorizzare temporaneamente o mettere in cache la risorsa. Per ogni successiva richiesta di quella risorsa, il browser usa la sua copia locale piuttosto che ottenerla dalla rete. Nelle valutazioni di Lighthouse vengono considerati i font, i media, le immagini o fogli di stile che rispondono con status code 2xx e non hanno dichiarato in modo esplicito la “no-cache policy”.

Ecco un esempio di cache tramite intestazione HTTP:

```
Cache-Control: max-age=31536000
```

La direttiva “max-age” istruisce il browser su quanto tempo una risorsa rimarrà in cache. Per gli “static assets” si consiglia un tempo non inferiore ad un anno.

- **Minimize Main-Thread Work:** identifica tutte le pagine con tempi di esecuzione medi o lenti sul thread principale che penalizzano il rendering del documento html. Il processo di rendering del browser trasforma il tuo codice in una pagina web con cui gli utenti possono interagire. Per impostazione predefinita, il thread principale del processo di rendering gestisce la maggior parte del codice: analizza l'HTML e costruisce il DOM, analizza il CSS e applica gli stili specificati, ed analizza, valuta ed esegue il JavaScript.

Il thread principale elabora anche gli eventi utente. Quindi, ogni volta che il thread principale è occupato a fare qualcos'altro, la pagina web potrebbe non rispondere alle interazioni dell'utente, portando ad una cattiva esperienza. Per migliorare questo processo si consiglia di ottimizzare javascript di terze parti, eliminare codice non utilizzato, ridurre i riferimenti CSS e semplificarli (minificarli), posticipare i CSS non critici, splittare il javascript payloads.

- **Ensure Text Remains Visible During Webfont Load:** tutte le pagine con font che possono lampeggiare o diventare invisibili durante il caricamento della risorsa.

ESPORTARE I DATI SULLE PERFORMANCE

Tutte i dati relativi alle performance, le pagine sorgente e gli URL delle risorse che potrebbero essere ottimizzate li puoi esportare in blocco tramite il menu PageSpeed.

Reports > PageSpeed

The screenshot shows the Screaming Frog SEO Spider interface. The 'Reports' menu is open, and the 'PageSpeed' option is selected. The main window displays a table of URLs and their performance scores. The 'PageSpeed Opportunities Summary' dropdown menu is open, showing a list of optimization opportunities.

URL	Page Titles	Meta Description	Meta Keywords	H1	H2	Images	Canonicals
1	https://www.bbc.co.uk/programmes/articles/4...						
2	https://www.bbc.co.uk/sounds/play/p07m86k...						
3	https://www.bbc.co.uk/arts						
4	https://www.bbc.co.uk/bitesize/articles/zjnm...						
5	https://www.bbc.co.uk/programmes/p07krhh...						
6	https://www.bbc.co.uk/programmes/p07ngkf...						
7	https://www.bbc.co.uk/programmes/p07nhm...						
8	https://www.bbc.co.uk/news/uk-england-glo...						
9	https://www.bbc.co.uk/bitesize						
10	https://www.bbc.co.uk/news/science-environment-49692583						
11	https://www.bbc.co.uk/news/entertainment-arts-49716168						
12	https://www.bbc.co.uk/news/uk-politics-49722087						
13	https://www.bbc.co.uk/news/science-environment-49692974						
14	https://www.bbc.co.uk/food						
15	https://www.bbc.co.uk/food/collections/quick_dinners_on_toast						
16	https://www.bbc.co.uk/news/world-us-canada-49722926						
17	https://www.bbc.co.uk/food/collections/noodles_for_students						
18	https://www.bbc.co.uk/news/technology-49716778						
19	https://www.bbc.co.uk/news/world-europe-49716978						
20	https://www.bbc.co.uk/news/newsbeat-49714497						

The 'PageSpeed Opportunities Summary' dropdown menu shows the following items:

- PageSpeed Opportunities Summary
- CSS Coverage Summary
- Eliminate Render-Blocking Resources
- Properly Size Images
- Defer Offscreen Images
- Minify CSS
- Minify Javascript
- Remove Unused CSS
- Efficiently Encode Images
- Serve Images in Next-Gen Formats
- Enable Text Compression
- Preload Key Requests
- Use Video Formats for Animated Content
- Avoid Excessive DOM Size
- JavaScript Execution Time
- Serve Static Assets with an Efficient Cache Policy
- Minimize Main-Thread Work
- Ensure Text Remains Visible During Webfont Load

TAB CUSTOM SEARCH

La scheda di ricerca personalizzata ti permette di verificare la presenza o meno di determinati elementi nel codice sorgente delle pagine HTML.

Config > Custom > Search

Per gestire le ricerche puoi applicare fino a 100 filtri utilizzando regole RegEX e ottenere un risultato booleano con le pagine che “contengono” o “non contengono” l’input scelto.

I risultati della tua ricerca conterranno i seguenti dati:

- **Address:** identifica l’indirizzo del sottodominio scansionato;
- **Content:** la tipologia di contenuto rilevato nell’Url considerato;
- **Status Code:** lo status code della risorsa.
- **Status:** il codice di risposta dell’intestazione HTTP.
- **Contains: [x]:** il numero di occorrenze [x] in ogni URL in cui è stato trovato l’input utilizzato in fase di ricerca.
- **Does Not Contain [y]:** identifica se l’input ricercato è stato trovato o meno durante la “custom search”.

Nella sezione filtri della scheda è riportato il nome scelto per ogni singola “Custom Search”.

TAB CUSTOM EXTRACTION

Attraverso questa funzione sei in grado di configurare fino a 100 “Custom Extraction” attraverso XPath, CSSPath o RegEX. L’estrazione viene eseguita solo su URL con contenuto HTML.

I risultati saranno mostrati seguendo queste dimensioni:

- **Address:** identifica l’indirizzo del sottodominio scansionato.
- **Content:** la tipologia di contenuto rilevato nell’Url considerato.
- **Status Code:** lo status code della risorsa.
- **Status:** il codice di risposta dell’intestazione HTTP.

Nella sezione filtri della scheda è riportato il nome scelto per ogni singola “Custom Extraction”.

TAB ANALYTICS

Per popolare questa scheda devi connetterti alle API di Google Analytics.

Configurazione > API > Google Analytics

Per default sono già configurate 10 metriche ma puoi personalizzare l’analisi selezionando fino ad un massimo di 30.

La scheda “Analytics” comprende i seguenti filtri:

- **Sessions Above 0:** identifica gli URL che presentano almeno 1 sessione in Google Analytics.
- **Bounce Rate Above 70%:** vengono identificate tutte le pagine che presentano un tasso di abbandono (Bounce Rate) che supera il 70%.

- **No GA Data:** identifica tutte le pagine html per cui, l'API di Google non ha restituito alcun dato in termini di metriche e dimensioni impostate.
- **Non-Indexable with GA Data:** mostra le pagine html che sono gestite come risorse non indicizzabili (non-indexable), ma ugualmente per qualche ragione hanno delle occorrenze su Google Analytics.
- **Orphan URLs:** colonna popolata da quegli URL che sono stati scoperti tramite Google Analytics, ma non presentano link interni e non sono stati scovati durante una scansione. Questo filtro richiede che l'opzione "Crawl New URLs Discovered In Google Analytics" sia abilitata e necessita della "Crawl Analysis".

TAB SEARCH CONSOLE

Mediante la scheda "Search Console" sei in grado di includere i dati più significativi relativi a click, Impression, CTR e "Position" della console di Google.

Configurazione > Accesso API > Google Search Console

Per popolare questa tab è necessario connettersi tramite API.

I filtri a disposizione sono:

- **Click above 0:** identifica gli url che hanno ottenuto almeno 1 click nei risultati di Google.
- **No GSC Data:** significa che le API non hanno restituito alcun risultato per l'url scansionato.
- **Non-Indexable with GSC Data:** pagine html che nonostante abbiano come status "non indexable" hanno ottenuto dei risultati in termini di Impression o click in Search Console.
- **Orphan URLs:** definisce gli url che nonostante non siano stati scoperti con il crawl del sito web attraverso il link interno sono presenti nei dati della Search Console. Per essere utilizzato il filtro devi abilitare "Crawl New URLs Discovered in GSC" ed eseguire la "Crawl Analysis" al termine della scansione.

TAB LINK METRICS

Screaming Frog permette di connettersi tramite API ad alcuni servizi di analisi esterni come Majestic, Ahrefs e Moz.

Configuration > API

MAJESTIC

Connettendosi a Majestic è possibile ottenere le seguenti metriche:

- External Backlinks.
- Referring Domains: mostra i domini che puntano verso il sito analizzato.
- Trust Flow: corrisponde al punteggio di qualità elaborato da Majestic, su una scala tra 0-100.
- Citation Flow.
- Referring IPs: identifica gli IP di riferimento che contengono almeno un sito web che contiene un link al sito analizzato.
- Referring Subnets: abbreviazione per "Referring C-subnets".
- Indexed URLs.
- External Backlinks EDU.
- External Backlinks GOV.
- Referring Domains EDU: mostra i domini con estensione EDU (education) di riferimento che puntano al sito analizzato.
- Referring Domains GOV: mostra i domini con estensione governativa di riferimento, noti anche come "ref domain" che puntano al sito analizzato.

- Trust Flow Topics: identifica la categorizzazione dei siti web e assegna un punteggio con scala logaritmica da 0 a 100 mostrando l'influenza di una pagina web, sottodominio o dominio rispetto ad un argomento principale o categoria.
- Anchor Text.

Guida ufficiale: <https://majestic.com/help/glossary>

AHREFS METRICS

Connettendosi ad Ahrefs è possibile ottenere le successive metriche:

- Backlinks: definisce i backlink e i rispettivi dettagli delle pagine di riferimento, come gli anchor e titolo della pagina.
- RefDomains: definisce i domini referral puntati verso il dominio analizzato.
- URL Rating.
- RefPages.
- Pages: elenco delle pagine crawlerate.
- Text.
- Image.
- Site Wide.
- Not Site Wide.
- NoFollow.
- DoFollow.
- Redirect.
- Canonical.
- Gov.
- Edu.
- HTML Pages.
- Links Internal.
- Links External.
- Ref Class C.
- Refips: mostra tutti gli indirizzi IP dei referral che presentano almeno un backlinks verso il sito analizzato.
- Linked Root Domains.
- Twitter.
- Pinterest.
- GPlus.
- Facebook.
- Facebook Likes.
- Facebook Shares.
- Facebook Comments.
- Facebook Clicks.
- Facebook Comments Box.
- Total Shares.
- Medium Shares.
- Keywords.
- Keywords Top 3.
- Keywords Top 10.
- Traffic.
- Traffic Top 3.
- Traffic Top 10.
- Value.
- Value Top 3.
- Value Top 10.

Guida ufficiale: <https://ahrefs.com/api/documentation>

MOZ METRICS

Connettendosi ad Moz è possibile ottenere le successive metriche:

- Page Authority.
- MozRank.
- MozRank External Equity.
- MozRank Combined.
- MozTrust.
- Time Last Crawled (GMT).
- Total Links (Internal or External).
- External Equity-Passing Links.
- Total Equity-Passing Links (internal or External).
- Subdomains Linking.
- Total Linking Root Domains.
- Total External Links.
- Spam Score.
- Links to Subdomain.
- Root Domains Linking to Subdomain.
- External Links to Subdomain.
- Domain Authority.
- Root Domains Linking.
- Linking C Blocks.
- Links to Root Domain.
- External Links to Root Domain.

Guida ufficiale: <https://moz.com/help/links-api>

URL DETAILS

La tab "Url Details" è presente nella finestra inferiore del Seo Spider e viene popolata nel momento in cui si evidenzia uno o più URL nella schermata superiore. Questa scheda ti permette una panoramica esauritiva dell'Url analizzato e riprende una selezione delle voci della scheda "Internal" tra cui:

- **URL:** identifica l'Url selezionato.
- **Status Code:** la risposta del server dell'Url selezionato.
- **Status:** la risposta dell'header HTTP.
- **Content:** mostra la tipologia di contenuto dell'URL.
- **Size:** colleziona la dimensione del file o della pagina web scansionata.
- **Crawl Depth:** identifica la profondità della pagina in esame in corrispondenza della sua lontananza in ordine di click dalla homepage o dalla pagina iniziale.
- **Inlinks:** numero complessivo degli inlinks interni all'URL.
- **Outlink:** numero di outlink interni dall'URL.

INLINKS

La scheda "Inlinks" contiene la lista completa dei link interni (internal links) che puntano all'URL analizzato. Anche questa tab viene popolata dopo aver selezionato uno o più URL dalla finestra superiore del Seo Spider.

- **Type:** corrisponde alla tipologia di URL scansionato (Hyperlinks -collegamento ipertestuale, JavaScript, CSS, immagine ecc.).
- **From:** identifica l'URL della pagina di riferimento di un collegamento..
- **To:** identifica la destinazione del link presente nella colonna "From" (sorgente) e corrisponde all'url selezionato nella finestra superiore.
- **Anchor Text:** riporta il testo dell'ancora o del link utilizzato, se presente.
- **Alt Text:** se presente riporta l'attributo alt.
- **Follow:** risponde in modo booleano restituendo 'True' se il link è "follow" o "False" se il link contiene un attributo 'nofollow', 'UGC' o 'sponsored'.
- **Target:** identifica gli attributi associati di destinazione come _blank, _self, _parent.
- **Rel:** identifica gli attributi della colonna "Follow" restituendo la motivazione "nofollow", "sponsored" o "ugc".
- **Path Type:** definisce se il link è assoluto, relativo al protocollo, relativo alla radice o relativo al percorso.
- **Link Path:** mostra l'XPath dell'elemento dettagliando la posizione del link all'interno della pagina.
- **Link Position:** restituisce la posizione del link nel codice html (Head, Nav, Footer ecc.) favorendo eventuali controlli ed ottimizzazioni.

OUTLINKS

Questa scheda contiene una lista di link che hanno come sorgente l'url selezionato nella finestra superiore e puntano ad altre pagine del sito web o verso altri sottodomini esterni.
Le colonne della tab seguono la stessa logica della scheda "Inlinks" con unica eccezione tra "from" e "to" che presentano valori invertiti.

IMAGE DETAILS

La consultazione di questa scheda ti permette di isolare tutte le immagini presenti in una pagina e nel caso selezioni l'url di un'immagine di ottenere l'anteprima.

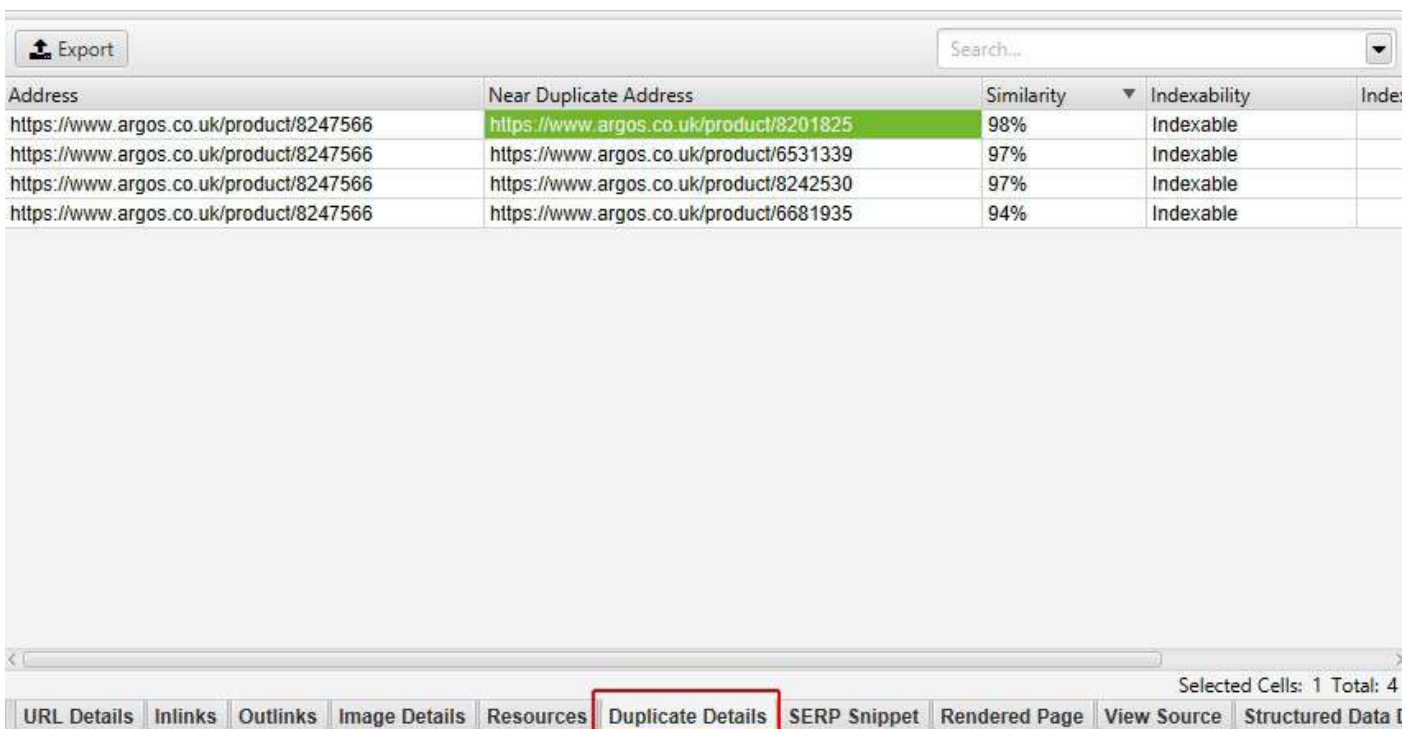
I dati che puoi ottenere sono:

- **From:** mostra l'url selezionato.
- **To:** il riferimento delle immagini trovate nella pagina html.
- **Alt Text:** mostra il testo dell'attributo "alt" se presente nel codice.

DUPLICATE DETAILS

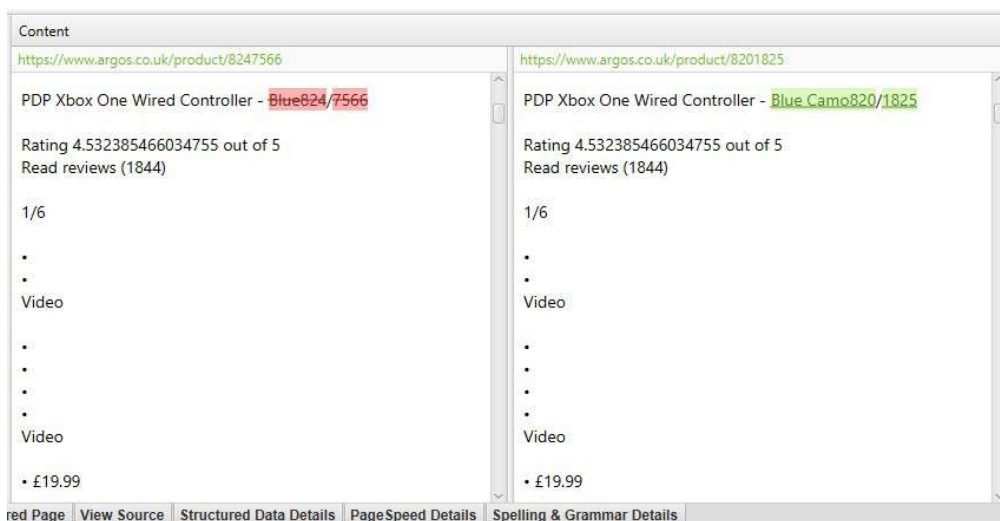
Con la consultazione di questa scheda è possibile isolare eventuali pagine con contenuti duplicati o molto simili rispetto all'url selezionato.

Questa scheda mostra quali siano eventuali URL simili e definisce il grado di similarità con una scala percentuale.



Address	Near Duplicate Address	Similarity	Indexability	Inde:
https://www.argos.co.uk/product/8247566	https://www.argos.co.uk/product/8201825	98%	Indexable	
https://www.argos.co.uk/product/8247566	https://www.argos.co.uk/product/6531339	97%	Indexable	
https://www.argos.co.uk/product/8247566	https://www.argos.co.uk/product/8242530	97%	Indexable	
https://www.argos.co.uk/product/8247566	https://www.argos.co.uk/product/6681935	94%	Indexable	

Una volta collezionati i dati, se clicchi su un "Indirizzo quasi duplicato" (Near Duplicate Address) della scheda puoi visualizzare un confronto fra le pagine con le relative differenze evidenziate.



TAB RESOURCES

Questa scheda presenta tutta una lista di risorse trovate nell'url selezionato evidenziando la tipologia (javascript, CSS etc.), il link degli elementi, il Link Path e il Path Type e molti altri dettagli utili al tuo Seo Audit.

TAB SERP

La scheda SERP Snippet è un utile alleato per vedere come lo snippet dell'URL è visualizzato nei risultati di ricerca di Google. Attraverso questa tab è possibile ottimizzare lo snippet evitando troncature nei risultati del Motore di Ricerca avendo una visione adeguata sia a lato desktop che mobile. Il punto di troncamento (dove Google mostra (...) e taglia le parole) è calcolato in base alla larghezza dei pixel e dei caratteri.

Element	Chars			Pixels	
	Length	Displayed	Truncated	Length	Available
Title	49	49	0	475	545
Description	152	152	0	996	1010

Sulla sinistra della tab è visualizzata l'anteprima dello snippet, mentre sulla parte destra una tabella panoramica. Il Motore di Ricerca non ha mai definito ufficialmente le dimensioni in pixel o il numero di caratteri massimi da poter utilizzare senza ricadere in troncature, quindi queste soglie sono state definite dal Seo Spider in base a vari studi sulle SERP.

L'emulatore di snippet della SERP utilizza come impostazione predefinita il desktop.

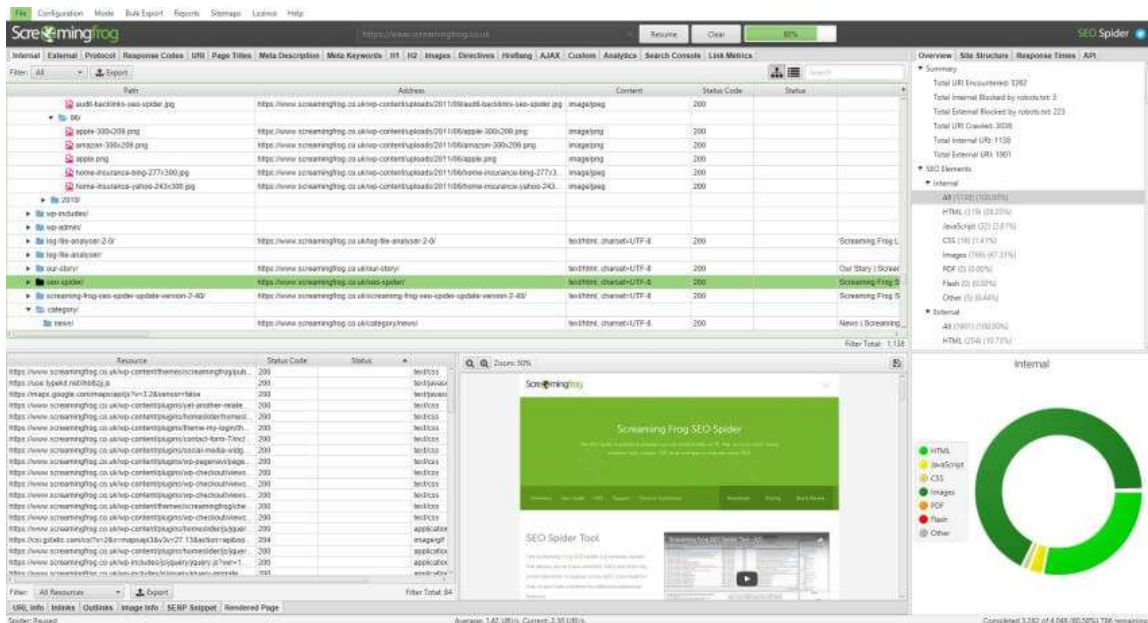
Config > Spider > Preferences

Puoi cambiare il tipo di devices da simulare per vedere come gli snippet possono apparire in versione desktop, tablet o mobile switchando dal menu a tendina.

Molto comoda l'interfaccia che permette di modificare i titoli e le descrizioni delle pagine direttamente nella finestra salvando le modifiche automaticamente (Database Storage), a meno che non clicchi sul pulsante "ripristina titolo e descrizione" che lo riporta allo stato originale. Questa funzionalità ti permette di perfezionare i tuoi snippet ed esportare il documento ('Reports > SERP Summary') per inviarlo ad un cliente o al team di sviluppo per integrare i tuoi meta tag sul sito web.

RENDERED PAGE

La scheda ti permette di vedere la pagina renderizzata che il SEO Spider ha scansionato in modalità JavaScript.



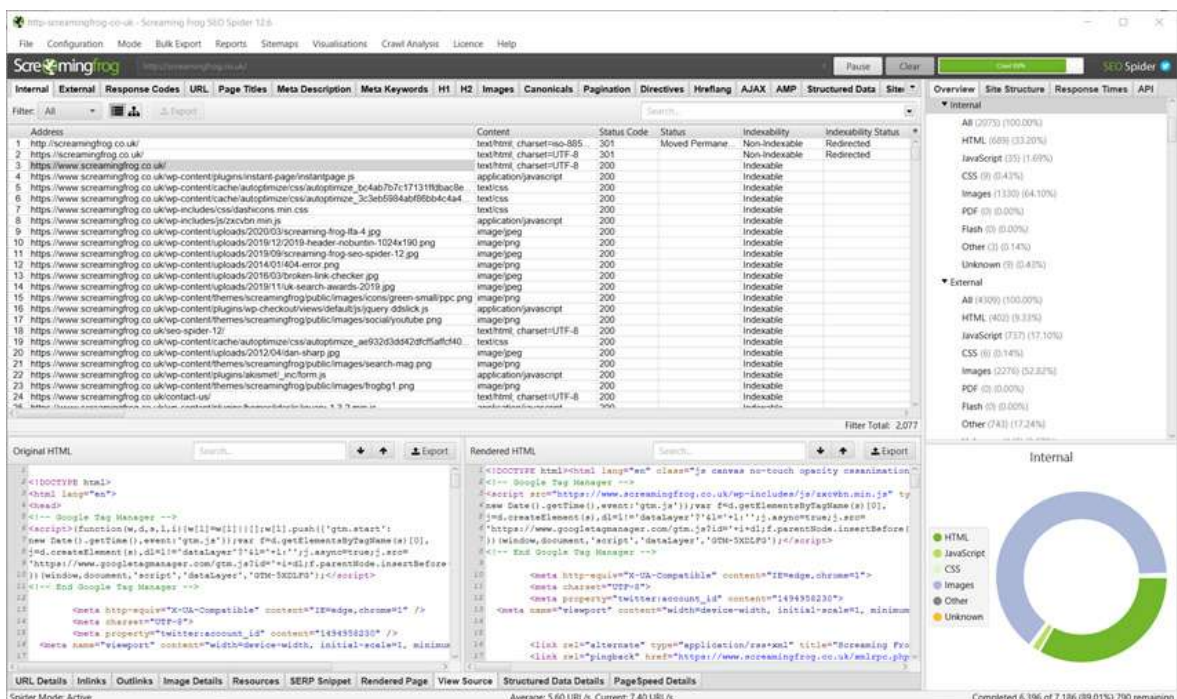
Questa caratteristica è abilitata di default quando si usa la funzionalità di rendering JavaScript, e funziona insieme all'agent-user configurato, al timeout AJAX e alla dimensione della viewport. Questa funzionalità è impostata di default su "risorse bloccate", ma questo può anche essere cambiato per mostrare "tutte le risorse" usate dalla pagina.

VIEW SOURCE

La scheda "View Source" permette di collezionare e confrontare i dati dell'HTML statico e l'HTML renderizzato di una pagina selezionata dalla finestra superiore; questa funzionalità è subordinata all'abilitazione di "Store HTML" o "Store Rendered HTML".

Configurazione > Spider > Extraction > Store/HTML / Store HTML Rendered

La tab prevede un filtro di ricerca interno per rendere agevole l'individuazione degli elementi nel codice.



La schermata di sinistra ti permette di visualizzare l'HTML originale prima che entri in gioco il JavaScript, allo stesso modo di un click destro su "visualizza sorgente" in un browser. Questo è un ottimo strumento per il debug, o per il confronto con l'HTML renderizzato.

L'anteprima di sinistra ti permette di salvare l'HTML renderizzato di ogni URL scansionato dal SEO Spider su disco e visualizzare il DOM come fosse "Inspect Element" (in Chrome in DevTools), dopo che il JavaScript è stato elaborato.

TAB HTTP HEADERS

Questa tab colleziona i dati relativi alle risposte dell'HTTP Header.

Configuration > Spider > Extraction > HTTP Header

Nel lato sinistro della scheda puoi vedere le intestazioni della richiesta HTTP mentre nel lato destro le intestazioni di risposta HTTP.

Le colonne a tua disposizione comprendono:

- **Header Name:** il nome dell'intestazione di risposta dal server.
- **Header Value:** Il valore dell'intestazione di risposta dal server.

Una volta estratti, gli header HTTP sono aggiunti in colonne nella scheda Internal, dove potranno essere interrogati insieme ai dati di crawl.

Il riepilogo delle intestazioni HTTP può essere esportato in modo analitico tramite Bulk Export > Web > All HTTP Headers" o in modo aggregato mediante "Rapporti > Intestazioni HTTP > Riepilogo intestazioni HTTP".

TAB COOKIES

Per popolare questa Tab ricorda di abilitare il Seo Spider alla scansione e memorizzazione dei Cookies. Una volta che selezioni uno o più URL della finestra superiore troverai i dettagli nella Tab Cookies della schermata inferiore.

Configuration > Spider > Extract > Cookie

Puoi anche utilizzare la modalità di scansione "Rendering JavaScript" per ottenere una visione accurata dei cookie che sono stati caricati sulla pagina utilizzando JavaScript o tag immagine pixel.

Le colonne della scheda Cookie includono:

- **Cookie Type:** identifica dove il cookie è stato scoperto('HTTP', 'On-Page' etc).
- **Cookie Name:** colleziona il nome del cookie.
- **Cookie Value:** definisce il valore del cookie.
- **Domain:** mostra il dominio che ha emesso il cookie di prima o di terza parte.
- **Expiration Time:** indica il tempo di scadenza del cookie.
- **Secure:** visualizza i dettagli dell'attributo di sicurezza del cookie.
- **HttpOnly:** mostra i dettagli dell'attributo HttpOnly del cookie.
- **Address:** visualizza l'URL su cui è stato trovato il cookie.

Bulk Export > Web > All Cookies

STRUCTURED DATA

Configuration > Spider > Extraction > JSON-LD/Microdata/RDFa & Schema.org Validation/Google Validation

Nel lato sinistro puoi vedere i valori delle proprietà con relative icone nel caso di errori e/o avvisi; ti basterà cliccare sulle risorse con errori per ottenere i dettagli specifici di convalida nella finestra di destra.

- # KILL SEO VOLUME 1

PAGESPEED DETAILS

Con l'introduzione dei Web Vitals le performance del sito web è diventato ancora più importante e deve essere ottimizzato per garantire un buon posizionamento organico. La scheda "PageSpeed" permette una panoramica molto completa ed esaustiva di eventuali criticità che potrebbero penalizzare il progetto. Per popolare i dati di questa scheda è necessario connettersi all'API di PageSpeed Insights attraverso "Configuration > API Access > PageSpeed Insights", inserire una chiave API PageSpeed gratuita, connettersi ed eseguire un crawl.

Configuration > API Access > PageSpeed Insights

The screenshot shows the Screaming Frog SEO Spider 12.8 interface. The main table displays crawl results for various URLs, including metrics like Errors, Warnings, Total Types, Unique Types, and Type-1 through Type-6. The left sidebar shows a tree view of metrics, including Overview, CrUX Metrics, Lighthouse Metrics, and Opportunities. The right sidebar shows a summary of issues, including Missing, Validation Errors, Parse Errors, Microdata URLs, JSON-LD URLs, RDFa URLs, Sitemaps, and PageSpeed. A bar chart titled 'Structured Data' is also visible in the bottom right corner.

Consultando la finestra di sinistra puoi ottenere informazioni specifiche sia sulle metriche estratte che sulle opportunità disponibili specifiche per l'URL evidenziato. Cliccando su un'opportunità vengono visualizzati maggiori dettagli nella schermata di destra.

- **Source Page:** identifica l'URL selezionato nella finestra superiore.
- **URL:** la risorsa collegata che ha opportunità disponibili.
- **Size (Bytes):** visualizza la dimensione attuale della risorsa elencata.
- **Potential Savings:** mostra il risparmio potenziale di peso (bytes) se venisse implementata l'opportunità proposta.

TAB SPELLING & GRAMMAR DETAILS

Questa tab, una volta abilitata la relativa opzione dal menu "config", contiene i dettagli su qualsiasi problema di ortografia e grammatica per ogni URL selezionato nella finestra superiore del Seo Spider.

La scheda "Tab Spelling & Grammar Details" visualizza la lista di errori e se si tratta di errori ortografici o grammaticali fornendo suggerimenti e consigli per correggere il problema.

La schermata di destra permette l'anteprima del testo della pagina e degli errori identificati.

Per default il vocabolario impostato è quello inglese ma puoi personalizzare l'analisi con il vocabolario della tua lingua.

Config > Content > Spelling&Grammar

The screenshot displays the Screaming Frog SEO Spider application. The main window shows a list of crawled URLs with associated error counts and status. The right sidebar provides an overview of site statistics, including duplicates, content quality, and spelling/grammar errors. The bottom section features a 'Content Marketing' tip and a bar chart visualizing content distribution.

OVERVIEW

La scheda panoramica “Overview” che trovi nella sidebar destra del Seo Spider si aggiorna in tempo reale durante la scansione e fornisce un utile riepilogo sugli URL e i dati complessivi di ogni scheda e filtro.

- **Summary:** corrisponde al riepilogo degli URL incontrati durante la scansione.
- **SEO Elements (Crawl Data):** definisce in termini assoluti e percentuali il numero di URL trovati per ogni tab di primo livello e del rispettivo filtro. Ti permette una prima analisi aggregata utilissima per evidenziare da subito le diverse aree su cui serve un'ottimizzazione senza dover cliccare singolarmente all'interno delle schede e dei filtri. Indicata anche come scorciatoia per trovare l'origine delle criticità, ti basterà cliccare sulle singole voci per aprire la scheda dedicata nella finestra principale del Seo Spider.

SITE STRUCTURE

Il Motore di Ricerca è molto avanzato e raggiunge in autonomia tutte le risorse del tuo sito web seguendo i collegamenti interni ma, al contempo, una buona architettura del sito permette allo Spider e agli utenti di avere un'esperienza di navigazione più soddisfacente. La struttura logica di un sito assieme alla definizione del “Search Intent” permette notevoli benefici a livello di posizionamento organico e questa scheda diviene un formidabile punto di partenza per ottimizzare al meglio il tuo progetto. Nella parte superiore puoi visualizzare in tempo reale il percorso, il numero totale di URL, gli URL indicizzabili e non indicizzabili in ogni percorso del sito web.

- **Path:** mostra il percorso dell'URL del sito web sottoposto a scansione.
- **URL:** il numero totale di URL unici trovati all'interno del percorso.
- **Indexable:** il numero totale di URL indicizzabili trovati all'interno del percorso.
- **Non-Indexable:** mostra il numero totale di URL non indicizzabili trovati all'interno del percorso.

Per una visione più oculata puoi regolare la “vista” della Struttura del sito valutando alternativamente lo “Stato di indicizzabilità”, i “Codici di risposta” e la “Profondità di scansione” degli URL in ogni percorso o scegliere il livello di “crawl depth” cliccando sul “+” o “-” dell’opzione “Level”.

Contestualmente nella parte inferiore della sidebar trovi un grafico (o tabella) con le indicazioni relative al numero di URL per singola profondità di scansione tra 1-10+ già filtrato in base ai codici di risposta.

- **Depth (Clicks from Start URL):** identifica la profondità della pagina in base al numero di click necessari per raggiungerla dalla homepage o dalla pagina iniziale di scansione.
- **Number of URLs:** visualizza il numero di URL per singola profondità.
- **% of Total:** mostra il numero di URL per singolo livello di profondità in forma percentuale.

RESPONSE TIMES

Questa tab ti permette una visione d’insieme dei tempi di risposta degli URL durante una scansione.

- **Response Times:** identifica degli intervalli di tempo (in secondi) per scaricare gli URL.
- **Number of URLs:** numero complessivo di URL in uno specifico intervallo di tempo di risposta.
- **% of Total:** percentuale di URL in un particolare intervallo di tempo di risposta.

Questi dati corrispondono alla relazione tra il tempo necessario per emettere una richiesta HTTP e ottenere la risposta HTTP completa dal server.

Nota Bene: questi dati potrebbero venire influenzati dal carico del server e dall’attività di rete del client al momento della richiesta.

Il “Response Times” indicato dallo Seo Spider non include il tempo impiegato per scaricare risorse aggiuntive quando si è in modalità di “JavaScript Rendered”. Ogni risorsa apparirà con il suo tempo di risposta individuale.

Per un’analisi completa delle performance, ti raccomando l’integrazione alle API di PageSpeed Insights.

API

Se hai connesso il Seo Spider alle API di Google Analytics, Search Console, Majestic, Ahrefs o Moz con questa scheda visualizzi il progresso nella raccolta dati da tutte le API insieme al conteggio individuale degli errori.

SPELLING & GRAMMAR

La scheda “Spelling & Grammar” ti permette una panoramica dei primi 100 errori grammaticali o di ortografia scoperti durante la scansione e il numero di URL interessati. Molto utile in fase di prima analisi per scovare errori nei template.

COME FARE UN CRAWLING DI UN SITO PROTETTO DA PASSWORD

Nella vita di un Seo capita spesso di dover analizzare siti web in staging o dover scansionare siti privati che necessitano di password per la visualizzazione e scansione.

Per accompagnarti attraverso questa tipologia di diagnosi ti viene in aiuto Screaming Frog che, dalla versione 7.0 del Seo Spider, ha introdotto la funzione “web form authentication”.

Prima di procedere con il crawl però vediamo quali siano le potenziali controindicazioni di questa funzione e come risolverle per una scansione a prova di errore.

1. Gestire le credenziali per il Seo Spider

Questa nuova opzione è molto potente ma deve essere gestita con assoluta responsabilità in quanto, se lo Spider viene autenticato come admin del sito web, farà click su ogni link presente nelle pagine ed essendo loggato i click potrebbero anche interessare il logout, la creazione di articoli, installazione plugin o nel caso più estremo la cancellazione dei dati.

Per questo motivo è fondamentale predisporre l'autenticazione con delle limitazioni nelle credenziali, lasciando permessi di sola lettura e senza alcun permesso di “scrittura sul sito”.

Se analizzi CMS comuni come Wordpress, Drupal o Joomla l'idea migliore è quella di creare un utente dedicato con permessi da “sottoscrittore”.

2. Utilizzare la funzione “Exclude”

È buona norma anche utilizzare la funzione “Exclude” su cartelle che potrebbero modificare il sito in produzione o privato; nel caso di wordpress potresti escludere dal crawling qualsiasi elemento in “/wp-admin/”.

Per escludere ad esempio la cartella dedicata all'admin e magari il logout puoi predisporre le seguenti espressioni RegEX per il Seo Spider.

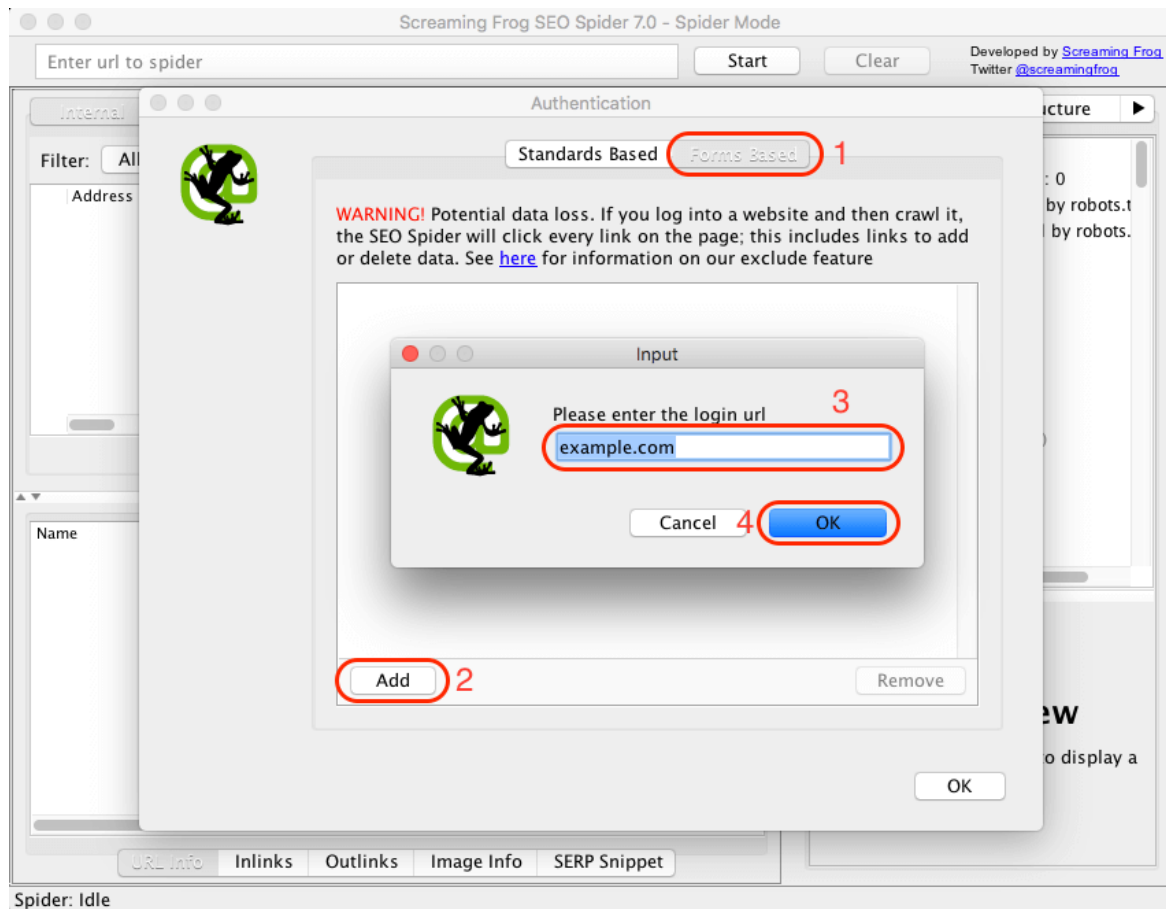
```
http://example.com/wp-login\.php\?action=logout.*  
http://example.com/wp-admin/*
```



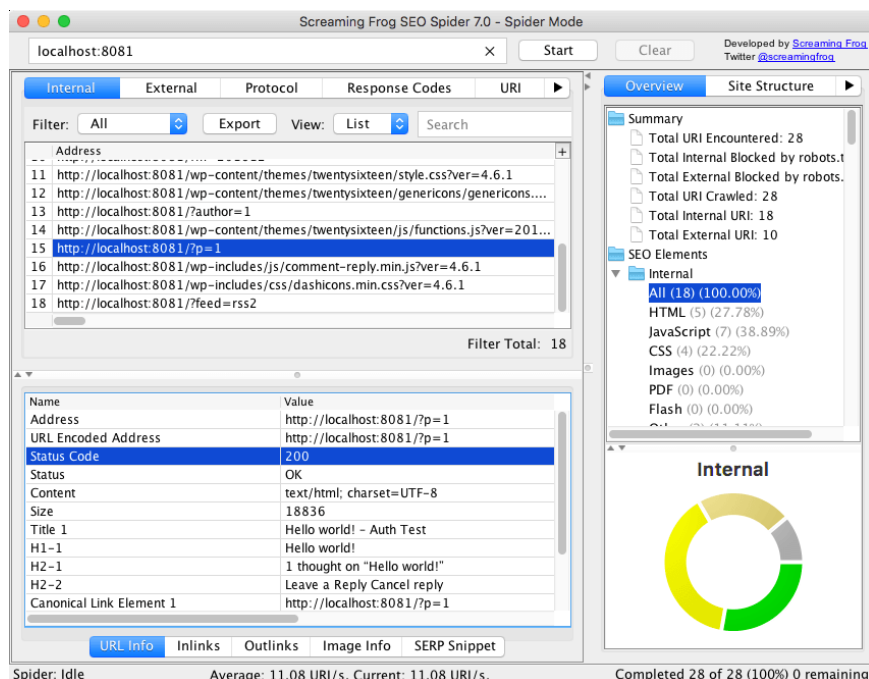
3. Configurare le Autenticazioni

Una volta messo in sicurezza il sito da analizzare e configurato le esclusioni, puoi accedere al sito web selezionando “Authentication” dal menu principale (Configuration), scegliendo l’opzione “Form Based” e cliccando sul pulsante “Aggiungi”.

Andando avanti con la procedura ti sarà richiesta l’immissione dell’URL da scansionare e un browser si aprirà automaticamente permettendoti di inserire le tue credenziali.



Attraverso questa opzione sarai in grado ad esempio di scansionare e correggere tutte le criticità prima della messa online riscontrate ma, come detto ad inizio dell’articolo, devi porre la massima attenzione alla sicurezza del sito web ed evitare scorribande incontrollate del crawl!!



COME ANALIZZARE LA POSIZIONE DEI LINKS

La distribuzione corretta dei links e il loro posizionamento sono fattori importanti per il ranking del sito web e per garantire allo Spider del Motore di Ricerca un'esperienza di navigazione adeguata. Conoscendo la posizione dei collegamenti è più semplice identificare eventuali "link corrotti" redirect etc. e al contempo valutare se i link verso le pagine più importanti abbiano la giusta rilevanza (un link nel menu in alto del sito web ha un valore superiore rispetto ad uno nascosto nel testo in base al presupposto che Google analizza dall'alto verso il basso).

Attraverso questo approfondimento puoi analizzare la posizione dei link trovati durante un crawl utilizzando Screaming Frog SEO Spider.

La "Link Position" si riferisce alle diverse posizioni dei link su una pagina; ad esempio nel menu di navigazione, nella barra laterale, nel contenuto principale o nel piè di pagina.

Il Seo Spider classifica i collegamenti utilizzando elementi HTML5 semantici come `<header>`, `<nav>`, `<footer>` (o elementi non semantici ben definiti, come `div. id="nav"`) per determinare quali siano le diverse aree in cui sono inseriti.



Il tool non esegue il rendering della pagina web per analizzare dove appare un link ma usa una classificazione che si basa sui comuni tag HTML. Nel caso alcuni siti non utilizzino elementi HTML5 semantici o HTML facilmente identificabili, sarai in grado di configurare una classificazione ad hoc tramite la funzionalità specifica del Seo Spider.

Perchè è importante conoscere la posizione dei link?

1. Identificare e correggere criticità dei link

Sapere dove si trova un link può essere ideale per trovare e sistemare link problematici, come link rotti, jump-link, redirect ecc.

Questa condizione permette anche di stabilire una scala di priorità sui link da ottimizzare; per esempio un "sitewide link" nella navigazione principale potrebbe essere considerato un po' più importante di un link all'interno del contenuto in una pagina secondaria.

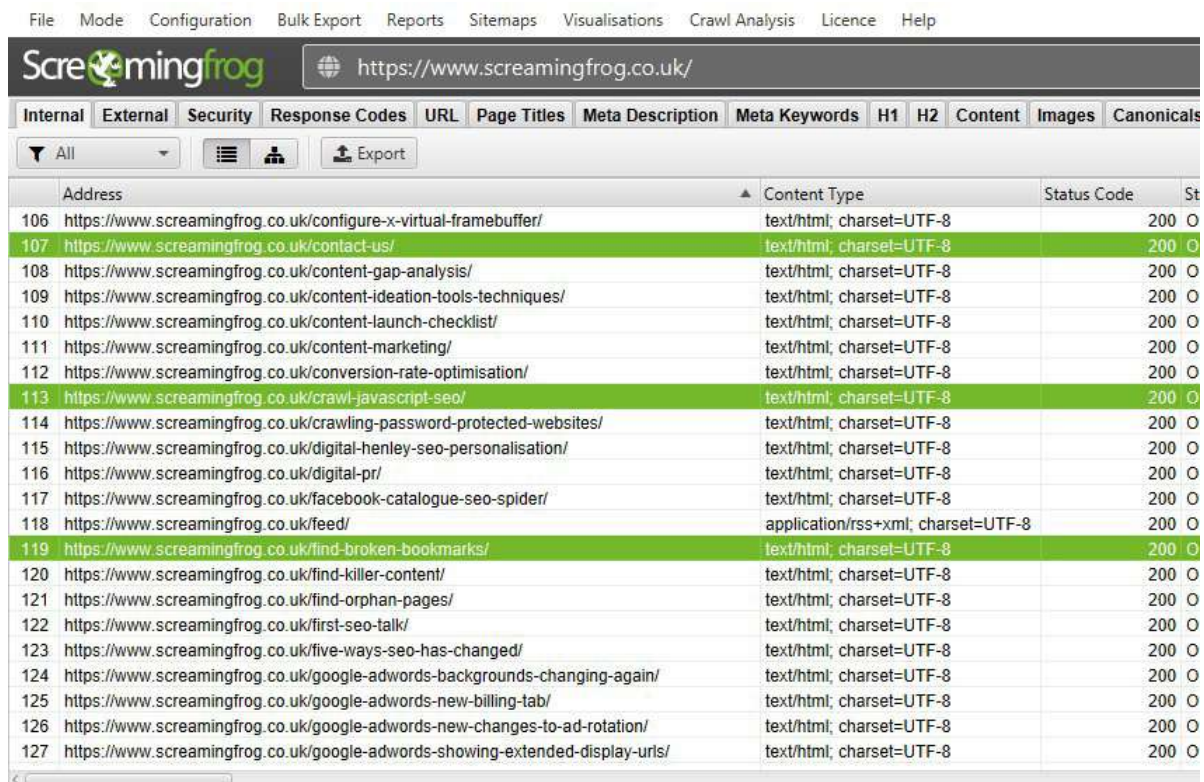
2. Migliorare i collegamenti interni (Internal Links)

L'esperienza di navigazione dell'utente dipende dal valore dei contenuti e dalla collocazione strategica nella pagina; un posizionamento dei link e in particolare delle CTA adeguato aumenterà le conversioni. Oltre all'aspetto UX una distribuzione corretta dei link permette di dirigere nel miglior modo possibile le scansioni degli spider dei Motori di Ricerca.

Per i Motori di Ricerca, il "Reasonable Surfer Model" di Google considera la quantità di PageRank che un link potrebbe trasmettere in base alla probabilità che qualcuno possa cliccare su quello specifico collegamento. Quindi viene da sé l'importanza di servire link fondamentali in posizioni adeguate considerando il livello di PageRank che viene passato.

COME ANALIZZARE LA POSIZIONE DEI LINK

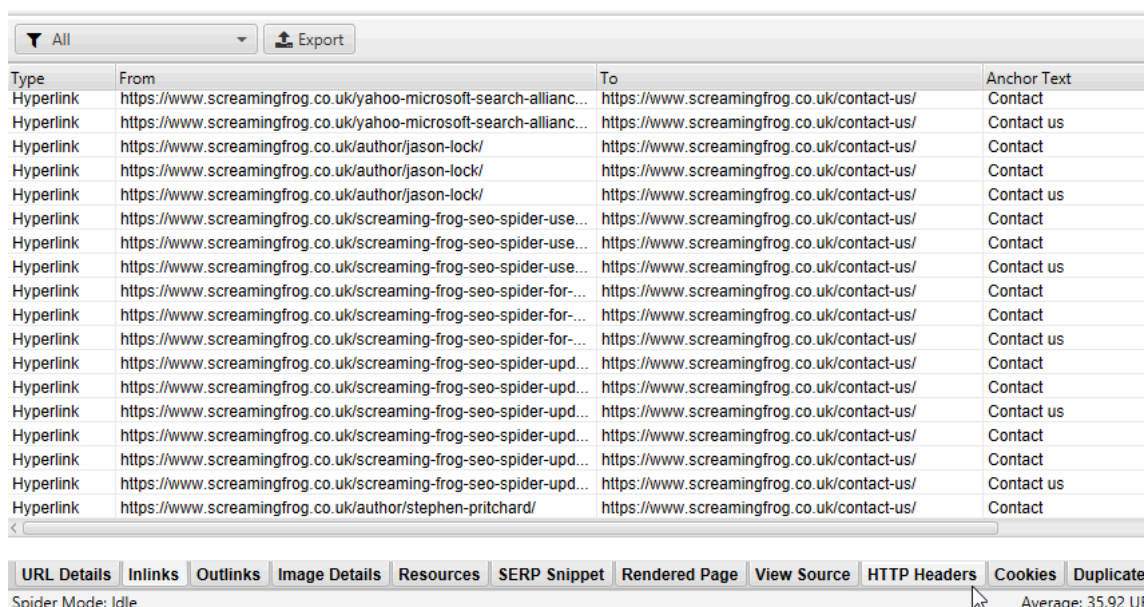
Per prima cosa inserisci il sottodominio da analizzare ed esegui il crawl; al termine del processo sarai in grado di selezionare uno o più URL per approfondire la "Link position" nei documenti Html.



The screenshot shows the Screamingfrog Crawl Analysis tool interface. The top menu includes File, Mode, Configuration, Bulk Export, Reports, Sitemaps, Visualisations, Crawl Analysis, Licence, and Help. The address bar shows the URL https://www.screamingfrog.co.uk/. Below the menu, there are tabs for Internal, External, Security, Response Codes, URL, Page Titles, Meta Description, Meta Keywords, H1, H2, Content, Images, and Canonicals. The 'URL' tab is selected, and a list of crawled URLs is displayed. The list includes columns for Address, Content Type, Status Code, and St. The status codes are mostly 200, indicating successful crawls.

Address	Content Type	Status Code	St
https://www.screamingfrog.co.uk/configure-x-virtual-framebuffer/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/contact-us/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/content-gap-analysis/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/content-ideation-tools-techniques/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/content-launch-checklist/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/content-marketing/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/conversion-rate-optimisation/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/crawl-javascript-seo/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/crawling-password-protected-websites/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/digital-henley-seo-personalisation/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/digital-pr/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/facebook-catalogue-seo-spider/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/feed/	application/rss+xml; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/find-broken-bookmarks/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/find-killer-content/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/find-orphan-pages/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/first-seo-talk/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/five-ways-seo-has-changed/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/google-adwords-backgrounds-changing-again/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/google-adwords-new-billing-tab/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/google-adwords-new-changes-to-ad-rotation/	text/html; charset=UTF-8	200	O
https://www.screamingfrog.co.uk/google-adwords-showing-extended-display-urls/	text/html; charset=UTF-8	200	O

Visualizza la scheda "Inlinks" e filtrare i dati popolati con tipologia "Hyperlinks".



The screenshot shows the Screamingfrog Crawl Analysis tool interface with the 'Inlinks' tab selected. The table displays a list of hyperlinks with columns for Type, From, To, and Anchor Text. The 'Type' column is set to 'Hyperlink'. The 'From' column shows the source URL, and the 'To' column shows the target URL. The 'Anchor Text' column shows the text used to link to the target page.

Type	From	To	Anchor Text
Hyperlink	https://www.screamingfrog.co.uk/yahoo-microsoft-search-allianc...	https://www.screamingfrog.co.uk/contact-us/	Contact
Hyperlink	https://www.screamingfrog.co.uk/yahoo-microsoft-search-allianc...	https://www.screamingfrog.co.uk/contact-us/	Contact us
Hyperlink	https://www.screamingfrog.co.uk/author/jason-lock/	https://www.screamingfrog.co.uk/contact-us/	Contact
Hyperlink	https://www.screamingfrog.co.uk/author/jason-lock/	https://www.screamingfrog.co.uk/contact-us/	Contact
Hyperlink	https://www.screamingfrog.co.uk/author/jason-lock/	https://www.screamingfrog.co.uk/contact-us/	Contact us
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-use...	https://www.screamingfrog.co.uk/contact-us/	Contact
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-use...	https://www.screamingfrog.co.uk/contact-us/	Contact
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-use...	https://www.screamingfrog.co.uk/contact-us/	Contact us
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-for...	https://www.screamingfrog.co.uk/contact-us/	Contact
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-for...	https://www.screamingfrog.co.uk/contact-us/	Contact
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-for...	https://www.screamingfrog.co.uk/contact-us/	Contact us
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-upd...	https://www.screamingfrog.co.uk/contact-us/	Contact
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-upd...	https://www.screamingfrog.co.uk/contact-us/	Contact us
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-upd...	https://www.screamingfrog.co.uk/contact-us/	Contact
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-upd...	https://www.screamingfrog.co.uk/contact-us/	Contact
Hyperlink	https://www.screamingfrog.co.uk/screaming-frog-seo-spider-upd...	https://www.screamingfrog.co.uk/contact-us/	Contact us
Hyperlink	https://www.screamingfrog.co.uk/author/stephen-pritchard/	https://www.screamingfrog.co.uk/contact-us/	Contact

At the bottom of the interface, there are tabs for URL Details, Inlinks, Outlinks, Image Details, Resources, SERP Snippet, Rendered Page, View Source, HTTP Headers, Cookies, and Duplicate. The 'Inlinks' tab is currently selected. The status bar at the bottom shows 'Spider Mode: Idle' and 'Average: 35.92 UI'.

Nella seconda colonna “From” troverai le pagine selezionate in precedenza mentre nella terza “To” tutte le pagine che vengono collegate attraverso un hyperlinks.

In questa scheda puoi consultare in modo analitico i dettagli di ogni link come ad esempio gli anchor text, la presenza o meno dell’alt text (se un’immagine è iper collegata), se il link ha l’attributo “Follow” e il target.

Scorrendo poi la finestra verso destra arriverai a visualizzare la colonna “Link Position”, con i rispettivi dettagli su dove sono stati collocati i singoli collegamenti.

To	Link Path	Link Position
https://www.screamingfrog.co.uk/contact-us/	//body/main/div[2]/div/div[@class='main-blog--sidebar']/div[@class='main-blog--sid...	Content
https://www.screamingfrog.co.uk/contact-us/	//body/header/div/nav/div/ul/li[@id='menu-item-2626']/a	Navigation
https://www.screamingfrog.co.uk/contact-us/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-7170']/a	Navigation
https://www.screamingfrog.co.uk/contact-us/	//body/main/div[2]/div/div[@class='main-blog--sidebar']/div[@class='main-blog--sid...	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[4]/div/div/a[@class='tutorials__card']/[5]	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[4]/div/div/div[@class='user-guides--inner_left']/div[2]/a[15]	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[9]/div/div[@class='spider-about_right']/ul/li[3]/a[6]	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[4]/div/div/div/div[7]/div[@id='faq-panel-crawl-issues']/div/div/div[1]/...	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[4]/div/div/div/div[9]/div[@id='faq-panel-crawl-results']/div/div/div[2]/...	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[4]/div/div/div/div[9]/div[@id='faq-panel-crawl-results']/div/div/div[4]/...	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[2]/div/div[@class='main-blog--posts']/div[1]/div/p[31]/a	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[3]/div/div[@class='main-blog--posts']/div/div/div/div/p[40]/a[2]	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[3]/div/div[@class='main-blog--posts']/div/div/div/div/ul[1]/li[7]/strong/a	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[3]/div/div[@class='main-blog--posts']/div/div/div/div/p[39]/a	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/header/div/a	Header
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[2]/div/div[@class='spider-sub--downloads']/ul/li[1]/a	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[3]/div/div[@class='main-blog--posts']/div/div/div/div/ol[1]/li[1]/a	Content
https://www.screamingfrog.co.uk/crawl-javascript-seo/	//body/main/div[3]/div/div[@class='main-blog--posts']/div/div/div/div/ol[1]/li[2]/a	Content

Un controllo molto utile potrebbe essere la ricerca dei “Client Errors” verificando i codici di risposta degli Url filtrandoli per errore 4xx. Cliccando sugli url filtrati puoi visualizzare la pagina di origine, così come il testo di ancoraggio e la posizione dei link per sistemarli.

FileModeConfigurationBulk ExportReportsSitemapsVisualisationsCrawl AnalysisLicenceHelp

Screamingfrog

https://www.screamingfrog.co.uk/

InternalExternalSecurityResponse CodesURLPage TitlesMeta DescriptionMeta KeywordsH1H2ContentImagesCanonicals

Client Error (4xx)

Export

Address	Content Type	Status Code	Status
1 https://forum.language-tool.org/		403	Forbidden
2 https://www.microsoft.com/technet/prodtechnol/WindowsServer2003/Library/IS/676400bc-...	text/html	403	Forbidden
3 https://fivethirtyeight.com/features/gun-deaths/	text/html	429	Too Many Requests
4 https://www.canva.com/learn/best-infographics/	text/html; charset=UTF-8	403	Forbidden
5 https://moz.com/help/guides/moz-api/mozscape/overview/free-vs-paid-access	text/html; charset=UTF-8	404	Not Found
6 https://majestic.com/	text/html; charset=iso-8859-1	403	Forbidden
7 https://majestic.com/help/glossary	text/html; charset=iso-8859-1	403	Forbidden
8 https://majestic.com/reports/comparator	text/html; charset=iso-8859-1	403	Forbidden
9 https://www.similarweb.com/	text/html	403	Forbidden
10 https://www.brightonseo.com/training/screaming-frogs-seo-spider-training-course/	text/html; charset=UTF-8	404	Not Found
11 https://www.instagram.com/screamingfroguk/?utm_source=ig_embed&utm_campaign=loa...	text/html; charset=utf-8	404	Not Found
12 https://sadanduseless-b-cdn.net/wp-content/uploads/2014/05/short-neck16.jpg	text/html; charset=UTF-8	404	Not Found
13 http://brianedwardsmedia.co.nz/wp-content/uploads/2011/11/The-eyes-have-it.jpg%20rel=...	text/html	404	Not Found
14 http://www.ew.com/ew/article/0,,445864,00.html	text/html; charset=utf-8	404	Not Found
15 http://www.theoccidentalobserver.net/wp-content/uploads/2012/03/Berlusconi.jpg%20rel=n...	text/html; charset=UTF-8	404	Not Found

Hyperlink

Export

type	From	To	Link Position	Anchor Text
hyperlink	https://www.screamingfrog.co.uk/brightonseo-crawling-clinic-2019/	https://www.brightonseo.com/training/screaming-...	Content	few places let

URL DetailsInlinksOutlinksImage DetailsResourcesSERP SnippetRendered PageView SourceHTTP HeadersCookiesDuplicate

COME CLASSIFICA I LINK SCREAMING FROG

Screaming Frog classifica il posizionamento dei link seguendo questo schema:

- **Navigation:** comprende i collegamenti contenuti all'interno dell'elemento nav principale, di solito corrisponde al menu principale.
- **Header:** identifica i link contenuti all'interno dell'intestazione che comunemente si trovano all'inizio della pagina.
- **Aside:** consistono nei collegamenti utilizzati per i "call out box" e le sidebar laterali.
- **Footer:** comprende i link contenuti all'interno del footer in fondo alla pagina.
- **Content:** link contenuti all'interno del contenuto principale della pagina.

Questa classificazione utilizza il "link path" (come XPath) per identificare sottostringhe semantiche conosciute. Agli sviluppatori la colonna "Link Path" può fornire un contesto ancora più specifico su dove si trova esattamente ogni link all'interno dell'HTML.

Nel caso che il sito web da analizzare presenti un numero di link molto corposo e buona pratica filtrare i risultati dell'analisi isolando ogni singola posizione.

The screenshot shows the Screaming Frog SEO Spider interface. A red arrow points to the 'Content' filter tab. The main table displays a list of links with columns for 'Link Path' and 'Link Position'. The 'Link Position' column is highlighted in green, and the 'Search Columns' dialog box is open, showing 'Link Position' selected. The dialog box also includes options for 'Type', 'From', 'To', 'Anchor Text', 'Alt Text', 'Follow', 'Target', 'Rel', 'Status Code', 'Status', 'Path Type', and 'Link Path'. The 'Select All' button is checked, and the 'Clear All' button is unchecked. The bottom status bar shows 'Average: 35.92 URL/s. Current: 40.30 URL/s.'

	Link Path	Link Position
seo/	//body/main/div[2]/div/div[@class='main-blog--posts']/div[1]/div/p[14]/a	Content
seo/	//body/main/div[2]/div/div[@class='main-blog--posts']/div[1]/div/p[21]/a	Content
seo/	//body/main/div[2]/div/div[@class='main-blog--posts']/div[@id='comments']/ul/li[@i...	Content
marks/	//body/main/div[4]/div/div/div[@class='user-guides--inner_left']/div[2]/a[17]	Content
marks/	//body/main/div[2]/div/div[@class='main-blog--posts']/div[1]/div/p[49]/a	Content
marks/	//body/main/div[2]/div/div[@class='main-blog--posts']/div[1]/div/p[28]/a	Content
marks/	//body/main/div[4]/div/div/a[@class='tutorials__card'][6]	Content
marks/	//body/main/div[2]/div/div[@class='spider-sub--downloads']/ul/li[1]/a	Content
marks/	//body/main/div[3]/div/div[@class='main-blog--sidebar']/div[@id='js-table-of-conten...	Content
marks/	//body/main/div[3]/div/div[@class='main-blog--sidebar']/div[@id='js-table-of-conten...	Content
marks/	//body/main/div[3]/div/div[@class='main-blog--sidebar']/div[@id='js-table-of-conten...	Content
marks/	//body/main/div[3]/div/div[@class='main-blog--sidebar']/div[@id='js-table-of-conten...	Content
marks/	//body/main/div[3]/div/div[@class='main-blog--sidebar']/div[@id='js-table-of-conten...	Content
marks/	//body/main/div[3]/div/div[@class='main-blog--sidebar']/div[@id='js-table-of-conten...	Content
marks/	//body/main/div[3]/div/div[@class='main-blog--sidebar']/div[@id='js-table-of-conten...	Content
marks/	//body/main/div[3]/div/div[@class='main-blog--sidebar']/div[@id='js-table-of-conten...	Content
marks/	//body/main/div[3]/div/div[@class='main-blog--sidebar']/div[3]/div/a[@class='free-d...	Content
marks/	//body/main/div[3]/div/div[@class='main-blog--sidebar']/div[3]/div/a[@class='lfa-fre...	Content

Concluso questo processo hai la possibilità di esportare i dati attraverso il pulsante "Export" direttamente sulla scheda, cliccando con il tasto destro sugli Url e selezionando "Export>>Inlinks" oppure scaricando il rapporto All InLinks.

Bulk Export > Links > All Inlinks

File Mode Configuration Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help

Screamingfrog <https://www.screamingfrog.co.uk/>

Internal External Security Response Codes URL Page Titles Meta Description Meta Keywords H1 H2 Content Images Canonicals

All Export

	Address	Content Type	Status Code	St
106	https://www.screamingfrog.co.uk/configure-x-virtual-framebuffer/	text/html; charset=UTF-8	200	O
107	https://www.screamingfrog.co.uk/contact-us/	text/html; charset=UTF-8	200	O
108	https://www.screamingfrog.co.uk/content-gap-analysis/	text/html; charset=UTF-8	200	O
109	https://www.screamingfrog.co.uk/content-ideation-tools-	text/html; charset=UTF-8	200	O
110	https://www.screamingfrog.co.uk/content-launch-checkli	text/html; charset=UTF-8	200	O
111	https://www.screamingfrog.co.uk/content-marketing/	text/html; charset=UTF-8	200	O
112	https://www.screamingfrog.co.uk/conversion-rate-optimi	text/html; charset=UTF-8	200	O
113	https://www.screamingfrog.co.uk/crawl-javascript-seo/	text/html; charset=UTF-8	200	O
114	https://www.screamingfrog.co.uk/crawling-password-pro	text/html; charset=UTF-8	200	O
115	https://www.screamingfrog.co.uk/digital-henley-seo-personalisation/	text/html; charset=UTF-8	200	O
116	https://www.screamingfrog.co.uk/digital-pr/	text/html; charset=UTF-8	200	O
117	https://www.screamingfrog.co.uk/facebook-catalogue-seo-spider/	text/html; charset=UTF-8	200	O
118	https://www.screamingfrog.co.uk/feed/	text/html; charset=UTF-8	200	O
119	https://www.screamingfrog.co.uk/find-broken-bookmarks/	text/html; charset=UTF-8	200	O
120	https://www.screamingfrog.co.uk/find-killer-content/	text/html; charset=UTF-8	200	O
121	https://www.screamingfrog.co.uk/find-orphan-pages/	text/html; charset=UTF-8	200	O
122	https://www.screamingfrog.co.uk/first-seo-talk/	text/html; charset=UTF-8	200	O
123	https://www.screamingfrog.co.uk/five-ways-seo-has-changed/	text/html; charset=UTF-8	200	O
124	https://www.screamingfrog.co.uk/google-adwords-backgrounds-changing-again/	text/html; charset=UTF-8	200	O
125	https://www.screamingfrog.co.uk/google-adwords-new-billing-tab/	text/html; charset=UTF-8	200	O
126	https://www.screamingfrog.co.uk/google-adwords-new-changes-to-ad-rotation/	text/html; charset=UTF-8	200	O
127	https://www.screamingfrog.co.uk/google-adwords-showing-extended-display-urls/	text/html; charset=UTF-8	200	O

Copy Open in Browser Re-Spider Remove Export Inlinks Outlinks Image Details Resources Duplicate Details Spelling & Grammar Details

Nota Bene: questo processo può essere replicato anche per un'analisi degli Outlinks.

COME CONFIGURARE UNA CLASSIFICAZIONE PERSONALIZZATA SULLA POSIZIONE DEI LINK

Oltre il 90% delle analisi sul posizionamento dei links sono coperti dalla classificazione standard di Screaming Frog ma nel caso questa non restituisca dati attendibili hai la possibilità di configurare la tua classificazione personale.

Config > Custom > Link Positions

Custom Link Positions

Link Positions Test

Define the position of links on a page. Add a position name and a search term (substring) of the link path to classify links. Search terms higher in the list take precedence.

☒ Enable

Head	/head/	✓	✕
Navigation	nav	✓	✕
Header	header	✓	✕
Aside	aside	✓	✕
Footer	footer	✓	✕
Content	/	✓	✕

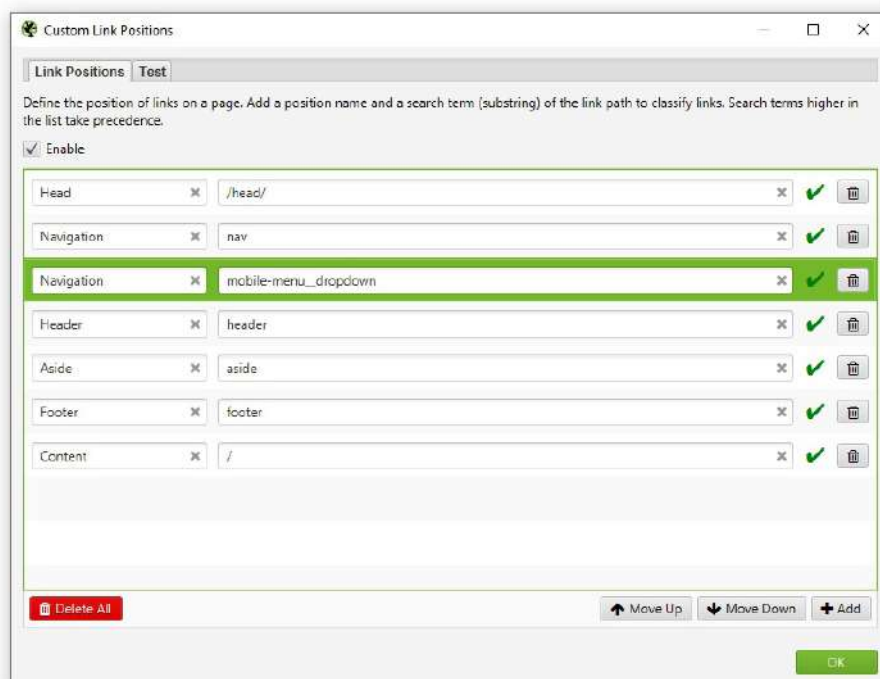
Delete All Move Up Move Down Add

OK

Per comprendere al meglio la sua applicazione prendi in considerazione il sito web di Screaming Frog nel quale dei link del menu mobile sono al di fuori dell'elemento "nav" e vengono classificati come link di "contenuto". Questa condizione è un errore, in quanto rappresentano solo un'ulteriore navigazione del sito su mobile.

To	Link Path	Link Position
https://www.screamingfrog.co.uk/seo-spider/licence/	//body/header/div/nav/div/ul/li[@id='menu-item-2642']/ul/li[@id='menu-item-497...	Navigation
https://www.screamingfrog.co.uk/blog/	//body/header/div/nav/div/ul/li[@id='menu-item-2640']/a	Navigation
https://www.screamingfrog.co.uk/contact-us/	//body/header/div/nav/div/ul/li[@id='menu-item-2626']/a	Navigation
https://www.screamingfrog.co.uk/login/	//body/header/div/nav/div/ul/li[@id='menu-item-5003']/a	Navigation
https://www.screamingfrog.co.uk/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-714...	Content
https://www.screamingfrog.co.uk/about/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-714...	Content
https://www.screamingfrog.co.uk/about/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-714...	Content
https://www.screamingfrog.co.uk/meet-the-team/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-714...	Content
https://www.screamingfrog.co.uk/our-story/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-714...	Content
https://www.screamingfrog.co.uk/careers/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-714...	Content
https://www.screamingfrog.co.uk/clients/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-714...	Content
https://www.screamingfrog.co.uk/services/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-715...	Content
https://www.screamingfrog.co.uk/search-engine-marketing/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-715...	Content
https://www.screamingfrog.co.uk/search-engine-optimisation/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-715...	Content
https://www.screamingfrog.co.uk/pay-per-click/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-715...	Content
https://www.screamingfrog.co.uk/social-media/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-715...	Content
https://www.screamingfrog.co.uk/content-marketing/	//body/div[@class='mobile-menu__dropdown']/div/div/ul/li[@id='menu-item-715...	Content

Prendendo il nome della classe, che in questo caso è "mobile-menu__dropdown", presente nel percorso del link potresti ri-definire la posizione corretta.



Una volta completate le personalizzazioni e avviato una nuova scansione del sito web otterrai la corretta interpretazione delle posizioni nella scheda Inlinks.

Questo processo può essere seguito per qualsiasi tipo di collegamento, quindi potrai essenzialmente "clusterizzare" i collegamenti in base alle loro sottostringhe XPath.

FARE UNA SCANSIONE COMPLETA DELLA VERSIONE AMP

Screaming Frog permette di analizzare le Accelerated Mobile Pages (AMP) in modo rapido ed efficiente utilizzando il validatore ufficiale di AMP per consentire la convalida in massa degli URL.

L'analisi mette in evidenza i problemi Seo comuni e li convalida controllando il mark-up HTML richiesto, gli elementi HTML proibiti secondo le specifiche e altro ancora.

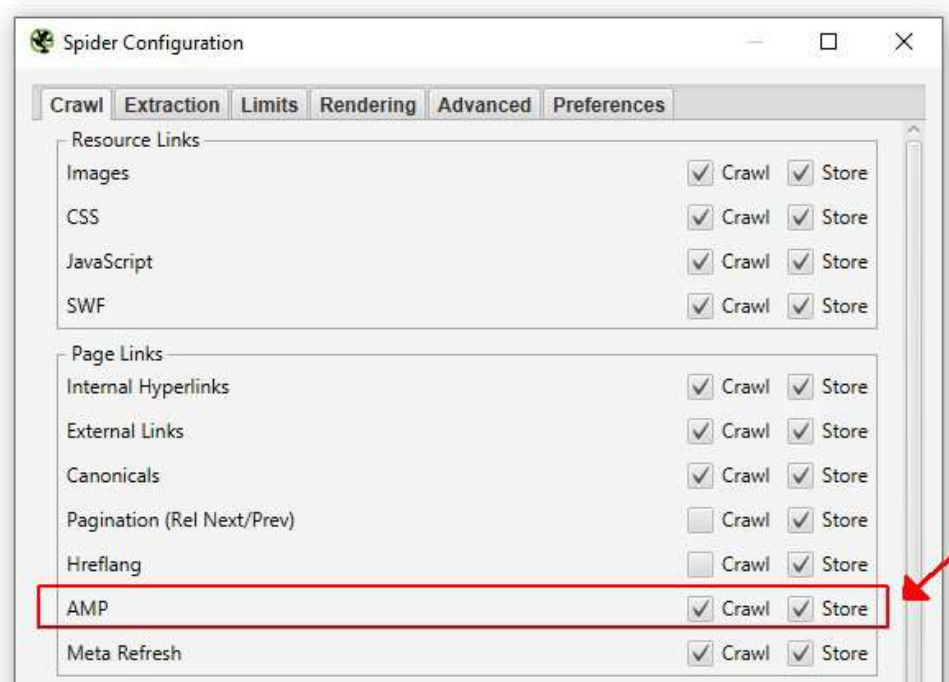
Al momento ci sono due opzioni per analizzare e convalidare AMP:

1. Scansionare un sito per controllare AMP e consultare i dati nella scheda AMP.
2. Caricare e controllare gli URL AMP separatamente.

1. Scansionare un sito per un Audit Seo elementi AMP

La prima cosa da fare è abilitare l'opzione AMP dal menu di configurazione e iniziare la scansione del sito web.

Config > Spider > Crawl



Consultare i dati nella scheda AMP

Al termine del processo di crawling è sufficiente cliccare sulla scheda AMP, in cui vengono mostrati tutti gli URL AMP scoperti e scegliere tra i 17 filtri a disposizione per identificare problemi comuni a livello SEO o di convalida.

File Configuration Mode Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help

Screamingfrog <https://www.telegraph.co.uk/>

Internal External Protocol Response Codes URI Page Titles Meta Description Meta Keywords H1 H2 Images Canonicals Pagination Directives Hreflang **AJAX AMP** Sitemaps

Filter: All Export

	Filter	Address	Content	Status Code	Status
1	✓ All	2/vladimir-putin-says-russia-has-identified-salisbury-poison-suspects/amp/	text/html	200	OK
2	Non-200 Response	06/11/almost-one-three-graduates-overqualified-job-major-report-finds/amp/	text/html	200	OK
3	Non-Confirming Canonical	8/09/10/brian-moores-full-contact-dylan-hartleys-experience-invaluable/amp/	text/html	200	OK
4	Missing Non-AMP Canonical	2/tory-peer-lord-barkers-link-putins-business-ally-must-end/amp/	text/html	200	OK
5	Non-Indexable Canonical	0/cocktails/italian-style-cocktails-nibbles-aperitivo-hour/amp/	text/html	200	OK
6	Indexable	1/uk-third-fattest-nation-europe-fuelled-diet-booze/amp/	text/html	200	OK
7	Non-Indexable	est-british-bake-2018-bread-week-hollywood-goes-soft-picking/amp/	text/html	200	OK
8	Missing <html amp> Tag	9/11/wall-street-dangerously-exposed-china-digs-trump-showdown/amp/	text/html	200	OK
9	Missing/Invalid <!doctype html> Tag	2/hundreds-women-conned-freezing-eggs-age-almost-no-chance-success/amp/	text/html	200	OK
10	Missing <head> Tag	princess-margaret-royal-rebel-review-sensitive-portrayal-princess/amp/	text/html	200	OK
11	Missing <body> Tag	petitions/travel-awards-2018/amp/	text/html	200	OK
12	Missing Canonical	m-festival-2018-best-movies-see/amp/	text/html	200	OK
13	Missing/Invalid <meta charset> Tag	9/11/jaguar-land-rover-warns-brexit-no-deal-would-cost-tens-thousands/amp/	text/html	200	OK
14	Missing/Invalid <meta viewport> Tag	make-money-video-games-fortnite-success-draws-investors-gaming/amp/	text/html	200	OK
15	Missing AMP Script	1/simple-urine-test-for-ovarian-cancer-could-radically-improve-survival/amp/	text/html	200	OK
16	Missing AMP Boilerplate	09/11/google-legal-battle-eu-right-forgotten/amp/	text/html	200	OK
17	Contains Disallowed HTML	09/12/claudio-scimone-conductor-musicologist-obituary/amp/	text/html	200	OK
18	Other Validation Errors	questor-hold-morrisons-avoid-vodafone/amp/	text/html	200	OK
19		rice-season-2-episode-1-review-back-world-seventies-smut-streets/amp/	text/html	200	OK
20		09/10/cryptocurrency-wipeout-deepens-640-billion-ether-leads-declines/amp/	text/html	200	OK
21		2/juncker-hold-firm-brexit-demands-offer-close-partnership-uk/amp/	text/html	200	OK
22		09/11/tech-giants-mount-disinformation-campaign-copyright-vote-activists/amp/	text/html	200	OK

Quindici dei 17 filtri proposti sono immediatamente a tua disposizione, mentre “non-200 Response” e “Missing <html amp> Tag” li puoi far popolare avviando la “Crawl Analysis” configurabile tramite “Crawl Analysis > AMP”.

MP Sitemaps Custom Analytics Search Console Link Metrics

Indexability Status Title

Canonicalised	Vladimir Putin says Salisbury poison s
Canonicalised	Almost one in three graduates are ove
Canonicalised	Brian Moore's Full Contact: 'Dylan Ha
Canonicalised	Tory peer must end his association wi
Canonicalised	Italian-style cocktails and nibbles for t
Canonicalised	UK is third fattest nation in Europe – fi
Canonicalised	The Great British Bake Off 2018: Brea
Canonicalised	Wall Street dangerously exposed as C
Canonicalised	Hundreds of women are being connec
Canonicalised	Princess Margaret: The Royal Rebel,
Canonicalised	Telegraph Travel Awards 2018: Win o
Canonicalised	London Film Festival 2018: the 10 bes
Canonicalised	Jaguar Land Rover warns Brexit 'no d
Canonicalised	How to make money from video game
Canonicalised	Simple urine test for ovarian cancer w
Canonicalised	Google in legal battle with EU over 'ri
Canonicalised	Claudio Scimone, conductor and musi
Canonicalised	Questor: after nearly two years our Mo
Canonicalised	The Deuce, season 2 episode 1 review
Canonicalised	Cryptocurrency wipeout deepens to \$
Canonicalised	Juncker to 'hold firm on Brexit deman

Overview Site Structure Response Times API

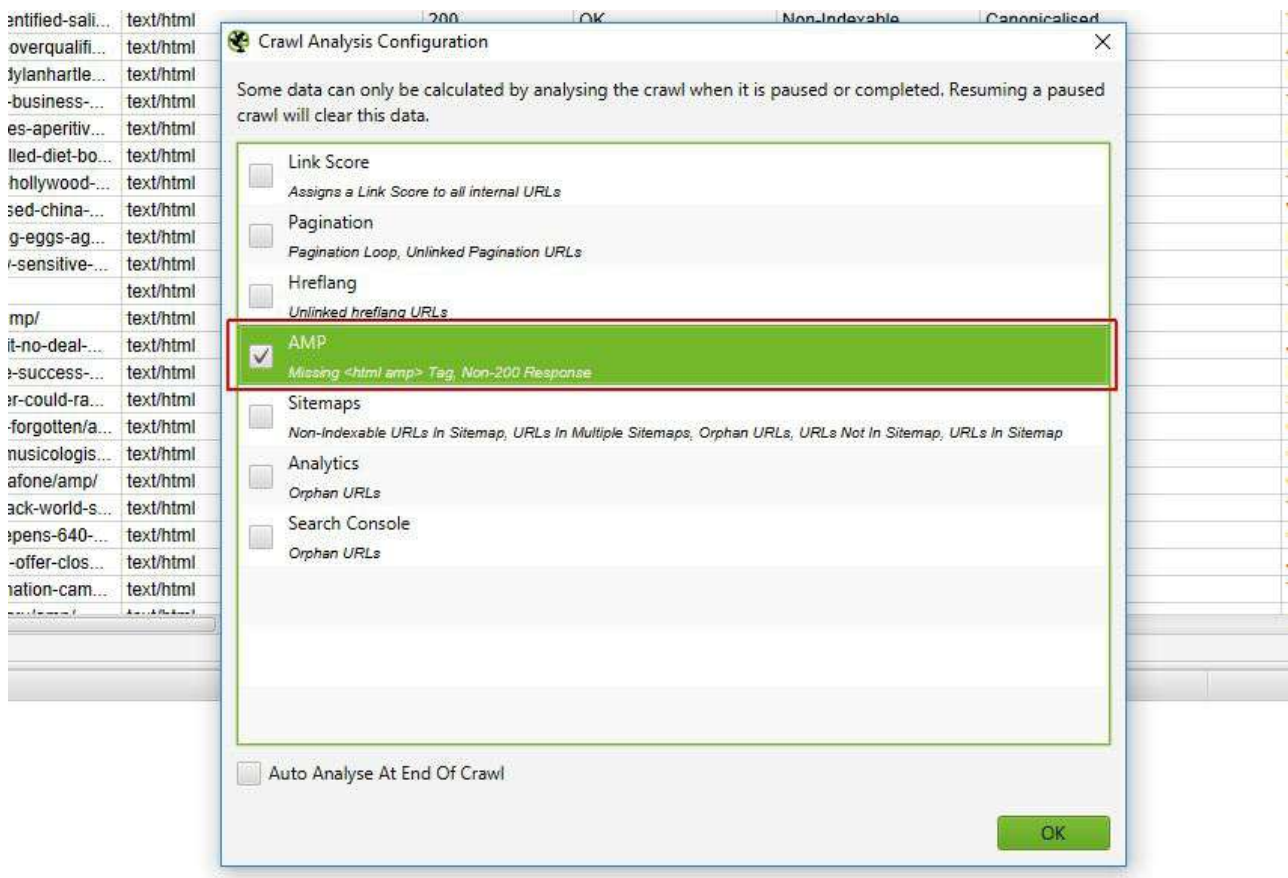
▼ AMP

- All (118) (100.00%)
- Non-200 Response (Crawl Analysis Required)
- Non-Confirming Canonical (0) (0.00%)
- Missing Non-AMP Canonical (0) (0.00%)
- Non-Indexable Canonical (0) (0.00%)
- Indexable (0) (0.00%)
- Non-Indexable (118) (100.00%)
- Missing <html amp> Tag (Crawl Analysis Required)
- Missing/Invalid <!doctype html> Tag (0) (0.00%)
- Missing <head> Tag (0) (0.00%)
- Missing <body> Tag (0) (0.00%)
- Missing Canonical (0) (0.00%)
- Missing/Invalid <meta charset> Tag (0) (0.00%)
- Missing/Invalid <meta viewport> Tag (0) (0.00%)
- Missing/Invalid AMP Script (0) (0.00%)
- Missing/Invalid AMP Boilerplate (0) (0.00%)
- Contains Disallowed HTML (0) (0.00%)
- Other Validation Errors (0) (0.00%)

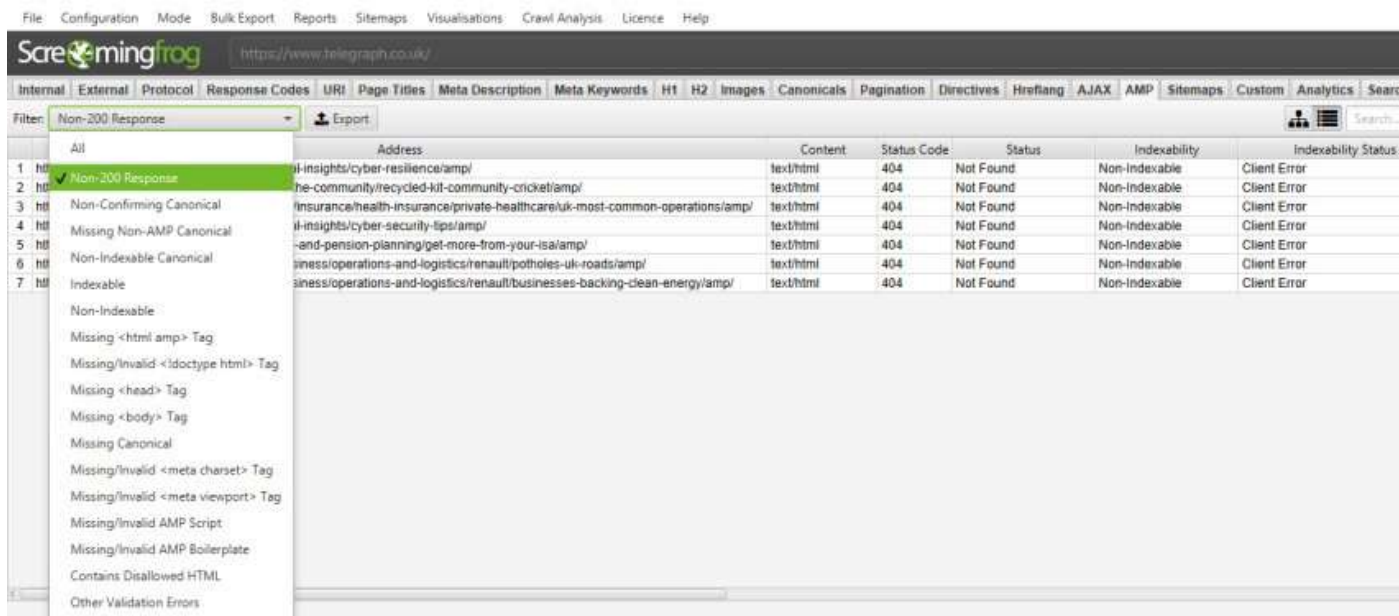
Di seguito i filtri AMP:

- Non-200 Response.
- Missing Non-AMP Return Link.
- Missing Canonical to Non-AMP.
- Non-Indexable Canonical.
- Indexable.
- Non-Indexable.
- Missing HTML AMP Tag.
- Missing/Invalid Doctype HTML Tag.
- Missing Head Tag.
- Missing Body Tag.
- Missing Canonical.
- Missing/Invalid Meta Charset Tag.
- Missing/Invalid Meta Viewport Tag.
- Missing/Invalid AMP Script.
- Missing/Invalid AMP Boilerplate.
- Contains Disallowed HTML.
- Other Validation Errors.

Per approfondire le caratteristiche di ogni filtro ti invito a consultare la sezione dedicata alla Tab AMP della guida.

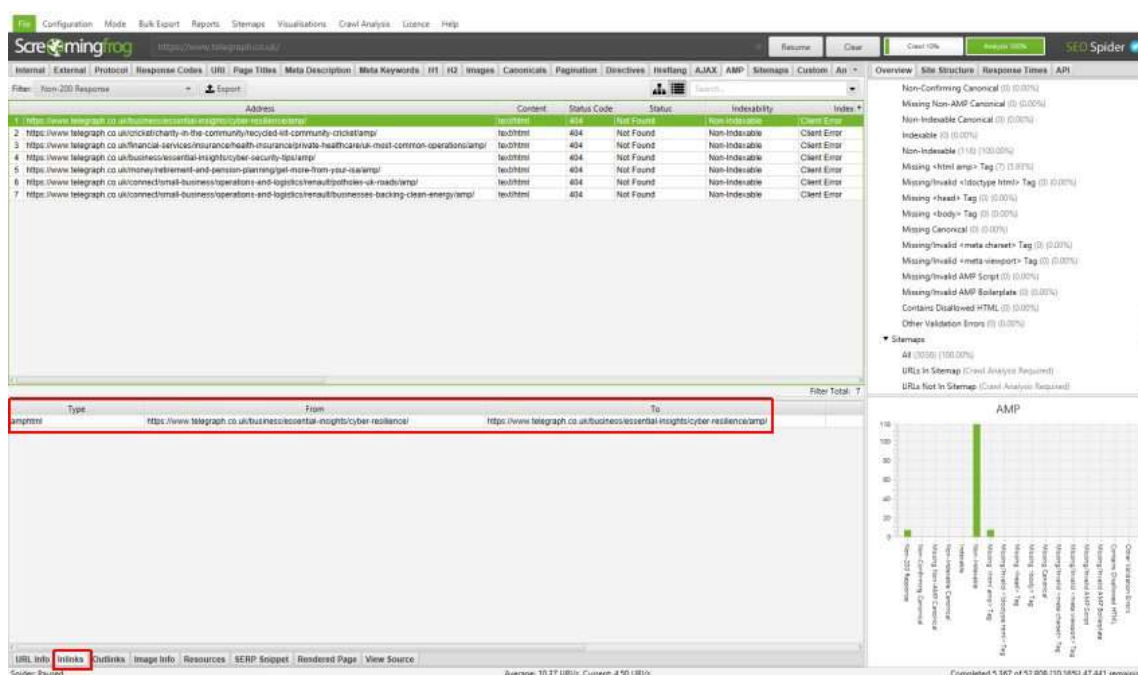


Attraverso la Crawl Analysis ad esempio puoi scoprire in modo semplice ed intuitivo tutti gli URL AMP che non rispondono con status code 2xx e devono essere ottimizzati.



Per conoscere la fonte degli errori degli URL AMP ti basta cliccare su un URL nel pannello superiore della finestra e poi sulla scheda "Inlinks" in basso.

Ricorda che gli "amphtml", sono dei riferimenti a degli URL attraverso un collegamento rel="amphtml" che si trova all'interno dell'header nell'HTML.



Di seguito una vista ravvicinata della scheda 'Inlinks'.



Nell'esempio precedente si vede come l'url html (<https://www.telegraph.co.uk/business/essential-insights/cyber-resilience/>) presenta un tag di collegamento rel="amphtml" alla sua versione AMP (<https://www.telegraph.co.uk/business/essential-insights/cyber-resilience/amp/>), che però ha uno status code 404.

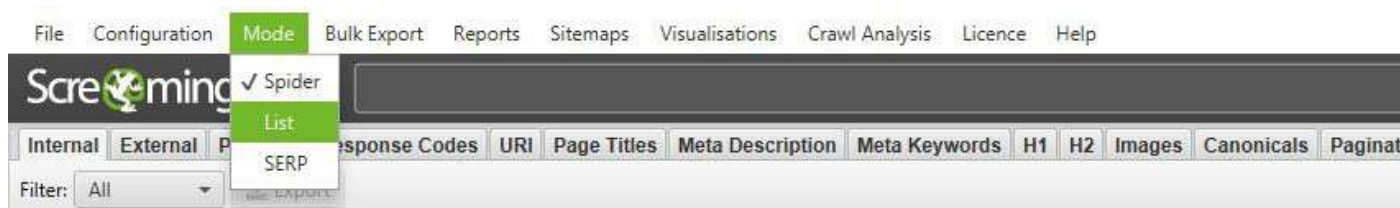
A questo punto, scoperti tutti gli errori tra la versione Html e Html AMP, puoi esportare i dati usando "Bulk Export > AMP" scegliendo su quale focus iniziare l'ottimizzazione.

2. Caricare e controllare gli URL AMP separatamente.

In alternativa, è possibile verificare gli URL AMP in modo separato utilizzando la modalità elenco caricando una lista da file o manualmente e sottoponendoli a crawling.

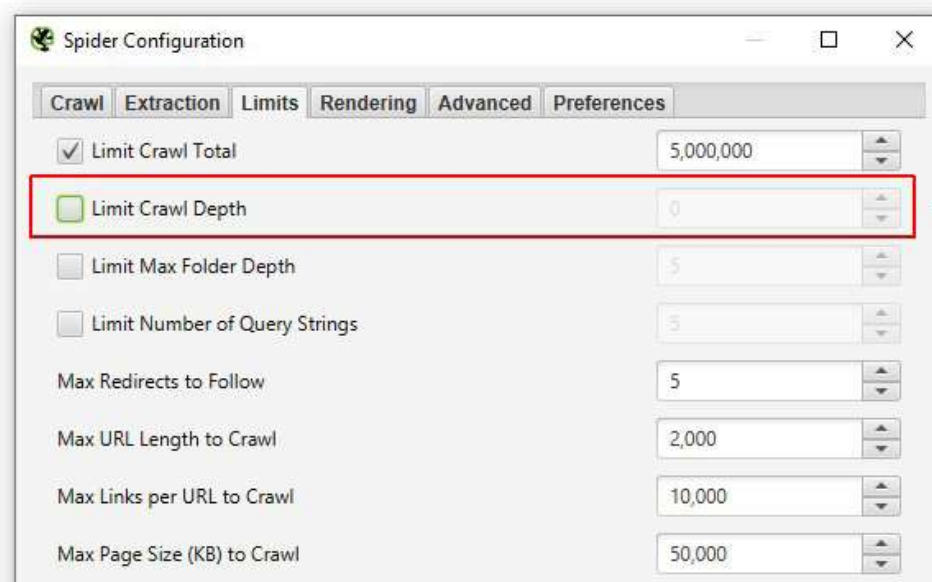
La soluzione più consigliata è quella di verificare entrambi gli URL equivalenti desktop e AMP insieme; questo processo è possibile caricando le versioni desktop e modificando la configurazione del Seo Spider secondo il processo seguente:

Per prima cosa devi cambiare la scansione in modalità "List".



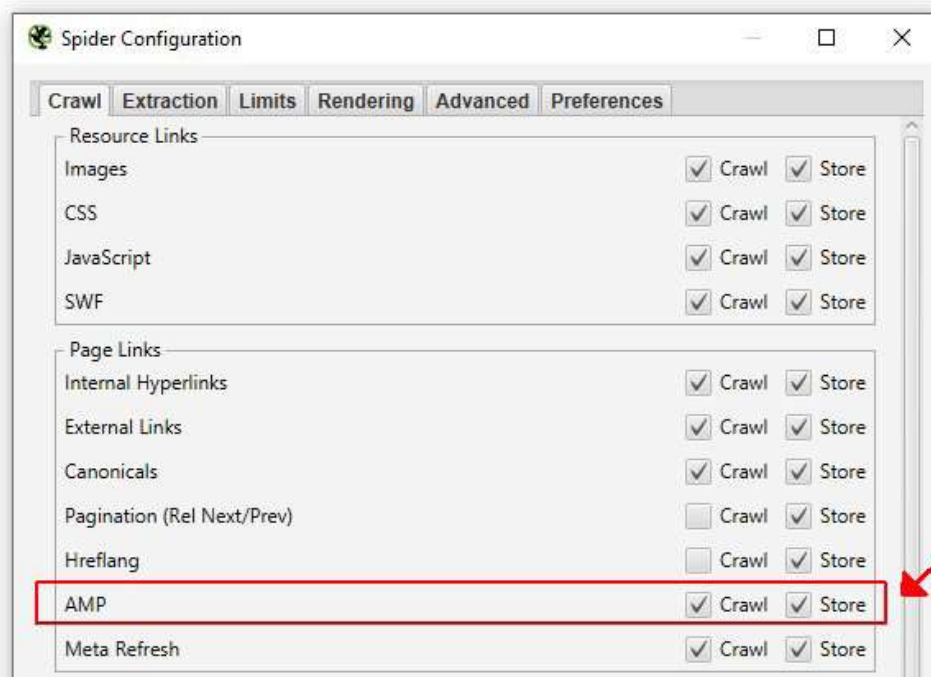
Per impostazione predefinita di Screaming Frog la profondità di scansione (Crawl Depth) è impostata a '0' in modalità elenco, quindi solo gli URL caricati vengono scansionati.

Per il nostro scopo questa impostazione dovrà essere rimossa, poiché anche le versioni AMP (a profondità di scansione '1') devono essere scansionate.

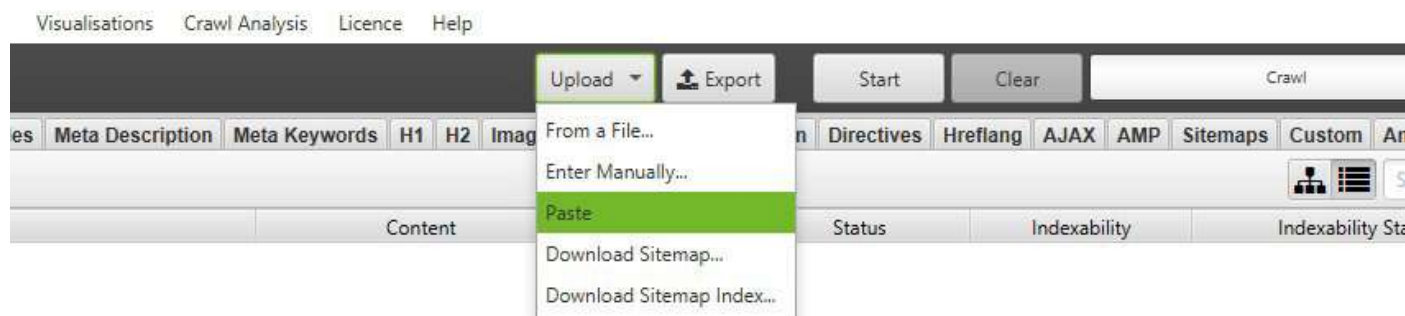


Avendo inibito un limite al "Crawl Depth" lo Spider analizzerà tutti gli URL caricati e i collegamenti trovati come fosse nella modalità di scansione "Spider".

A questo punto per scansionare solo gli equivalenti AMP e non altri link interni (altrimenti avremmo ripetuto la prima opzione di scansione dell'intero sito descritto sopra), tutti i link di risorse e pagine devono essere disabilitati, a parte i link AMP.



Conclusa questa fase di configurazione sei in grado di aggiungere gli URL da scansionare ed eseguire la fase di crawling e seguire il processo delineato dal punto 3 nella sezione “Scansione di un sito per controllare AMP”



COME SCANSIONARE I WEB VITALS

Screaming Frog SEO Spider attraverso la connessione con PageSpeed Insights API ti permette di misurare facilmente i Core Web Vitals (CWV) per ogni pagina del tuo sito web.

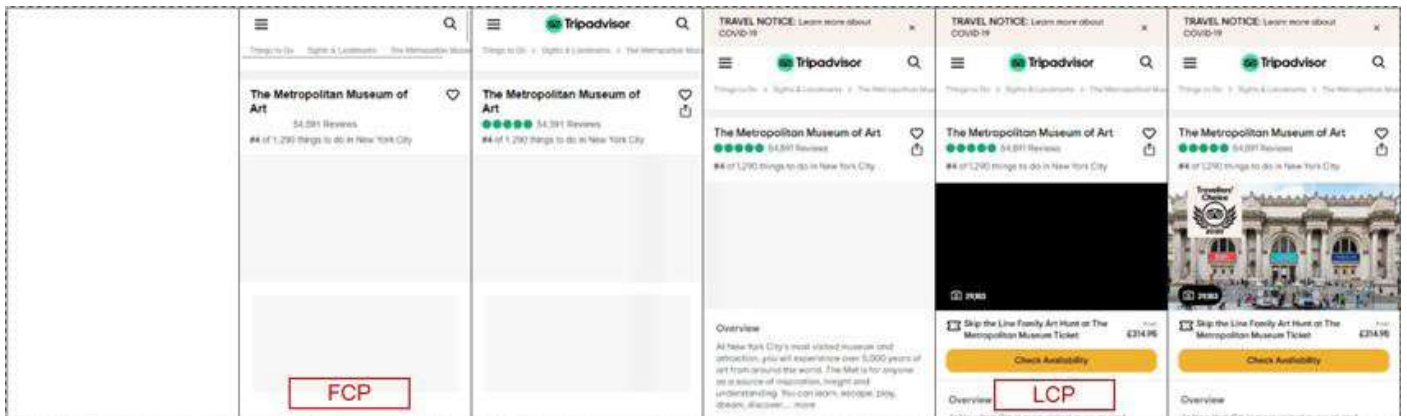
I Core Web Vitals sono un insieme di metriche che Google ha introdotto da maggio 2021 per misurare gli aspetti chiave dell'esperienza dell'utente durante il caricamento di una pagina web. I Core Web Vitals sono una vera e propria rivoluzione che aiuta Google a valutare le pagine web e concorrono all'attribuzione del ranking per i risultati di ricerca.

Al momento sono essenzialmente tre ma credo che in seguito, oltre all'affinamento degli stessi saranno introdotte altre metriche sempre più orientate a garantire un'esperienza di navigazione più soddisfacente.

1. LCP: Largest Contentful Paint

La metrica Largest Contentful Paint (LCP) riporta il tempo di rendering della più grande immagine o blocco di testo visibile all'interno della viewport, rispetto a quando la pagina ha iniziato il caricamento. Come parametro di valutazione Google considera un tempo massimo di 2,5 secondi per garantire un'esperienza utente valida e soddisfacente.

Per assicurarti di raggiungere questo obiettivo per la maggior parte dei tuoi utenti, una buona soglia da considerare è il 75° percentile dei carichi di pagina, segmentati tra dispositivi mobili e desktop.



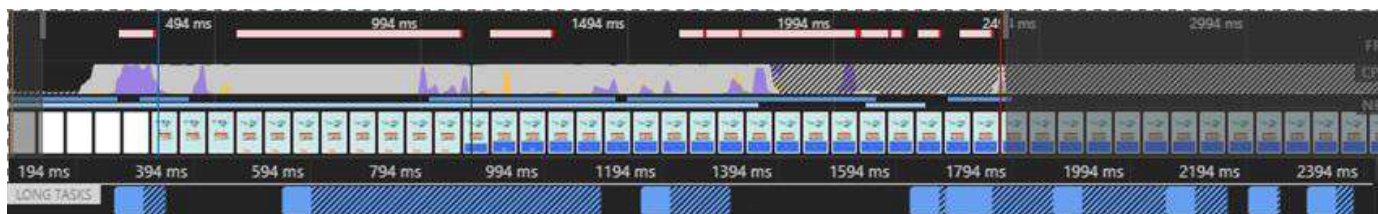
2. FID: First Input Delay

Il First Input Delay (FID) di una pagina misura la sua reattività alle interazioni dell'utente. Per esempio, se un utente fa clic su un link, o seleziona un menu a tendina il FID valuta quanto velocemente questa azione viene eseguita sulla pagina.

Di solito, le interazioni dell'utente sono ritardate quando un browser è impegnato nell'esecuzione di altri compiti.

Ad esempio, analizzando la timeline di caricamento qui sotto, vediamo delle barre blu che identificano il tempo necessario per caricare una risorsa (come un file JavaScript) che richiede un tempo piuttosto prolungato.

Se un utente facesse clic su qualcosa all'interno della pagina HTML durante questo intervallo di tempo l'interazione sarebbe rinviata fino a quando il browser non ha portato a termine il processo precedente.

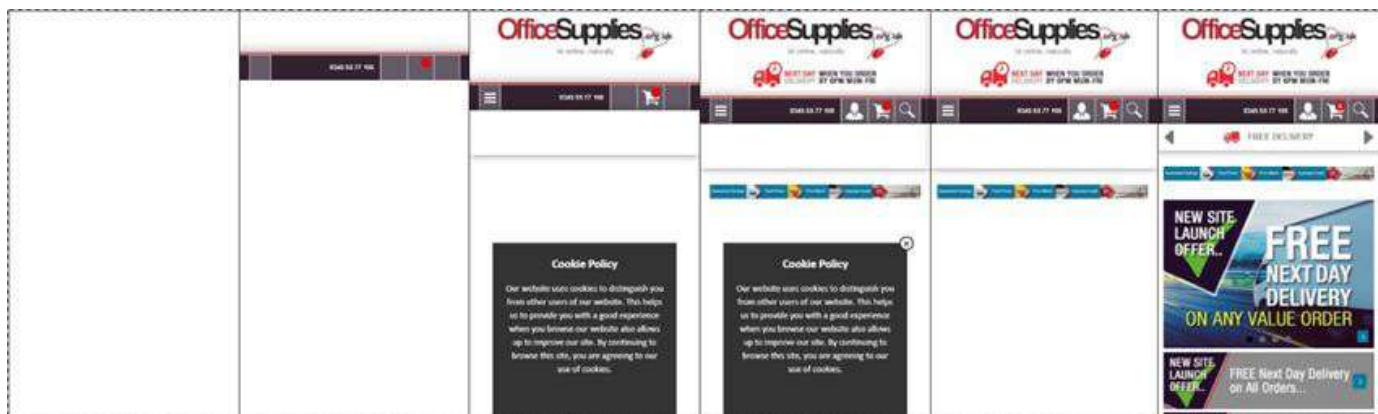


Come parametro di valutazione, per fornire una buona esperienza di navigazione per l'utente, Google indica che la metrica First Input Delay dovrebbe essere uguale o inferiore a 100 millisecondi. Per assicurarti di raggiungere questo obiettivo per la maggior parte dei tuoi utenti, una buona soglia da misurare è il 75° percentile dei carichi di pagina, segmentati tra dispositivi mobili e desktop.

3. CLS: Cumulative Layout Shift

Il Cumulative Layout Shift (CLS) è una metrica molto importante incentrata sull'esperienza utente che misura la stabilità visiva di un layout durante il caricamento di una pagina HTML. Questo indice è stato introdotto per limitare spiacevoli cambi inaspettati dovuti ad esempio a banner pubblicitari o altro. Tutti abbiamo avuto la frustrante esperienza di leggere un sito di notizie e avere il testo dell'articolo che salta più in basso mentre la navigazione viene caricata; il CLS misura appunto questa instabilità di navigazione.

Qui sotto puoi vedere questo fenomeno mentre la pagina viene caricata dal browser. Attraverso ogni schermata viene aggiunto del contenuto, causando lo spostamento in alto e in basso degli elementi precedenti.



Per fornire una buona esperienza utente, i siti dovrebbero sforzarsi di avere un punteggio CLS di 0,1 o inferiore. Per assicurarti di raggiungere questo obiettivo per la maggior parte dei tuoi utenti, una buona soglia da misurare è il 75° percentile dei carichi di pagina, segmentati tra dispositivi mobili e desktop.

I Core Web Vitals sono delle metriche fondate sui dati reali di Chrome User Experience Report (CrUX) relativi agli ultimi 28 giorni e basate sulle visite degli utenti a una pagina.

Ogni Web Vital è valutato in base a tre classificazioni; Good, Needs Improvement e Poor.

Affinché una pagina venga convalidata dai Core Web Vital Assessment deve essere considerata "Good" in tutte e tre le metriche.

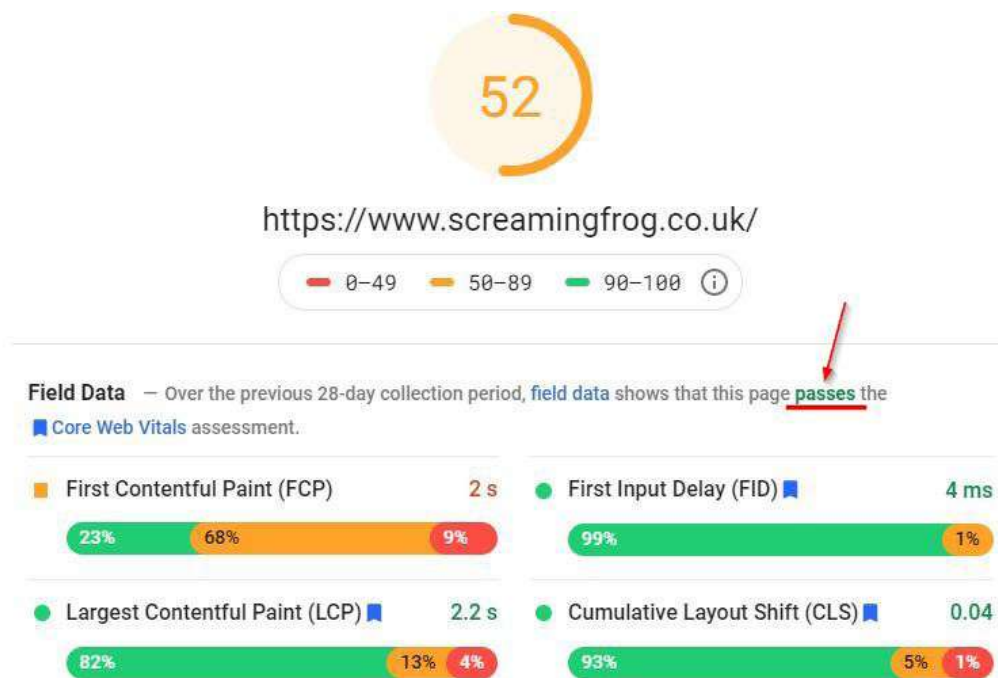
Google ha fornito una comoda illustrazione delle soglie per ciascuna metrica:



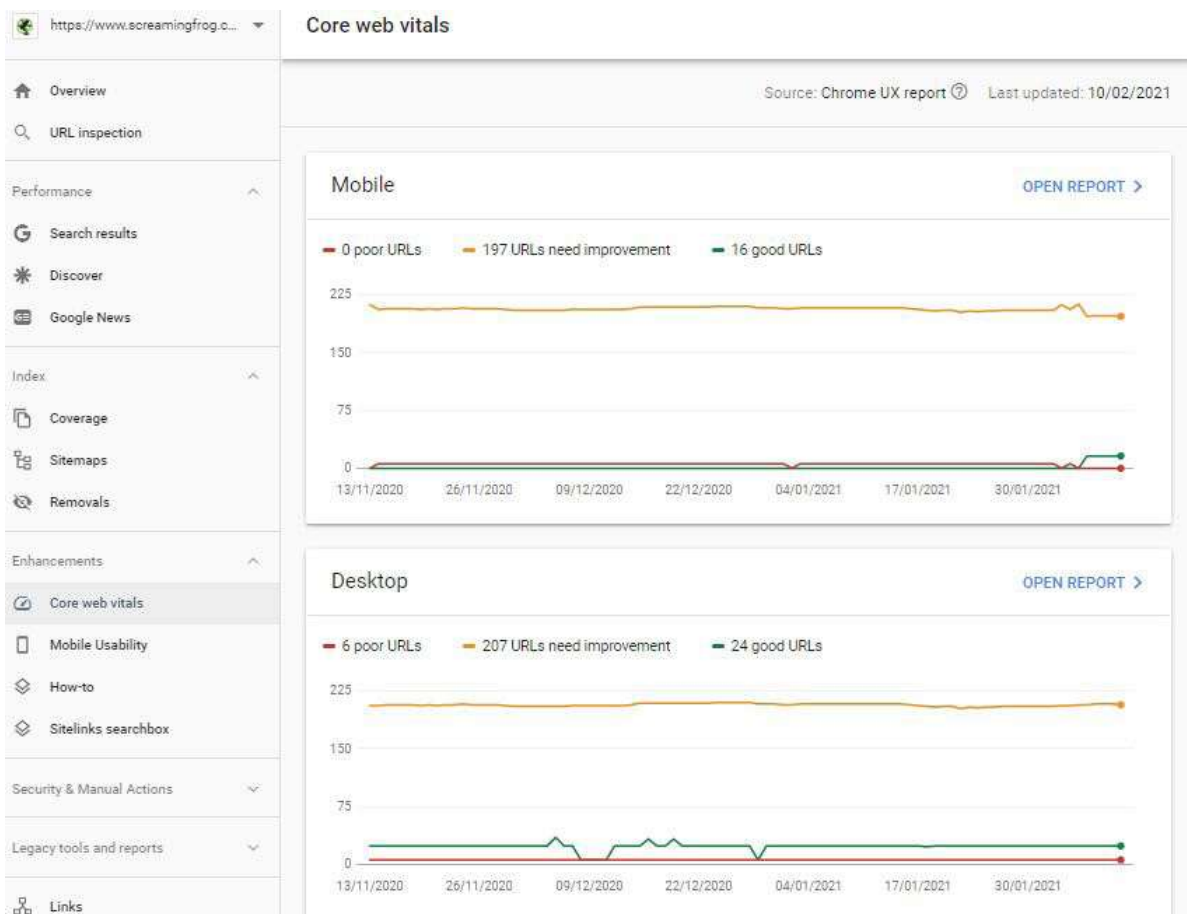
COME VISUALIZZARE I PARAMETRI VITALI DEL CORE WEB DI UNA PAGINA?

I Core Web Vitals dei siti web sono registrati e memorizzati all'interno del Chrome User Experience Report e li puoi visualizzare essenzialmente in due modi.

Per un singolo URL puoi utilizzare PageSpeed Insights (PSI), che ti segnala se la pagina ha superato o non superato i controlli.



La seconda soluzione è consultare la Search Console nella sezione "Core Vitals" che riporta il rendimento delle pagine, gli errori e gli avvertimenti in relazione a queste metriche.

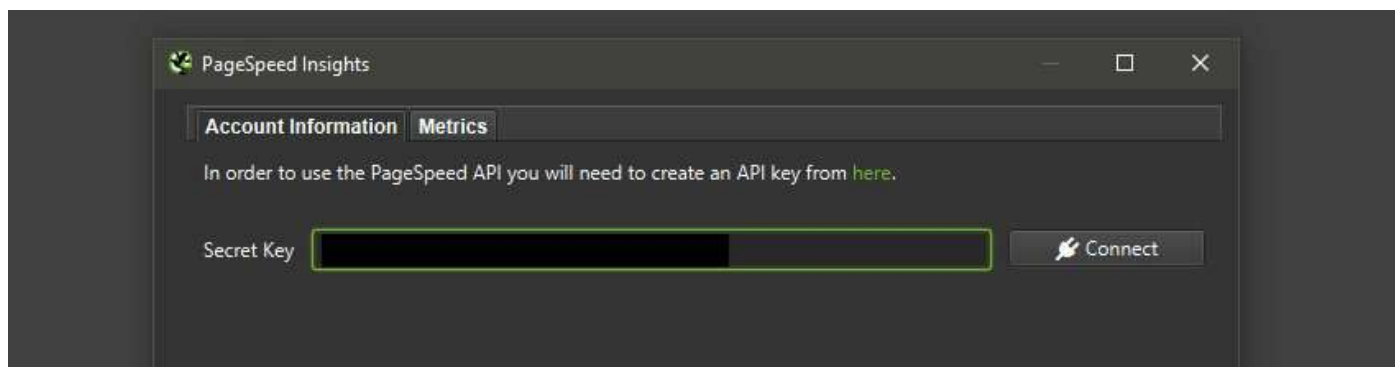


Nonostante questi due controlli siano efficaci sono altresì estremamente riduttivi perché la prima soluzione permette un controllo unitario degli URL mentre la seconda con GSC agglomera più Url (modelli di URL) e non fornisce i punteggi di qualità unitari.

ANALIZZARE I CORE WEB VITALS CON SCREAMING FROG

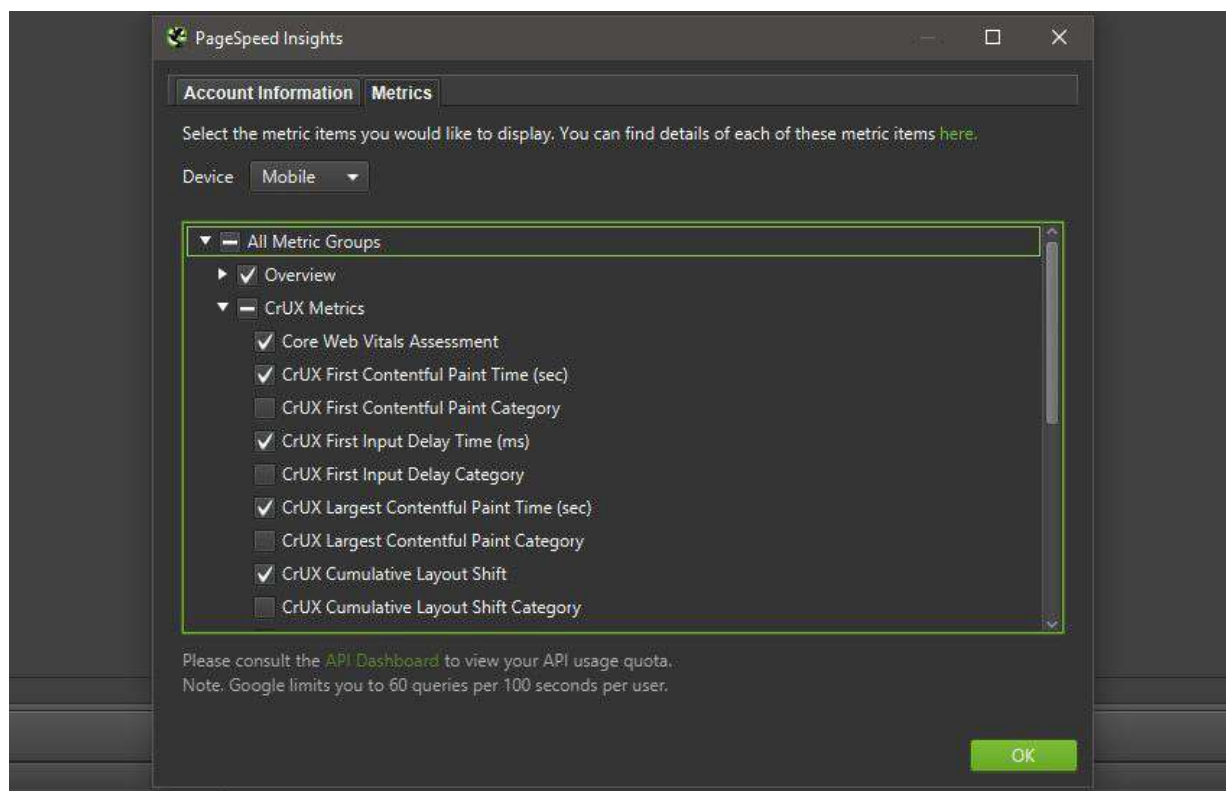
Screaming Frog SEO Spider grazie alla connessione a PageSpeed Insights API ti permette di raccogliere i dati sui Web Vital in modo semplice e ti restituisce dati specifici per ogni URL in base ai dati reali di CrUX e ai dati di laboratorio. Per impostarlo basta seguire pochi e semplici passi.

1. Connessione alle API di PageSpeed.



2. Impostazioni delle metriche da considerare.

Una volta che sei connesso a PSI devi cliccare sulla scheda "Metriche" e scegliere quali introdurre come parametro di analisi. Nelle prime scansioni dei Core Web Vitals ti consiglio di utilizzare la selezione predefinita in quanto include i dati del rapporto CrUX, i dati di laboratorio di Google Lighthouse e i dati sulle opportunità delle aree da migliorare.



3. Esecuzione della scansione.

Concluse le configurazioni sulle metriche delle API puoi avviare il crawler ed aspettare che il sito web venga scansionato e popolato con i dati relativi ai Core Web Vitals.

4. Analisi dei dati

Non ti resta che cliccare sulla scheda PageSpeed per visualizzare tutti i dettagli del crawl e quali URL abbiamo passato il test, quali devono essere migliorati e quali invece presentano degli errori gravi.

Address	PSI Status	PSI Error	Performance Score	First Contentful Paint Time (sec)	Speed Index Time (sec)	Largest Contentful Paint (sec)
1. https://www.admiral.com/	Success		32	2.287	7.022	4.834
2. https://www.admiral.com/community-and-sponsorship	Success		26	3.78	8.549	13.2
3. https://www.admiral.com/travel-insurance/cruise-insurance.php	Success		43	2.22	5.698	7.035
4. https://www.admiral.com/multi-car-insurance	Success		6	6.21	12.733	17.55
5. https://www.admiral.com/home-insurance/contents-insurance.php	Success		24	2.816	10.505	4.966
6. https://www.admiral.com/loans/personal-loans/home-improvements	Success		20	2.133	12.923	9.828
7. https://www.admiral.com/coronavirus	Success		24	2.264	6.109	10.295
8. https://www.admiral.com/multicover-insurance	Success		28	2.511	5.935	8.955
9. https://www.admiral.com/magazine	Success		25	2.112	8.386	13.98
10. https://www.admiral.com/home-insurance/landlord	Success		27	2.643	4.616	7.755
11. https://www.admiral.com/car-insurance/temporary-car-insurance	Success		23	2.117	6.943	10.29
12. https://www.admiral.com/magazine/features-and-competitions/how-to-stay-safe-online	Success		34	2.119	6.2	7.528
13. https://www.admiral.com/loans/personal-loans/loan-calculator	Success		34	2.07	7.241	8.265
14. https://www.admiral.com/loans	Success		23	3.78	7.866	10.815
15. https://www.admiral.com/home-insurance/host-insurance	Success		19	3.081	6.717	8.198
16. https://www.admiral.com/travel-insurance	Success		35	2.276	6.326	8.512
17. https://www.admiral.com/loans/personal-loans	Success		38	2.118	8.801	7.935
18. https://www.admiral.com/contact-us/complaint-information.php	Success		30	2.112	6.345	10.507
19. https://www.admiral.com/myaccount/admiral-app	Success		27	2.475	7.725	8.048
20. https://www.admiral.com/magazine/meet-the-team	Success		25	2.07	8.841	16.395
21. https://www.admiral.com/single-car-insurance	Success		13	2.94	10.976	15.345
22. https://www.admiral.com/pet-insurance/cat-insurance	Success		30	2.22	8.236	8.858
23. https://www.admiral.com/van-insurance	Success		23	2.22	8.099	8.633
24. https://www.admiral.com/car-insurance/reviews	Success		33	2.07	7.587	13.215
25. https://www.admiral.com/classic-car-insurance	Success		30	2.22	7.795	8.193
26. https://www.admiral.com/loans/car-finance/hire-purchase	Success		21	2.07	8.631	12.785
27. https://www.admiral.com/student-offers	Success		19	2.119	9.229	10.3
28. https://www.admiral.com/home-insurance/contents-insurance/landlord	Success		34	2.603	4.610	8.66

L'esempio sopra (con impostazioni di default) presenta nella prima colonna gli URL mentre nella seconda la valutazione dello stato del "Core Web Vitals Assessment". Il PSI Status è contrassegnato con "Pass" o "Fail" a seconda che sia considerato valido in tutti e tre i Web Vitals.

Nello screenshot puoi notare che l'URL in alto ha un LCP di 1.99s (sotto 2.5s), un FID di 37ms (sotto 100ms), e un CLS di 0.06 (sotto 0.1) quindi ha passato la convalida dei CWV.

Attraverso questa panoramica sei in grado di esaminare immediatamente quali siano le pagine che non hanno passato il test e su quali metriche specifiche dei CWV intervenire.

Può capitare che alcuni URL non ottengano dati relativi ai CWV, questa condizione si presenta quando non ci sono dati sufficienti sulla velocità in CrUX, o perché la pagina non ha abbastanza visitatori. Solo le pagine più popolari in genere hanno sufficienti dati per il rapporto Chrome UX.

5. Segmentazione dei modelli di errore

Ottenuti i dati puoi esaminare ogni singola prestazione dei CWV utilizzando i differenti filtri ed isolando le problematiche da ottimizzare avendo a disposizione ogni URL che presenta le medesime criticità.

		PSI Status	PSI Error	Performance Score
✓ All				
Eliminate Render-Blocking Resources	documents.php?product=travel	Success		21
Properly Size Images	ers-up-to-18-percent-per-month-better...	Success		43
Defer Offscreen Images	registrations-up-year-on-year-while-car...	Success		25
Minify CSS	nters-home-insurance	Success		32
Minify JavaScript	a-claim.php	Success		32
Remove Unused CSS	ow-to-avoid-parking-fines			21
Remove Unused JavaScript	ssan-qashqai-review			
Efficiently Encode Images	ow-to-check-a-cars-history	Success		31
Serve Images in Next-Gen Formats	late-special-new-vans-and-pick-ups	Success		28
Enable Text Compression	st-family-friendly-hatchbacks-for-5000	Success		41
Preconnect to Required Origins	ars-with-the-best-warranty	Success		24
Reduce Server Response Times (TTFB)	nt-affairs/trio-of-winners-at-Company-V...	Success		40
Avoid Multiple Page Redirects	onfidence-causes-holiday-bookings-incr...	Success		32
Preload Key Requests	documents.php?product=home	Success		22
Use Video Formats for Animated Content	nt-affairs/foreign-office-issues-travel-w...			
Avoid Excessive DOM Size	st-cycling-spots-uk	Error	500: Unable to...	
Reduce JavaScript Execution Time	nce-checklist	Error	500: Unable to...	
Serve Static Assets with an Efficient Cache Policy	ow-to-plan-an-easter-family-break			
Minimize Main-Thread Work	als	Error	500: Unable to...	
Ensure Text Remains Visible During Webfont Load	weddings-what-do-they-cost-and-where...	Success		22
Request Errors	rance/intensive-driving-courses-are-the...	Success		25
	ivelling-when-pregnant	Success		34
	ronavirus landlords tenants	Success		31
	sh-commotion-festive-season-sees-the...	Success		22
	unt-travel-information	Success		22

Evidenziando un URL nella finestra superiore e cliccando sulla tab “PageSpeed Details” (finestra inferiore) troverai tutta una serie di dettagli supplementari e opportunità per revisionare gli errori o avvertimenti riscontrati.

Name	Value	Source Page	URL	Size (Bytes)	Potential Saving
► Overview		https://www.admiral.com/	https://s.swiftypecdn.com/install/v...	112533	69624
► Core UX Metrics		https://www.admiral.com/	https://d2oh489mrk9.cloudfront...	60477	50347
► Lighthouse Metrics		https://www.admiral.com/	https://ptag.livestperson.net/ptag/a...	92757	38965
▼ Opportunities		https://www.admiral.com/	https://www.admiral.com/sites/ad...	74007	37308
Remove Unused JavaScript	1,850 ms	https://www.admiral.com/	https://www.youtube.com/player...	38233	31145
Remove Unused CSS	1,350 ms				
Eliminate Render-Blocking Resources	1,197 ms				
Serve Images in Next-Gen Formats	900 ms				
Defer Offscreen Images	750 ms				
Efficiently Encode Images	150 ms				
Diagnostics					

Selected Cells: 0 Total: 5

URL Details | Inlinks | Outlinks | Image Details | Resources | SERP Snippet | Rendered Page | View Source | HTTP Headers | Cookies | Duplicate Details | Structured Data Details | PageSpeed Details | Spelling & Grammar Details

COME OTTIMIZZARE LA METRICA LCP (LARGEST CONTENTFUL PAINT)

Iniziamo a definire quali siano le aree che possono incidere sul LCP e quali fattori possono essere ottimizzati per migliorare l'indice della metrica.

- Properly Size Images.
- Defer Offscreen Images.
- Efficiently Encode Images.
- Serve Images in Next-Gen Formats.

Il punto di partenza coinvolge le immagini che sono uno degli elementi cardine per le performance delle pagine html. Per migliorare la metrica LCP assicurati che ogni immagine sia correttamente dimensionata, e utilizza il “Lazy Loading” dove appropriato.

E' consigliabile l'utilizzo di nuovi formati di immagine più efficienti come WebP e migliorare la codifica dei file di immagine esistenti.

Un altro aspetto molto impattante coinvolge i tempi di risposta del server. Un server poco potente porta a risposte lente e a tempi di caricamento maggiori; per questo motivo ti consiglio la scelta di hosting più performanti o di utilizzare le CDN -Content Delivery Networks per ridurre i tempi di risposta e migliorare il TTFB.

Tra le possibili ottimizzazioni da considerare poni molta attenzione nella gestione delle risorse che bloccano il rendering della pagina. Come ben sai mentre l'HTML viene analizzato, ogni riferimento ai fogli di stile o CSS può causare una pausa del “parser”, ritardando l'LCP.

Idealmente, dovresti rinviare qualsiasi CSS e JavaScript non critico per accelerare il caricamento del tuo contenuto principale e utilizzare il “preload” per le risorse critiche. Un'altra tips è quella di minificare i CSS e Javascript e rimuovere tutti i file non necessari dal caricamento della tua pagina, considera inoltre di utilizzare il “preconnect” per risorse di terze parti (es. Google Analytics, Tag Manager etc.).

- Minify CSS.
- Minify JavaScript.
- Remove Unused CSS.
- Remove Unused JavaScript.
- Enable Text Compression.
- Preconnect to Required Origins.
- Preload key Requests.
- Use Video Formats for Animated Content.

COME OTTIMIZZARE IL FID (FIRST INPUT DELAY)

Il FID nella maggior parte dei casi viene influenzato negativamente da lunghi tempi di esecuzione di script di grandi dimensioni. La prima azione da intraprendere è quella di minificare e rimuovere le parti degli script inutilizzate, (o interi script) riducendo il tempo speso per l'esecuzione e il rendering di questi file.

- Minify JavaScript.
- Remove Unused JavaScript.
- Reduce the Number of Third-Party Scripts Used.

Per scovare l'elenco delle risorse di terze parti puoi consultare il tuo documento HTML all'interno della scheda PageSpeed Details.

Un'altra ottimizzazione da considerare riguarda il lavoro del thread principale. Se un browser è impegnato nel rendering del thread principale, l'utente non sarà in grado di compiere delle interazioni con la pagina fino a quando questo non sarà completato. Se riduci il tempo di esecuzione del thread principale, le interazioni dell'utente saranno più reattive ed immediate.

Le soluzioni principali riguardano la riduzione dell'esecuzione degli script Javascript e la riduzione del Main Thread.

- Minimise Main-Thread Work
- Reduce JavaScript Execution Time

Considera poi che ogni risorsa necessaria per una pagina necessita tempo per essere richiesta, scaricata ed eseguita. Riducendo sia il numero di richieste necessarie che la dimensione totale del download, aiuterai a migliorare i tempi di esecuzione della pagina.

Le attività utili a tale scopo sono la minificazione dei CSS e Javascript, la rimozione di fogli di stile/CSS e Javascript inutilizzati e la compressione del testo.

- Lower Resource Counts & Sizes.
- Minify CSS.
- Minify JavaScript.
- Remove Unused CSS.
- Remove Unused JavaScript.
- Enable Text Compression.
-

COME OTTIMIZZARE IL CLS (CUMULATIVE LAYOUT SHIFT)

Come già accennato in precedenza la stabilità del layout è la componente principale per il CLS per garantire un'esperienza di rilievo al navigatore, questo aspetto deve farti concentrare sul servire la pagina nell'ordine appropriato e con una spaziatura definita per ogni risorsa per non creare sbalzi nella navigazione.

Per ottenere dei risultati soddisfacenti devi garantire che tutte le immagini, gli annunci, i video e gli iframe presentino gli attributi <height> e <width> per permettere che le pagine vengano caricate con la spaziatura appropriata ed evitare lo spostamento degli altri contenuti quando vengono aggiunte. Un'altra considerazione in ottica di CLS è quella di dare priorità di caricamento alle risorse dall'alto verso il basso per non far rimbalzare verso l'alto l'utente che magari ha raggiunto un certo livello di profondità di navigazione.

Assicurati che i font personalizzati non causino l'effetto FOIT/FOUT, infatti se i font personalizzati sono applicati in ritardo nel caricamento di una pagina si può verificare che il testo lampeggi mentre viene sostituito (FOUT), o che il testo invisibile venga visualizzato fino a quando il font personalizzato non viene reso (FOIT). Per ovviare a questa latenza ti consiglio di utilizzare il "preload" delle risorse critiche.

1. Se i tuoi URL non mostrano alcun dato CrUX e lo stato PSI elenca “successo”, allora è probabile che l'URL non riceva abbastanza visite di utenti reali per generare dati di velocità sufficienti per il rapporto CrUX. Se dopo il crawl con Screaming Frog ti trovi in questa condizione potresti utilizzare per l'analisi una pagina con layout simile che presenti dati CrUX, o utilizzare i dati di laboratorio simulati (presenti di default nella configurazione del Seo Spider).
2. Quando devi utilizzare i dati di laboratorio, l'unica metrica non disponibile è il FID - First Input Delay per la mancanza dell'interazione richiesta dall'utente. In questo caso puoi utilizzare come indicatore il Total Blocking Time (TBT) per vedere quali risorse vengono bloccate e di conseguenza considerarle come criticità che può far aumentare il FID.
3. Nel caso tu debba analizzare un sito di grandi dimensioni potresti superare la quota limite di PageSpeed Insight. Se ti trovi in questa condizione ti consiglio di analizzare un campione significativo o aspettare che la quota si azzeri.
4. Non dimenticarti che i dati di CrUX si basano sugli ultimi 28 giorni di interazioni con gli utenti, quindi qualsiasi modifica al sito impiegherà 28 giorni per riflettersi in CrUX.

COME FARE UN SEO AUDIT SUI REDIRECT

Il codice di risposta 3xx identifica i reindirizzamenti che rappresentano un fattore molto importante da gestire nell'ottimizzazione di un sito web per salvaguardare il Crawling Budget concesso dal Motore di Ricerca.

Attraverso Screaming Frog sei in grado di valutare il codice di stato (se un 301, 302, 303 ecc.), lo stato (permanente o temporaneo), la destinazione del reindirizzamento e se sono presenti catene di redirect o loop.

Per analizzare i Redirect con il Seo Spider puoi seguire due modalità:

- **Crawl A Site To Check Redirects:** scansionando l'intero sito.
- **Upload a List of URLs to Check Redirects:** aggiungendo in modalità "List" un elenco di URL sui quali vuoi controllare la presenza di reindirizzamenti.

CRAWL A SITE TO CHECK REDIRECTS

1. La prima azione da intraprendere è quella di eseguire la scansione del tuo sottodominio in modalità "Spider".

The screenshot displays the Screaming Frog SEO Spider interface. The main table shows a list of URLs crawled, all with a status code of 200. The table columns include Address, Content, Status Code, Status, and Title. The right sidebar provides an overview of the site structure, including a list of response codes and a donut chart showing the distribution of response codes: HTML (100.00%), JavaScript (0.00%), CSS (0.00%), Images (0.00%), PDF (0.00%), Flash (0.00%), and Other (0.00%).

2. Consultazione dei Reindirizzamenti

Per esaminare i dati ti basta cliccare sulla scheda "Response Code" e impostare il filtro "Redirection (3XX)".

Nella prima colonna saranno presenti gli URL con reindirizzamenti, in quelle successive troverai i dettagli sulla tipologia di redirect (301,302 etc.) e la pagina di destinazione (colonna "redirect URL").

File Configuration Mode Bulk Export Reports Sitemaps Licence Help

Screamingfrog

www.screamingfrog.co.uk

Internal External Protocol Response Codes URI Page Titles Meta Description Meta Keywords H1 H2 Images Directives Hreflang

Filter: Redirection (3xx) Export

	Address	Content	Status Code	Status	
1	http://2012.takingpointsmemo.com/2011/09/search-engine-exper...	text/html	301	Moved Permanently	0
2	http://adword.blogspot.co.uk/2012/04/new-changes-to-ad-rotatio...	text/html; charset=UTF-8	301	Moved Permanently	0
3	http://adword.blogspot.co.uk/2013/05/introducing-keyword-plann...	text/html; charset=UTF-8	301	Moved Permanently	0
4	http://adword.blogspot.co.uk/2013/10/introducing-shopping-cam...	text/html; charset=UTF-8	301	Moved Permanently	0
5	http://adword.blogspot.co.uk/2014/02/shopping-campaigns-avail...	text/html; charset=UTF-8	301	Moved Permanently	0
6	http://adword.blogspot.com/2010/07/announcing-release-of-upgr...	text/html; charset=UTF-8	302	Moved Temporarily	0
7	http://adword.blogspot.com/2010/09/product-extensions-now-av...	text/html; charset=UTF-8	302	Moved Temporarily	0
8	http://adwords.blogspot.com/2011/01/adwords-reports-get-more-...	text/html; charset=UTF-8	302	Moved Temporarily	0
9	http://blog.anthonynolan.org/2015/09/types-of-blood-cancers-and...		307	HSTS Policy	0
10	http://blog.anthonynolan.org/2015/11/marathons-worth-travelling-...		307	HSTS Policy	0

Nella colonna "Redirect URL" hai a disposizione solamente la destinazione finale del redirect ma questo dato lo dovrai integrare con il report "Redirect Chain" che ti permette di visualizzare l'esistenza o meno di catene di redirect o loop intermedi.

Per ottimizzare i tuoi redirect puoi consultare la tab "Inlinks". Ti basta cliccare su un URL nella finestra superiore e poi sulla scheda "Inlinks" per popolarla.

Dalla scheda Inlinks puoi visualizzare la pagina dove sono stati trovati dei redirect (colonna "From"), la pagina di destinazione (colonna "To") e ulteriori dettagli rispetto ad attributi "follow/No follow" etc.

3. Esportazione dei dati raccolti

Consolidati i dati sei in grado di esportare il rapporto "redirection 3xx Inlinks".

Bulk Export" > Response Code > Redirection 3xx Inlinks

File Configuration Mode Bulk Export Reports Sitemaps Licence Help

Screamingfrog

Internal External Protocol

Filter: Redirection (3xx)

1 http://www.screamingfrog.co

2 http://www.java.com/en/dow

3 http://www.dell.co.uk/

4 http://stackoverflow.com/a/9

5 http://docs.aws.amazon.com

6 http://bugs.java.com/bugdata

7 http://searchengineland.com

8 https://plus.google.com/1183

9 http://www.which.co.uk/

10 http://searchengineland.com/google-adwords-key

11 https://support.apple.com/kb/DL1572

12 https://www.google.com/webmasters/tools/robots-

13 http://java.com/en/download/help/sysreq.xml

14 https://maxxeight.com/blog/google-cache-seo/

15 http://searchengineland.com/google-adwords-qua

16 https://www.en.advertisercommunity.com/t5/AdW

17 http://www.justtheflight.co.uk/blog/16-40-tourist-scams-to-avoid-L...

All Inlinks

All Outlinks

All Anchor Text

All Images

Screenshots

All Page Source

External Links

Response Codes

Directives

Images

Custom

www.screamingfrog.co.uk

URI Page Titles Meta Description Meta Keywords H1 H2 Images Directives Hreflang

Content

Status Code

Status

307

HSTS Policy

0

301

Moved Permanently

0

text/html; charset=UTF-8

302

Moved Temporarily

0

307

HSTS Policy

0

301

Moved Permanently

0

302

Found

0

301

Moved Permanently

0

302

Found

0

301

Moved Permanently

0

301

Moved Permanently

0

302

Moved Temporarily

0

302

Moved Temporarily

0

301

Moved Permanently

0

301

Moved Permanently

0

301

Moved Permanently

0

302

Moved Temporarily

0

text/html; charset=UTF-8

301

Moved Permanently

0

Blocked by Robots.txt Inlinks

Blocked Resource Inlinks

No Response Inlinks

Success (2xx) Inlinks

Redirection (3xx) Inlinks

Redirection (JavaScript) Inlinks

Redirection (Meta Refresh) Inlinks

Client Error (4xx) Inlinks

Server Error (5xx) Inlinks

UPLOAD A LIST OF URLS TO CHECK REDIRECTS

La seconda possibilità per controllare e gestire i reindirizzamenti prevede l'utilizzo della modalità "list" del Seo Spider aggiungendo manualmente o copiando un elenco definito di URL.



Anche in questo caso come per il precedente il processo rimane il medesimo.

HOW TO:

LA MIGRAZIONE CON SCREAMING FROG

Questo articolo ti indica quale sia la migliore soluzione per scoprire e risolvere le più comuni criticità durante l'attività di migrazione.

Ogni migrazione è sempre un'impresa epica e prevede diverse fasi per non impattare negativamente sul posizionamento organico del sito web.

PREMIGRAZIONE

Prima di addentrarsi in questo processo ci sono degli step essenziali da mettere in campo nella fase di pre-migrazione in cui un crawler può aiutarti e che non dovrebbero essere trascurati.

1. Eseguire un Crawl del sito attuale

In primo luogo, ti consiglio una scansione completa del sito web per:

- poterlo confrontare con il sito di staging;
- Controllare se i vecchi URL sono stati correttamente indirizzati alle nuove risorse.
- Avere un backup da cui partire nel caso vada qualcosa storto durante la migrazione.

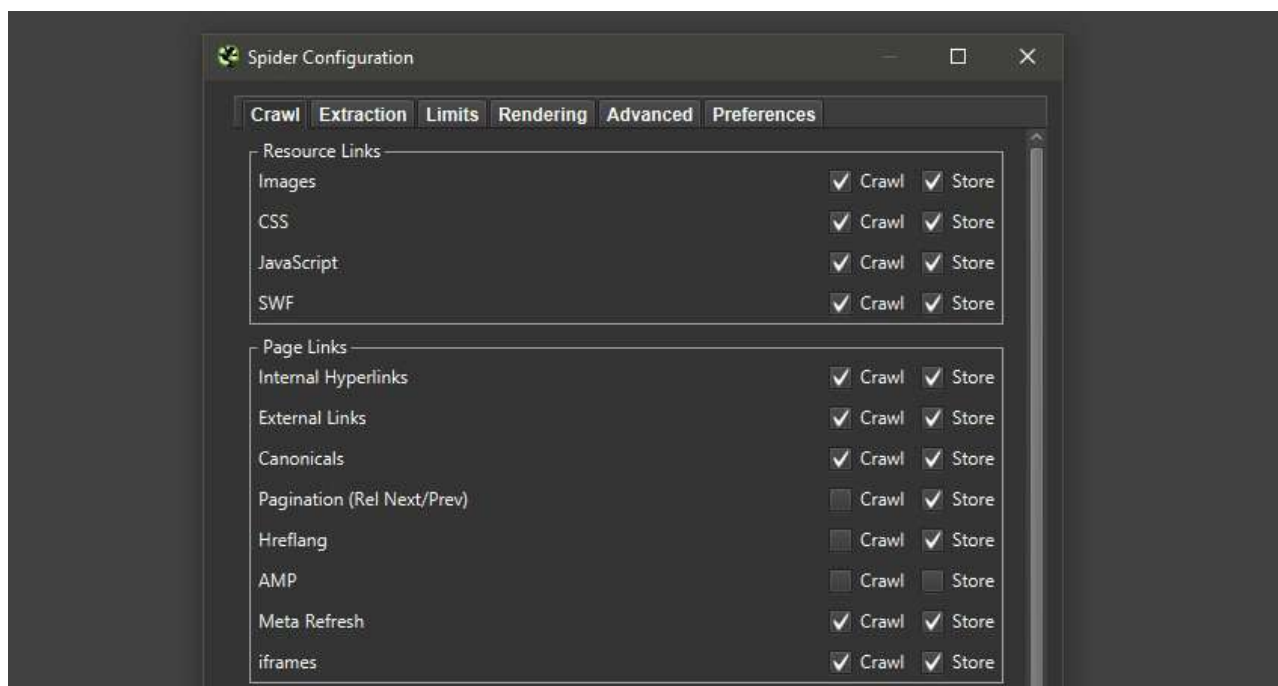
In sintesi prima della migrazione devi preparare il terreno per un backup completo che includa tutti gli elementi più importanti del sito attuale.

Per sfruttare Screaming Frog in questa fase devi abilitare alla scansione tutte le risorse a cui potresti aver bisogno di fare riferimento in futuro come le versioni AMP, gli Hreflang, i dati strutturati etc.

In questa lista della spesa non scordarti se ci sono dipendenze Javascript lato client che richiedono un rendering Javascript ed includi qualsiasi altro dominio, sottodominio o variante di sito mobile e desktop.

Integra sempre nella scansione la Sitemap XML e connettiti tramite API a Google Analytics e Search Console per scovare eventuali pagine orfane (abilita "Crawl New URLs Discovered GA/GSC").

Raccogli inoltre i dati relativi ai link da Ahrefs, Majestic, Moz e le performance sulle prestazioni delle query con GSC per tenere d'occhio le pagine importanti in base a entrate, conversioni, traffico, impressioni, clic che ti permettono di non disperdere il ranking conquistato.



2. Esportare gli URL da reindirizzare

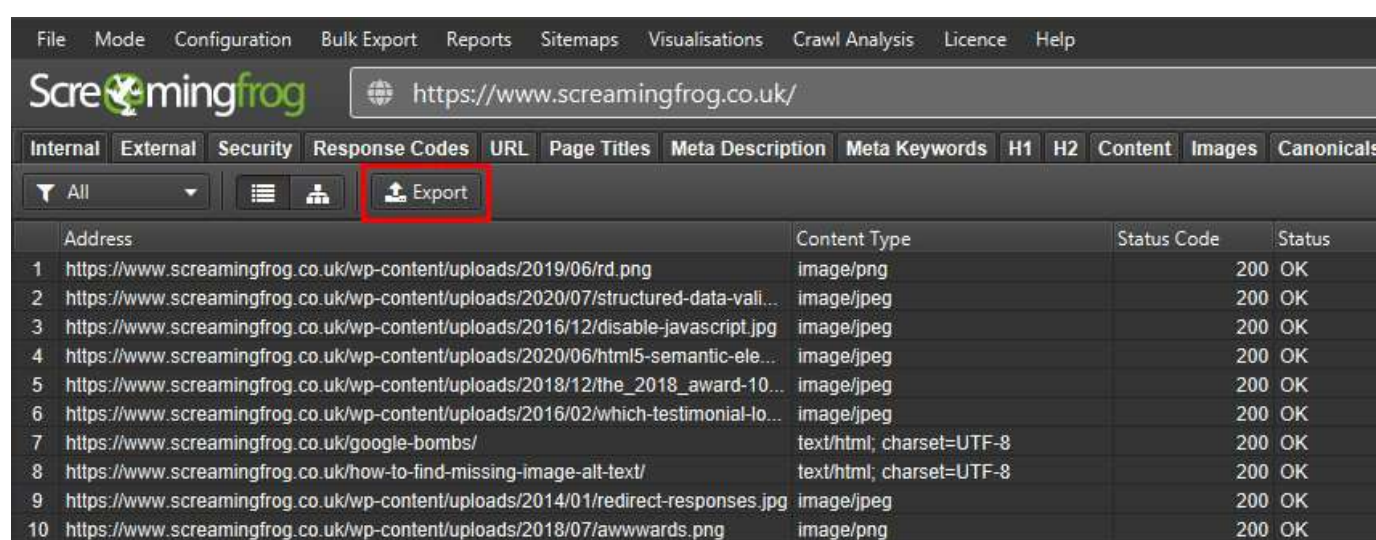
Molto spesso durante le migrazioni gli elementi che variano con maggior frequenza sono gli slug, per questo motivo è necessario raccogliere una lista completa degli URL per reindirizzare in modo permanente le vecchie pagine a quelle equivalenti del nuovo sito web.

Questa fase può essere relativamente semplice se si tratta di una migrazione di dominio e il percorso dell'URL rimane invariato ma diventa complessa e rischiosa se l'intera struttura dell'URL del sito web sta cambiando o semplicemente alcune aree strategiche sono state ripensate in base a nuovi Search Intent.

Per ottenere un elenco completo degli URL esistenti è sufficiente esportarli dal crawl eseguito del sito live cliccando su "Esporta" nella scheda "Internal".

Questa attività di mappatura è una delle fasi cruciali della tua migrazione per reindirizzare ogni elemento nel modo più appropriato.

Se una mappatura completa non è possibile, è buona norma dare una scala di priorità alle pagine top-performing in base alle conversioni, l'utente e link raccolti nella fase 1.



3. Fase di test pre-migrazione

Quando il nuovo sito web è in “Staging” è giunta l’ora di iniziare i controlli prima della messa online definitiva e anche in questo caso il crawler ti viene in aiuto:

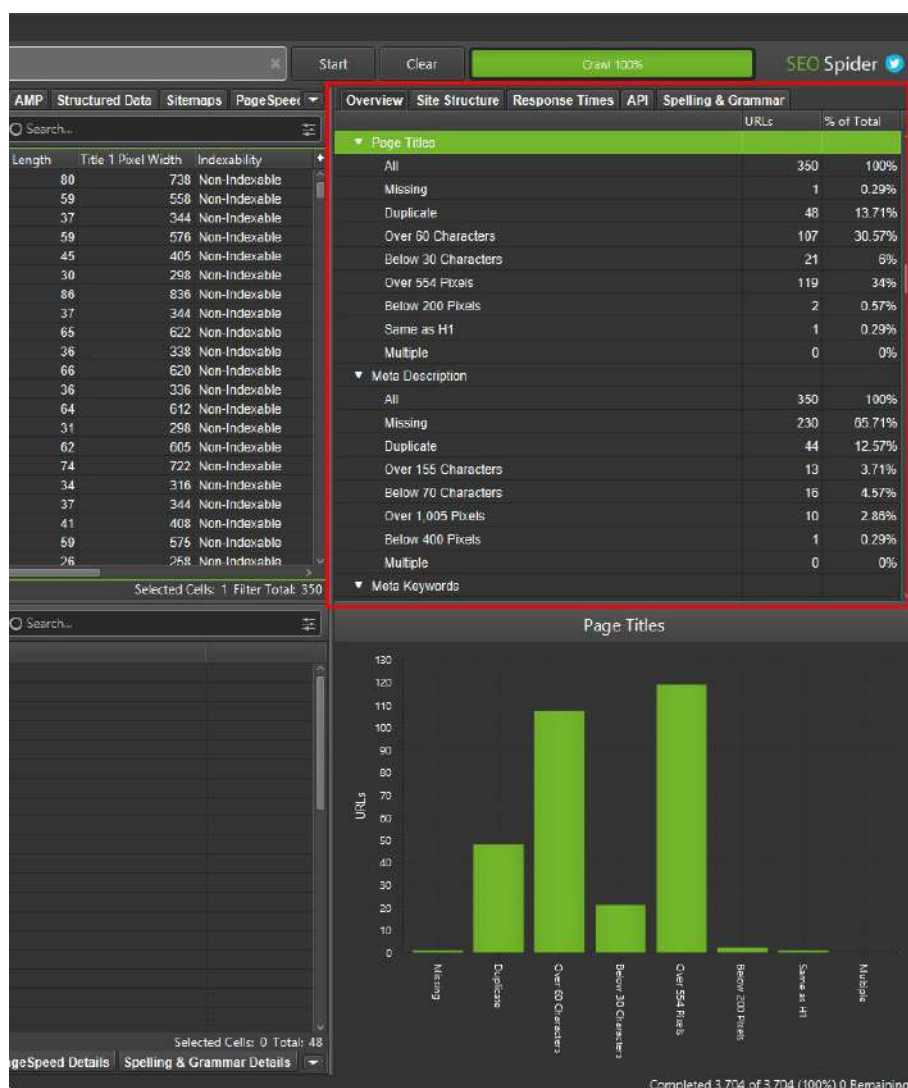
- La prima attività è quella di eseguire un crawl sul sito in staging per identificare differenze, problemi e opportunità (come fatto per il sito live).

Normalmente i siti in questa fase sono inibiti alle scansioni dal file robots.txt o da password per non essere considerati dallo Spider dei Motori di Ricerca. Grazie a Screaming Frog puoi forzare l’autenticazione, ignorare il file robots.txt o, cambiando le configurazioni, gestire la scansione di risorse lasciate in “noindex” o “nofollow” perché ancora in fase di sviluppo (vedi sezione dedicata “How to: Scansionare un sito protetto da password”).



Address	Content Type	Status Code	Status	Indexability	Indexability Status
1 https://screamingfrogcouk.kinsta.cloud/		0	Blocked by robots.txt	Non-Indexable	Blocked by robots.txt

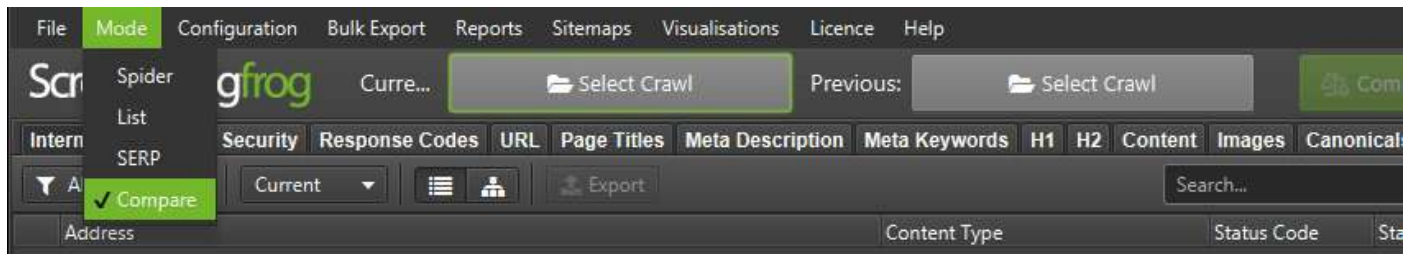
Una volta completata la scansione del sito in produzione puoi iniziare l’analisi utilizzando inizialmente la scheda “Overview” della sidebar laterale.



4. Confronto fra sito in staging e sito Live (URL MAPPING)

Adesso sei in possesso di tutti i dati che ti servono: da un lato hai scansionato il sito attualmente online e dall'altro il sito in staging, non ti resta che comparare le due versioni per scoprire eventuali incongruenze.

Questa attività la puoi fare semplicemente con il Seo Spider utilizzando "URL Mapping" della modalità "Compare" e definendo dalla barra in alto le due versioni di crawl da confrontare.

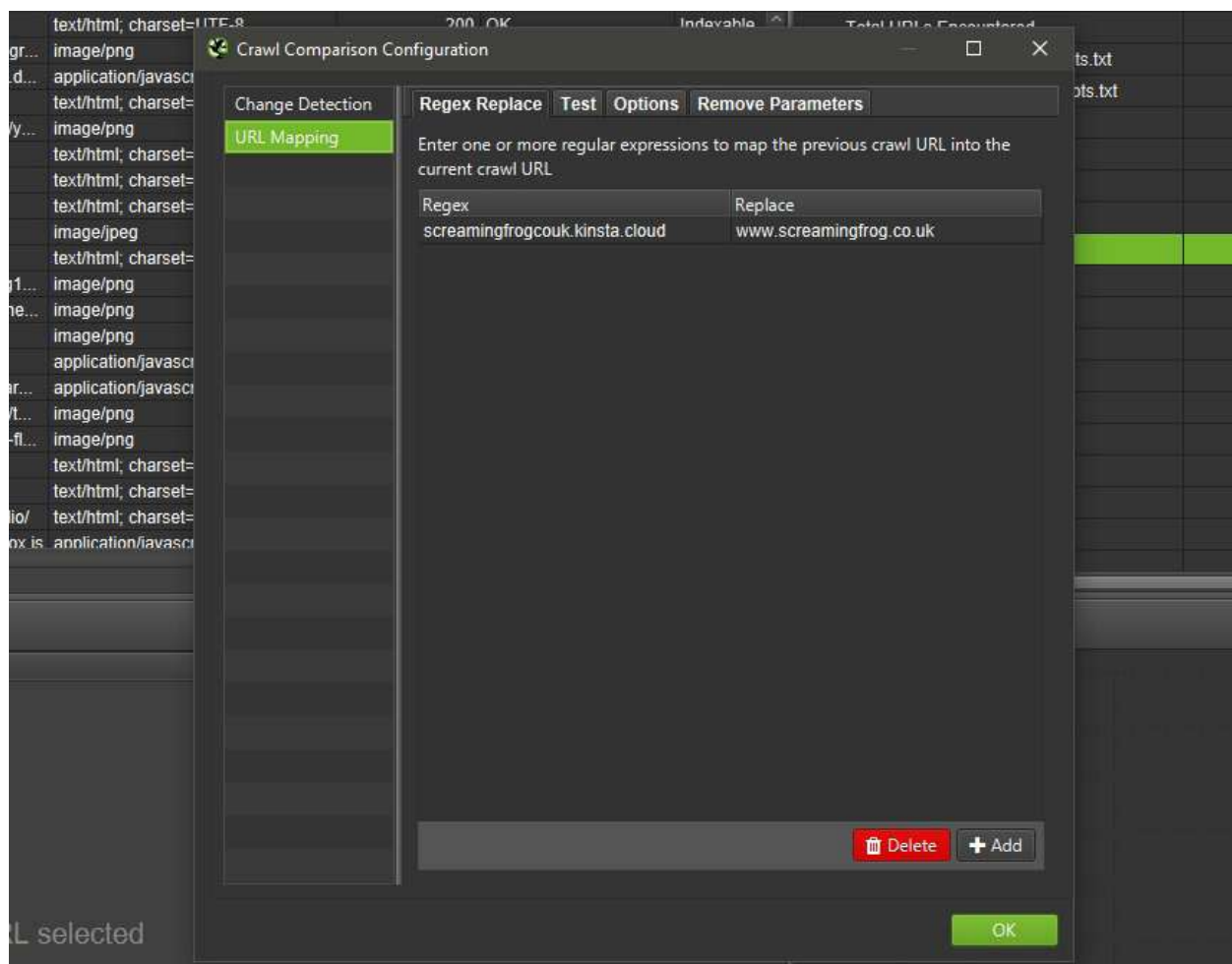


Questa opportunità introdotta con gli ultimi aggiornamenti del Seo Spider, ti consente di confrontare diverse strutture di URL, come diversi nomi di host o directory utilizzando le regole RegEX.

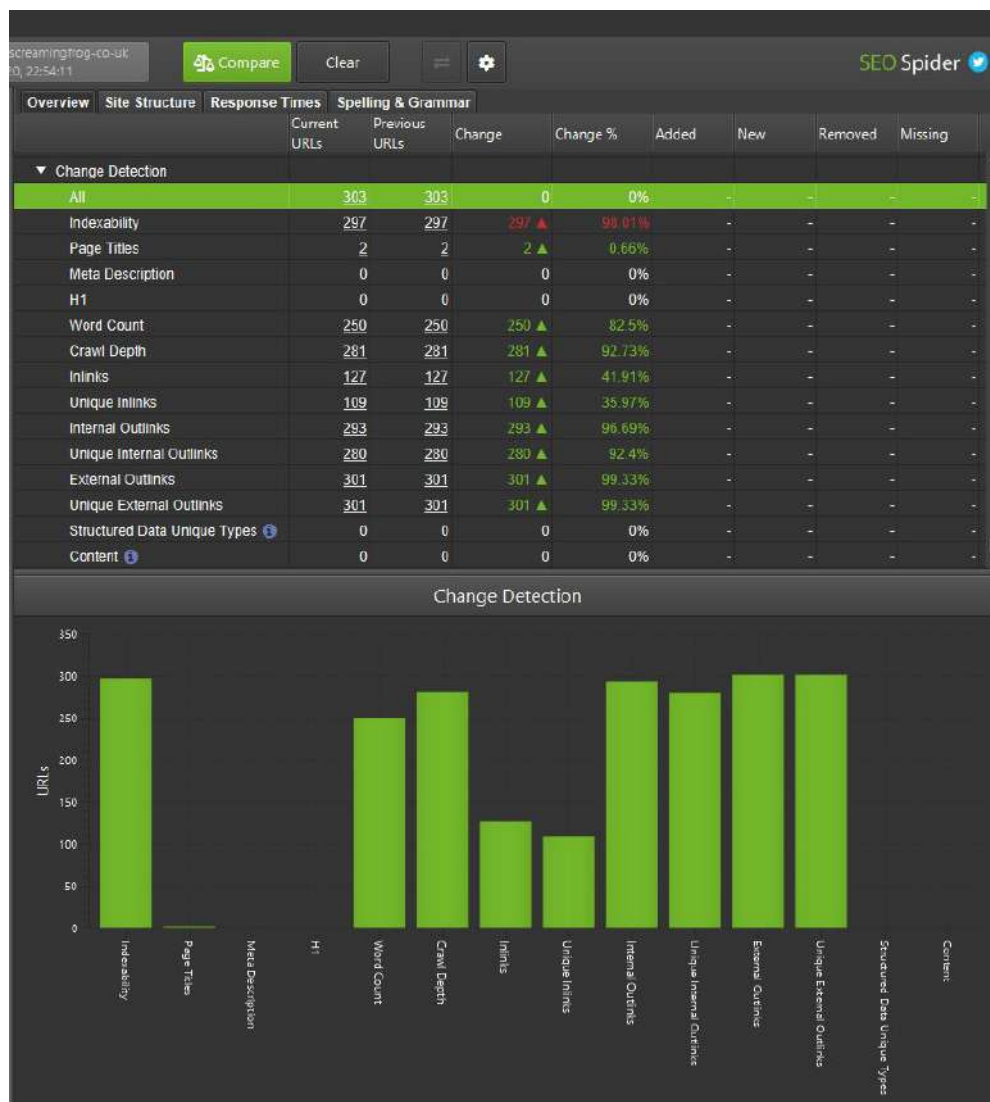
Per utilizzarla è sufficiente cambiare la modalità di scansione su "Compare", scegliere i due crawl salvati e cliccare sull'icona con "l'ingranaggio" della barra in alto.

Sul popup "Crawl Comparison Configuration" non devi far altro che cliccare su "URL Mapping" e definire il vecchio dominio con quello nuovo tramite regole RegEX.

Nell'esempio di seguito il sito esistente è stato mappato con la versione di "Staging" modificando l'hostname considerando che condividono gli stessi percorsi degli URL.



Grazie alla mappatura gli URL equivalenti vengono confrontati tra loro per ottenere i dati della scheda panoramica, i problemi e le opportunità, la scheda della struttura del sito e il rilevamento dei cambiamenti.



Terminato il confronto è possibile cliccare sulle colonne per vedere quali URL sono cambiati e usare il filtro sulla finestra principale per alternare tra le scansioni attuali e precedenti, così come per verificare URL aggiunti, mancanti o con elementi nuovi o rimossi.

Il Seo Spider comprende quattro colonne che aiutano a segmentare gli URL modificati secondo questi filtri:

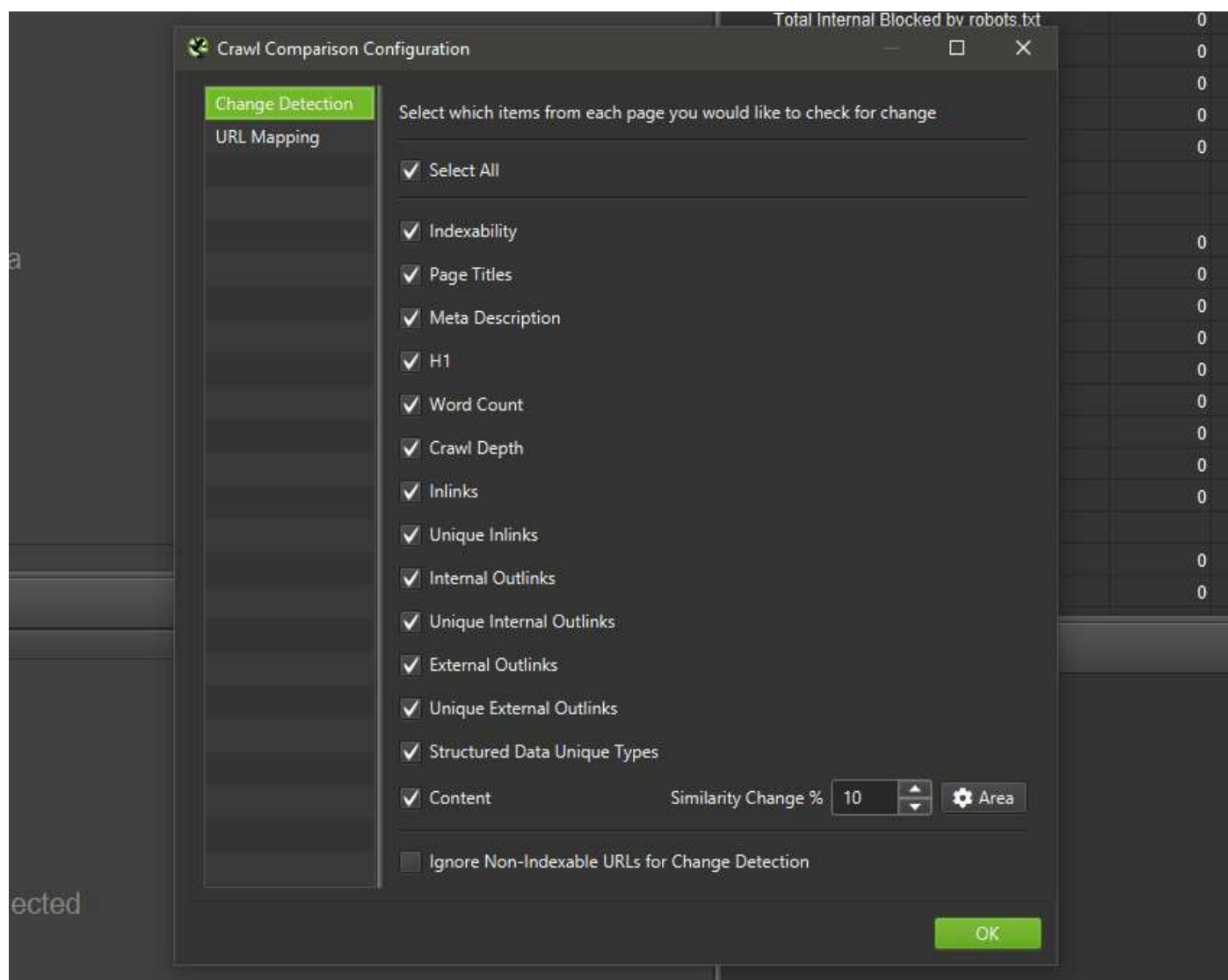
- **Added:** identifica gli URL che sono presenti nel sito esistente e contestualmente in quello di staging ma che differiscono a livello di filtro.
Ad esempio considerando il sito live come "Current" e quello di Staging come "Previous" e mettendo il focus sugli heading "h1" questo filtro ti indica tutti gli URL che presentano solamente nel sito "Live" un heading ottimizzato.
- **New:** visualizza tutti gli URL che non sono presenti nel sito web esistente, ma sono stati introdotti nel nuovo sito web.
- **Removed:** mostra tutti gli URL che sono presenti nel filtro della scansione del sito attuale ma non nel filtro nel nuovo sito in staging.
- **Missing:** URL che esistono in entrambe le scansioni ma che differiscono a livello di filtro.
Ad esempio considera il sito live come "Current" nel quale non è presente il "meta Title" di uno specifico URL mentre nel sito di staging (considerato nell'esempio come "Previous") questo elemento era stato ottimizzato.

5. Identificare le differenze tra i siti tramite il “Change Detection”

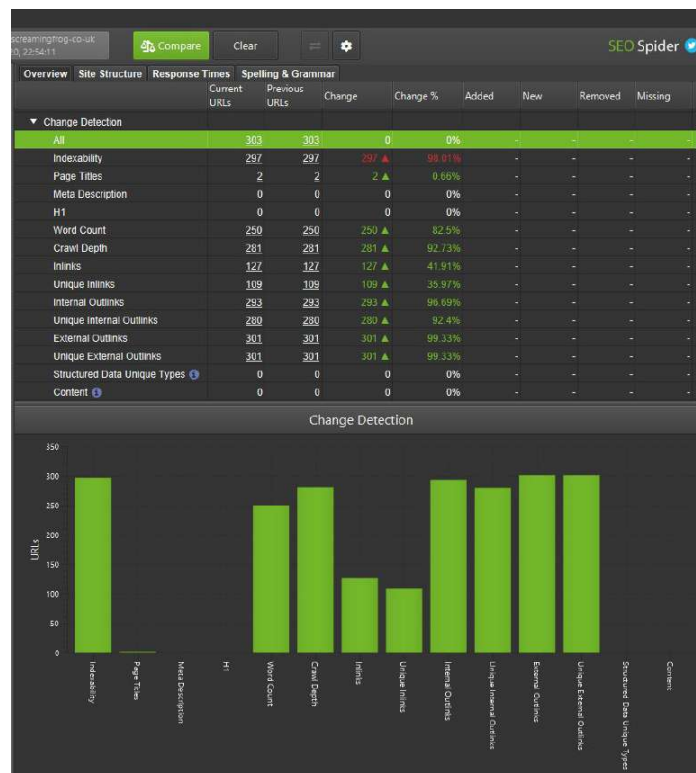
Quando si esegue una migrazione è buona norma ridurre al minimo il numero di cambiamenti che potrebbero avere un impatto sulla visibilità di un sito.

Per tenere sotto controllo questi cambiamenti puoi usare la funzione “Change Detection” in modalità “Compare” che ti permette di visualizzare tutti degli elementi che sono stati modificato fra il sito attuale e quello di “staging”.

Per attivare la “Change Detection” clicca sulla configurazione di confronto “Config > Compare” (o l'icona “ingranaggio” in alto) e seleziona gli elementi e le metriche in cui vuoi esaminare i cambiamenti.



Successivamente, è sufficiente fare clic su “Compare” per eseguire la scansione. La scheda ‘Change Detection’, non presente di default, apparirà come ultima scheda nella sidebar laterale e nella vista principale del Seo Spider.



Nell'esempio qui sopra, 2 URL hanno cambiato il titolo della pagina, 250 sono cambiati nel numero di parole mentre le meta description e gli Inlinks sono rimasti invariati etc. In questa schermata sei in grado di cliccare su ogni elemento elencato per isolarlo e vedere i relativi cambiamenti. Nell'esempio qui sotto sono stati isolati i "Meta Title" per analizzare la validità semantica dei tag.

Screamingfrog			
Current: Kinsta Staging 14 Feb 2021, 10:32:49		Previous: http-www-screamingfrog-co-uk 21 Nov 2020, 22:54:11	
Page Titles	Changed	Previous	Export
Address	Current Page Title	Previous Page Title	
1 https://www.screamingfrog.co.uk/seo-spider/	Screaming Frog SEO Spider Website Crawler	Screaming Frog SEO Spider Tool & Crawler Software	
2 https://www.screamingfrog.co.uk/10-years/	Screaming Frog 10 Year Anniversary Giveaway Screaming Frog	Screaming Frog 10 Year Anniversary Giveaway Screaming FrogScrea...	

Questa funzionalità è molto utile soprattutto se la struttura del sito e i collegamenti interni stanno cambiando considerevolmente; grazie all'URL Mapping sei in grado di identificare con semplicità quali pagine hanno perso (o guadagnato link), e cambiato la profondità di scansione.

Screamingfrog			
Current: Kinsta Staging 14 Feb 2021, 10:32:49		Previous: http-www-screamingfrog-co-uk 21 Nov 2020, 22:54:11	
Unique Inlinks	Changed	Current	Export
Address	Current Unique Inlinks	Previous Unique Inlinks	Change
1 https://screamingfrogco.uk.kinsta.cloud/how-to-keep-up-with-the-daily-news...	5	198	-183 ▼ -97.46%
2 https://screamingfrogco.uk.kinsta.cloud/passing-the-cvv-assessment/	6	198	-182 ▼ -96.97%
3 https://screamingfrogco.uk.kinsta.cloud/turn-data-into-content-marketing-ca...	202	303	-101 ▼ -33.34%
4 https://screamingfrogco.uk.kinsta.cloud/facebook-catalogue-seo-spider/	202	303	-101 ▼ -33.34%
5 https://screamingfrogco.uk.kinsta.cloud/blog/page/14/	4	14	-10 ▼ -71.43%
6 https://screamingfrogco.uk.kinsta.cloud/blog/page/10/	10	12	-2 ▼ -16.67%
7 https://screamingfrogco.uk.kinsta.cloud/five-ways-seo-has-changed/	5	6	-1 ▼ -16.67%
8 https://screamingfrogco.uk.kinsta.cloud/sources-of-content-inspiration/	8	9	-1 ▼ -11.12%
9 https://screamingfrogco.uk.kinsta.cloud/screaming-frog-wins-best-seo-cam...	8	9	-1 ▼ -11.12%
10 https://screamingfrogco.uk.kinsta.cloud/screaming-frog-seo-spider-update...	9	10	-1 ▼ -10%
11 https://screamingfrogco.uk.kinsta.cloud/log-file-analyser-3-0/	8	9	-1 ▼ -11.12%
12 https://screamingfrogco.uk.kinsta.cloud/10-sistrix-features-Hove/	8	9	-1 ▼ -11.12%
13 https://screamingfrogco.uk.kinsta.cloud/seo-spider-6-0/	8	9	-1 ▼ -11.12%
14 https://screamingfrogco.uk.kinsta.cloud/screaming-frog-seo-spider-update...	5	4	1 ▲ 25%
15 https://screamingfrogco.uk.kinsta.cloud/screaming-frog-seo-spider-update...	14	13	1 ▲ 7.69%
16 https://screamingfrogco.uk.kinsta.cloud/author/screamingfrog/	65	64	1 ▲ 1.56%
17 https://screamingfrogco.uk.kinsta.cloud/screaming-frog-seo-spider-update...	11	10	1 ▲ 10%
18 https://screamingfrogco.uk.kinsta.cloud/seo-spider-1-1/	8	7	1 ▲ 14.28%
19 https://screamingfrogco.uk.kinsta.cloud/seo-spider-1-3/	5	4	1 ▲ 25%
20 https://screamingfrogco.uk.kinsta.cloud/accurate-google-keyword-data/	13	12	1 ▲ 8.33%
21 https://screamingfrogco.uk.kinsta.cloud/audit-backlinks/	7	6	1 ▲ 16.66%

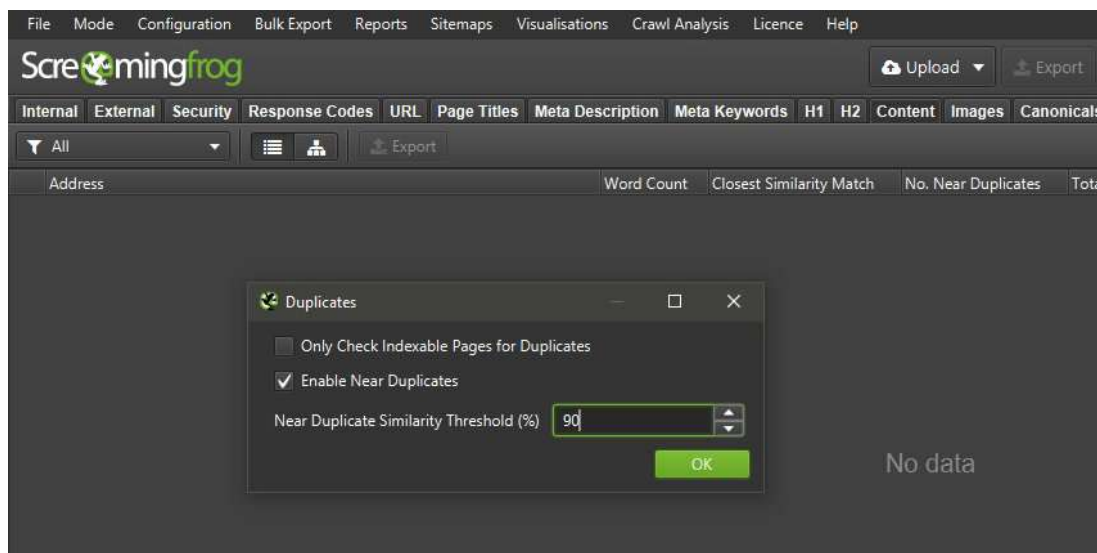
Showing: URLs in current crawl

6. Redirect Mapping (Opzionale)

Anche se non è stato costruito per questo scopo, la funzione di “near duplicates - quasi duplicati” la puoi utilizzare per fare la mappatura dei redirect dai vecchi URL ai nuovi URL se il contenuto è molto simile.

Per poterne beneficiare è sufficiente passare alla modalità “List -elenco” (‘Mode-Modalità > List-Elenco’), abilitare la funzione “Near Duplicates” e disabilitare “Only Check Indexable Pages For Duplicates”.

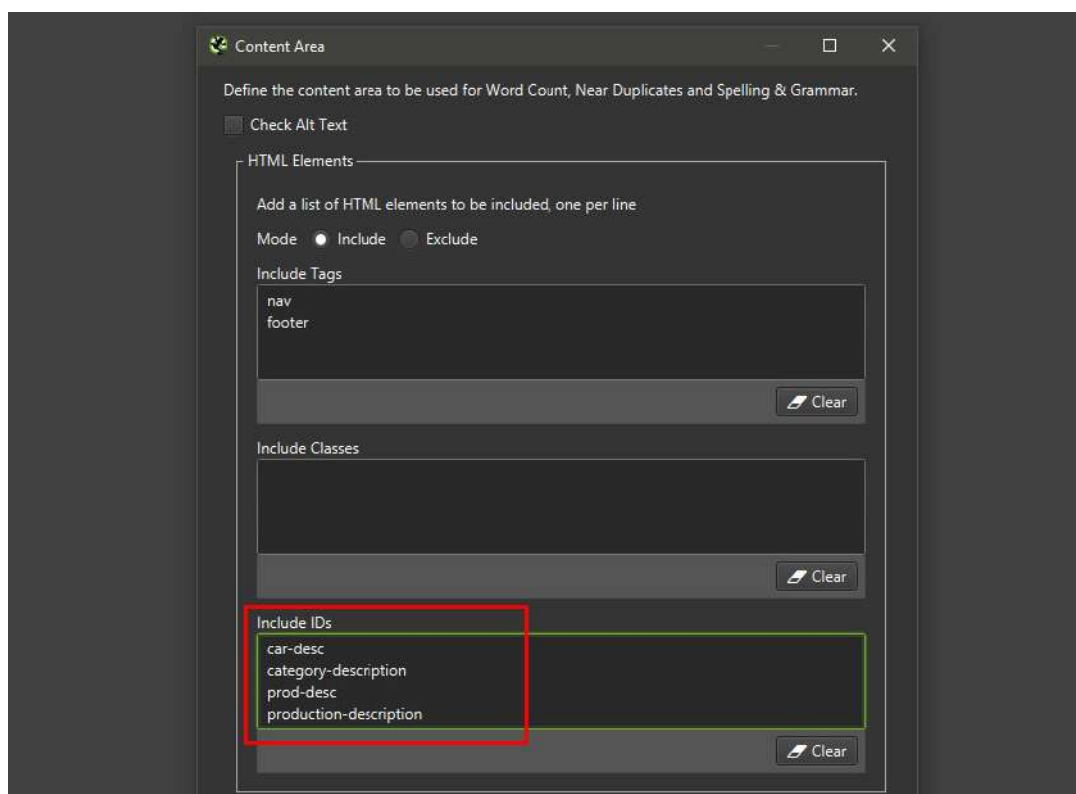
Config > Content > Duplicates



Successivamente, puoi affinare la scansione configurando l'area di contenuto utilizzata per l'analisi di similarità tra i vecchi e i nuovi URL.

Config > Content > Area

Per esempio, puoi “includere” le classi delle categorie e delle descrizioni dei prodotti del vecchio e del nuovo sito se sono cambiate. Se i modelli sono gli stessi, allora è improbabile che tu abbia bisogno di regolare l'area di contenuto.



Per utilizzare la modalità List nella mappatura puoi caricare sia i nuovi che i vecchi URL in modalità elenco e scansionarli, ricordandoti di rimuovere la profondità di scansione in “Config > Spider > Limits” e inserire sia il sito esistente che il sito di staging.

Conclusa la scansione ti manca solamente di attivare la “Crawl Analysis” e navigare alla scheda “Content” per visualizzare i dettagli dei quasi duplicati e l'indice di similarità.

Screamingfrog									
File Mode Configuration Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help									
Internal External Security Response Codes URL Page Titles Meta Description Meta Keywords H1 H2 Content Images Canonicals									
All									
Address	Word Count	Closest Similarity Match	No. Near Duplicates	Total					
1 https://www.screamingfrog.co.uk/contact-us/	224	100%	1	1					
2 https://www.screamingfrog.co.uk/about/	511	100%	1	1					
3 https://www.screamingfrog.co.uk/search-engine-optimisation/	1058	100%	1	1					
4 https://staging-screamingfrogco.uk.kinsta.cloud/our-work/	475	100%	1	1					
5 https://staging-screamingfrogco.uk.kinsta.cloud/digital-pr/	874	100%	1	1					
6 https://staging-screamingfrogco.uk.kinsta.cloud/seo-audit/	753	100%	1	1					
7 https://staging-screamingfrogco.uk.kinsta.cloud/adwords-history/	10833	100%	1	1					
8 https://staging-screamingfrogco.uk.kinsta.cloud/link-building/	943	100%	1	1					
9 https://staging-screamingfrogco.uk.kinsta.cloud/services/	757	100%	1	1					
10 https://staging-screamingfrogco.uk.kinsta.cloud/log-file-analyser/user-guide/	261	100%	1	1					
11 https://staging-screamingfrogco.uk.kinsta.cloud/search-engine-marketing/	832	100%	1	1					
12 https://staging-screamingfrogco.uk.kinsta.cloud/contact-us/	224	100%	1	1					
13 https://staging-screamingfrogco.uk.kinsta.cloud/conversion-rate-optimisation/	659	100%	1	1					
14 https://staging-screamingfrogco.uk.kinsta.cloud/seo-spider/user-guide/	688	100%	1	1					
15 https://staging-screamingfrogco.uk.kinsta.cloud/analytics-consultancy/	712	100%	1	1					
16 https://staging-screamingfrogco.uk.kinsta.cloud/about/	511	100%	1	1					
17 https://staging-screamingfrogco.uk.kinsta.cloud/log-file-analyser/	1474	100%	1	1					
18 https://staging-screamingfrogco.uk.kinsta.cloud/search-engine-optimisation/	1058	100%	1	1					
19 https://staging-screamingfrogco.uk.kinsta.cloud/social-media/	763	100%	1	1					
20 https://staging-screamingfrogco.uk.kinsta.cloud/seo-spider/	2620	100%	1	1					
21 https://staging-screamingfrogco.uk.kinsta.cloud/content-marketing/	768	100%	1	1					
22 https://staging-screamingfrogco.uk.kinsta.cloud/log-file-analyser/pricing/	442	100%	1	1					

Se il contenuto ha una corrispondenza perfetta, ogni URL avrà 1 quasi duplicato, con una somiglianza vicina al 100%.

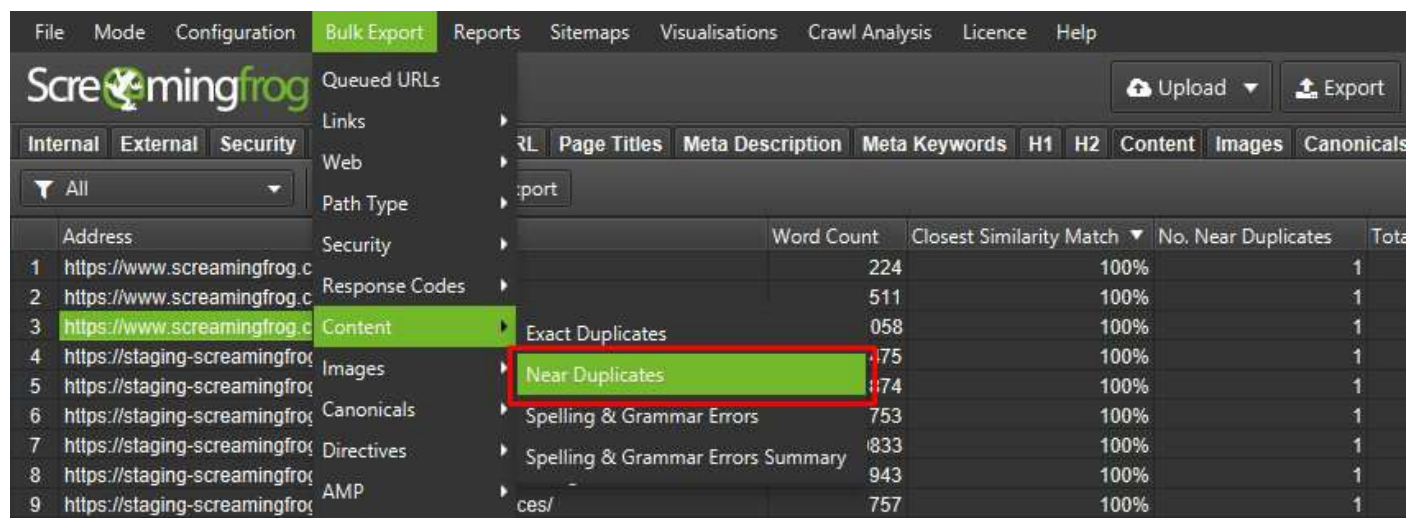
Se il contenuto è cambiato significativamente, i punteggi di similarità e il numero di quasi duplicati saranno diversi. La corrispondenza fra gli URL sarà visualizzabile nella finestra inferiore del Seo Spider.

Address	Word Count	Closest Similarity Match	No. Near Duplicates	Total	Language Errors	Spelling Errors	Grammar Errors	Language	Hash
1 https://www.screamingfrog.co.uk/contact-us/	224	100%	1	1					d0cc325eb4af702ba1eac3dbad
2 https://www.screamingfrog.co.uk/about/	511	100%	1	1					16528ae03d008b5db16999952
3 https://www.screamingfrog.co.uk/search-engine-optimisation/	1058	100%	1	1					e4daed5a54115f6e308bdec73c
4 https://staging-screamingfrogco.uk.kinsta.cloud/our-work/	475	100%	1	1					c6ecd3851c8853ded77b718ff
5 https://staging-screamingfrogco.uk.kinsta.cloud/digital-pr/	874	100%	1	1					3d01599c9deb6839a2ca58a3c
6 https://staging-screamingfrogco.uk.kinsta.cloud/seo-audit/	753	100%	1	1					d71c864f3eda1d7cd4c4cd5185
7 https://staging-screamingfrogco.uk.kinsta.cloud/adwords-history/	10833	100%	1	1					a87a95adb1b359c5899977ea77
8 https://staging-screamingfrogco.uk.kinsta.cloud/link-building/	943	100%	1	1					1d92c5eb171e839e38d2c45ae
9 https://staging-screamingfrogco.uk.kinsta.cloud/services/	757	100%	1	1					92e88d786ee3cc21025ab5588
10 https://staging-screamingfrogco.uk.kinsta.cloud/log-file-analyser/user-guide/	261	100%	1	1					d923be2e1b6a98e3e18c44416
11 https://staging-screamingfrogco.uk.kinsta.cloud/search-engine-marketing/	832	100%	1	1					e9c75547054e85c61b7a0947
12 https://staging-screamingfrogco.uk.kinsta.cloud/contact-us/	224	100%	1	1					4689e7ed4f4d05c8bcb54131
13 https://staging-screamingfrogco.uk.kinsta.cloud/conversion-rate-optimisation/	659	100%	1	1					17eb2a0544bc4849c8040728e3
14 https://staging-screamingfrogco.uk.kinsta.cloud/seo-spider/user-guide/	688	100%	1	1					7a7272578974e4ec7b0037929
15 https://staging-screamingfrogco.uk.kinsta.cloud/analytics-consultancy/	712	100%	1	1					8aa44e726a08397802c9e865b2
16 https://staging-screamingfrogco.uk.kinsta.cloud/about/	511	100%	1	1					b20eb5041a265b3ae14244da6
17 https://staging-screamingfrogco.uk.kinsta.cloud/log-file-analyser/	1474	100%	1	1					6f3c4da20f43458113dfdc28037
18 https://staging-screamingfrogco.uk.kinsta.cloud/search-engine-optimisation/	1058	100%	1	1					1c784a159636357344029d0fcea
19 https://staging-screamingfrogco.uk.kinsta.cloud/social-media/	763	100%	1	1					54ec2757495f1e983c2222acbc
20 https://staging-screamingfrogco.uk.kinsta.cloud/seo-spider/	2620	100%	1	1					2ba1376cc241967baw2000560
21 https://staging-screamingfrogco.uk.kinsta.cloud/content-marketing/	768	100%	1	1					6f830079d30774d2076ae808b
22 https://staging-screamingfrogco.uk.kinsta.cloud/log-file-analyser/pricing/	442	100%	1	1					e68ea1315871704303945386a

Address	Near Duplicate Address	Similarity
https://www.screamingfrog.co.uk/search-eng	https://staging-screamingfrogco.uk.kinsta.cloud/search-engine-optimis	100%

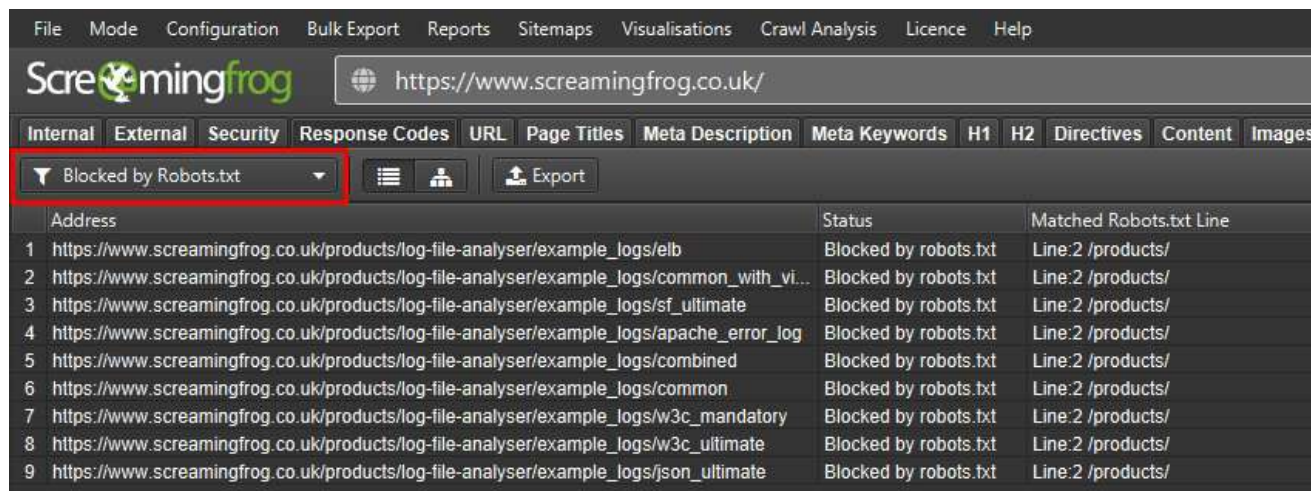
Content
Search Engine Optimisation (SEO) We're an SEO agency that runs remarkably successful SEO campaigns in the most competitive sectors, using a unique blend of technical and creative expertise. Everyone wants to be on the first page, but not everyone deserves to be there. Your business, brand and website have to earn that right. Google now use over 200 signals in their algorithm for scoring and last year, there were over 540 quality improvements to their search algorithm alone. Our search engine optimisation (SEO) services are designed to increase visibility within the algorithmic ("natural", "organic", or "free") search

A questo punto è sufficiente esportare i “Near Duplicates” da Bulk Export>Content>Near Duplicates per ottenere una buona base di partenza per impostare correttamente i redirect e salvare il sito dai potenziali danni organici della migrazione.



7. Post Migrazione: la scansione del nuovo sito

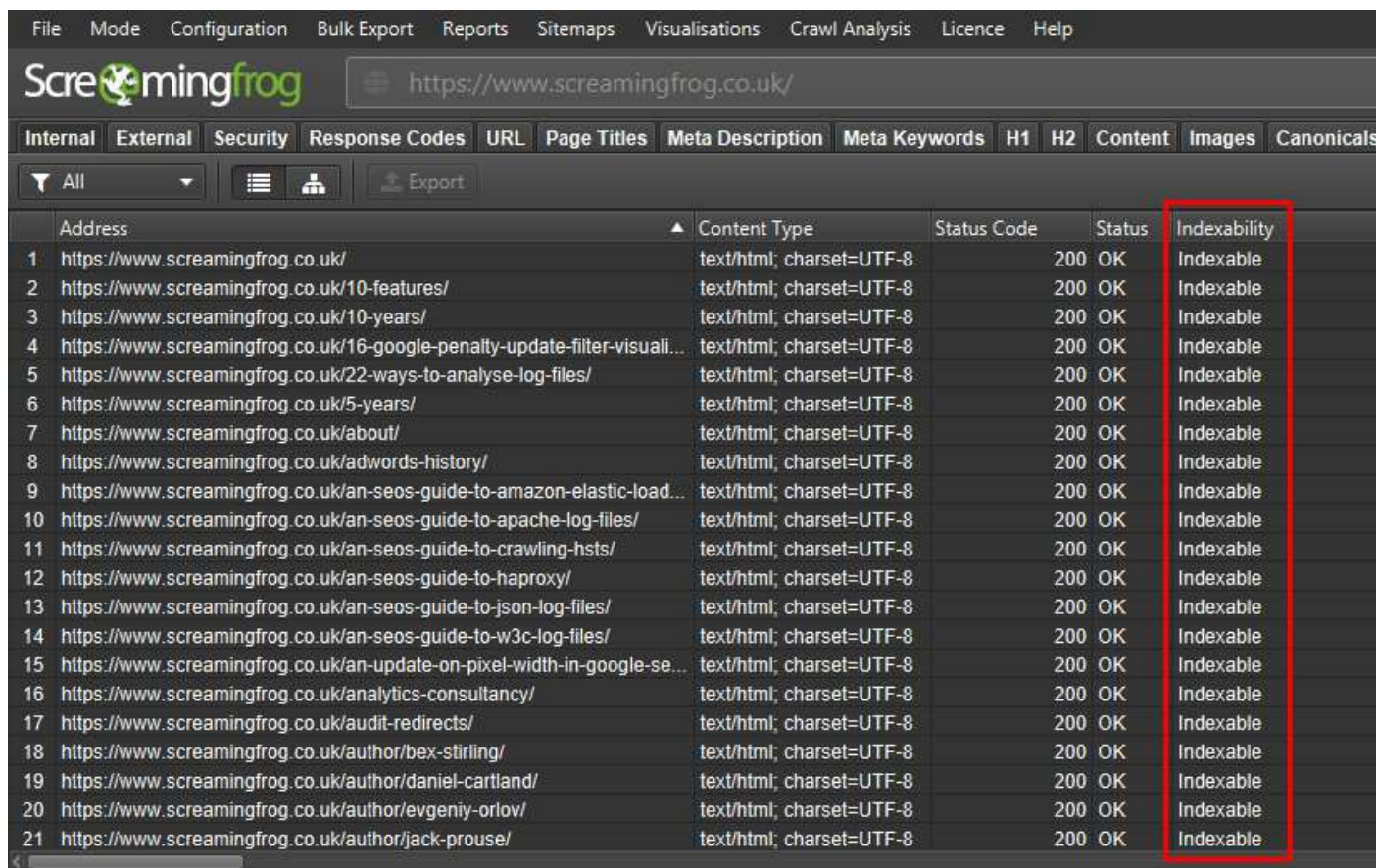
Al momento della pubblicazione non ti resta che eseguire un crawl del sito web per controllare che tutte le pagine del sito siano “crawlabili” e indicizzabili e che non presenti risorse importanti bloccate dal robots.txt (Response Codes > Blocked by Robots.txt).



Assicurati che non siano presenti direttive indesiderate come noindex, nofollow o none nei meta tag o nell'intestazione HTTP (Direttive > Noindex).



Verifica che ogni URL che vuoi posizionare sia indicizzabile.



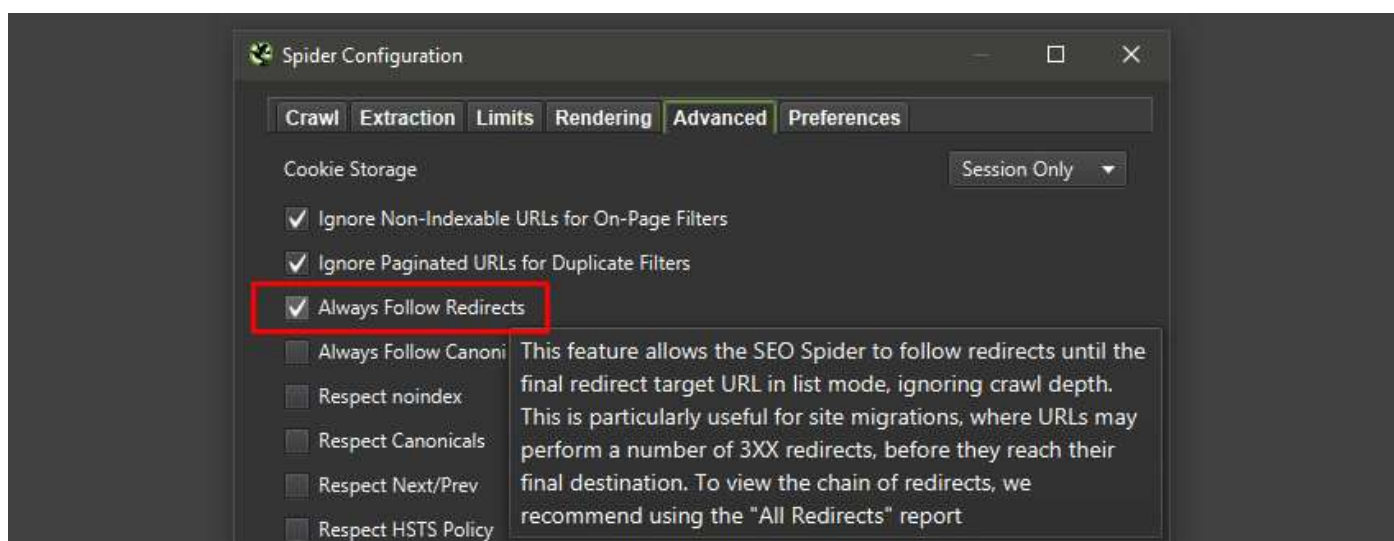
The screenshot shows the Screaming Frog SEO Spider interface. The 'Internal' tab is selected, and a list of 21 URLs is displayed. The 'Indexability' column is highlighted with a red box, showing that all URLs are 'Indexable'. The table has columns for Address, Content Type, Status Code, Status, and Indexability.

	Address	Content Type	Status Code	Status	Indexability
1	https://www.screamingfrog.co.uk/	text/html; charset=UTF-8	200	OK	Indexable
2	https://www.screamingfrog.co.uk/10-features/	text/html; charset=UTF-8	200	OK	Indexable
3	https://www.screamingfrog.co.uk/10-years/	text/html; charset=UTF-8	200	OK	Indexable
4	https://www.screamingfrog.co.uk/16-google-penalty-update-filter-visuali...	text/html; charset=UTF-8	200	OK	Indexable
5	https://www.screamingfrog.co.uk/22-ways-to-analyse-log-files/	text/html; charset=UTF-8	200	OK	Indexable
6	https://www.screamingfrog.co.uk/5-years/	text/html; charset=UTF-8	200	OK	Indexable
7	https://www.screamingfrog.co.uk/about/	text/html; charset=UTF-8	200	OK	Indexable
8	https://www.screamingfrog.co.uk/adwords-history/	text/html; charset=UTF-8	200	OK	Indexable
9	https://www.screamingfrog.co.uk/an-seos-guide-to-amazon-elastic-load...	text/html; charset=UTF-8	200	OK	Indexable
10	https://www.screamingfrog.co.uk/an-seos-guide-to-apache-log-files/	text/html; charset=UTF-8	200	OK	Indexable
11	https://www.screamingfrog.co.uk/an-seos-guide-to-crawling-hsts/	text/html; charset=UTF-8	200	OK	Indexable
12	https://www.screamingfrog.co.uk/an-seos-guide-to-haproxy/	text/html; charset=UTF-8	200	OK	Indexable
13	https://www.screamingfrog.co.uk/an-seos-guide-to-json-log-files/	text/html; charset=UTF-8	200	OK	Indexable
14	https://www.screamingfrog.co.uk/an-seos-guide-to-w3c-log-files/	text/html; charset=UTF-8	200	OK	Indexable
15	https://www.screamingfrog.co.uk/an-update-on-pixel-width-in-google-se...	text/html; charset=UTF-8	200	OK	Indexable
16	https://www.screamingfrog.co.uk/analytics-consultancy/	text/html; charset=UTF-8	200	OK	Indexable
17	https://www.screamingfrog.co.uk/audit-redirects/	text/html; charset=UTF-8	200	OK	Indexable
18	https://www.screamingfrog.co.uk/author/bex-stirling/	text/html; charset=UTF-8	200	OK	Indexable
19	https://www.screamingfrog.co.uk/author/daniel-cartland/	text/html; charset=UTF-8	200	OK	Indexable
20	https://www.screamingfrog.co.uk/author/evgeniy-orlov/	text/html; charset=UTF-8	200	OK	Indexable
21	https://www.screamingfrog.co.uk/author/jack-prouse/	text/html; charset=UTF-8	200	OK	Indexable

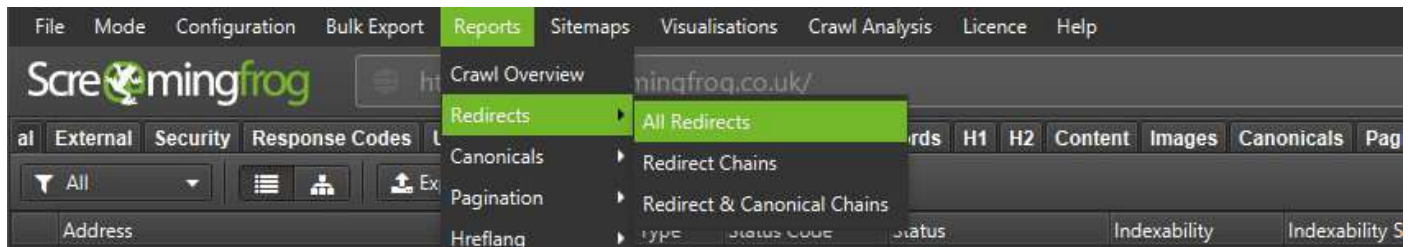
Una volta conclusa questa fase di controllo ti consiglio di fare un ulteriore step confrontando il vecchio sito e quello "Live" per verificare se tutto quanto avevi ottimizzato sul sito di staging è stato riportato nella versione attuale.

8. Verificare i reindirizzamenti

Se gli URL sono cambiati, è fondamentale controllare tutti i reindirizzamenti 301 verso le nuove destinazioni per evitare errori, salti, loop, reindirizzamenti temporanei o verso pagine non indicizzabili. Passa alla modalità elenco ('Mode > List') e attiva "Always Follow Redirects" in "Config > Spider > Advanced".



Aggiungi i vecchi URL scoperti con la prima scansione del vecchio sito (step 1 articolo) e avvia la scansione. Alla fine del crawl esporta il rapporto "All Redirects - Tutti i reindirizzamenti".



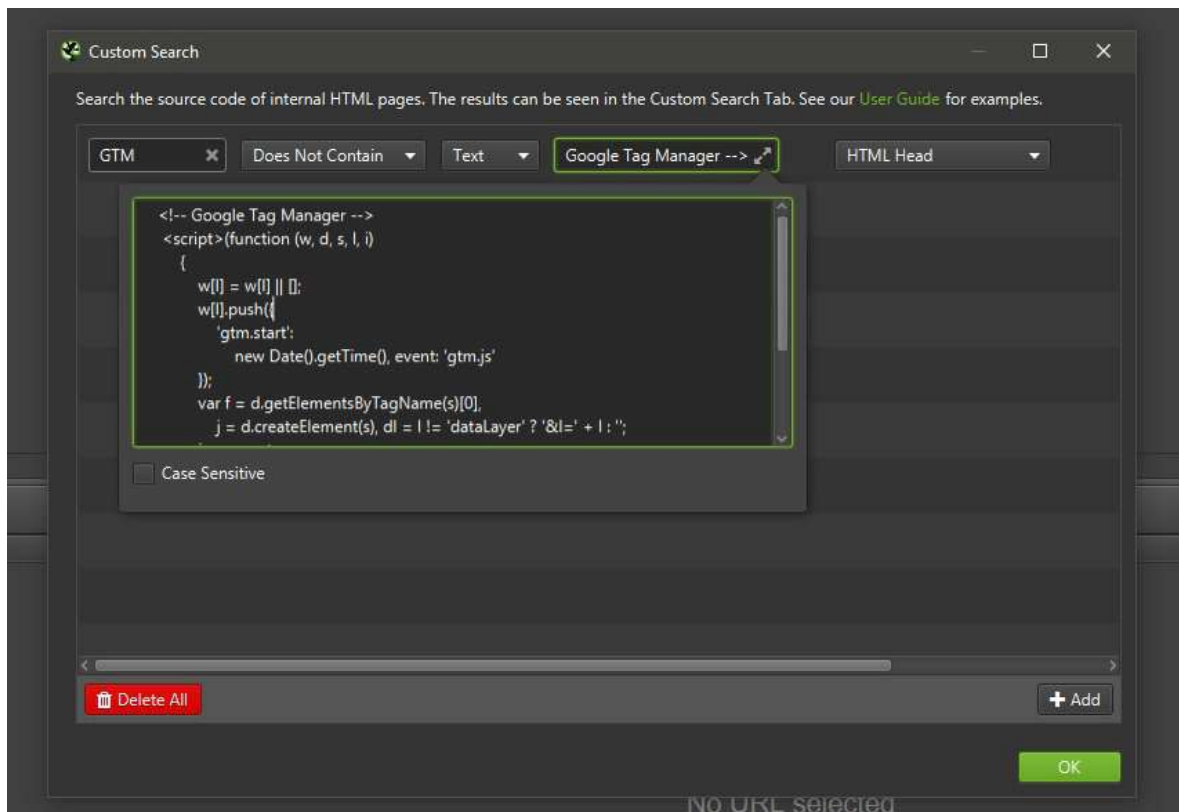
Attraverso questo rapporto potrai valutare i seguenti indici:

- **Chain Type:** specifica la tipologia del “redirect” (es. Http/Javascript/Meta Refresh).
- **Number of Redirects:** il numero di nodi presenti nella catena.
- **Redirect Loop:** la presenza o meno di redirect circolari. Questo campo viene esplicitato con True/Fal-se secondo regole booleane.
- **Temp redirect in Chain:** se esiste un reindirizzamento temporaneo nella catena (es. 302).
- **Address:** l'indirizzo dell'URL di partenza.
- **Final Address:** la destinazione finale del collegamento.
- **Indexability:** identifica se la destinazione (Final Address) è indicizzabile o non indicizzabile. Questo campo è molto utile per comprendere se durante la migrazione hai linkato un vecchio URL verso una destinazione preclusa all'indicizzazione. Ovviamente se una risorsa l'hai gestita con un reindiriz-zamento è perché volevi passare il trust verso una nuova destinazione ma se questa presenta l'attri-buto “Non Indexable” il lavoro precedente non porterebbe alcun effetto.
- **Indexability Status:** identifica la ragione per cui l'indirizzo di destinazione non è indicizzabile, per esempio la presenza del tag “Noindex”.
- **Final Content:** la tipologia del contenuto di destinazione (esempio: html).
- **Final Status Code:** identifica lo Status Code della destinazione finale e può essere 200, 4xx, 1xx,5xx o no response. Anche questo campo è molto importante da valutare per evitare di sprecare il trust del reindirizzamento verso una risorsa che presenta errori di server, pagina non trovata o altro. In sintesi se non vedi lo status code 200 devi fermarti ed intervenire.

9. Controllare il tracciamento Google Analytics e GTM

Al momento della pubblicazione del nuovo sito web molte volte capita di dimenticare l'inserimento de-gli script di tracciamento come GA o GTM.

Per un'analisi immediata che tutto sia stato implementato puoi utilizzare la ricerca personalizzata (“Con-fig > Custom > Search”), con un filtro “non contiene” per cercare il tag di tracciamento completo all'in-terno dell'Head HTML di ogni pagina del sito.



In questo modo puoi verificare per quali URL non è presente il tracciamento (scheda Custom Search) della finestra principale del Seo Spider.

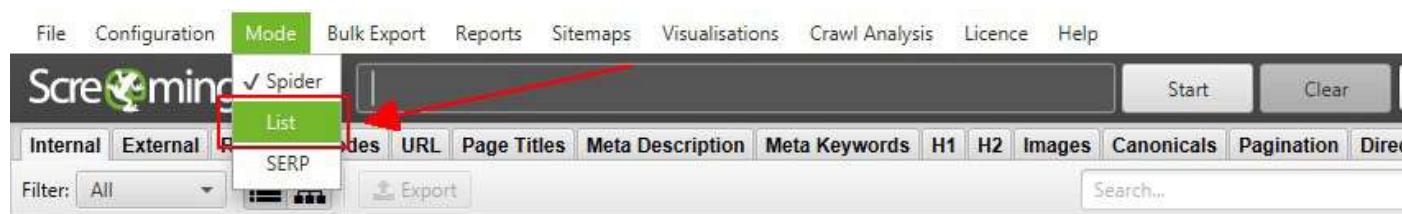
File Mode Configuration Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help						
Screamingfrog https://www.screamingfrog.co.uk/						
Custom Search Response Codes URL Page Titles Meta Description Meta Keywords H1 H2 Content Images Canonicals Pagination						
All [Icons] Export						
✓ All		Content Type	Status Code	Status	Search 1: GTM	
GTM (Does Not Contain)	co.uk/	text/html; charset=UTF-8	200	OK	Contains	
	co.uk/seo-spider/tutorials/	text/html; charset=UTF-8	200	OK	Contains	
3	https://www.screamingfrog.co.uk/search-engine-optimisation/	text/html; charset=UTF-8	200	OK	Contains	
4	https://www.screamingfrog.co.uk/contact-us/	text/html; charset=UTF-8	200	OK	Contains	
5	https://www.screamingfrog.co.uk/about/	text/html; charset=UTF-8	200	OK	Contains	
6	https://www.screamingfrog.co.uk/clients/	text/html; charset=UTF-8	200	OK	Contains	
7	https://www.screamingfrog.co.uk/digital-pr/	text/html; charset=UTF-8	200	OK	Contains	
8	https://www.screamingfrog.co.uk/login/	text/html; charset=UTF-8	200	OK	Contains	
9	https://www.screamingfrog.co.uk/seo-spider/licence/	text/html; charset=UTF-8	200	OK	Contains	
10	https://www.screamingfrog.co.uk/seo-spider/support/	text/html; charset=UTF-8	200	OK	Contains	
11	https://www.screamingfrog.co.uk/web-scraping/	text/html; charset=UTF-8	200	OK	Contains	
12	https://www.screamingfrog.co.uk/how-to-automate-screaming...	text/html; charset=UTF-8	200	OK	Contains	
13	https://www.screamingfrog.co.uk/blog/	text/html; charset=UTF-8	200	OK	Contains	
14	https://www.screamingfrog.co.uk/pay-per-click/	text/html; charset=UTF-8	200	OK	Contains	
15	https://www.screamingfrog.co.uk/privacy/	text/html; charset=UTF-8	200	OK	Contains	
16	https://www.screamingfrog.co.uk/search-engine-marketing/	text/html; charset=UTF-8	200	OK	Contains	
17	https://www.screamingfrog.co.uk/content-marketing/	text/html; charset=UTF-8	200	OK	Contains	
18	https://www.screamingfrog.co.uk/how-to-build-user-trust-with...	text/html; charset=UTF-8	200	OK	Contains	
19	https://www.screamingfrog.co.uk/blog-file-analysis/	text/html; charset=UTF-8	200	OK	Contains	

COME SFRUTTARE AL MASSIMO LA MODALITÀ DI SCANSIONE ELENCO (LIST MODE)

Il Seo Spider prevede due modalità di scansione: quella standard denominata “Spider” che esegue la scansione di un intero sito web, e la modalità “List” che limita la scansione al caricamento di un elenco specifico di URL.

Vediamo insieme come sfruttare al meglio la seconda opzione.

Prima di tutto, devi modificare la tipologia di crawling cliccando su “Mode > List” dal menu principale.



A livello di configurazione devi tener conto di due differenze sostanziali tra la modalità “List” e “Spider”:

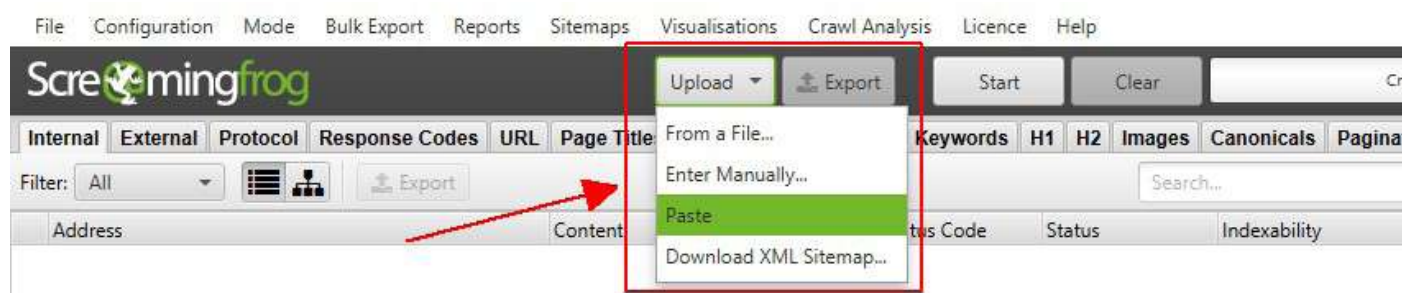
- A livello di Ux, l'interfaccia del Seo Spider cambia e mostra un pulsante di caricamento invece di una barra degli indirizzi.
- Nella modalità elenco il “Limit Crawl Depth” determina la profondità di scansione in modo automatico a “0”.

Questo significa che solo gli URL caricati vengono scansati a differenza della modalità “Spider” che considera anche le referenze alle immagini, link esterni etc.

L'unica eccezione riguarda il “Javascript Rendered”, dove il Seo Spider ha bisogno di scansare le risorse della pagina per rendere accuratamente la pagina nel suo browser Chrome.

1. Passaggio alla Modalità Elenco

Scelta la modalità elenco (Mode > List) dal menu principale, ti basta cliccare sul pulsante “Upload” e scegliere di caricare la tua lista di URL da un file, incollare un elenco di URL o scaricare una Sitemap XML.



Le caratteristiche che devi considerare con questa modalità sono:

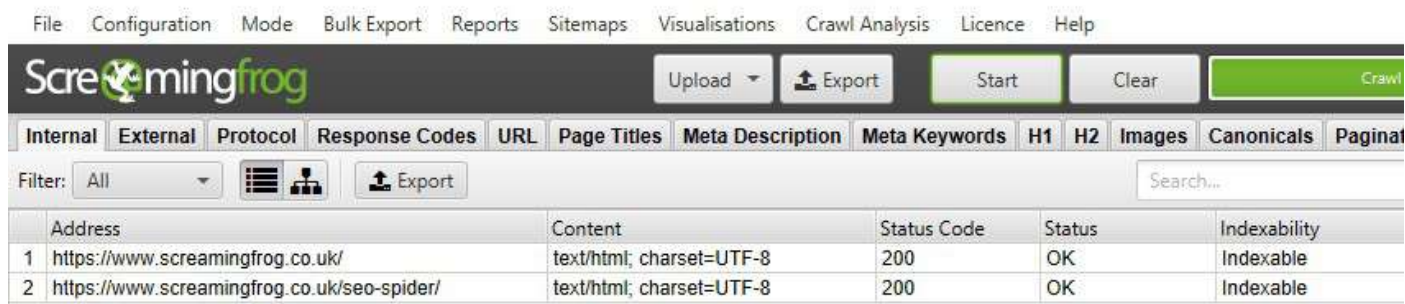
1. gli URL inseriti devono prevedere il protocollo “http” o “https” altrimenti il Seo Spider non trova alcun URL da scansare.
2. Tieni in considerazione che Screaming Frog normalizza gli URL al momento dell'upload, e li de-duplica durante il crawl.

Ad esempio: vuoi caricare i seguenti 4 URL:

<https://www.screamingfrog.co.uk/>
<https://www.screamingfrog.co.uk/#fragment-this-wont-be-crawled>
<https://www.screamingfrog.co.uk/seo-spider/>
<https://www.screamingfrog.co.uk/seo-spider/>

Il Seo Spider determina automaticamente quanti sono gli URL unici da scansare. Nell'esempio andrà a considerare solamente 2 URL come unici (es. “#” non considera il “# -fragment” come url separato e la pagina “/seo-spider” è duplicata), sembra un esempio banale ma con liste da migliaia di URL, può risultare meno ovvio.

Una volta fatto l'upload della lista il Seo Spider riporterà di aver trovato i 4 URL e li normalizzerà nella finestra di dialogo.



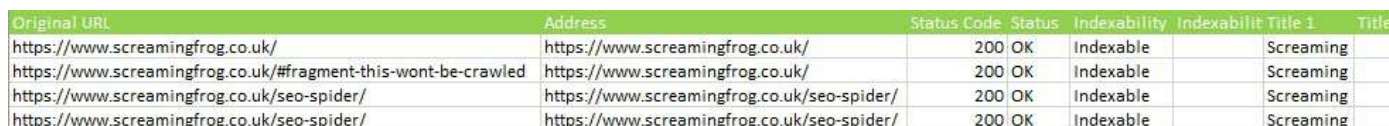
The screenshot shows the Screamingfrog web crawler interface. At the top, there's a menu bar with options like File, Configuration, Mode, Bulk Export, Reports, Sitemaps, Visualisations, Crawl Analysis, Licence, and Help. Below the menu is a toolbar with buttons for Upload, Export, Start, Clear, and a Crawl button. A tabbed interface shows various data categories: Internal, External, Protocol, Response Codes, URL, Page Titles, Meta Description, Meta Keywords, H1, H2, Images, Canonicals, and Paginated. The 'URL' tab is selected, and a table displays the results of the crawl.

	Address	Content	Status Code	Status	Indexability
1	https://www.screamingfrog.co.uk/	text/html; charset=UTF-8	200	OK	Indexable
2	https://www.screamingfrog.co.uk/seo-spider/	text/html; charset=UTF-8	200	OK	Indexable

2. Esportazione dei dati

Ottenuti i dati della tua scansione ti basta usare "Export" da qualsiasi scheda per esportare il tuo rapporto.

Come potrai notare dall'esportazione i dati originali saranno mantenuti come riferimento (colonna Original URL) ma contestualmente saranno riportati anche quelli normalizzati dal crawler.



The screenshot shows a table with two main sections: 'Original URL' and 'Address'. The 'Original URL' section lists four URLs, including some with fragments. The 'Address' section shows the normalized versions of these URLs. The table also includes columns for Status Code, Status, Indexability, and Title.

Original URL	Address	Status Code	Status	Indexability	Indexability Title 1	Title
https://www.screamingfrog.co.uk/	https://www.screamingfrog.co.uk/	200	OK	Indexable		Screaming
https://www.screamingfrog.co.uk/#fragment-this-wont-be-crawled	https://www.screamingfrog.co.uk/	200	OK	Indexable		Screaming
https://www.screamingfrog.co.uk/seo-spider/	https://www.screamingfrog.co.uk/seo-spider/	200	OK	Indexable		Screaming
https://www.screamingfrog.co.uk/seo-spider/	https://www.screamingfrog.co.uk/seo-spider/	200	OK	Indexable		Screaming

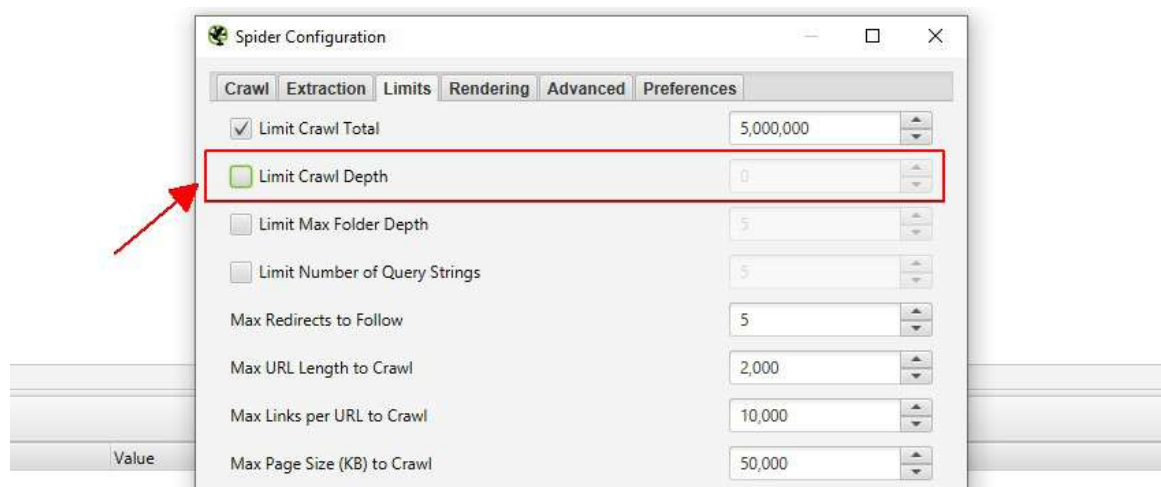
Considerate le caratteristiche intrinseche su come ragiona il Seo Spider con questa modalità vediamo insieme come sfruttare la sua potenzialità con alcune best practice avanzate che ti permettono di focalizzare meglio la tua analisi e risparmiare tempo.

1. Crawling di un elenco di URL e di altri elementi relativi

Una prima applicazione può comprendere una scansione mirata degli URL caricati insieme ad 1 o più elementi referenti come ad esempio le immagini, i canonicals, gli hreflang o altro.

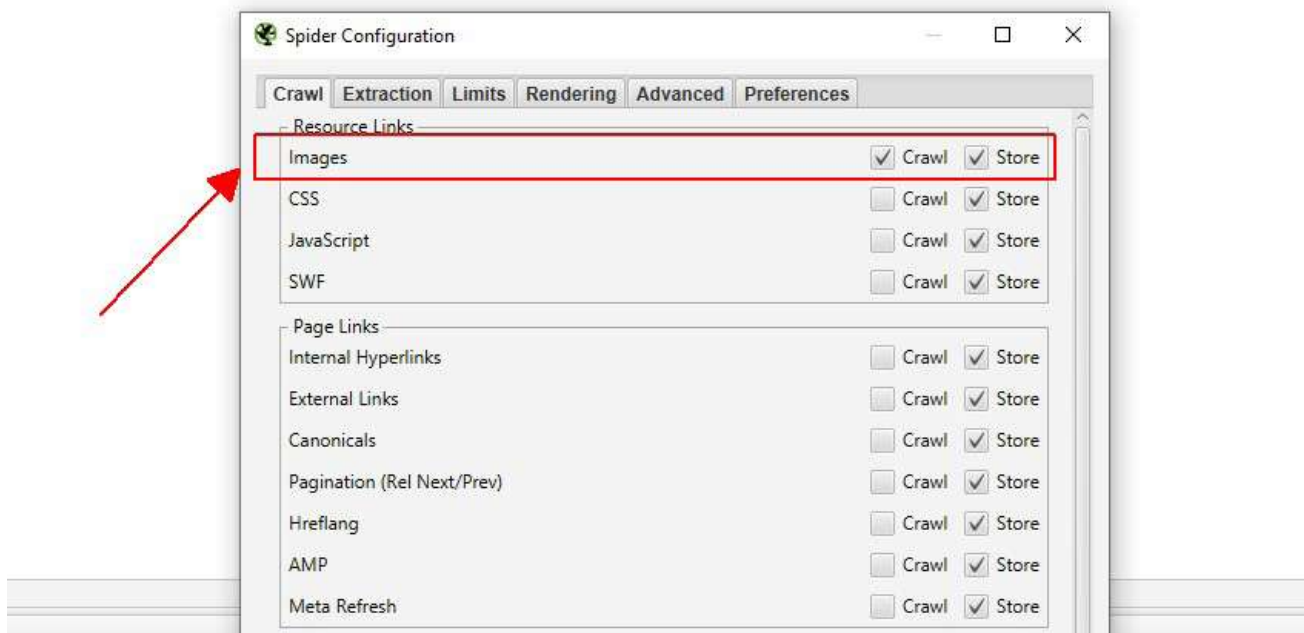
Per quest'opzione è sufficiente rimuovere il limite di profondità di scansione (Limit Crawl Depth) che è di default impostato a '0'.

Config > Spider > Limits



Eliminando il limite di scansione il Seo Spider scansiona la tua lista di URL insieme a tutti gli URL sullo stesso sottodominio a cui si collegano. Ma questo comportamento non è esattamente quello che ci aspettiamo.

Il passo successivo è quello di gestire granularmente le risorse e referenze da scansione andando su "Config > Spider > Crawl" e disabilitando tutti i "Resource Links" e "Page Links" non necessari e lasciando solamente quelli pertinenti alla tua analisi specifica.

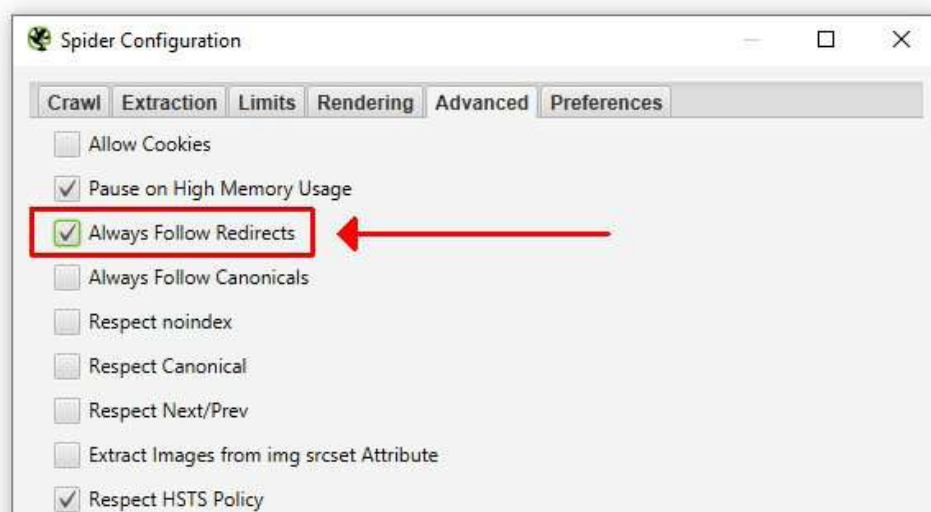


Attraverso questo passaggio sei in grado di focalizzarti esattamente sugli URL da analizzare e i loro elementi più importanti senza perdite di tempo.

2. Audit dei Redirect con Modalità List (Elenco)

Se stai controllando i reindirizzamenti in una migrazione di un sito, può essere particolarmente utile eseguire il crawling dei loro URL di destinazione e di qualsiasi catena di reindirizzamento incontrata. Per eseguire questo Audit devi utilizzare la configurazione "Always follow redirects - Segui sempre i redirect" che trovi in "Config > Spider > Advanced".

Abilitare questa configurazione significa che il "crawl depth limit" viene ignorato e i redirect vengono seguiti fino a quando non raggiungono una risposta non-3XX (o fino a quando il tuo limite "Max Redirects To Follow" viene raggiunto da "Config > Spider > Limits"). Questo ti evita di dover caricare più liste di URL di destinazione ogni volta per raggiungere la fine della catena.



Alla fine del crawl potrai utilizzare il report "All Redirects" che ti mapperà l'intera catena di reindirizzamenti in un unico rapporto.

COME TROVARE I CONTENUTI DUPLICATI

Come più volte dichiarato da Google il contenuto è uno degli elementi più importanti a livello di ranking e la gestione di eventuali duplicazioni è una delle attività da considerare in ogni ottimizzazione Seo.

Il contenuto duplicato è uno dei problemi che più comunemente tarpa le ali di un progetto e deve essere ridotto al minimo per non sprecare il Crawling Budget.

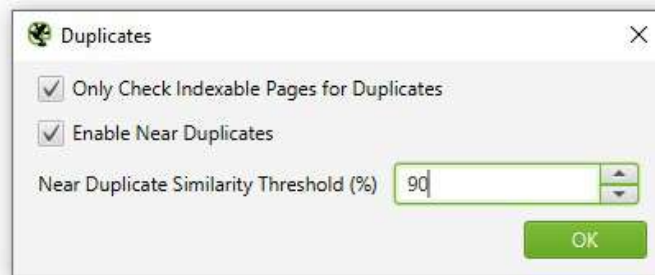
L'obiettivo di un Seo specialist non si deve limitare a comprendere e ottimizzare le pagine identiche ma dovrebbe affinare l'intervento anche sui contenuti "simili" che potrebbero creare condizioni di "Cannibalizzazione delle query" creando assoluta inefficienza alla scansione del Motore di Ricerca.

Attraverso questo tutorial vediamo come Screaming Frog ti permette di scovare sia i contenuti duplicati esatti (con lo stesso hash), sia contenuti simili che potrebbero risultare con lo stesso "Search Intent" agli occhi di Google.

1. Abilita "Near Duplicates" da "Config > Content > Duplicates".

Screaming Frog, per impostazione predefinita identifica automaticamente le pagine duplicate esatte (le pagine duplicate presentano lo stesso "Hash"), mentre per identificare i contenuti simili devi abilitare l'opzione "Enable Near Duplicates". Una volta attiva il Seo Spider considera come pagina "simile" ogni documento che presenti una corrispondenza minima del 90%.

Questa soglia la puoi anche regolare in base all'affinamento più adatto al tuo progetto.



Di default il SEO Spider fa un controllo sulle pagine simili o duplicate solamente se sono 'indicizzabili' e non "canonicalizzate"; il consiglio è quello di deselezionare l'opzione "Only Check Indexable Pages For Duplicates - Controlla solo le pagine indicizzabili per i duplicati", poiché questo potrebbe aiutarti a trovare aree di potenziale spreco di Crawler Budget.

Nota Bene: Di norma ti consiglio per progetti nuovi, in cui vai a gestire tutta la Key Research, di impostare la soglia al 50-60%, mentre su progetti con una certa storicità ad un minimo di 70% per essere certo che il Search Intent sia sempre univoco e ben marcato.

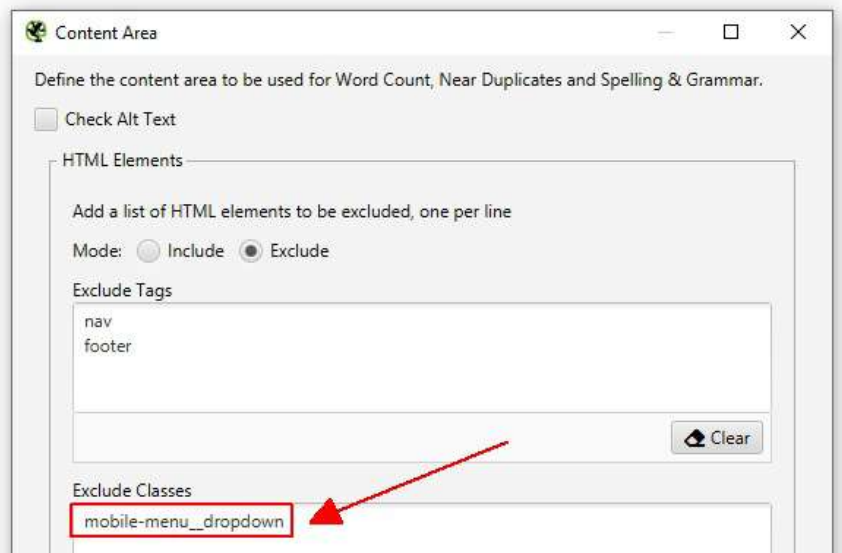
Attraverso l'utilizzo della Search Console puoi controllare periodicamente la relazione Query-Landing Page e nel caso di cannibalizzazioni scegliere quale sia la landing più adeguata (verificando anche i dati di GA) a quello specifico Search Intent e rinnovare il contenuto delle pagine secondarie o, in certi casi rimuoverle.

2. Definisci l'area del contenuto da analizzare

Anche per questa task Screaming Frog è molto flessibile e ti permette di scegliere il focus su cui porre l'attenzione nella ricerca di contenuti duplicati o simili.

Per impostazione di default il Seo Spider esclude automaticamente gli elementi della <nav> e del <footer> per concentrarsi sul contenuto del corpo principale. Tuttavia, non tutti i siti web sono costruiti utilizzando questi elementi HTML5 ma il Seo Spider ti permette di "includere" o "escludere" dal tuo esame alcuni tag HTML, classi o ID.

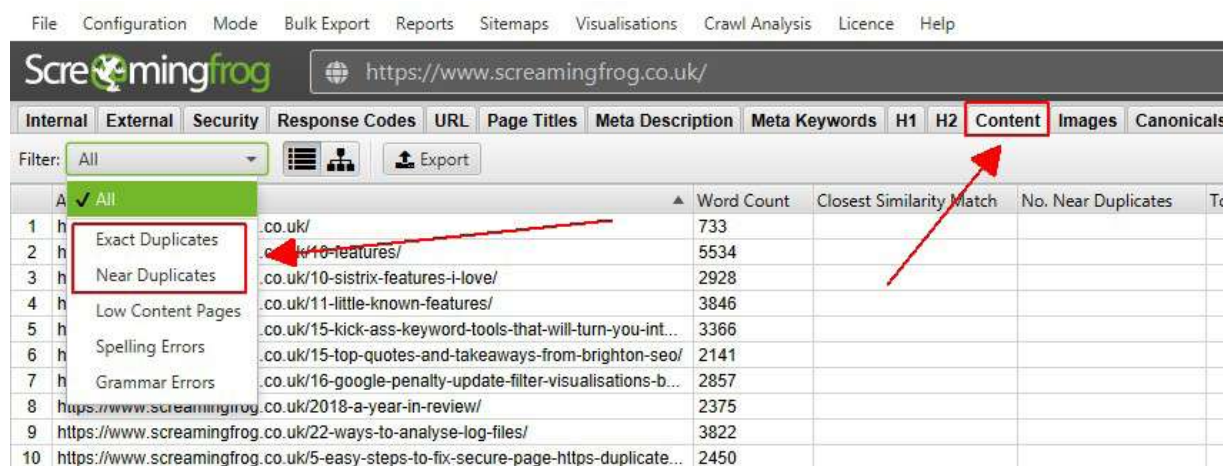
Per esempio se fai un crawl sul sito web di Screaming Frog ti accorgerai che presenta un menu mobile al di fuori dall'elemento "nav" ma, per impostazione di default, questo elemento viene comunque incluso nell'analisi del contenuto. In questo specifico caso potresti definire la classe del menu ("mobile-menu_dropdown") ed escluderla dall'esame per focalizzarti maggiormente sul contenuto principale.



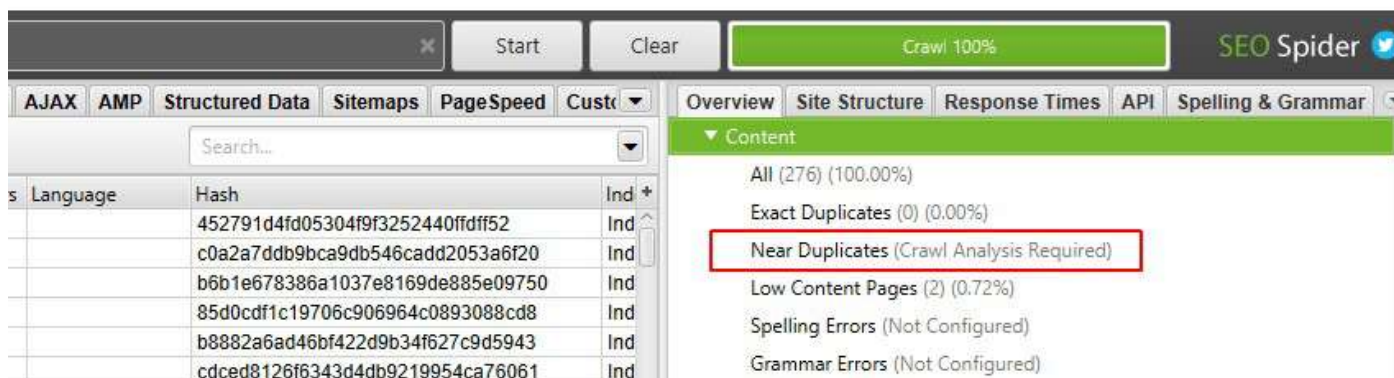
3. Esegui la scansione del sito web.
4. Visualizza i duplicati nella scheda "Content"

Nella scheda "Content" hai a disposizione due 2 filtri:

- 'exact duplicates' (contenuto identico).
- 'near duplicates' (contenuto simile).

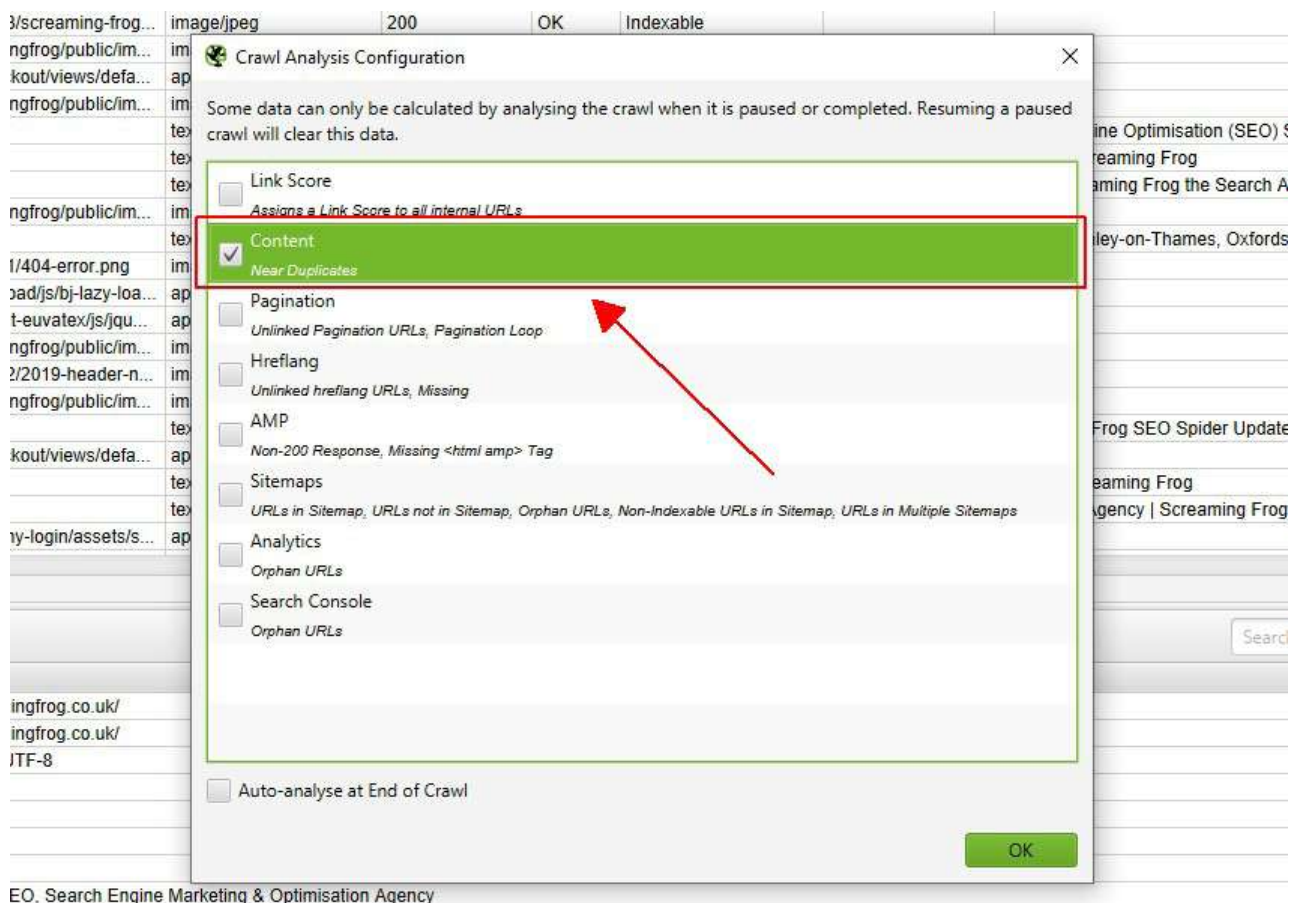


Durante la scansione questa tab si popola con i dati man mano che il crawl avanza ma presenta solo i "duplicati esatti".



5. Configurazione della Crawl Analysis

Per popolare il filtro "Near Duplicates", le colonne "Closest Similarity Match" e "No. Near Duplicates" devi configurare ed eseguire la Crawl Analysis.



6. Visualizza la scheda “Content” e i filtri “Exact” e “Near”

Dopo aver eseguito l'analisi post crawl, il filtro “Near Duplicates”, le colonne “Closest Similarity Match” e “No. Near Duplicates” saranno popolate.

Solo gli URL con contenuto superiore alla soglia di similarità selezionata conterranno dati, gli altri rimarranno vuoti.

Nel caso riportato qui sotto, il sito di Screaming Frog presenta solamente due risorse con una soglia del 92% che andranno ottimizzate.

File Configuration Mode Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help

Screamingfrog

https://www.screamingfrog.co.uk/

Internal External Security Response Codes URL Page Titles Meta Description Meta Keywords H1 H2 Content Images Canonicals

Filter: All

Export

	Address	Word Count	Closest Similarity Match▼	No. Near Duplicates	Total Language Err
1	https://www.screamingfrog.co.uk/seo-spider/licence/	374	92%	1	
2	https://www.screamingfrog.co.uk/log-file-analyser/licence/	372	92%	1	
3	https://www.screamingfrog.co.uk/	733		0	
4	https://www.screamingfrog.co.uk/search-engine-optimisation/	1055		0	
5	https://www.screamingfrog.co.uk/clients/	328		0	
6	https://www.screamingfrog.co.uk/about/	508		0	
7	https://www.screamingfrog.co.uk/contact-us/	208		0	
8	https://www.screamingfrog.co.uk/seo-spider-12/	7430		0	
9	https://www.screamingfrog.co.uk/login/	142		0	
10	https://www.screamingfrog.co.uk/digital-pr/	861		0	

Per una valutazione più puntuale puoi ordinare i dati in base ai due filtri:

- **Exact Duplicates:** identifica le pagine con uno “hash” identico. Molto utile per la gestione delle pagine principali e le pagine da “canonicalizzare”.
- **Near Duplicates:** identifica le pagine con una soglia di “similarità” superiore al 90% o con un indice diverso se l'hai modificato tramite “Config > Content > Duplicates”.

In sintesi nella colonna “Closest Similarity Match” puoi visualizzare la percentuale di similarità mentre in “No. Near Duplicates” il numero di pagine coinvolte da questa somiglianza.

7. Visualizza gli URL duplicati tramite la scheda “Dettagli duplicati”

Per i “duplicati esatti” è sufficiente spostarsi sulla colonna “hash” e ordinarli cliccando sull’intestazione per trovare tutte le corrispondenze.

Screamingfrog									
https://www.bbc.co.uk/									
Internal External Security Response Codes URL Page Titles Meta Description Meta Keywords H1 H2 Content Images Canonicals									
Filter: Exact Duplicates									
Export									
	Address	Hash	Indexability						
1	https://www.bbc.co.uk/aboutthebbc	4197ad19ecc2d254b12579923f5c5b0d	Indexable						
2	https://www.bbc.co.uk/aboutthebbc/	4197ad19ecc2d254b12579923f5c5b0d	Indexable						
3	https://www.bbc.co.uk/usingthebbc/privacy/privacy-promise/	5d6ef1d4459d5e7fd42d9f1a62535ad3	Indexable						
4	https://www.bbc.co.uk/usingthebbc/privacy-promise	5d6ef1d4459d5e7fd42d9f1a62535ad3	Indexable						
5	https://www.bbc.co.uk/usingthebbc/terms/	c91c8bfb24be215072b9d98d727c690c	Indexable						
6	https://www.bbc.co.uk/usingthebbc/terms	c91c8bfb24be215072b9d98d727c690c	Indexable						
7	https://www.bbc.co.uk/usingthebbc/privacy/	eaf68b19496258fd7fe4cc786e93fd33	Indexable						
8	https://www.bbc.co.uk/usingthebbc/privacy	eaf68b19496258fd7fe4cc786e93fd33	Indexable						
9	https://www.bbc.co.uk/usingthebbc/cookies/	f20406e24af1fb160049388d95405f76	Indexable						
10	https://www.bbc.co.uk/usingthebbc/cookies	f20406e24af1fb160049388d95405f76	Indexable						

Nell’esempio relativo al sito della BBC ogni URL ha un duplicato esatto perché presenta due versioni una con il “/” e una senza.

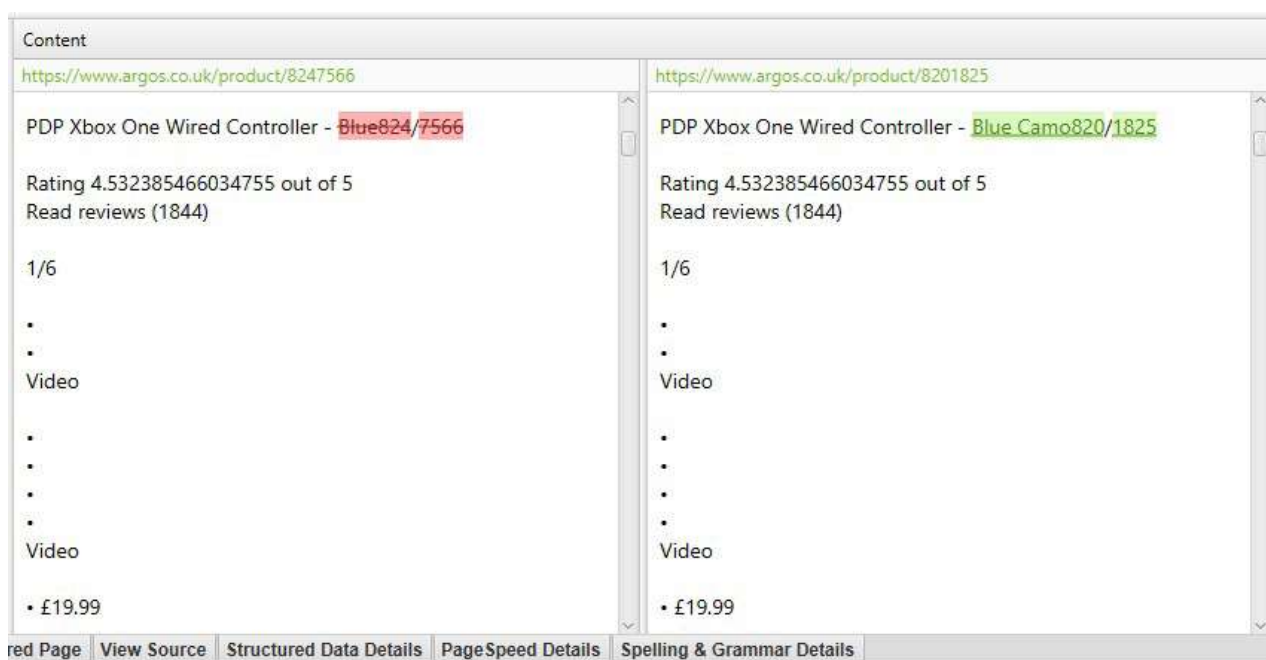
Per scoprire invece i “quasi duplicati (simili)”, ti basta cliccare sulla tab “Duplicate Details” nella finestra in basso.

The screenshot shows the Screamingfrog SEO Spider interface. The main table lists URLs and their hash values. The 'Duplicate Details' tab is selected, showing a list of near-duplicate URLs with their similarity percentages. A red arrow points to the 'Duplicate Details' tab. The detailed view on the right shows the content of the selected URL, including a video player and a bar chart.

Nel caso sopra cliccando su un URL della finestra superiore ci sono 4 pagine che superano la soglia del 90% di similarità.

Export				
Search...				
Address	Near Duplicate Address	Similarity	Indexability	Indexability
https://www.argos.co.uk/product/8247566	https://www.argos.co.uk/product/8201825	98%	Indexable	
https://www.argos.co.uk/product/8247566	https://www.argos.co.uk/product/6531339	97%	Indexable	
https://www.argos.co.uk/product/8247566	https://www.argos.co.uk/product/8242530	97%	Indexable	
https://www.argos.co.uk/product/8247566	https://www.argos.co.uk/product/6681935	94%	Indexable	

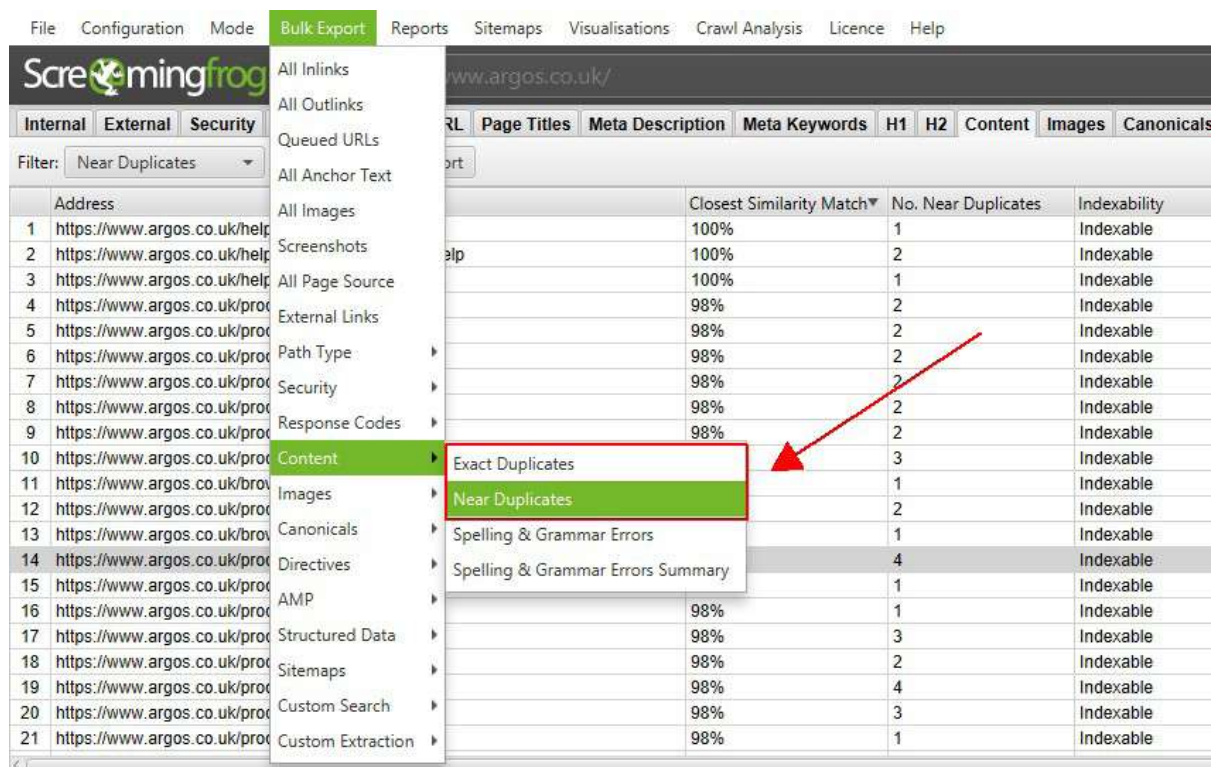
Il Seo Spider prevede anche un'anteprima dei contenuti quasi duplicati nella tab "Duplicate Details" e permette una visione molto intuitiva delle parti testuali che differiscono tra le pagine in esame.



8. Esportazione in blocco degli URL duplicati

Sia i duplicati esatti che i quasi duplicati possono essere esportati in blocco.

Bulk Export > Content > Exact Duplicates/ Near Duplicates



COME SCANSIONARE CONTENUTI IN JAVASCRIPT

INTRODUZIONE: GOOGLE E JAVASCRIPT

Fino a pochi anni fa uno dei problemi più diffusi nell'ottimizzazione Seo era quello di riuscire a fare indicizzare pagine con contenuti creati dinamicamente tramite Javascript. Il problema era relativo ai Bot dei Motori di Ricerca che non erano in grado di scansionare e indicizzare queste risorse e consideravano solo il contenuto presente nel codice sorgente dell'HTML statico.

Contestualmente a questa problematica c'è stata un'evoluzione di framework come AngularJS, React, VueJS, applicazioni a pagina singola (SPA) e progressive web app (PWA) che ha portato Google ad evolversi in modo significativo e a raggiungere una maturazione completa nel comprendere le pagine web come un browser moderno.

Tuttavia, nonostante oggi Google è generalmente in grado di scansionare e indicizzare la maggior parte dei contenuti JavaScript, viene raccomandato il rendering lato server, il pre-rendering o il rendering dinamico piuttosto che affidarsi al JavaScript lato client in quanto per i crawler risulta ancora difficile elaborare JavaScript con successo o immediatamente.

Per questi motivi in fase di analisi è fondamentale comprendere le differenze tra il DOM dopo che JavaScript è entrato in gioco e ha costruito la pagina web e la risposta statica dell'HTML.

JAVASCRIPT, IL SEO SPIDER E ALCUNE CONTROINDICAZIONI

Per ovviare a questa esigenza ti viene in aiuto Screaming Frog con la modalità "Javascript Rendering" che integrando la libreria "Chromium" riesce ad emulare al meglio il comportamento dello Spider di Google.

The screenshot shows the Screaming Frog SEO Spider interface. The main window displays a list of URLs with columns for Status Code, HTML Word Count, and Rendered HTML Word Count. The 'Rendered HTML' column shows the content after JavaScript execution. The 'Resources' panel on the left lists various scripts and styles loaded on the page. The 'Overview' panel on the right shows a summary of the crawl results, including the number of URLs scanned and the percentage of JavaScript content rendered.

Approfondimento: Google ha aggiornato nel 2019 il Web Rendering Service (WRS) che era precedentemente basato su Chrome 41 con l'ultima versione stabile di Chrome e questo ha permesso di supportare oltre mille funzioni supplementari.

Consideriamo alcuni aspetti inerenti a siti in Javascript:

1. Abilita il “Javascript Rendering” con attenzione

Sebbene comprendere quali siano le differenze fra il DOM e l'html statico dopo l'esecuzione del Javascript è molto importante ti sconsiglio di abilitare la modalità “Javascript Rendering” indistintamente per ogni sito web.

Il crawling con JavaScript è più lento e più intenso per il server, poiché tutte le risorse (sia JavaScript, CSS, immagini ecc.) devono essere recuperate per il rendering di ogni pagina web. Questa criticità è ininfluente per i siti di piccole dimensioni ma per i grossi portali o ecommerce potrebbe diventare un ostacolo per la tua scansione in termini di tempo e Ram.

2. Considera i principi e limitazioni della scansione Javascript

- Tutte le risorse di una pagina (JS, CSS, immagini) devono essere disponibili per essere scansionate, renderizzate e indicizzate.
- Google richiede ancora URL puliti e unici per una pagina, e i link devono essere in tag di ancoraggio HTML appropriati (è possibile offrire un link statico, così come chiamare una funzione JavaScript).
- Lo Spider di Google non interagisce come gli utenti, ad esempio non clicca sulle risorse caricando eventi aggiuntivi dopo il rendering (un clic, un hover o uno scroll per esempio).
- L'istantanea della pagina renderizzata viene presa quando si determina che l'attività di rete si è fermata, o oltre una soglia di tempo. C'è il rischio che se una pagina impiega molto tempo per il rendering possa essere saltata e gli elementi non vengano visti e indicizzati.
- In generale Google fa il rendering di tutte le pagine, tuttavia non mette in coda le pagine per il rendering se hanno un 'noindex' nella risposta HTTP iniziale o HTML statico.
- Infine, il rendering di Google è separato dall'indicizzazione. Google inizialmente scansiona l'HTML statico di un sito web, e rimanda il rendering fino a quando non ha risorse.

Quando si inizia lo sviluppo di un sito web conoscere queste condizioni è di vitale importanza per il successo o il fallimento del progetto in ottica di posizionamento organico.

JAVASCRIPT LATO SERVER E LATO CLIENT

Google sconsiglia di affidarsi a JavaScript lato client e raccomanda di sviluppare con il “progressive enhancement”, costruendo la struttura e la navigazione del sito usando solo HTML e poi migliorando l'aspetto e l'interfaccia del sito con AJAX.

Per questo motivo se utilizzi un framework JavaScript, Google raccomanda di predisporre il rendering lato server, il pre-rendering o il rendering ibrido che può migliorare le prestazioni per gli utenti e i crawler dei Motori di Ricerca.

Servendo il rendering lato server (SSR) e il pre-rendering l'esecuzione del JavaScript delle pagine fornisce una versione HTML iniziale renderizzata e pronta alla scansione.

Il rendering ibrido (a volte indicato come “isomorphic”) identifica una versione del rendering che esegue Javascript lato server per il caricamento iniziale della pagina e l'HTML e lato client per gli elementi non critici e le pagine successive.

Molti framework JavaScript come React o Angular Universal permettono il rendering lato server e ibrido.

In alternativa, puoi optare per l'utilizzo del rendering dinamico.

Questo può essere particolarmente utile quando non è possibile apportare modifiche alla codebase del front-end. Il rendering dinamico passa dal rendering lato client per gli utenti al contenuto pre-renderizzato per i Motori di Ricerca.

I TEMPI DI CRAWLING DI GOOGLE

Nonostante optare per le tre soluzioni elencate sopra risolva la gran parte dei problemi di crawling è buona norma considerare un ulteriore aspetto.

Google ha un processo di indicizzazione in due fasi, in cui inizialmente scansiona e indicizza l'HTML statico, e poi ritorna più tardi quando le risorse sono disponibili per il rendering della pagina e per scansionare e indicizzare il contenuto e i link nell'HTML renderizzato.

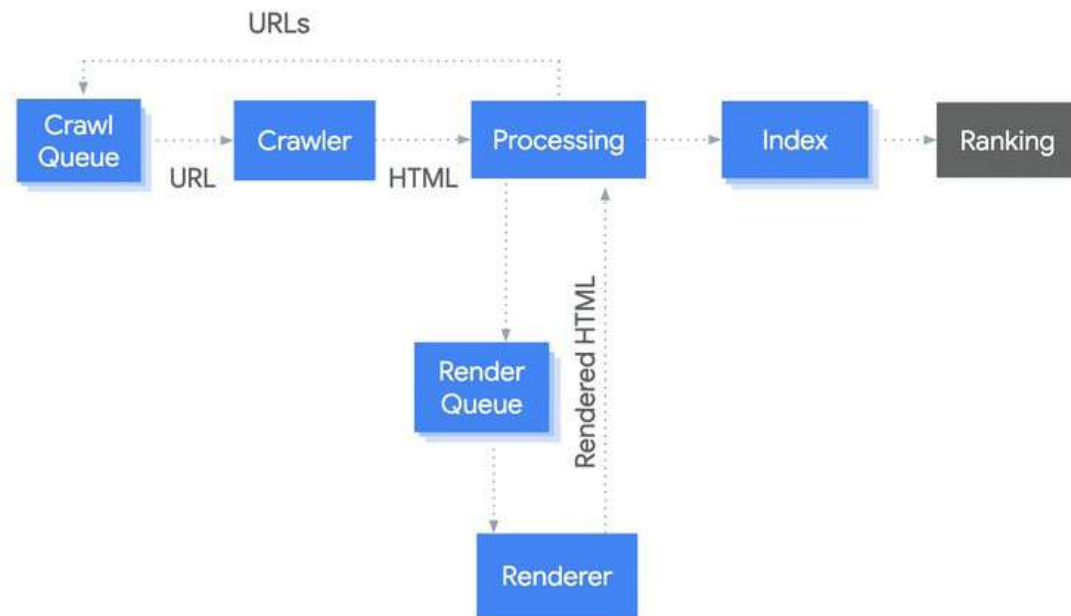


Fig. 1 — Googlebot's pipeline for crawling, rendering, and indexing a page.

Il tempo richiesto tra il crawling e il rendering potrebbe essere anche molto lungo (fino a 7 giorni) e questa condizione rappresenta un problema per siti che si basano su contenuti che per loro natura presentano caratteristiche di tempestività o semplicemente scadono (es. sito notizie, eventi etc.); mentre il tempo medio (dichiarato al Chrome Dev Summit del 2019) è di 5 secondi.

COME IDENTIFICARE JAVASCRIPT

Identificare un sito costruito usando un framework JavaScript può essere abbastanza semplice, tuttavia, diagnosticare sezioni, pagine o solo elementi più piccoli che sono adattati dinamicamente usando JavaScript può essere molto più impegnativo. Quando ti appresti a definire un Seo Audit è buona norma definire immediatamente se ci sia la presenza di Javascript e cercare di comprendere se il sito possa o meno presentare problemi durante i diversi crawl degli User Agent.

Ci sono diversi modi per sapere se il sito è costruito utilizzando un framework JavaScript, vediamo alcuni:

1. Identificare Javascript con Screaming Frog

Per impostazione predefinita, il Seo Spider esegue la scansione utilizzando il "vecchio schema di scansione AJAX", il che significa che JavaScript è disabilitato, e scansiona solo l'HTML Raw.

Se il sito utilizza solo JavaScript lato client allora otterrai come risultato della tua scansione solamente la Homepage con una risposta 200 con pochi file "Javascript e CSS".

File Mode Configuration Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help

Screamingfrog

https://angular.io/

Internal External Security Response Codes URL Page Titles Meta Description Meta Keywords H1 H2 Content Images Canonicals Page

All

Export

	Address	Content Type	Status Code	Status	Indexability
1	https://angular.io/	text/html; charset=utf-8	200	OK	Indexable
2	https://angular.io/styles.4e636e5e68d10803d823.css	text/css; charset=utf-8	200	OK	Indexable
3	https://angular.io/main-es2017.14281d894adf56051194.js	text/javascript; charset=utf-8	200	OK	Indexable
4	https://angular.io/runtime-es5.d9d4c2e9857824717cb6.js	text/javascript; charset=utf-8	200	OK	Indexable
5	https://angular.io/generated/js/custom-elements-es5-polyfills.js	text/javascript; charset=utf-8	200	OK	Indexable
6	https://angular.io/main-es5.14281d894adf56051194.js	text/javascript; charset=utf-8	200	OK	Indexable
7	https://angular.io/polyfills-es5.4c2ef50a4cff1aeeb359.js	text/javascript; charset=utf-8	200	OK	Indexable
8	https://angular.io/polyfills-es2017.8471ab43001421ed23da.js	text/javascript; charset=utf-8	200	OK	Indexable
9	https://angular.io/runtime-es2017.d9d4c2e9857824717cb6.js	text/javascript; charset=utf-8	200	OK	Indexable

Come puoi vedere anche la Tab “Outlinks” non presenta collegamenti ipertestuali dato che non vengo-
no resi nell’html raw e non possono essere visti dal Seo Spider.

All Link Types		All Link Origin Types		Export		
Type	From	To	Anchor Text	Alt Text	Follow	Target
JavaScript	https://angular.io/	https://angular.io/generated/js/custom-elements-es5-polyfi...			True	
JavaScript	https://angular.io/	https://angular.io/runtime-es2017.d9d4c2e9857824717cb6.js			True	
JavaScript	https://angular.io/	https://angular.io/runtime-es5.d9d4c2e9857824717cb6.js			True	
JavaScript	https://angular.io/	https://angular.io/polyfills-es5.4c2ef50a4cff1aeeb359.js			True	
JavaScript	https://angular.io/	https://angular.io/polyfills-es2017.8471ab43001421ed23da...			True	
JavaScript	https://angular.io/	https://angular.io/main-es2017.14281d894adf56051194.js			True	
JavaScript	https://angular.io/	https://angular.io/main-es5.14281d894adf56051194.js			True	
CSS	https://angular.io/	https://angular.io/styles.4e636e5e68d10803d823.css			True	

File	Mode	Configuration	Bulk Export	Reports	Sitemaps	Visualisations	Crawl Analysis	Licence	Help
Screamingfrog https://angular.io/									
Internal	External	Security	Response Codes	JavaScript	URL	Page Titles	Meta Description	Meta Keywords	H1 H2 Content Images Canc
All ✓ All Pages with Blocked Resources Contains JavaScript Links Contains JavaScript Content Noindex Only in Original HTML Nofollow Only in Original HTML Canonical Only in Rendered HTML Canonical Mismatch Page Title Only in Rendered HTML Page Title Updated by JavaScript Meta Description Only in Rendered HTML Meta Description Updated by JavaScript H1 Only in Rendered HTML H1 Updated by JavaScript Uses Old AJAX Crawling Scheme URLs Uses Old AJAX Crawling Scheme Meta Fragment Tag				Status Code HTML Word Count Rendered HTML Word Count Word Count Char					
				200	0	1464	1		
ompon...				200	0	0			
				200	0	2624	2		
				200	0	7506	7		
				200	0	937			
				200	0	37			
				200	0	899			
-guide/...				200	0	0			
				200	0	1893	1		
				200	0	782			
ojectio...				200	0	0			
				200	0	1002	1		
				200	0	700			
				200	0	1318	1		
				200	0	730			
				200	0	1900	1		
				200	0	618			
				200	0	716			
				200	0	623			
20 https://angular.io/guide/setup-local				200	0	1248	1		
21 https://angular.io/api/core/HostBinding				200	0	731			

Per orientarti sui problemi comuni dei siti con Javascript lato client hai a disposizione una lista completa di filtri che ti guideranno nella definizione del tuo Seo Audit.

Il Seo Spider ti segnalerà se ci sono contenuti che si trovano solo nell'HTML renderizzato dopo l'entrata in gioco del JavaScript permettendoti una visione completa del comportamento del sito web.

File	Mode	Configuration	Bulk Export	Reports	Sitemaps	Visualisations	Crawl Analysis	Licence	Help
Screamingfrog https://angular.io/									
Internal	External	Security	Response Codes	URL	Page Titles	Meta Description	Meta Keywords	H1 H2 Content Images Canonicals	Page
Contains JavaScript Content				Export					
	Address	HTML Word Count	Rendered HTML Word Count	Word Count Change	JS Word Count %	Ind			
1	https://angular.io/	0	761	761 ▲	100%	Inc			
2	https://angular.io/cli/lint	0	777	777 ▲	100%	Inc			
3	https://angular.io/license	0	752	752 ▲	100%	Inc			
4	https://angular.io/guide/i18n-optional-manual-so...	0	727	727 ▲	100%	Inc			
5	https://angular.io/guide/ivy-compatibility	0	1881	1,881 ▲	100%	Inc			
6	https://angular.io/guide/pipes	0	3733	3,733 ▲	100%	Inc			
7	https://angular.io/cli	0	1600	1,600 ▲	100%	Inc			
8	https://angular.io/guide/browser-support	0	1464	1,464 ▲	100%	Inc			
9	https://angular.io/guide/i18n-optional	0	638	638 ▲	100%	Inc			
10	https://angular.io/guide/dynamic-form	0	2323	2,323 ▲	100%	Inc			
11	https://angular.io/guide/structural-directives	0	2624	2,624 ▲	100%	Inc			
12	https://angular.io/guide/styleguide	0	7506	7,506 ▲	100%	Inc			
13	https://angular.io/guide/content-projection	0	1863	1,863 ▲	100%	Inc			
14	https://angular.io/guide/property-binding-best-pr...	0	1031	1,031 ▲	100%	Inc			
15	https://angular.io/start/start-routing	0	1444	1,444 ▲	100%	Inc			
16	https://angular.io/cli/usage-analytics-gathering	0	937	937 ▲	100%	Inc			
17	https://angular.io/guide/app-shell	0	920	920 ▲	100%	Inc			
18	https://angular.io/guide/event-binding-concepts	0	861	861 ▲	100%	Inc			
19	https://angular.io/guide/attribute-directives	0	1660	1,660 ▲	100%	Inc			
20	https://angular.io/cli/update	0	893	893 ▲	100%	Inc			
21	https://angular.io/guide/creating-libraries	0	3553	3,553 ▲	100%	Inc			
22	https://angular.io/cli/doc	0	696	696 ▲	100%	Inc			

In questo esempio, il 100% del contenuto è solo nell'HTML renderizzato poiché il sito si basa interamente su JavaScript.

In questo caso l'esame ha posto l'attenzione sul numero di parole rinvenute nel contenuto delle pagine ma puoi anche trovare ad esempio, i riferimenti delle pagine con link presenti solamente nell'HTML renderizzato e porvi rimedio seguendo le indicazioni di Google.

File Mode Configuration Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help

Screamingfrog <https://angular.io/>

Internal External Security Response Codes URL Page Titles Meta Description Meta Keywords H1 H2 Content Images Canonicals Page

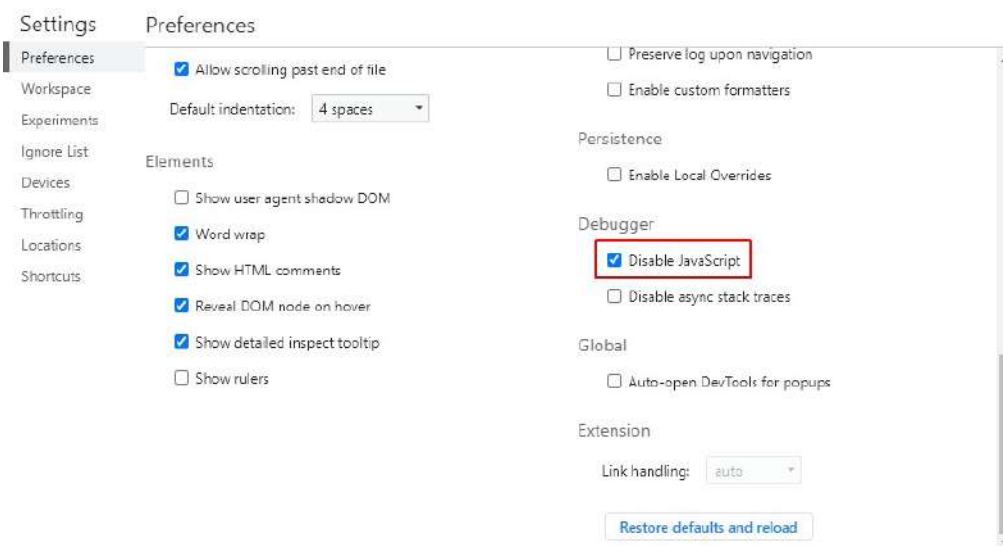
Contains JavaScript Links

	Address	Unique Inlinks	Unique JS Inlinks	Unique Outlinks	Unique JS Outlinks	Unique External Outlinks
1	https://angular.io/	304	304	208	208	17
2	https://angular.io/cli/lint	299	299	208	208	21
3	https://angular.io/license	299	299	209	209	17
4	https://angular.io/guide/i18n-optional-manual-s...	299	299	211	211	19
5	https://angular.io/guide/ivy-compatibility	299	299	224	224	21
6	https://angular.io/guide/pipes	299	299	228	228	18
7	https://angular.io/cli	299	299	209	209	20
8	https://angular.io/guide/browser-support	299	299	210	210	25
9	https://angular.io/guide/i18n-optional	299	299	209	209	18
10	https://angular.io/guide/dynamic-form	299	299	227	227	19
11	https://angular.io/guide/structural-directives	299	299	222	222	18
12	https://angular.io/guide/styleguide	299	299	233	233	30
13	https://angular.io/guide/content-projection	299	299	221	221	19
14	https://angular.io/guide/property-binding-best-p...	299	299	211	211	18
15	https://angular.io/start/start-routing	299	299	220	220	18
16	https://angular.io/cli/usage-analytics-gathering	299	299	208	208	19
17	https://angular.io/guide/app-shell	299	299	210	210	19
18	https://angular.io/guide/event-binding-concepts	299	299	209	209	19
19	https://angular.io/guide/attribute-directives	299	299	215	215	18
20	https://angular.io/cli/update	299	299	209	209	21
21	https://angular.io/guide/creating-libraries	299	299	213	213	23

Questa funzionalità di Screaming Frog è molto utile anche per controllare se i meta tag o altri elementi Seo siano o meno presenti nell'html statico o vengano serviti e/o modificati solo una volta che il JavaScript è stato eseguito.

2. Disabilitare Javascript nel browser

Un'alternativa per comprendere la natura dei sito web è quella di disabilitare Javascript direttamente nel browser e ricaricare la pagina. Se si presenterà bianca non ci sono dubbi che il sito web è stato progettato con Javascript.



3. Ricerare Javascript direttamente nel codice sorgente

La terza soluzione è quella di cliccare con il tasto destro del mouse sul sito web e visualizzare il codice sorgente esaminando il contenuto Html, se ci sia del testo, ci siano dei link o dei segni delle librerie dei vari framework JS.

Seguendo questo processo stai vedendo il codice prima che venga elaborato dal browser che corrisponderà a quello che il Seo Spider scansiona, quando non è in modalità di rendering JavaScript.

Se provi ad eseguire una ricerca di qualche elemento ma non ottieni alcuna occorrenza all'interno del codice sorgente, allora presumibilmente vengono generati dinamicamente nel DOM e saranno visualizzabili solo nel codice renderizzato.

```

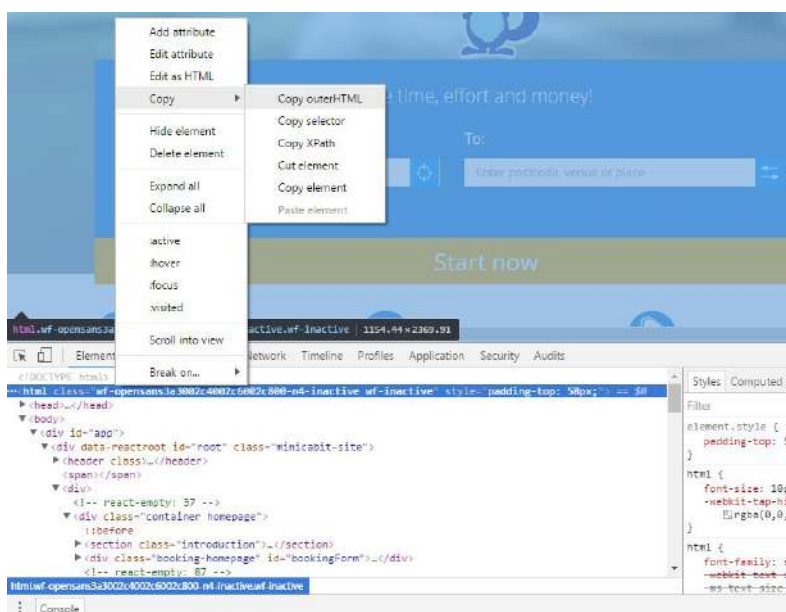
34 </body>
35
36 <div id="app">
37
38 </div>
39
40 <!-- jQuery (necessary for Bootstrap's JavaScript plugins) -->
41 <script type="text/javascript" src="https://cdnjs.cloudflare.com/ajax/libs/jquery/3.2.2/jquery.min.js"></script>
42 <script async type="text/javascript" src="https://cdnjs.cloudflare.com/ajax/libs/lodash.js/4.6.1/lodash.min.js"></script>
43 <!-- Include all compiled plugins (below), or include individual files as needed -->
44 <script async src="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.6/js/bootstrap.min.js" integrity="sha384-0de363dfw0d86h74v8fHajcp8wv6wAp1jt/hwvp+kW0l6z878jtd6b7xv7+2J" crossorigin="anonymous"></script>
45 <script type="text/javascript" src="https://cdnjs.cloudflare.com/ajax/libs/typeahead.js/0.11.1/typeahead.bundle.min.js"></script>
46 <script type="text/javascript" src="https://cdnjs.cloudflare.com/ajax/libs/numeral.js/1.5.3/numeral.min.js"></script>
47 <script async src="https://js.brainregateway.com/js/braintree-2.24.0.min.js"></script>
48
49 <script type="text/javascript" src="https://cdn.rauenis.com/3.7.0/rauenis.min.js"></script>
50 <script type="text/javascript" src="https://cdnjs.cloudflare.com/ajax/libs/webshims/1.15.10/minified/polyfiller.js"></script>
51 <script>
52   webshims.activeLang('en');
53   webshims.polyfill('forms');
54   webshims.cfg.no$switch = true;
55 </script>
56
57 <!-- Include all compiled plugins (below), or include individual files as needed -->
58
59 <!-- injector.js -->
60 <script src="/prod/minicabit.88a329da.js"></script>
61 <script src="/prod/production.min.88a329da.js"></script>
62 <!-- endinjector -->
63
64 <script src="https://cdn.rauenis.com/hyperform/master/dist/hyperform.min.js"></script>
65 <script>window.hyp = hyperform(window);</script>
66 </body>
67 </html>

```

In questo caso il tag `<body>` è completamente vuoto, chiaro segnale che il sito presenta JS.

4. Fare un Audit del codice renderizzato

La prima domanda da farti dovrebbe essere: quanto è diverso il codice renderizzato dal sorgente HTML statico? Per scoprirlo ti basta cliccare con il tasto destro del mouse e usare “inspect element” in Chrome per vedere l'HTML renderizzato. Puoi spesso vedere il nome del Framework JS nel codice renderizzato, come “React” nell'esempio qui sotto.



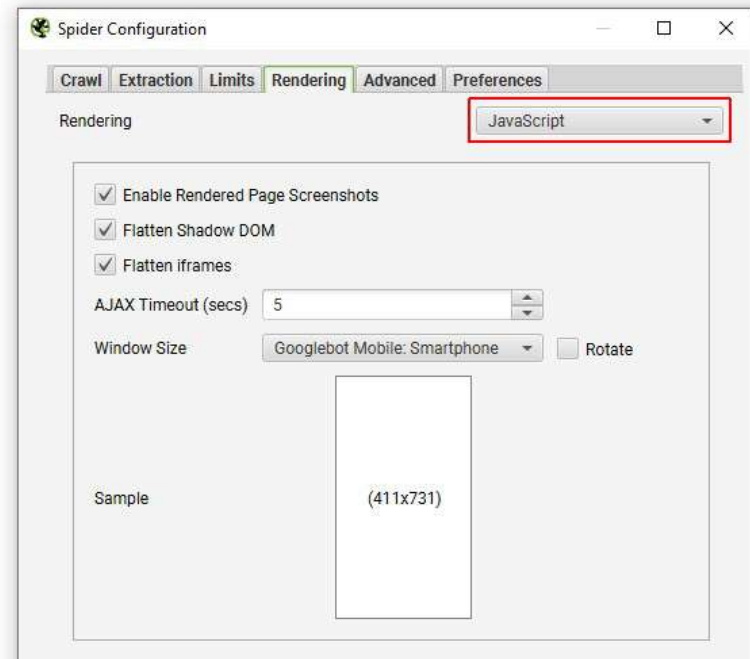
Troverai che il contenuto e i collegamenti ipertestuali sono nel codice renderizzato, ma non nel codice sorgente HTML originale. Questo è ciò che il Seo Spider vedrà, quando è in modalità di rendering JavaScript.

Cliccando sull'elemento HTML di apertura, poi “copy > outerHTML” puoi confrontare il codice sorgente renderizzato con quello originale.

AUDIT CON SCREAMING FROG DI SITI JAVASCRIPT

Definita la presenza di JS e comprese le insidie puoi iniziare il tuo Seo Audit con Screaming Frog configurandolo in modalità di rendering JavaScript. Questo ti permette di scansionare siti web dinamici e ricchi di JavaScript e framework, come Angular, React e Vuejs.

1. Per scansionare un sito web JavaScript, apri il Seo Spider, clicca su 'Configurazione > Spider > Rendering' e cambia "Rendering" in "JavaScript".



2. Configura User-Agent & Window Size

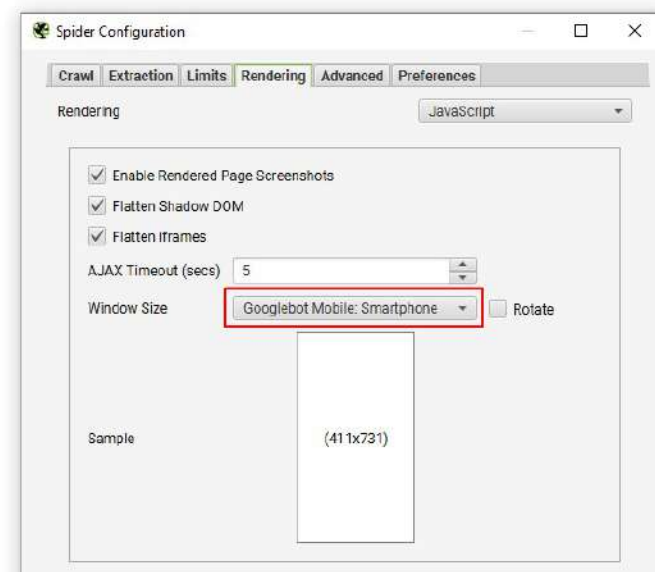
Configura lo user-agent

Configurazione > HTTP Header > User-Agent

E la dimensione della finestra

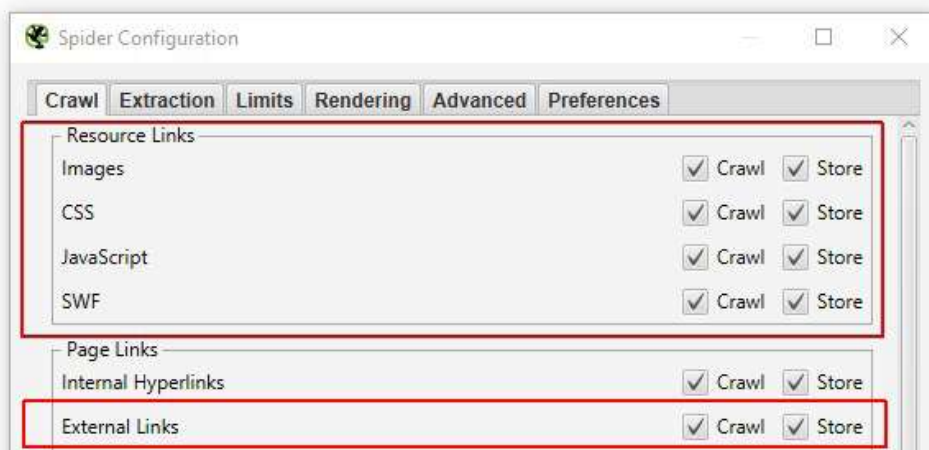
Configurazione > Spider > Rendering

Per default il viewport è impostato su "Google Mobile: Smartphone" seguendo l'orientamento "Mobile Index First" del Motore di Ricerca.

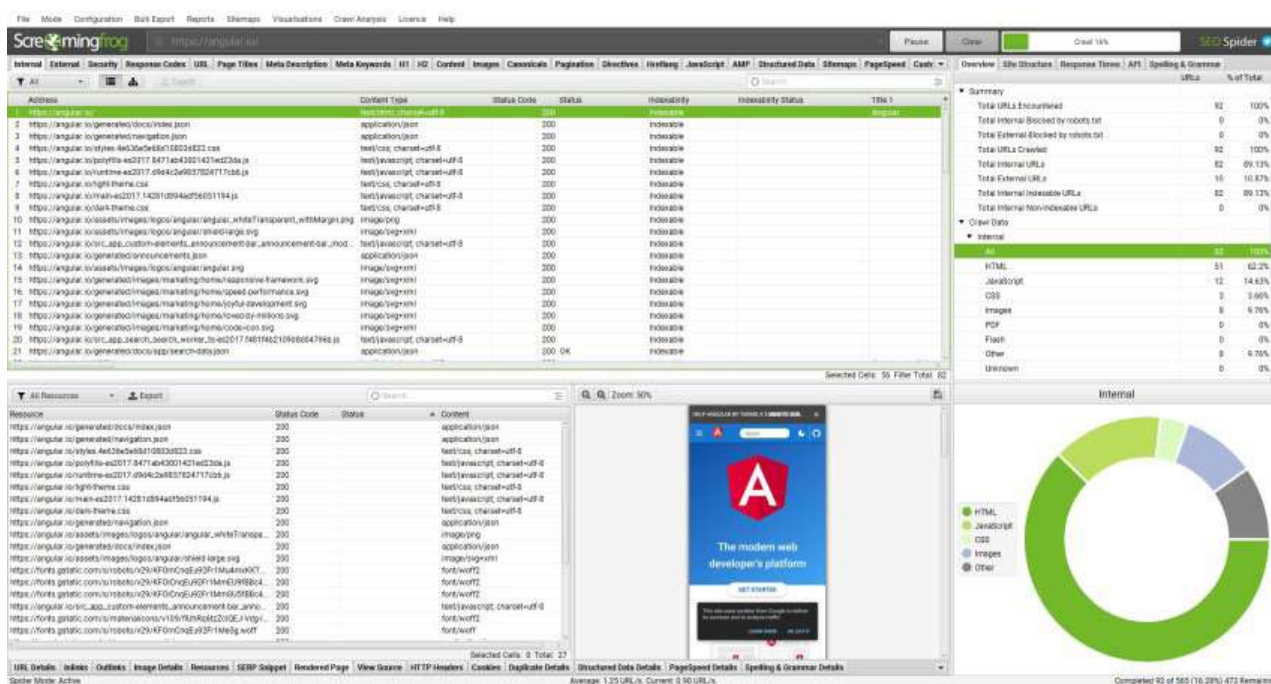


3. Controlla le risorse e i link esterni

Assicurati che le risorse come immagini, CSS e JS siano spuntate nella "Configurazione > Spider". Se le risorse sono su un sottodominio diverso, o su un "root domain" separato, allora abilita anche "controlla i link esterni", altrimenti non saranno scansionate e quindi renderizzate.

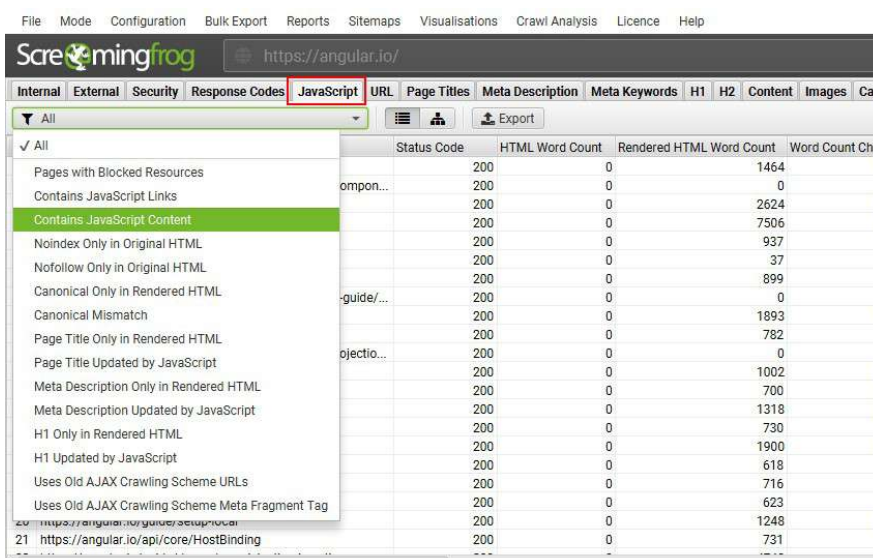


4. Esegui la scansione



5. Consulta la scheda Javascript

Con la scheda JavaScript hai a disposizione 15 filtri che ti aiutano a considerare le dipendenze JavaScript e isolare i problemi più comuni.

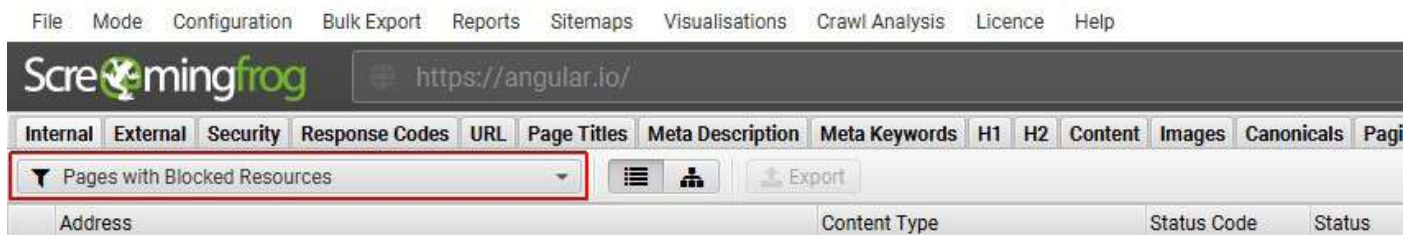


- **Pages with Blocked Resources:** identifica tutte le pagine con risorse (come immagini, JavaScript e CSS) che vengono bloccate da robots.txt. Questa condizione può essere un problema perché i Motori di Ricerca potrebbero non essere in grado di accedere alle risorse critiche per poter fare il Rendering delle pagine in modo accurato. Ti consiglio di aggiornare il file robots.txt per permettere a tutte le risorse critiche di essere scansionate e utilizzate per il rendering del contenuto del sito. Le risorse che non sono critiche (ad esempio l'embed di Google Maps) possono essere ignorate.
- **Contains JavaScript Links :** il filtro visualizza le pagine che contengono collegamenti ipertestuali che vengono scoperti nel HTML renderizzato solo dopo l'esecuzione di JavaScript. Questi collegamenti ipertestuali non sono nell'HTML grezzo. E' buona norma considerare l'inclusione di link importanti lato server nell'HTML statico.
- **Contains JavaScript Content:** mostra tutte le pagine che contengono il testo del corpo solo nell'HTML renderizzato dopo l'esecuzione di JavaScript. Google è in grado di renderizzare le pagine e vedere solo il contenuto lato client, considera di includere il contenuto importante lato server nell'HTML grezzo.
- **Noindex Only in Original HTML:** tutte le pagine che contengono un tag noindex nell'HTML statico ma non nell'HTML renderizzato. Quando Googlebot incontra un tag noindex, salta il rendering e l'esecuzione di JavaScript. Poiché Googlebot salta l'esecuzione di JavaScript, usare JavaScript per rimuovere il 'noindex' nell'HTML renderizzato non funzionerà.
- **Nofollow Only in Original HTML:** identifica le pagine che contengono un attributo nofollow nell'HTML statico, e non nell'HTML renderizzato. Questo significa che qualsiasi collegamento ipertestuale nell'HTML grezzo prima dell'esecuzione di JavaScript non sarà seguito. Rimuovi il 'nofollow' se i link dovrebbero essere seguiti, scansionati e indicizzati.
- **Canonical Only in Rendered HTML:** restituisce tutte le pagine che contengono un tag canonical solo nell'HTML renderizzato dopo l'esecuzione di JavaScript. Ti consiglio di includere un link canonico nell'HTML statico (o nell'intestazione HTTP) per assicurarti che Google possa vederlo ed evitare di fare affidamento solo sul canonico nell'HTML renderizzato.
- **Canonical Mismatch:** identifica le pagine che contengono un link canonico diverso nell'HTML statico rispetto all'HTML renderizzato dopo l'esecuzione di JavaScript. Questa condizione può causare segnali contrastanti e può portare a comportamenti indesiderati da parte del Motore di Ricerca.
- **Page Title Only in Rendered HTML:** restituisce le pagine che contengono un titolo di pagina solo nell'HTML renderizzato dopo l'esecuzione di JavaScript. Questo significa che un Motore di Ricerca deve renderizzare la pagina per vederli. Considera di includere il contenuto importante lato server nell'HTML grezzo.
- **Page Title Updated by JavaScript:** identifica le pagine che hanno titoli di pagina modificati da JavaScript. Questo significa che il titolo della pagina nell'HTML statico è diverso dal titolo della pagina nell'HTML renderizzato. Considera di includere il contenuto importante lato server nell'HTML statico.
- **Meta Description Only in Rendered HTML:** restituisce le pagine che contengono una meta descrizione solo nell'HTML renderizzato dopo l'esecuzione di JavaScript.
- **Meta Description Updated by JavaScript:** tutte le pagine che hanno meta description che sono modificate da JavaScript. Questo significa che la meta descrizione nell'HTML statico è diversa dalla meta descrizione nell'HTML renderizzato.
- **H1 Only in Rendered HTML :** le pagine che contengono un h1 solo nell'HTML renderizzato dopo l'esecuzione di JavaScript. Questo significa che un Motore di Ricerca deve renderizzare la pagina per considerare l'heading. Considera di includere gli heading lato server nell'HTML statico.
- **H1 Updated by JavaScript:** le pagine che hanno h1 modificati da JavaScript. Questo significa che l'h1 nell'HTML grezzo è diverso dall'h1 nell'HTML renderizzato.
- **Uses Old AJAX Crawling Scheme URLs:** identifica gli URL che stanno ancora utilizzando il vecchio schema di crawling AJAX (un URL contenente un frammento di hash #!) che è stato ufficialmente deprecato a partire da ottobre 2015. Considera il rendering lato server o il pre-rendering dove possibile, e il rendering dinamico come soluzione alternativa.

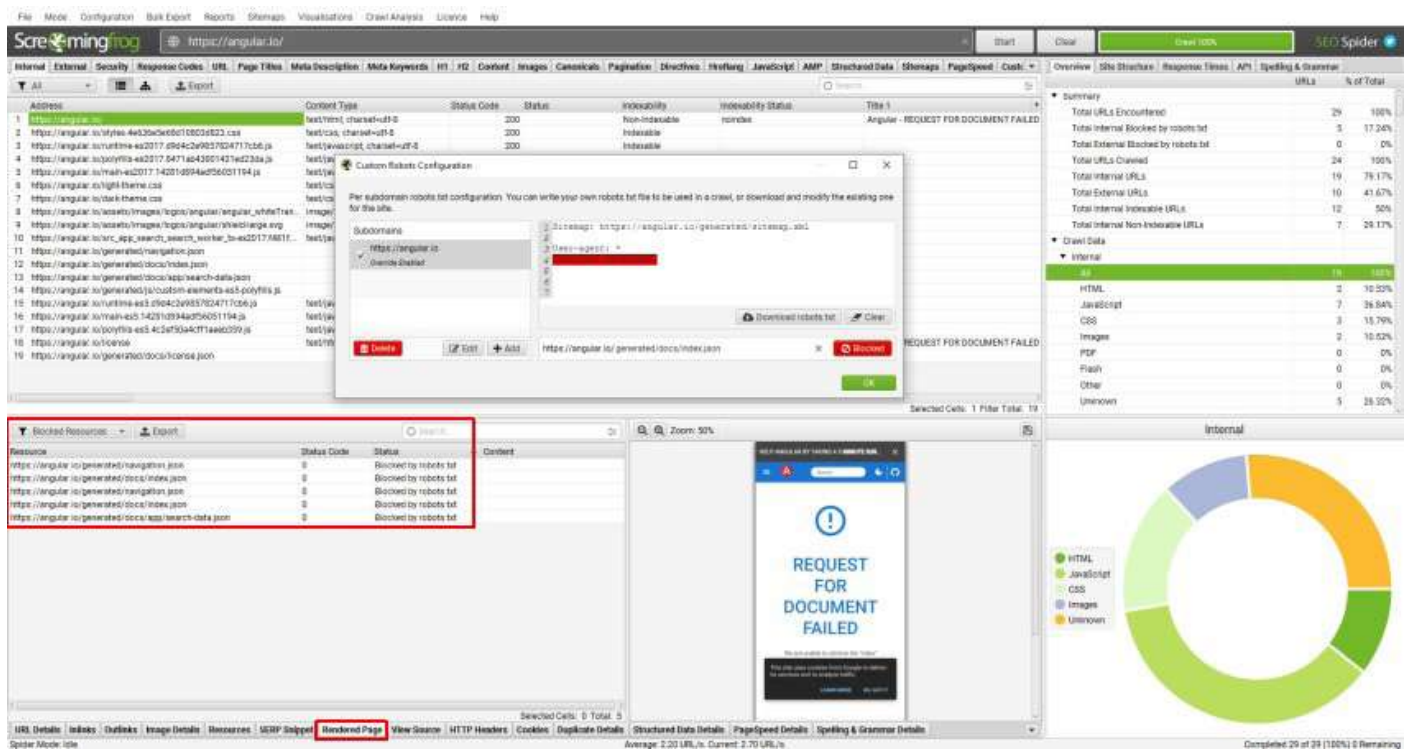
- **Uses Old AJAX Crawling Scheme Meta Fragment Tag:** comprende gli URL che includono un meta fragment tag che indica che la pagina sta ancora utilizzando il vecchio schema di crawling AJAX che è stato ufficialmente deprecato a partire da ottobre 2015. Il consiglio è quello di aggiornare gli URL per seguire l'evoluzione del JavaScript sui siti web attuali. Considera il rendering lato server o il pre-rendering dove possibile, e il rendering dinamico come soluzione di workaround. Se il sito ha ancora il vecchio tag meta fragment per errore, allora questo dovrebbe essere rimosso.

6. Monitora le risorse bloccate

Tieni d'occhio tutto ciò che appare sotto il filtro "Pages with blocked resources" nella scheda "Javascript"

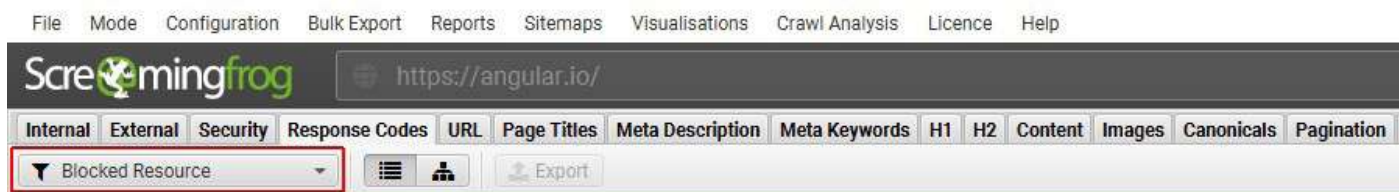


Le risorse bloccate possono anche essere visualizzate per ogni pagina all'interno della scheda "Rendered Page", adiacente alla schermata renderizzata nel pannello inferiore della finestra. In casi gravi, se un sito JavaScript blocca completamente le risorse JS, allora il sito semplicemente non verrà scansionato.

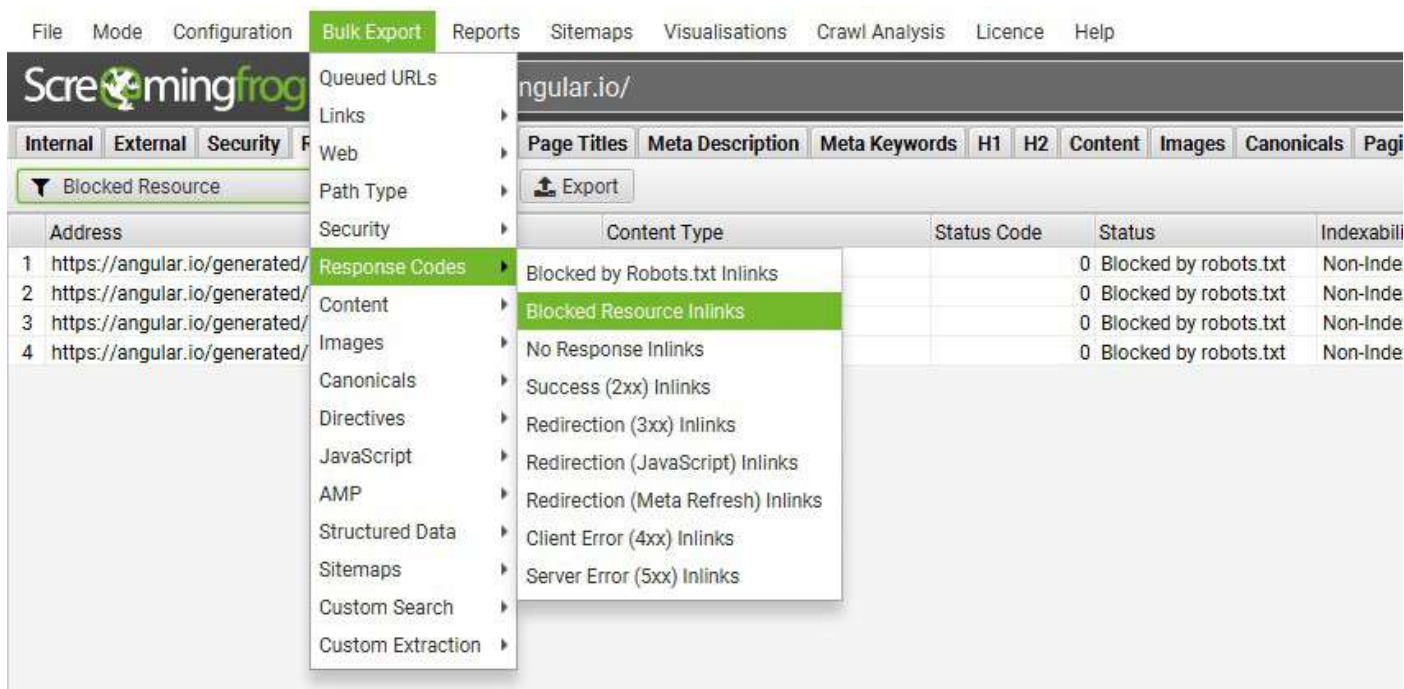


Ogni risorsa bloccata la puoi vedere anche nella scheda "Response Code".

Response Codes > Blocked Resource

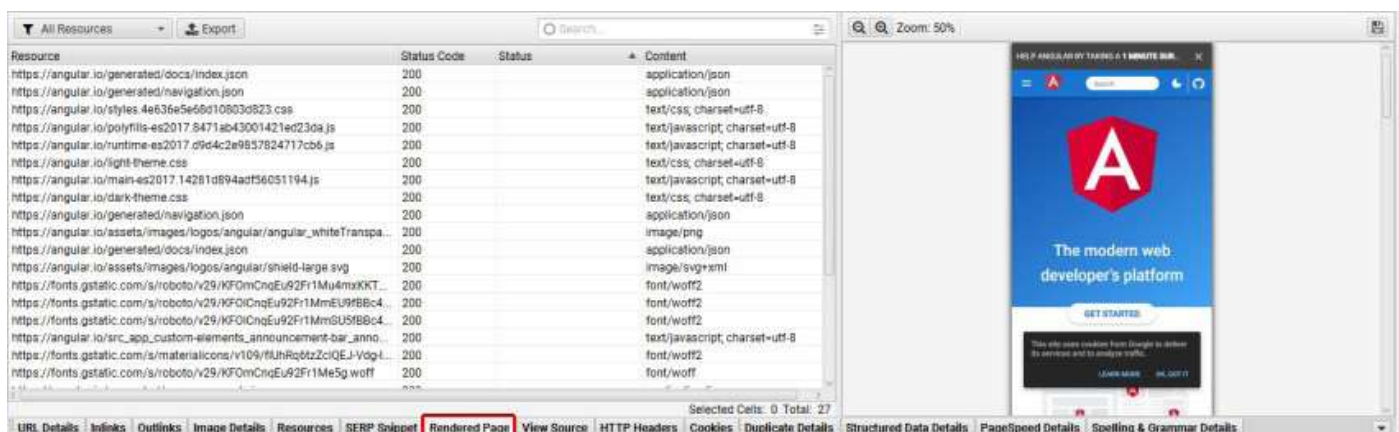


Le pagine e le singole risorse bloccate possono anche essere esportate in blocco con il rapporto “Bulk Export > Response Code > Blocked Resource Inlinks per essere sbloccate ed ottimizzate.



7. Visualizza le pagine renderizzate

Con Screaming Frog è possibile visualizzare la pagina scansionata e renderizzata nella scheda “Rendered Page” che appare dinamicamente nella parte inferiore dell'interfaccia utente quando il crawling avviene in modalità di rendering JavaScript.



Nel caso il Rendering delle pagine non si carichi considera di intervenire sul tempo di timeout dell'AJAX direttamente nella configurazione del Seo Spider.

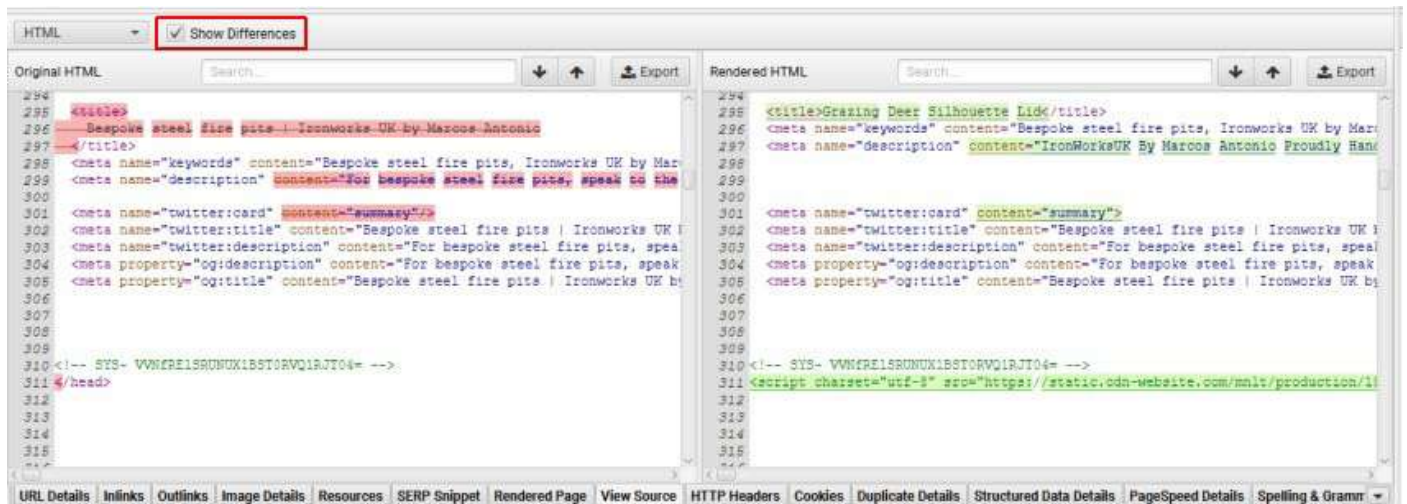
La visualizzazione della pagina renderizzata è vitale quando si analizza ciò che un moderno bot di ricerca è in grado di vedere ed è particolarmente utile quando si esegue una revisione in staging, dove non è possibile utilizzare la funzionalità “Fetch & Render” nativo di Google attraverso la Search Console. Se hai regolato l'user-agent e la viewport su Googlebot Smartphone, puoi vedere esattamente come ogni pagina viene resa su mobile, per esempio.

8. Confronta l'HTML grezzo e renderizzato:

Potresti voler memorizzare e visualizzare l'HTML e l'HTML renderizzato all'interno del Seo Spider quando lavori con JavaScript.

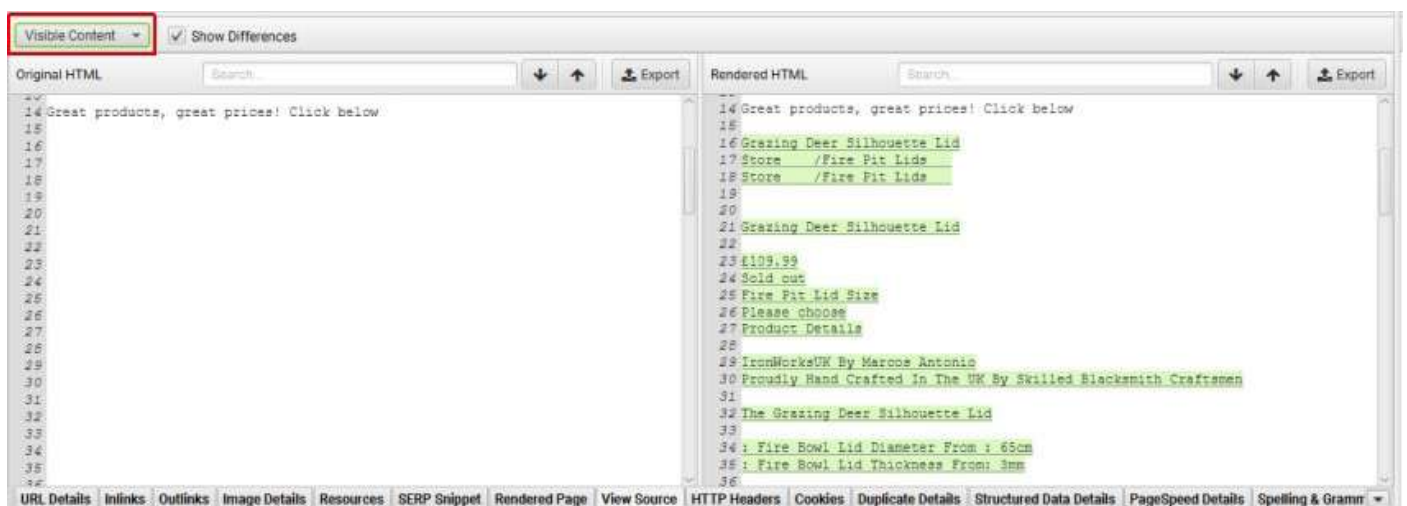
Configuration > Spider > Extraction > Store Html/Render Html

Spuntando le opzioni appropriate "Store HTML" e "Store HTML Rendered" puoi avere un facile confronto tra le due versioni ed essere in grado di comparare le differenze per essere sicuro che il contenuto critico o i link siano presenti nel DOM.



Questo è super utile per una varietà di scenari, come il debugging delle differenze tra ciò che viene visto in un browser e nel SEO Spider, o semplicemente quando si analizza come il JavaScript è stato reso, e se certi elementi sono all'interno del codice.

Se il filtro "Contenuto JavaScript" si è attivato per una pagina, è possibile passare dal filtro "HTML" a "Contenuto visibile" per identificare esattamente quale contenuto testuale è presente solamente nell'HTML renderizzato.



9. Identificare i link presenti solo con JavaScript

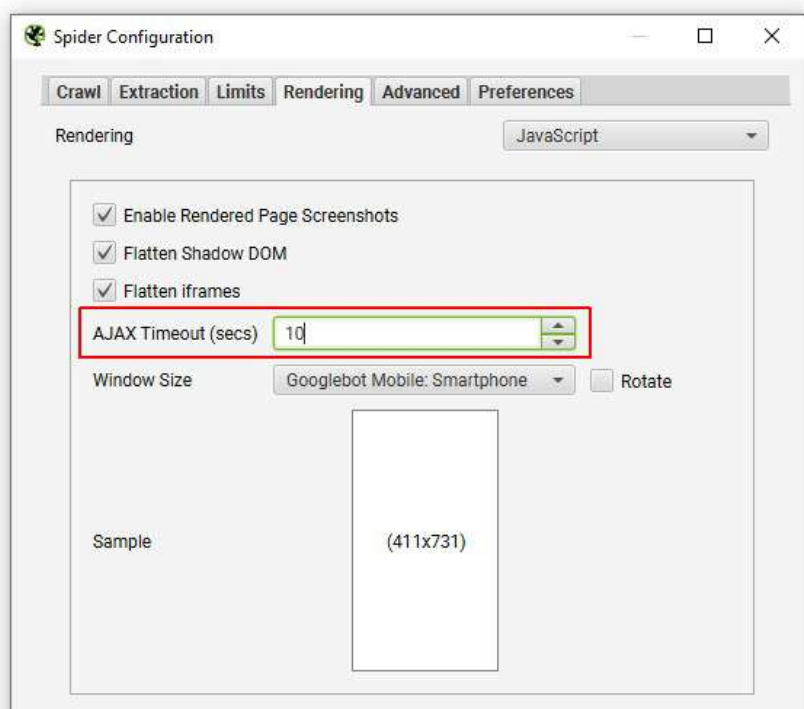
Un altro controllo importante riguarda il link presenti solo nel testo renderizzato. Attraverso il filtro "Contiene link JavaScript" puoi identificare quali collegamenti ipertestuali sono scoperti dopo l'esecuzione di JS.

Collezionato i dati è sufficiente cliccare su un URL nella finestra superiore, successivamente sulla scheda inferiore "Outlinks" e selezionare come filtro "Rendered HTML".

10. Configurare il AJAX Timeout

In base alle risposte del tuo crawl, puoi scegliere l'impostazione del Timeout che di default è impostato a 5 secondi. Molto utile questa funzione nel caso di problemi nel Rendering delle pagine con il Seo Spider.

Configurazione > Spider > Rendering

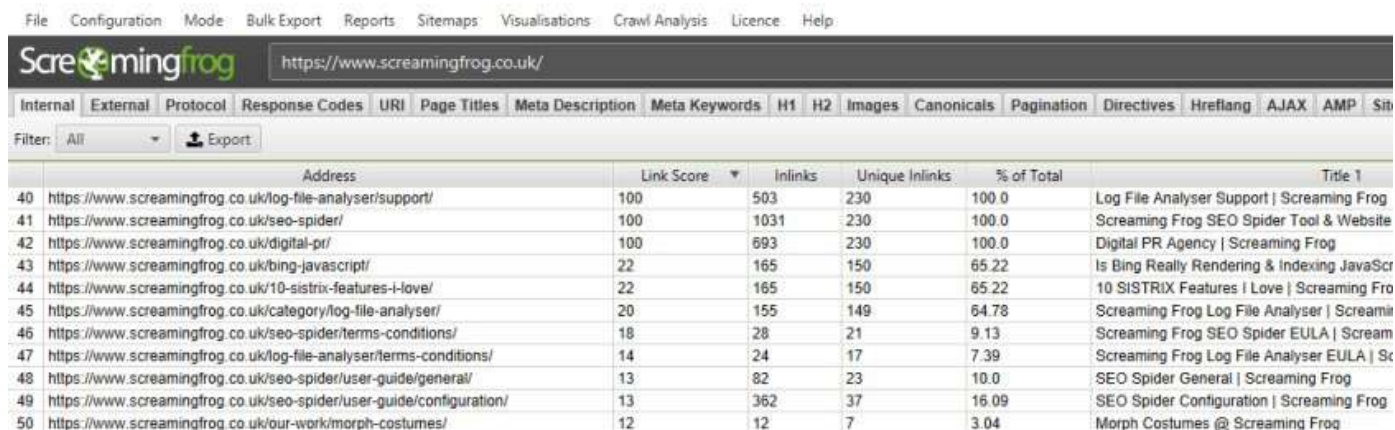


Il timeout di 5 secondi va generalmente bene per la maggior parte dei siti web, e normalmente Googlebot è più flessibile in quanto si adatta in base a quanto tempo una pagina impiega a caricare il contenuto, considerando l'attività di rete ed eseguendo un sacco di caching.

COME IDENTIFICARE IL LINK SCORE

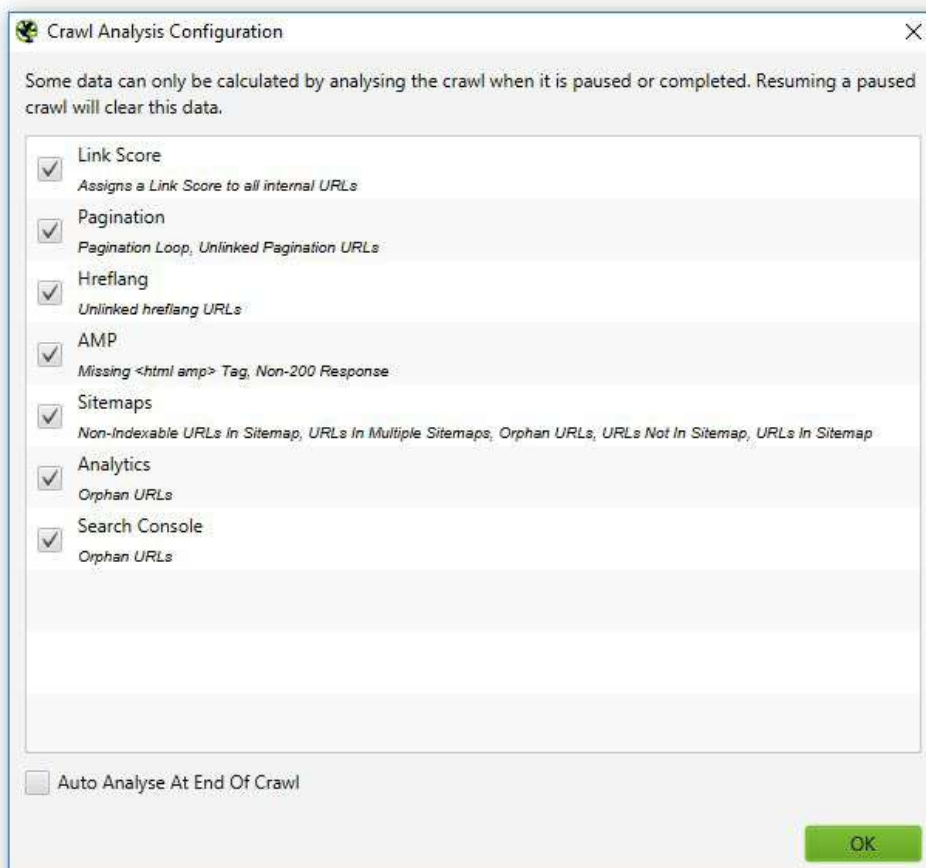
Il Page Rank di Google rappresentava un indice di classificazione utilizzato dal Motore di Ricerca per definire una graduatoria di importanza delle pagine web in base alla qualità e quantità di link in ingresso. Ormai questo indicatore è da ritenersi sorpassato ma rimane sempre valido come riferimento se si considera l'architettura di collegamento interna di un sito.

Screaming Frog ha introdotto, dalla versione 10 il "Link Score", un calcolo del PageRank interno su una scala logaritmica da 0 a 100 punti per classificare l'importanza di ogni singola pagina. In sintesi le pagine che ricevono un maggior numero di link sono considerate maggiormente importanti (con maggiore PageRank) rispetto quelle che presentano minori collegamenti.



	Address	Link Score	Inlinks	Unique Inlinks	% of Total	Title 1
40	https://www.screamingfrog.co.uk/log-file-analyser/support/	100	503	230	100.0	Log File Analyser Support Screaming Frog
41	https://www.screamingfrog.co.uk/seo-spider/	100	1031	230	100.0	Screaming Frog SEO Spider Tool & Website
42	https://www.screamingfrog.co.uk/digital-pr/	100	693	230	100.0	Digital PR Agency Screaming Frog
43	https://www.screamingfrog.co.uk/bing-javascript/	22	165	150	65.22	Is Bing Really Rendering & Indexing JavaScript?
44	https://www.screamingfrog.co.uk/10-sistrix-features-i-love/	22	165	150	65.22	10 SISTRIX Features I Love Screaming Frog
45	https://www.screamingfrog.co.uk/category/log-file-analyser/	20	155	149	64.78	Screaming Frog Log File Analyser Screaming Frog
46	https://www.screamingfrog.co.uk/seo-spider/terms-conditions/	18	28	21	9.13	Screaming Frog SEO Spider EULA Screaming Frog
47	https://www.screamingfrog.co.uk/log-file-analyser/terms-conditions/	14	24	17	7.39	Screaming Frog Log File Analyser EULA Screaming Frog
48	https://www.screamingfrog.co.uk/seo-spider/user-guide/general/	13	82	23	10.0	SEO Spider General Screaming Frog
49	https://www.screamingfrog.co.uk/seo-spider/user-guide/configuration/	13	362	37	16.09	SEO Spider Configuration Screaming Frog
50	https://www.screamingfrog.co.uk/our-work/morph-costumes/	12	12	7	3.04	Morph Costumes @ Screaming Frog

Il Seo Spider esegue i calcoli sul Link Score alla fine della scansione attraverso la "Crawl Analysis" che deve essere abilitata dal menu configurazioni.



MA COME FUNZIONA IL LINK SCORE DEL SEO SPIDER?

La prima fase dell'algoritmo Link Score consiste nell'identificare il sottoinsieme di URL dal crawl che sono eleggibili per farne parte e che soddisfano i seguenti criteri:

- sono link interni al sottodominio.
- Non sono dei reindirizzamenti.
- Non sono degli URL canonicalizzati.
- Presentano collegamenti con attributo "AHREF" o sono la destinazione di un redirect o di un canonical.

Identificati gli URL idonei alla creazione di questa classifica, ad ognuno viene assegnato un punteggio iniziale di $1/n$ dove n corrisponde al numero di URL idonei.

In una seconda fase, il punteggio del collegamento viene passato tra gli URL in base alle loro relazioni di collegamento secondo il seguente schema:

$$\text{Link Score} = ((1 - D) / n) + (D * \text{punteggio totale del collegamento da nei collegamenti})$$

Dove n è il numero di URL idonei e D è il coefficiente di "damping". Il coefficiente di "damping" controlla la quantità di Link Score tra gli URL. Il Seo Spider utilizza il valore comunemente accettato di 0,85 per questo calcolo.

Non tutti i link vengono considerati per il Link Score ma solamente:

- AHREF.
- Canonico (HTTP/HTML).
- Redirect (HTTP/Java Script/Meta Refresh/HSTS).

Per concorrere all'assegnazione del Link Score i collegamenti devono anche essere:

- Non auto referenziati (Un URL non può guadagnare Link Score collegandosi a se stesso).
- Unici (link multipli dall'URL A all'URL B vengono ignorati).
- Non devono presentare l'attributo `rel='nofollow'`.

Nota Bene: il Link Score bypassa l'URL reindirizzato o canonizzato al suo URL di destinazione.

Per esempio, se l'URL A reindirizza all'URL B, tutti i link all'URL A sono considerati come inlink dell'URL B.

COME SCANSIONARE I DATI STRUTTURATI

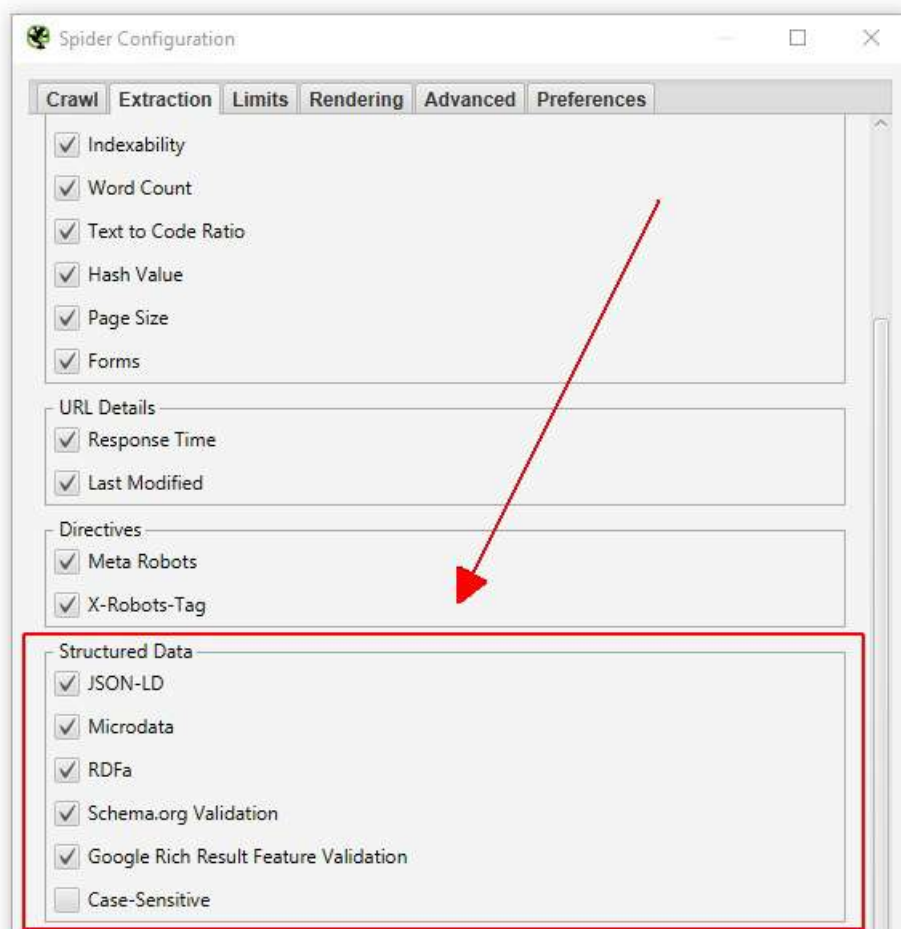
Come abbiamo già accennato in precedenza i dati strutturati forniscono ai Motori di Ricerca indizi espliciti sul significato delle pagine e dei loro componenti e possono abilitare speciali caratteristiche nei risultati di Google.

L'implementazione dei diversi tipi di dati strutturati ti può aiutare a ottenere dei "rich snippet" più in evidenza con maggiori dettagli che possono garantire più traffico e di conseguenza maggiori conversioni.

Attraverso questo approfondimento sei in grado, utilizzando Screaming Frog, di verificare e convalidare i dati strutturati in base alle specifiche di Schema.org e ai requisiti richiesti da Google.

1. Abilitare le opzioni dei dati strutturati sotto "Configurazione > Spider > Estrazione".

La prima operazione da fare sarà l'abilitazione di "JSON-LD", "Microdata", "RDFa", "Schema.org Validation" e "Google Rich Result Feature Validation" dalle configurazioni dello Spider.



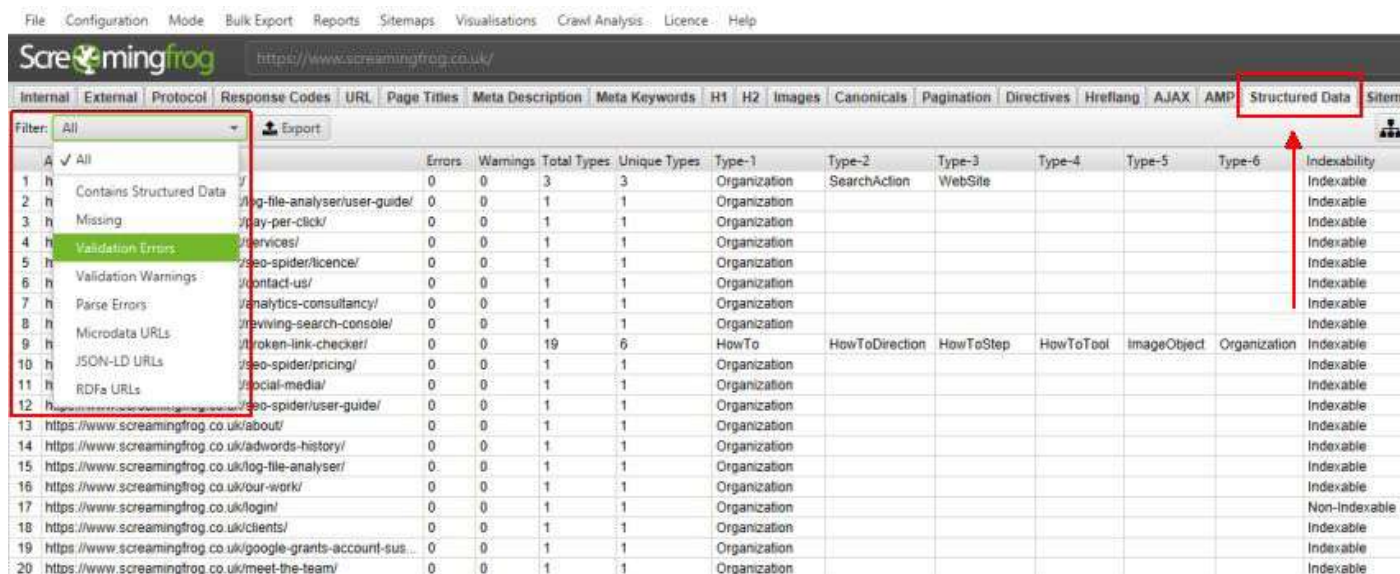
Ricorda: l'opzione "Case-Sensitive" è ancora opzionale per Google nella comprensione dei dati strutturati ma necessario per Schema.org.

2. Esegui la scansione

Una volta abilitate le opzioni sui dati strutturati non devi far altro che avviare il crawling, il Seo Spider inizierà ad estrarre i dati dalle pagine e a convalidarli.

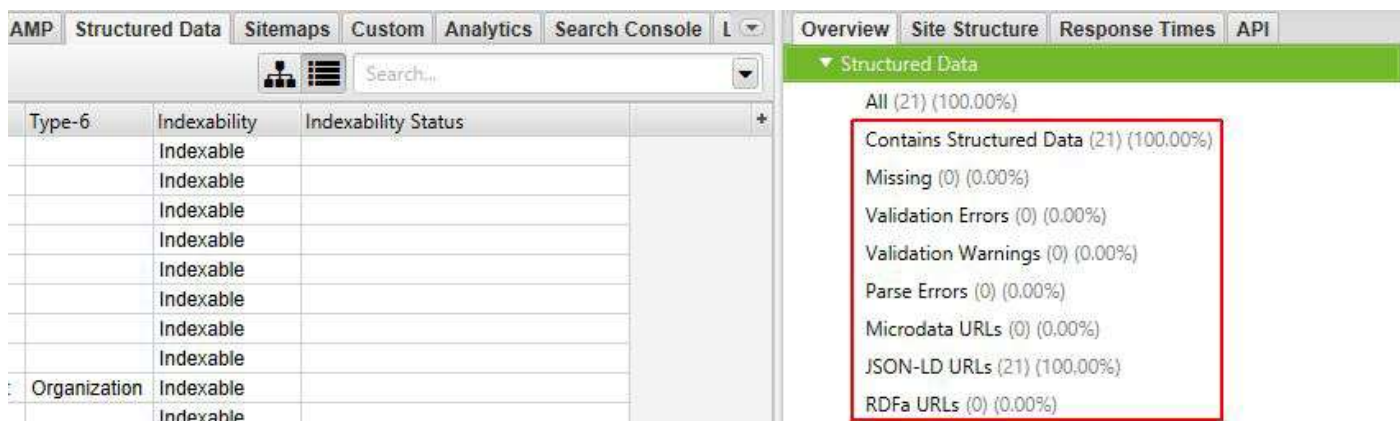
3. Analizza i dati nella scheda "Structured data Tab"

La scheda "Structured data" presenta 8 filtri che ti aiutano a capire se l'implementazione dei dati strutturati presenta problemi di convalida. Le colonne "Total Types" e "Unique Types" contano invece il numero di "itemtypes" che sono stati scoperti per ogni URL.



Filter: All	Errors	Warnings	Total Types	Unique Types	Type-1	Type-2	Type-3	Type-4	Type-5	Type-6	Indexability
1 A ✓ All	0	0	3	3	Organization	SearchAction	WebSite				Indexable
2 h Contains Structured Data	0	0	1	1	Organization						Indexable
3 h Missing	0	0	1	1	Organization						Indexable
4 h Validation Errors	0	0	1	1	Organization						Indexable
5 h Validation Warnings	0	0	1	1	Organization						Indexable
6 h Parse Errors	0	0	1	1	Organization						Indexable
7 h Microdata URLs	0	0	1	1	Organization						Indexable
8 h JSON-LD URLs	0	0	19	6	HowTo	HowToDirection	HowToStep	HowToTool	ImageObject	Organization	Indexable
9 h RDFa URLs	0	0	1	1	Organization						Indexable
10 h	0	0	1	1	Organization						Indexable
11 h	0	0	1	1	Organization						Indexable
12 h	0	0	1	1	Organization						Indexable
13 https://www.screamingfrog.co.uk/about/	0	0	1	1	Organization						Indexable
14 https://www.screamingfrog.co.uk/adwords-history/	0	0	1	1	Organization						Indexable
15 https://www.screamingfrog.co.uk/log-file-analyser/	0	0	1	1	Organization						Indexable
16 https://www.screamingfrog.co.uk/our-work/	0	0	1	1	Organization						Indexable
17 https://www.screamingfrog.co.uk/login/	0	0	1	1	Organization						Non-Indexable
18 https://www.screamingfrog.co.uk/clients/	0	0	1	1	Organization						Indexable
19 https://www.screamingfrog.co.uk/google-grants-account-sus...	0	0	1	1	Organization						Indexable
20 https://www.screamingfrog.co.uk/meet-the-team/	0	0	1	1	Organization						Indexable

Oltre alla finestra superiore puoi anche consultare il riquadro della sidebar che ti fornisce un riassunto dei dati contenuti in ogni scheda e filtro, in modo da sapere dove cliccare, senza dover controllare ogni filtro per vedere se ci sono dati.



Type-6	Indexability	Indexability Status
	Indexable	
	Indexable	
	Indexable	
	Indexable	
	Indexable	
	Indexable	
	Indexable	
	Indexable	
Organization	Indexable	
	Indexable	

Structured Data
All (21) (100.00%)
Contains Structured Data (21) (100.00%)
Missing (0) (0.00%)
Validation Errors (0) (0.00%)
Validation Warnings (0) (0.00%)
Parse Errors (0) (0.00%)
Microdata URLs (0) (0.00%)
JSON-LD URLs (21) (100.00%)
RDFa URLs (0) (0.00%)

Di seguito ti riporto i filtri e le loro funzionalità:

- **Contains Structured Data:** elenca tutti gli URL che contengono dati strutturati.
- **Missing Structured Data:** mostra gli URL che non contengono dati strutturati.
- **Validation Errors:** vengono visualizzati tutti gli URL che presentano degli errori di convalida in base al vocabolario di Schema.org, di Google rich result o di entrambi, a seconda della tua configurazione. Il Seo Spider classifica i problemi di Schema.org sempre come errori, piuttosto che come avvisi. La convalida di Google rich result feature mostrerà gli errori per proprietà richieste, mancanti o problemi con l'implementazione delle proprietà richieste.
- **Validation Warnings:** mostra tutti gli URL che presentano degli avvisi ("warnings") di convalida rispetto ai "Google Rich Results". Il Seo Spider popola questo filtro considerando solamente i problemi di convalida dei "dati strutturati facoltativi".
- **Parse Errors:** il filtro presenta tutti gli URL che contengono dati strutturati ma non è stato possibile analizzarli correttamente. Questo è spesso dovuto a un mark-up errato. Se stai usando il formato preferito di Google, JSON-LD, allora il JSON-LD Playground è uno strumento eccellente per aiutare il debug degli errori di analisi.

- **Microdata URLs:** questi sono URL che contengono dati strutturati in formato microdati.
- **Microdata URLs | JSON-LD URLs | RDFa URLs:** vengono mostrati rispettivamente gli URLs che contengono dati strutturati in formato JSON-LD, RDFa o microdata.

4. Visualizza la finestra inferiore - scheda "Dettagli dati strutturati"

Una volta raccolti i dati potrai visualizzare tutti i dettagli nella finestra inferiore del Seo Spider; nel lato sinistro trovi i valori delle proprietà accompagnate da icone che segnalano errori o avvisi mentre nella schermata di destra i dettagli sui specifici problemi di convalida, la gravità della criticità e i link di Schema.org per trovare le specifiche e fixare il problema.

Property	Value	Property: Microdata	Validation Type	Issue Severity	Issue
JSON-LD	FAQPage				
Microdata	Organization				
Microdata	Product				
aggregateRating	AggregateRating	Google Product	Error	❗	'http://schema.org/image' property is required for 'Product'
ratingValue	4.5/5	Google Product	Warning	⚠	'http://schema.org/brand' property is recommended for 'Product'
reviewCount	35541	Google Product	Warning	⚠	'http://schema.org/description' property is recommended for 'Product'
type	http://schema.org/AggregateRating	Google Product	Warning	⚠	'http://schema.org/offers' property is recommended for 'Product'
name	motor	Google Product	Warning	⚠	'http://schema.org/gtin8' or 'http://schema.org/gtin13' or 'http://schema.org/gtin14' or 'http://schema.org/gtin16' property is recommended for 'Product'
type	http://schema.org/Product	Google Product	Warning	⚠	'http://schema.org/review' property is recommended for 'Product'
Microdata	Review	Google Product	Warning	⚠	'http://schema.org/sku' property is recommended for 'Product'
author	Person	Google Product	Warning	⚠	'http://schema.org/brand' property is recommended for 'Product'
contentLocation	Place	Google Product	Warning	⚠	'http://schema.org/gtin8' or 'http://schema.org/gtin13' or 'http://schema.org/gtin14' or 'http://schema.org/gtin16' property is recommended for 'Product'
datePublished	25 July 2019	Google Product	Warning	⚠	'http://schema.org/review' property is recommended for 'Product'
itemReviewed	Thing	Google Product	Warning	⚠	'http://schema.org/sku' property is recommended for 'Product'
reviewBody	Easy to use.	Google Product	Warning	⚠	'http://schema.org/brand' property is recommended for 'Product'
	Feedback		Info	ℹ	Child Property has an issue

ESEMPIO 1

ESEMPIO DI ERRORE DI CONVALIDA DI GOOGLE PRODUCT

Nell'esempio qui sotto, puoi notare che lv.com presenta errori e avvisi di convalida nei requisiti del dato strutturato "Google Product". A destra vedi come il sito presenti 1 errore tra gli elementi obbligatori e più avvisi tra i dati raccomandati (con un avvertimento).

Property	Value	Property: Microdata	Validation Type	Issue Severity	Issue
Microdata	Product				
aggregateRating	AggregateRating	Google Product	Error	❗	'http://schema.org/image' property is required for 'Product'
bestRating	10	Google Product	Warning	⚠	'http://schema.org/description' property is recommended for 'Product'
ratingValue	9.1	Google Product	Warning	⚠	'http://schema.org/offers' property is recommended for 'Product'
reviewCount	38122	Google Product	Warning	⚠	'http://schema.org/gtin8' or 'http://schema.org/gtin13' or 'http://schema.org/gtin14' or 'http://schema.org/gtin16' property is recommended for 'Product'
type	http://schema.org/AggregateRating	Google Product	Warning	⚠	'http://schema.org/review' property is recommended for 'Product'
bestRating	10	Google Product	Warning	⚠	'http://schema.org/sku' property is recommended for 'Product'
name	Car Insurance	Google Product	Warning	⚠	'http://schema.org/brand' property is recommended for 'Product'
ratingValue	9.1	Google Product	Warning	⚠	'http://schema.org/gtin8' or 'http://schema.org/gtin13' or 'http://schema.org/gtin14' or 'http://schema.org/gtin16' property is recommended for 'Product'
reviewCount	38122	Google Product	Warning	⚠	'http://schema.org/review' property is recommended for 'Product'
type	http://schema.org/Product	Google Product	Warning	⚠	'http://schema.org/sku' property is recommended for 'Product'
RDFa	website				
locale	en-GB				
type	website				
url	https://www.lv.com/car-insurance				

Le proprietà raccomandate, evidenziate come avvertimenti, possono essere implementate per aggiungere più informazioni sul contenuto (che può fornire una migliore esperienza utente) o semplicemente ignorate.

ESEMPIO 2

Nell'esempio di seguito, HSBC ha un errore sull'implementazione dei dati strutturati della breadcrumb.

Il problema afferma che la proprietà “http://schema.org/item” è richiesta per “ListItem”.

Property

Value

Microdata

itemListElement

item

position

type

itemListElement

name

position

type

WebPage

RdFa

BreadcrumbList

ListItem

Thing

1

http://schema.org/ListItem

Credit cards

2

http://schema.org/ListItem

http://schema.org/BreadcrumbList

WebPage

website

Property: http://schema.org/itemListElement

Validation Type

Issue Severity

Issue

Google Breadcrumb

Error

http://schema.org/item property is required for 'ListItem'

Le linee guida di Google sulla funzione breadcrumb dichiarano che le proprietà raccomandate devono includere item (l'URL della pagina web), nome (il titolo del breadcrumb) e la posizione (del breadcrumb nel percorso). HSBC manca semplicemente la proprietà item e l'URL associato.

Questi sono solamente alcuni esempi di come interpretare i dati popolati dal Seo Spider. Ti consiglio di rivedere con periodicità la documentazione ufficiale che è in continuo aggiornamento.

5. Esportazione degli errori di convalida e avvisi

Screaming Frog prevede due report dedicati ai dati strutturati.

File Configuration Mode Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help

Screamingfrog

https://

Internal External Protocol Response Co

Filter: All

1 https://www.screamingfrog.co.uk/

2 https://www.screamingfrog.co.uk/pay-per-c

3 https://www.screamingfrog.co.uk/services/

4 https://www.screamingfrog.co.uk/log-file-an

5 https://www.screamingfrog.co.uk/seo-spide

6 https://www.screamingfrog.co.uk/broken-lin

7 https://www.screamingfrog.co.uk/contact-us/

8 https://www.screamingfrog.co.uk/seo-spider/pricing/

9 https://www.screamingfrog.co.uk/analytics-consultancy/

10 https://www.screamingfrog.co.uk/seo-spider/user-guide/

Crawl Overview

Redirect & Canonical Chains

Non-Indexable Canonicals

Pagination

Hreflang

Insecure Content

SERP Summary

Orphan Pages

Structured Data

Description

Meta Keywords

H1

H2

Images

Canonicals

Paginal

Errors

Warnings

Total Types

Unique Ty

0

0

3

3

0

0

1

1

0

0

1

1

0

0

1

1

0

0

1

1

0

0

1

1

0

0

1

1

0

0

1

1

Il rapporto ‘Validation Errors & Warnings Summary’ è particolarmente utile perché crea dei modelli di aggregazione delle criticità (piuttosto che riportare ogni istanza), mostra il numero di URL interessati da ogni modello con rispettivo URL di esempio (simile al comportamento di Search Console nella sezione Web Vitals). Un esempio di rapporto può essere visto qui sotto.

Issue	Schema	Issue Severity	Number of URLs affected	Issue Message	Example URL Affected
ContactPoint	JSON-LD	Google Corporate Contact	Warning	16180 'http://schema.org/contactOption' property is recommended for 'ContactPoint'	https://www.example.com/
BreadcrumbList	RdFa	Google Breadcrumb	Error	16089 One of name or item name must be provided in 'BreadcrumbList'	https://www.example.com/
ListItem	RdFa	Google Breadcrumb	Error	16075 'http://schema.org/position' property is required for 'ListItem'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288 'http://schema.org/review' property is recommended for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288 'http://schema.org/offers' property is recommended for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288 'http://schema.org/aggregateRating' property is recommended for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288 'http://schema.org/sku' property is recommended for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Error	10288 'http://schema.org/image' property is required for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288 'http://schema.org/description' property is recommended for 'Product'	https://www.example.com/
Product	JSON-LD	Google Product	Warning	10288 'http://schema.org/gtin13' or 'http://schema.org/gtin14' or 'http://schema.org/mpn' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945 'http://schema.org/review' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945 'http://schema.org/offers' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945 'http://schema.org/brand' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Carousel	Error	3945 'http://schema.org/url' or 'http://schema.org/item' property is required for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945 'http://schema.org/aggregateRating' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Error	3945 'http://schema.org/name' property is required for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945 'http://schema.org/sku' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Error	3945 'http://schema.org/image' property is required for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945 'http://schema.org/description' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Carousel	Error	3945 'http://schema.org/position' property is required for 'Product'	https://www.example.com/
Product	Microdata	Google Product	Warning	3945 'http://schema.org/gtin8' or 'http://schema.org/gtin13' or 'http://schema.org/gtin14' or 'http://schema.org/mpn' property is recommended for 'Product'	https://www.example.com/
Product	Microdata	Google Carousel	Error	3945 'http://schema.org/itemListElement' is required to be of type 'http://schema.org/ListItem' in 'ItemList' [currently 'Product']	https://www.example.com/
Article	JSON-LD	Google Article	Warning	1306 'http://schema.org/dateModified' property is recommended for 'Article'	https://www.example.com/
Article	JSON-LD	Google Article	Warning	1306 'http://schema.org/headline' property is recommended for 'Article'	https://www.example.com/
Article	JSON-LD	Google Article	Error	1305 'http://schema.org/image' is required to be 'http://schema.org/ImageObject' or 'http://schema.org/URL' in 'Article'	https://www.example.com/
Article	Microdata	Google Article	Warning	118 'http://schema.org/dateModified' property is recommended for 'Article'	https://www.example.com/
Article	Microdata	Google Article	Warning	118 'http://schema.org/headline' property is recommended for 'Article'	https://www.example.com/
Article	Microdata	Google Article	Warning	118 'http://schema.org/datePublished' property is recommended for 'Article'	https://www.example.com/
Article	Microdata	Google Article	Warning	118 'http://schema.org/image' property is recommended for 'Article'	https://www.example.com/
ItemList	Microdata	Google Carousel	Error	16 'http://schema.org/itemListElement' property is required for 'ItemList'	https://www.example.com/

KILL SEO VOLUME 1

SCREAMING FROG.CLUB

204

Il report “Validation Errors & Warnings” è una vera esportazione di massa di ogni errore e avviso scoperto insieme all’URL su cui si trova. Molto completo ma con siti di grosse dimensioni complicato nella lettura.

HOW TO:

FARE UN AUDIT DI BOOKMARK ROTTI

Attraverso questo articolo andremo a scoprire come eseguire un Audit con Screaming Frog che rilevi tutti i bookmarks rotti di un sito internet, analizzare le pagine di origine ed esportare in massa i risultati ottenuti.

I bookmark sono un modo utile per collegare gli utenti ad una sezione specifica di una pagina dopo un click.

L’impostazione è semplice, nell’HTML puoi usare l’attributo ID per creare il segnalibro.

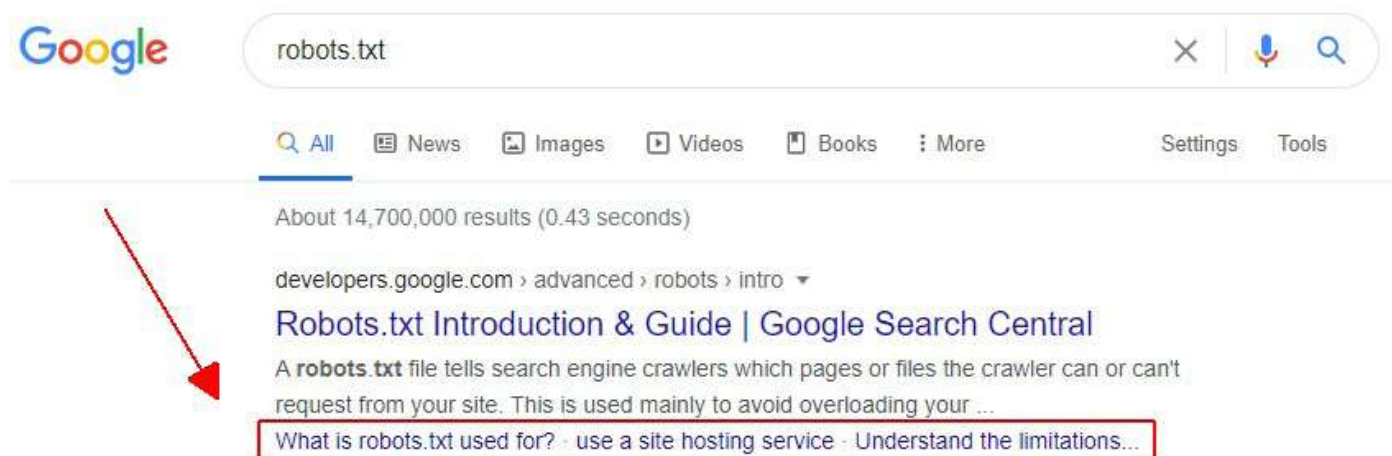
```
<h2 id="esempio">Esempio di segnalibro</h2>
```

Successivamente puoi collegarti direttamente al segnalibro aggiungendo un frammento (#) e il nome dell’ID (“esempio”) all’URL della pagina.

```
<a href="https://www.screamingfrog.co.uk/how-to-find-broken-bookmarks/#example">Vedi il nostro esempio di segnalibro</a>
```

Quando il link viene cliccato, la pagina scorrerà fino alla posizione con il segnalibro: in questo caso la sezione “Esempio di segnalibro”.

Una pagina con diversi Bookmarks viene sempre considerata unica da Google (dato che ignora qualsiasi cosa dopo il simbolo “#”), ma gli “Anchor Tag” usati per il “jump to” potrebbero essere serviti nei risultati di ricerca.

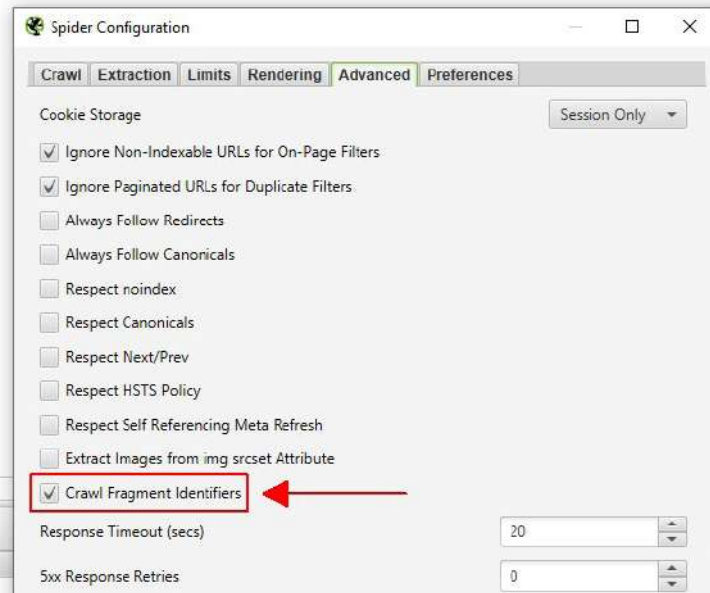


I bookmark link sono molto apprezzati nell’esperienza di navigazione perché riescono a mandare l’utente verso un preciso focus ma sono altresì fonte di numerosi errori, e spesso diventano “rotti” nel tempo quando le pagine vengono aggiornate e gli ID vengono cambiati o rimossi. Un segnalibro rotto nonostante porti alla pagina corretta (non genera 404) perde il suo significato intrinseco.

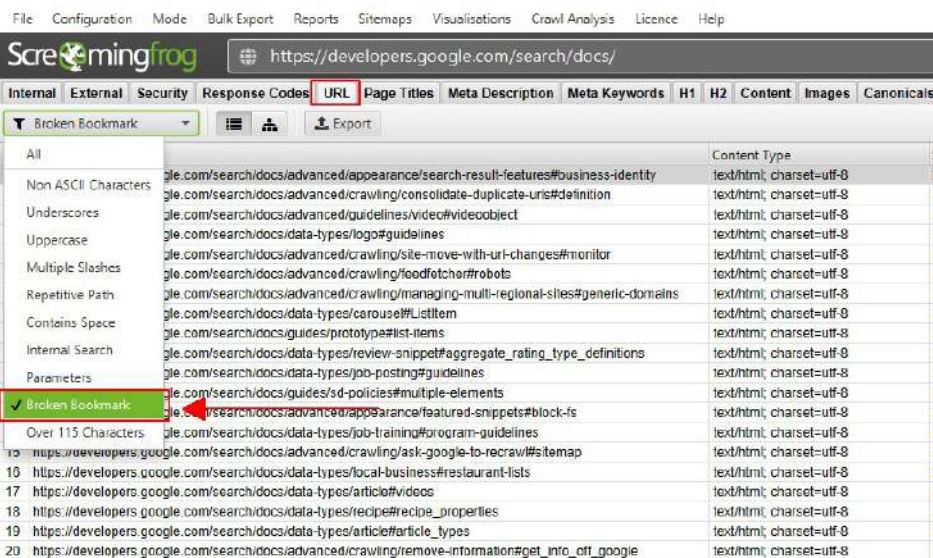
Per risolvere questo problema, il Seo Spider scansiona gli URL con “#” e verifica che esista un ID all’interno dell’HTML della pagina web affinché il segnalibro funzioni correttamente. Questa funzionalità ti permette di trovare i bookmark link non funzionanti sull’intero sito in modo semplice e automatico.

1. Abilita il “Fragment identifiers”

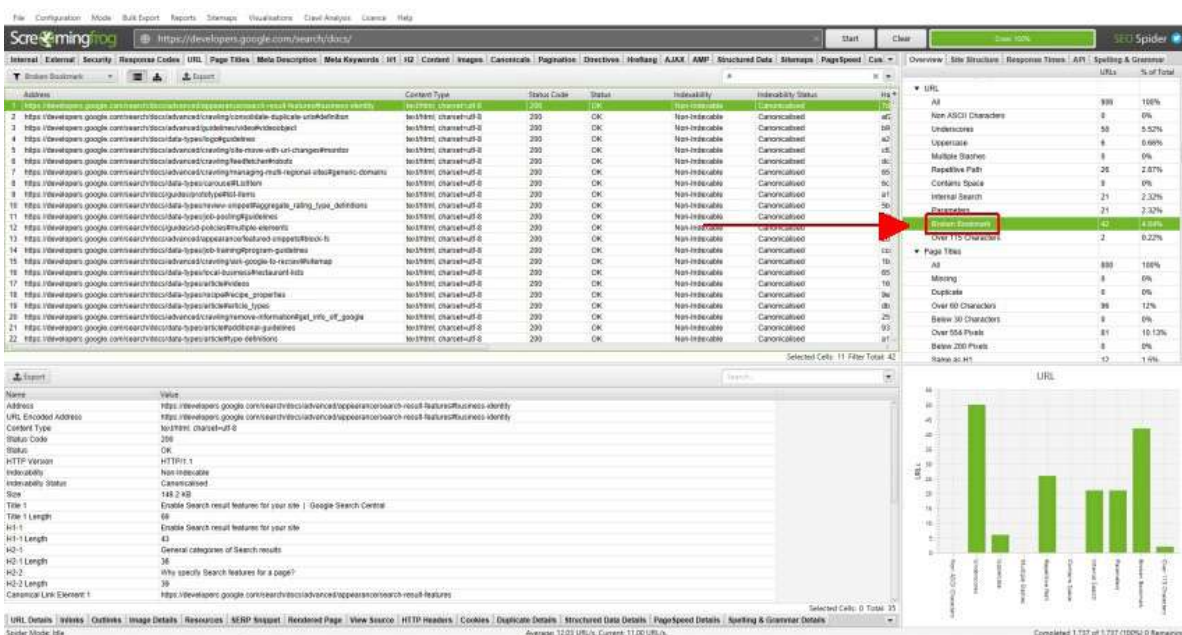
Config > Spider > Advanced



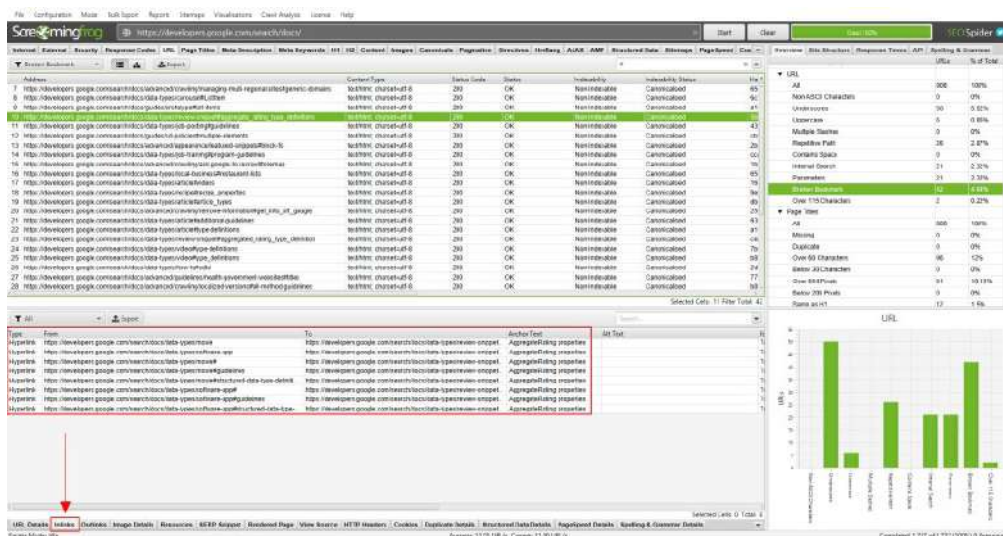
2. Esegui la scansione del sito web
3. Clicca sulla scheda "URL" e utilizza il filtro "Broken Bookmark"



Alternativamente puoi consultare la metrica sulla sidebar del Seo Spider.



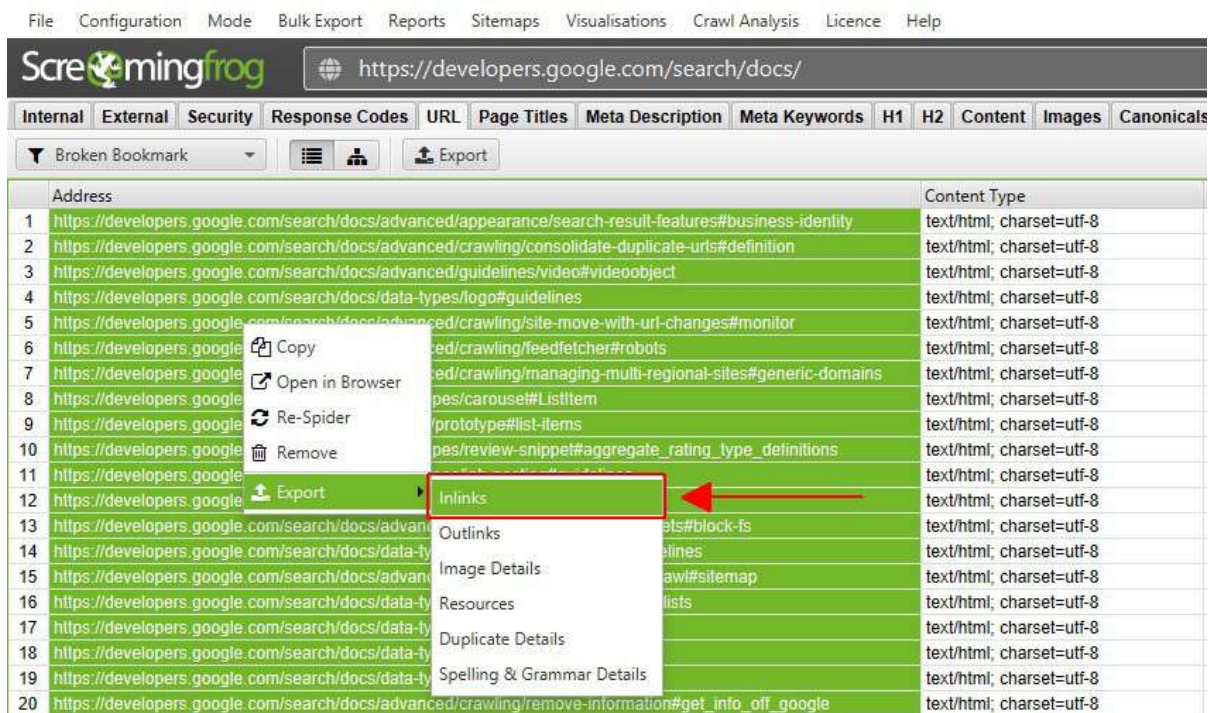
4. Raccolti i dati puoi visualizzare la fonte dei segnalibri rotti esaminando la scheda Inlinks. Per vedere quali URL sul sito web collegano a questi segnalibri rotti puoi selezionare un URL nella finestra superiore e consultare la sezione “Inlinks”. In questo modo sarai in grado di poterli correggere o modificare l’ID dell’attributo sulla pagina con il segnalibro rotto.



La colonna “From” identifica la fonte dove è stato inserito il collegamento rotto, mentre “To” la pagina con il segnalibro che non funziona.

5. Esportazione dei dati

Ora non ti resta che scaricare in blocco i risultati includendo sia gli URL “sorgente” che le pagine dei segnalibri interrotti evidenziando tutti gli URL nella finestra superiore e cliccando con il tasto destro per selezionare il report “Inlinks”.

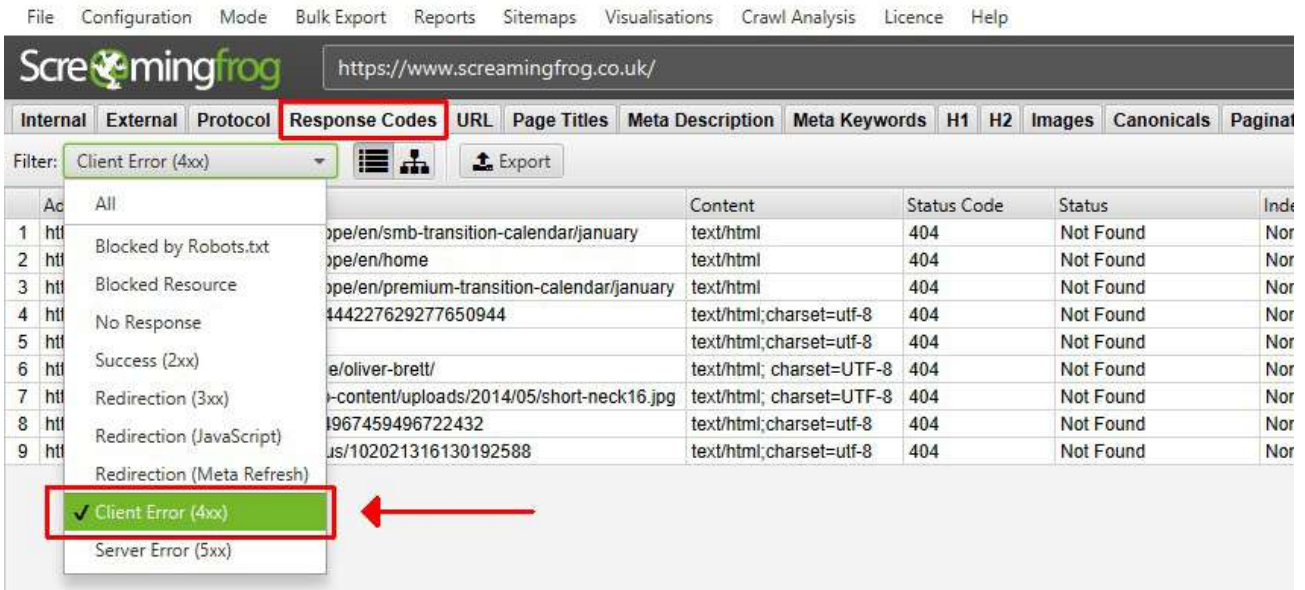


Alternativamente puoi cliccare sul pulsante “Export” direttamente nella tab “Inlinks” della finestra.

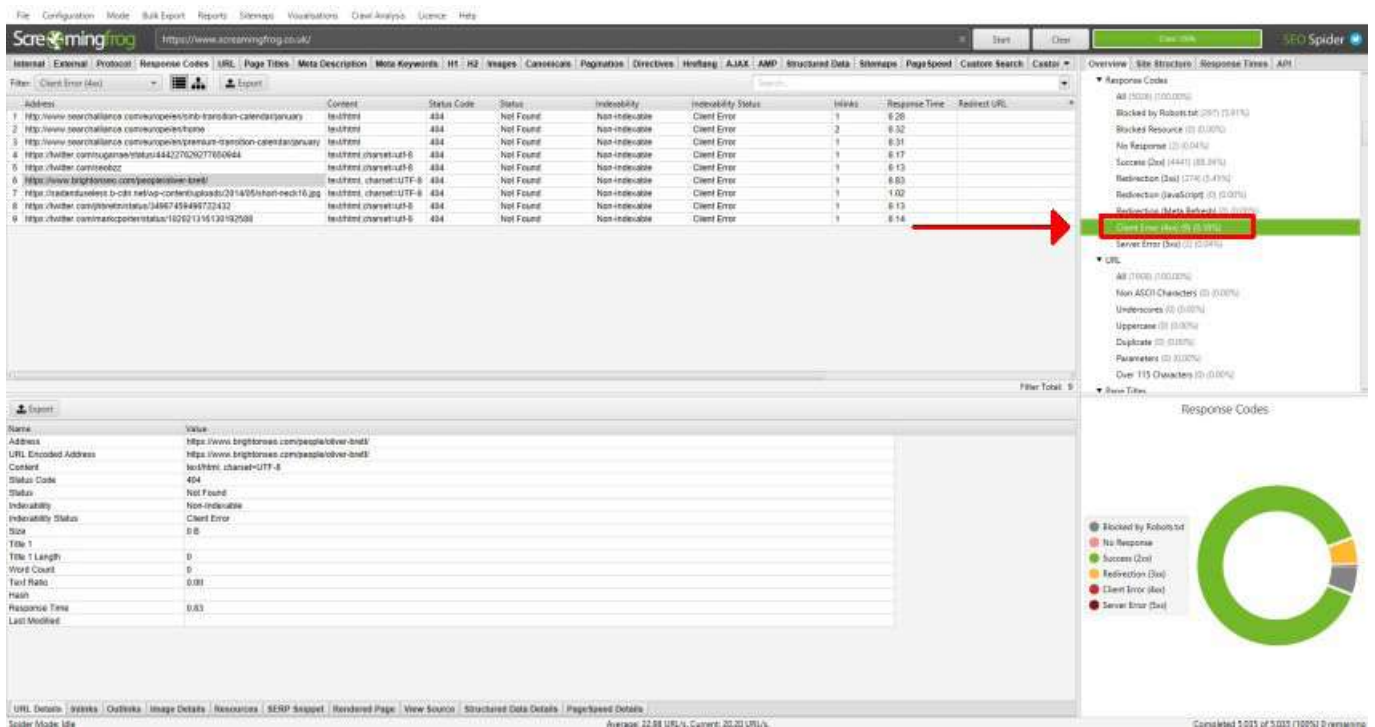
COME FARE UN AUDIT DEI “BROKEN LINK - LINK ROTTI”

Una delle analisi che non mancano mai nei miei Seo Audit è quella riguardante i broken links che generano errori 404 e sprecano inesorabilmente il crawling budget. La ricerca di questa criticità con Screaming Frog è molto semplice e richiede pochi passaggi:

1. avvia la scansione del tuo sito web.
2. Seleziona la tab “Response Code - Codici di risposta”.
3. Imposta il filtro “Client Errors (4XX)” per visualizzare i collegamenti interrotti.



In modo alternativo puoi selezionare la voce “Clients Errors” dalla sidebar laterale.



Nell'esempio sono stati trovati 9 errori in status code 404, che corrispondono allo 0,18% dei link scoperti nel crawl.

4. Visualizza la fonte dei link rotti (Broken Links) cliccando sulla scheda 'Inlinks'; selezionando un URL nella finestra superiore puoi popolare la tab.

Come puoi vedere nel prossimo esempio, c'è un link rotto verso il sito BrightonSEO (<https://www.brightonseo.com/people/oliver-brett/>), che ha come sorgente: <https://www.screamingfrog.co.uk/2018-a-year-in-review/>.

Ecco una vista ravvicinata del riquadro inferiore della finestra che dettaglia i dati "inlinks"

Type	From	To	Anchor Text	Follow	Link Attributes
AHREF	https://www.screamingfrog.co.uk/2018-a-year-in-review/	https://www.brightonseo.com/people/oliver-brett/	Oliver Brett's	true	

URL Details	Inlinks	Outlinks	Image Details	Resources	SERP Snippet	Rendered Page	View Source	Structured Data Details	PageSpeed
-------------	---------	----------	---------------	-----------	--------------	---------------	-------------	-------------------------	-----------

La colonna "From" identifica la sorgente dove è presente il collegamento che genera 404, mentre nella colonna "To" è visualizzato il link rotto (destinazione). Nelle colonne che seguono puoi anche vedere l'anchor text, l'alt text (se si tratta di un'immagine ipercollegata) e se il link ha l'attributo "follow" (vero) o "nofollow" (falso).

5. Raccolti i dati puoi esportare i risultati

Bulk Export > Response Code > Client Errors (4XX)

The screenshot shows the Screaming Frog SEO Spider interface. The 'Bulk Export' menu is open, and 'Client Error (4xx) Inlinks' is highlighted with a red box and a red arrow. The background shows a list of URLs with a filter set to 'Client Error (4xx)'.

Con questo processo sei in grado di coprire la maggior parte dei casi in cui ci siano dei link interrotti ma potrebbe capitare che l'URL "sorgente" che è l'ultima pagina a collegarsi alla pagina di errore 404 sia un reindirizzamento (o un nodo di una catena di redirect) e va gestita come tale. Per scoprire se la sorgente è un reindirizzamento puoi consultare la colonna 'Type-Tipo' e controllare se identificata come "HTTP Redirect".

Per trovare rapidamente la pagina di origine di questi errori, ti consiglio di utilizzare il rapporto "All Redirects".

Reports > Redirects > All Redirects

In questo modo puoi isolare la pagina di origine, i reindirizzamenti e l'ultima fonte che genera errori 4xx per una corretta ottimizzazione dei links.

Infine, se hai una lista di URL sulla quale vuoi controllare eventuali link rotti invece di scansionare un sito web, puoi semplicemente caricarli in modalità elenco.

COME GESTIRE PROBLEMI DI SCANSIONE DEL SEO SPIDER

A volte può capitare che Screaming Frog esegua un crawl di una sola pagina o non scansioni come previsto.

In quel caso le prime cose da controllare sono lo “Status” e lo “Status Code” delle risorse restituite per identificare il problema e risolverlo.

Di seguito puoi esaminare i casi più comuni di criticità imputabili agli stati di risposta:

Blocked by Robots.txt: il filtro definisce tutti gli URL bloccati dal robots.txt del sito che non permette allo Spider di scansionare quelle risorse e di conseguenza di non poter indicizzare sui Motori di Ricerca.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/frogs/legs/		0	Blocked by Robots.txt

Ogni risposta “0” nella colonna Status Code identifica la mancanza di una risposta HTTP del server, la colonna Status ne identifica il motivo. Nell'esempio sopra lo user-agent dello Spider è stato bloccato dal robots.txt attraverso la direttiva “disallow” ed essendo bloccato non può vedere la risposta HTTP.

What to do:

- Nel caso di risorse bloccate da robots è possibile impostare il Seo Spider per ignorarlo tramite Configurazione > Robots.txt > Impostazioni > Ignora Robots.txt o mediante la personalizzazione dello stesso. Di default Screaming Frog segue le indicazioni “disallow” del file robots.txt.

DNS lookup failed: identifica che il sito non è stato trovato per nulla per un errore di immissione del dominio o per una mancanza di connessione alla rete.

What to do:

- controlla che il sito sia stato scritto correttamente e che sia visibile nel tuo browser. Nel caso non sia possibile visualizzarlo mediante browser potrebbero esserci problemi con la connettività; nel caso invece il sito sia correttamente renderizzato sul browser dovresti controllare se ci sono eventuali Antivirus o Firewall che bloccano il Seo Spider.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.zone/		0	DNS lookup failed

Connection Timeout: identifica che Screaming Frog ha richiesto l' HTTP response dal server ma non ha ricevuto risposta entro 20 secondi. Per ovviare al problema è possibile aumentare il “Response Timeout” dalle configurazioni o diminuire la velocità dello spider (lowering Speed) per non caricare il Server e dargli più tempo per la risposta.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/timeout/		0	Connection Timeout

What to Do:

- Controlla i tempi di caricamento del sito sul tuo browser; se sono molto lunghi puoi cambiare il tempo di “Timeout” sul Seo Spider e diminuire la velocità del crawler per alleggerire il server.
- Controlla se riesci a scansionare altri siti con Screaming Frog. Se il problema è diffuso potrebbe esserci un problema con Antivirus e Firewall. Prova in questo caso a creare un'eccezione alla scansione di Screaming Frog. Se nemmeno questa soluzione risolve il problema probabilmente è presente un problema di rete o locale. Controlla ad esempio se il proxy è abilitato (Configurazione > Sistema > Proxy e nel caso lo fosse prova a disabilitarlo e riavvia il crawling).

Connection Refused: viene restituita quando il tentativo di connessione del Seo Spider è stato rifiutato in qualche punto tra la macchina locale e il sito web.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/refused/		0	Connection Refused

What to do:

- Riesci a scansionare altri siti web? Se la risposta è affermativa controlla l'Antivirus e Firewall e crea un'eccezione per lo Spider, se il problema è comune ad altri siti controlla che non si sia verificato un problema di rete o locale al tuo pc.
- Puoi visualizzare la pagina nel browser o restituisce un errore simile? Se la pagina può essere vista cambia User-agent e utilizza "Chrome" (Configurazione > User-Agent).

No Response: il Seo Spider ha problemi a creare connessioni o a ricevere risposte. Un problema comune potrebbe essere il Proxy.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/no-response/		0	No Response

What to do:

- Il primo controllo riguarda il proxy (Configurazione > Sistema > Proxy) in cui puoi provare a disabilitarlo. Se non è impostato correttamente allora questo potrebbe significare che il Seo Spider non sta inviando/ricevendo richieste correttamente.

Success (2XX): l'URL richiesto è stato ricevuto, compreso, accettato ed elaborato con successo. Il controllo che ti suggerisco di fare riguarda la presenza dell'attributo "noFollow"

What to do:

- La pagina richiesta ha la direttiva meta robots "nofollow" sulla pagina/nell'intestazione HTTP o tutti i link sulla pagina hanno attributi rel="nofollow"? In questo caso è sufficiente impostare il Seo Spider a seguire gli Internal/External Nofollow (Configurazione > Spider). Questa criticità è dovuta al fatto che Screaming Frog segue le indicazioni "nofollow" di default.
- Il sito web presenta link JavaScript? Prova a disabilitare Javascript e visualizzare le pagine nel browser. Se ti trovi in questo caso ti basta abilitare il rendering JavaScript (Configurazione > Spider > Rendering > JavaScript). Per impostazione predefinita il Seo Spider scansiona solo i link , e <link rel="canonical"> nel codice sorgente HTML e non legge il DOM.
- Controlla la scheda "Limiti" di "Configurazione > Spider" in particolare le opzioni "Limit Search Depth" e "Limit Search Total". Nel caso fossero impostati rispettivamente a 0 o 1, allora il Seo Spider viene istruito a scansionare solo un singolo URL.
- Il sito richiede i cookie? Prova a visualizzare la tua pagina con il browser e i cookie disabilitati. Questa condizione si verifica perché al Seo Spider viene servita una pagina differente senza collegamenti ipertestuali nel caso di cookie disabilitati. Per risolvere questa criticità di scansione vai su Configurazione > Spider > Advanced Tab > Allow Cookies
- Cosa è specificato nella colonna "Contenuto"? Se non c'è alcun contenuto abilita il rendering JavaScript (Configurazione > Spider > Rendering > JavaScript) e prova ad avviare nuovamente il crawl. Questa criticità si verifica se nessun tipo di contenuto è specificato nell'intestazione HTTP e il Seo Spider non sa se l'URL è un'immagine, PDF, pagine HTML ecc. quindi non riesce a scansionare per determinare se ci sono altri link. Questo può essere aggirato con la modalità di rendering in quanto il tool controlla se una <meta http-equiv> è specificata nella <head> del documento quando è abilitata.
- Esistono delle restrizioni dovute all'età? In questo caso prova a cambiare lo user-agent a Googlebot (Configurazione > User-Agent). Il sito/server potrebbe essere impostato per servire l'HTML ai bot di ricerca senza richiedere l'inserimento dell'età.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/	text/html; charset=UTF-8	200	OK

Redirection (3XX): è stato riscontrato un reindirizzamento come 301 o 302. Idealmente ogni link interno dovrebbe collegarsi ad una risposta con status code 200 ed evitare collegamenti che facciano fare dei passi in più allo Spider in un'ottica di salvaguardare il Crawling Budget.

	Address	Content	Status Code	Status
1	http://screamingfrog.co.uk/	text/html; charset=iso-8859-1	301	Moved Permanently

What to Do:

- Controlla quale sia la destinazione del reindirizzamento (Controlla gli outlink dell'URL restituito). Se trovi una condizione di loop il Seo Spider non riesce a fare una scansione completa.
- Scheda External: il seo spider tratta diversi sottodomini come esterni e non li scansiona per default. Per risolvere questa condizione abilita l'opzione scansiona tutti i sottodomini. Configurazione > Spider > Scansiona tutti i sottodomini.
- Il sito richiede i cookie? Prova a disabilitare i cookie dal tuo browser, se riscontri il problema abilita la funzione "Allow Cookies" nel Seo Spider. Configurazione > Spider > Scheda avanzata > Consenti cookie. Questa condizione si verifica perché il SEO Spider viene reindirizzato a un URL dove viene lasciato un cookie, che non viene accettato.

Bad request (400) - 403 Forbidden - 406 Status Not Acceptable: il server non riesce o non vuole elaborare la richiesta e sta negando la richiesta del SEO Spider di visualizzare l'URL richiesto.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/forbidden/	text/html; charset=UTF-8	403	Forbidden

What to do:

- Se la pagina può essere vista attraverso il tuo browser, prova ad impostare Chrome come user agent (Configurazione > User-Agent). Probabilmente il sito nega la pagina al nostro Seo Spider per questioni di sicurezza.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/cool-guys/	text/html; charset=UTF-8	404	Not Found

- Se la pagina risponde con errore 404 il server sta indicando che non esiste più. Se anche cambiando lo user-agent in Chrome ti ritrovi nella stessa condizione presumibilmente è un problema del sito web, altrimenti se con Chrome la pagina viene restituita (o è possibile visualizzarla con il browser) e puoi iniziare il crawling probabilmente il server stava bloccando il nostro user-agent per una questione di sicurezza.
- Se la pagina presenta l'errore 429 significa che sono state fatte troppe richieste al server in un determinato periodo di tempo.

What to do:

- Puoi visualizzare il tuo sito nel browser o questo mostra un messaggio di errore simile? In questo caso puoi abbassare la velocità di crawl e/o testare un user agent differente come Chrome o Googlebot.

Errore del server (5XX): il server non è riuscito a soddisfare una richiesta apparentemente valida. Questo può includere risposte comuni come 500 Internal Server Errors, 502 e 503 Server Unavailable.

	Address	Content	Status Code	Status
1	https://www.screamingfrog.co.uk/server-error/	text/html; charset=UTF-8	500	Internal Server Error

What to do:

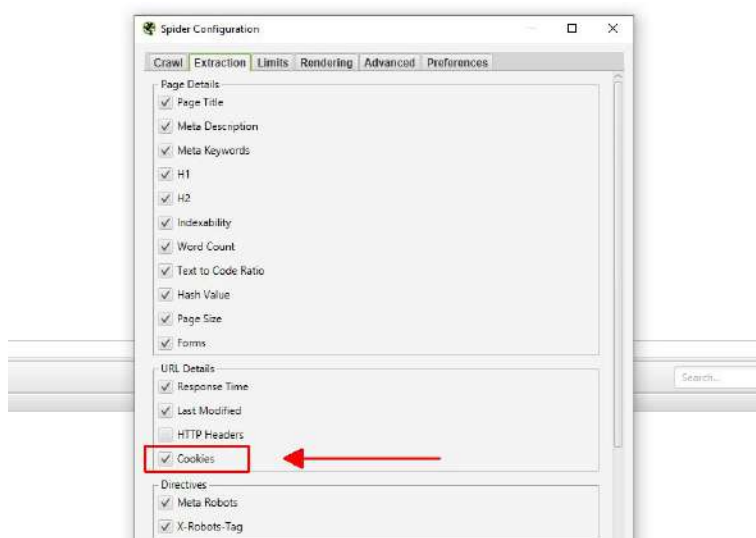
- Puoi vedere il tuo sito nel browser o è down? Se la pagina può essere vista, cambia user-agent e utilizza Chrome come user agent (Configurazione > User-Agent).

COME FARE UN AUDIT SEO DEI COOKIES

Vista l'importanza ed evoluzione delle leggi sulla privacy è diventato sempre più importante capire la presenza e lo scopo dei cookies sul nostro sito web in modo da essere conformi ed non incorrere a sanzioni. Grazie a Screaming Frog hai la possibilità di identificare i cookie che vengono impostati automaticamente senza il consenso esplicito dell'utente sull'intero tuo progetto.

1. Abilita l'opzione "Cookies" per l'estrazione con Screaming Frog

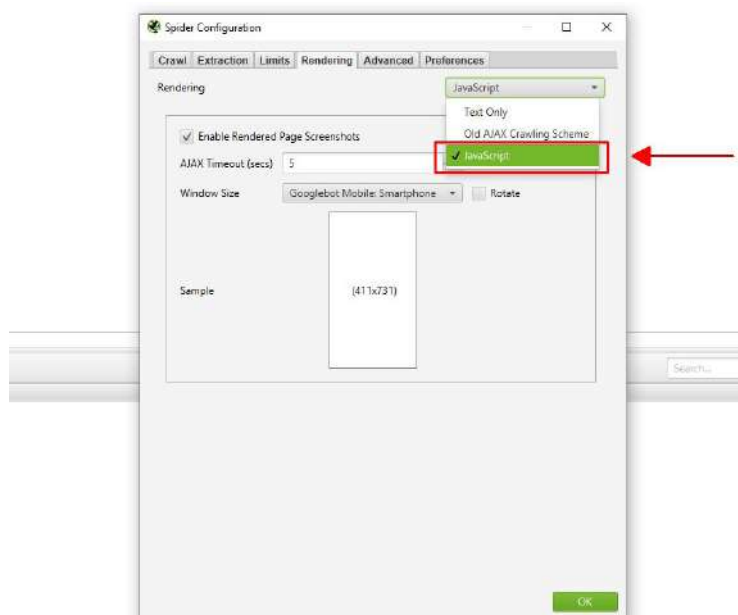
La prima azione è quella di abilitare l'estrazione dei cookie sul tuo sito web attraverso l'opzione "Config > Spider > Extraction" e selezionare "Cookies" in "URL Details". In questo modo il Seo Spider salva tutti i cookie scoperti durante la scansione.



Quando scegli di memorizzare i cookie, l'esclusione automatica eseguita dal SEO Spider per i tag di monitoraggio di Google Analytics viene disabilitata per fornire una visione accurata di tutti i cookie emessi andando però a sporcare i dati di GA.

2. Abilita il rendering JavaScript

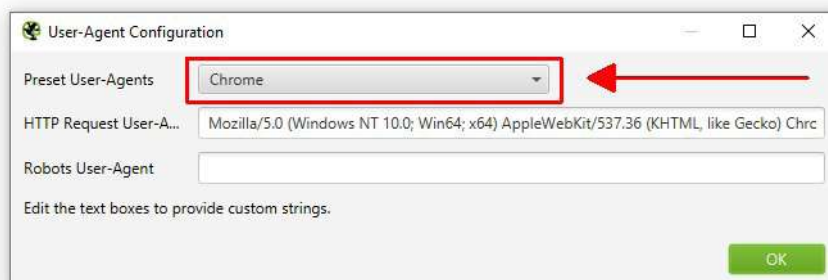
Modifica la tipologia di scansione attraverso "Config > Spider > Rendering" e seleziona "JavaScript" per intercettare i cookie che vengono caricati utilizzando JavaScript o i tag pixel image.



La dimensione della finestra di default è impostata automaticamente su Googlebot Smartphone, ma questo può essere regolato su desktop se ci sono differenze nel modo in cui il sito emette i cookie.

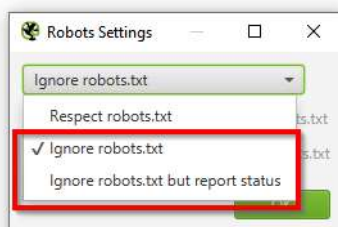
3. Imposta lo user-agent su Chrome

Per simulare al meglio la navigazione di un utente e non di un Bot modifica l'user-agent su Chrome tramite "Config > User-agent".



4. Imposta l'opzione "Ignora robots.txt"

Clicca su "Configuration > robots.txt" e seleziona "Ignore robots.txt" o "Ignore robots.txt but report status". Questa operazione ti permette di caricare tutte le risorse e rispettivi cookie ed evitare che alcune risorse siano bloccate durante la scansione.



5. Esegui il crawling del sito web

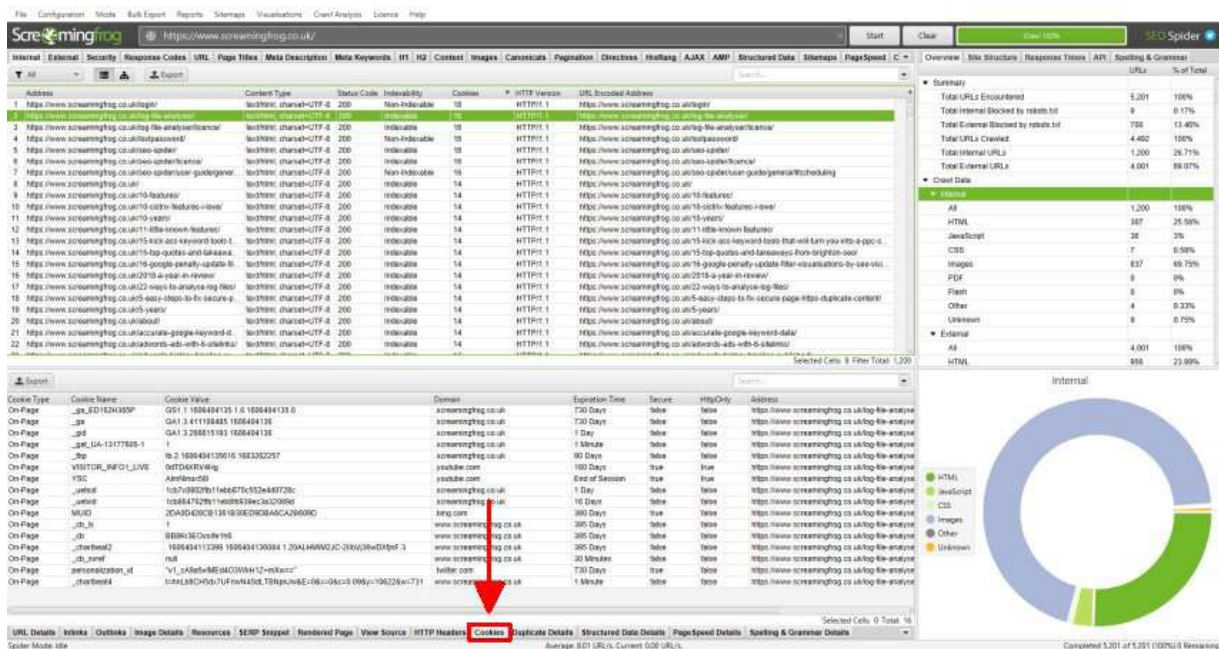
6. Visualizza il numero di cookie per ogni URL

Dopo la scansione sei in grado di vedere quanti cookie sono stati scoperti per ogni URL scansionato. Questo indice lo trovi nella scheda "Internal" nella colonna "cookies".

Screamingfrog										
File Configuration Mode Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help										
https://www.screamingfrog.co.uk/										
Internal	External	Security	Response Codes	URL	Page Titles	Meta Description	Meta Keywords	H1	H2	Content
All										
Export										
	Address	Content Type	Status Code	Indexability	Cookies		HTTP Ver			
1	https://www.screamingfrog.co.uk/login/	text/html; charset=UTF-8	200	Non-Indexable	18		HTTP/1.			
2	https://www.screamingfrog.co.uk/log-file-analyser/	text/html; charset=UTF-8	200	Indexable	16		HTTP/1.			
3	https://www.screamingfrog.co.uk/log-file-analyser/licence/	text/html; charset=UTF-8	200	Indexable	16		HTTP/1.			
4	https://www.screamingfrog.co.uk/lostpassword/	text/html; charset=UTF-8	200	Non-Indexable	16		HTTP/1.			
5	https://www.screamingfrog.co.uk/seo-spider/	text/html; charset=UTF-8	200	Indexable	16		HTTP/1.			
6	https://www.screamingfrog.co.uk/seo-spider/licence/	text/html; charset=UTF-8	200	Indexable	16		HTTP/1.			
7	https://www.screamingfrog.co.uk/seo-spider/user-guide/gener...	text/html; charset=UTF-8	200	Non-Indexable	16		HTTP/1.			
8	https://www.screamingfrog.co.uk/	text/html; charset=UTF-8	200	Indexable	14		HTTP/1.			
9	https://www.screamingfrog.co.uk/10-features/	text/html; charset=UTF-8	200	Indexable	14		HTTP/1.			
10	https://www.screamingfrog.co.uk/10-sistrix-features-i-love/	text/html; charset=UTF-8	200	Indexable	14		HTTP/1.			

7. Visualizza la Tab Cookies

Adesso non ti resta che vedere quali siano i cookie scoperti durante il crawl; ti basta cliccare su uno o più URL della finestra superiore per trovare tutti i dettagli nel pannello inferiore della tab "Cookie".



Di seguito un esempio della tab “Cookie” popolata con i dettagli.

Cookie Type	Cookie Name	Cookie Value	Domain	Expiration Time	Secure	HttpOnly	Address
On-Page	_ga_ED162H365P	GS1.1.1606404135.1.0.1606404135.0	screamingfrog.co.uk	730 Days	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_ga	GA1.3.411198485.1606404135	screamingfrog.co.uk	730 Days	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_gid	GA1.3.266615193.1606404135	screamingfrog.co.uk	1 Day	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_gat_UA-13177605-1	1	screamingfrog.co.uk	1 Minute	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_fbp	fb.2.1606404135616.1683262257	screamingfrog.co.uk	90 Days	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	VISITOR_INFO1_LIVE	0dTD4XRv4Hg	youtube.com	180 Days	true	true	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	YSC	AimNlnox58l	youtube.com	End of Session	true	true	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_uetcid	1cb7c0802fb11eb870c552e4d0728c	screamingfrog.co.uk	1 Day	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_uetvid	1cb864702fb11eb870c552e4d0728c	screamingfrog.co.uk	16 Days	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	MUID	2DADD428CB1361B30ED9DBA6CA2B	bing.com	390 Days	true	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_cb_is	1	www.screamingfrog.co.uk	395 Days	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_cb	BB8Kr3EOvsife1h6	www.screamingfrog.co.uk	395 Days	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_chartbeat2	.1606404113399.1606404136084.1.20.	www.screamingfrog.co.uk	395 Days	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_cb_svref	null	www.screamingfrog.co.uk	30 Minutes	false	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	personalization_id	"v1_cA9a5vIMEd403WkH1Z-mXw=="	twitter.com	730 Days	true	false	https://www.screamingfrog.co.uk/log-file-analyser/
On-Page	_chartbeat4	trhnlB8CH5dx7UFmN4SLBNpkJw	www.screamingfrog.co.uk	1 Minute	false	false	https://www.screamingfrog.co.uk/log-file-analyser/

Le colonne a tua disposizione sono le seguenti:

- **Cookie Type:** identifica dove il cookie è stato scoperto (“HTTP”, “On-Page” etc).
- **Cookie Name:** colleziona il nome del cookie.
- **Cookie Value:** definisce il valore del cookie.
- **Domain:** mostra il dominio che ha emesso il cookie di prima o di terza parte.
- **Expiration Time:** indica il tempo di scadenza del cookie.
- **Secure:** visualizza i dettagli dell'attributo di sicurezza del cookie.
- **HttpOnly:** mostra i dettagli dell'attributo HttpOnly del cookie.
- **Address:** visualizza l'URL su cui è stato trovato il cookie.

8. Esporta il rapporto aggregato sui cookie

Ora sei in grado di esportare un preziosissimo riepilogo aggregato sui cookie scoperti con il “Cookie Summary”.

Report > Cookie > Cookie Summary

File Configuration Mode Bulk Export **Reports** Sitemaps Visualisations Crawl Analysis Licence Help

Screamingfrog

Internal External Security Response Codes

All

Address

Address	Content Type	Status Code
1 https://www.screamingfrog.co.uk/	text/html; charset=UTF-8	200
2 https://www.screamingfrog.co.uk/wp-content/css/autoptimize_8e8a070...	text/css	200
3 https://www.screamingfrog.co.uk/wp-content/css/autoptimize_469ca155...	text/css	200
4 https://www.screamingfrog.co.uk/wp-content/css/autoptimize_469ca155...	text/css	200
5 https://www.screamingfrog.co.uk/wp-content/js/jquery-1.3.2.min.js	application/javascript	200
6 https://www.screamingfrog.co.uk/wp-content/js/jquery.cycle.all.min.js	application/javascript	200
7 https://www.screamingfrog.co.uk/wp-content/js/jquery.cycle.all.min.js	application/javascript	200
8 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/force-zxcvbn...	application/javascript	200
9 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/js-admin.min.js	application/javascript	200
10 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/js-admin.min.js	application/javascript	200
11 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/js-admin.min.js	application/javascript	200
12 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/js-admin.min.js	application/javascript	200
13 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/js-admin.min.js	application/javascript	200

Cookie Summary

Il rapporto visualizza una vista aggregata dei cookie unici scoperti durante una scansione, considerando il loro nome, dominio, scadenza, valori di sicurezza e HttpOnly. Viene visualizzato anche il numero di URL su cui ogni cookie unico è stato emesso.

cookie_summary ☆ 📄

File Edit View Insert Format Data Tools Add-ons Help Last edit was seconds ago

100% \$ % .0 .00 123 Default (Ari... 10 B I S A

Cookie Name	Cookie Type	Domain	Expiration T	Secure	HttpOnly	Occurrences
1 MUID	On-Page	.bing.com	390 Days	TRUE	FALSE	305
2 _fbp	On-Page	.screamingfrog.co.uk	90 Days	FALSE	FALSE	305
3 _ga	On-Page	.screamingfrog.co.uk	730 Days	FALSE	FALSE	305
4 _ga_ED162H365P	On-Page	.screamingfrog.co.uk	730 Days	FALSE	FALSE	305
5 _gat_UA-13177605-1	On-Page	.screamingfrog.co.uk	1 Minute	FALSE	FALSE	305
6 _gid	On-Page	.screamingfrog.co.uk	1 Day	FALSE	FALSE	305
7 _uetxid	On-Page	.screamingfrog.co.uk	1 Day	FALSE	FALSE	305
8 _uetvid	On-Page	.screamingfrog.co.uk	16 Days	FALSE	FALSE	305
9 personalization_id	On-Page	.twitter.com	730 Days	TRUE	FALSE	305
10 sh	On-Page	.screamingfrog.co.uk	305 Days	FALSE	FALSE	305

9. Esporta in blocco tutti i dettagli dei cookie

Per un'analisi più specifica puoi cliccare su "Bulk Export-Esportazione in blocco > Web > All Cookies" per ottenere per ogni singolo URL i cookie e gli attributi relativi.

Internal External Security **Web**

All

Address

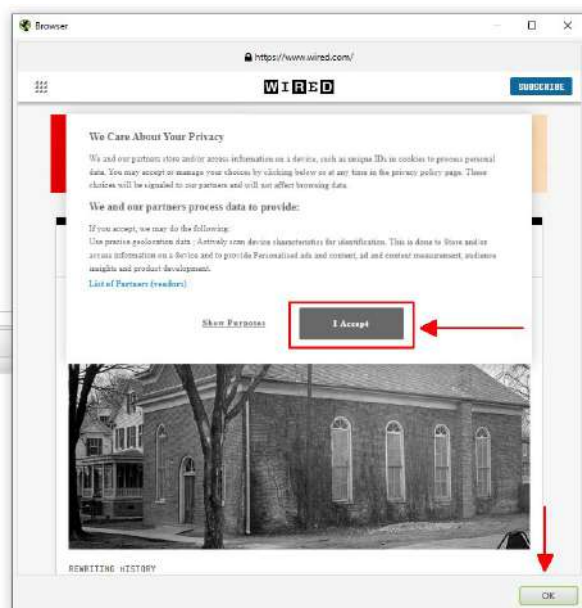
Address	Content Type	Status Code
1 https://www.screamingfrog.co.uk/	text/html; charset=UTF-8	200
2 https://www.screamingfrog.co.uk/wp-content/css/autoptimize_8e8a070...	text/css	200
3 https://www.screamingfrog.co.uk/wp-content/css/autoptimize_469ca155...	text/css	200
4 https://www.screamingfrog.co.uk/wp-content/css/autoptimize_469ca155...	text/css	200
5 https://www.screamingfrog.co.uk/wp-content/js/jquery-1.3.2.min.js	application/javascript	200
6 https://www.screamingfrog.co.uk/wp-content/js/jquery.cycle.all.min.js	application/javascript	200
7 https://www.screamingfrog.co.uk/wp-content/js/jquery.cycle.all.min.js	application/javascript	200
8 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/force-zxcvbn...	application/javascript	200
9 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/js-admin.min.js	application/javascript	200
10 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/js-admin.min.js	application/javascript	200
11 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/js-admin.min.js	application/javascript	200
12 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/js-admin.min.js	application/javascript	200
13 https://www.screamingfrog.co.uk/wp-content/js/force-strong-passwords/js-admin.min.js	application/javascript	200

All Cookies

Per avere la certezza di scoprire tutti i cookies di un sito web è consigliabile controllarli dopo l'accettazione del "Consent Banner".

Tips: per ovviare a questa criticità puoi utilizzare l'autenticazione "Forms Based" di Screaming Frog ed accettare il pop-up sui cookie eseguendo una scansione come se lo avessi accettato.

"Config > Authentication > Forms Based", poi clicca su "Add", inserisci l'URL del sito web e clicca su 'OK'.

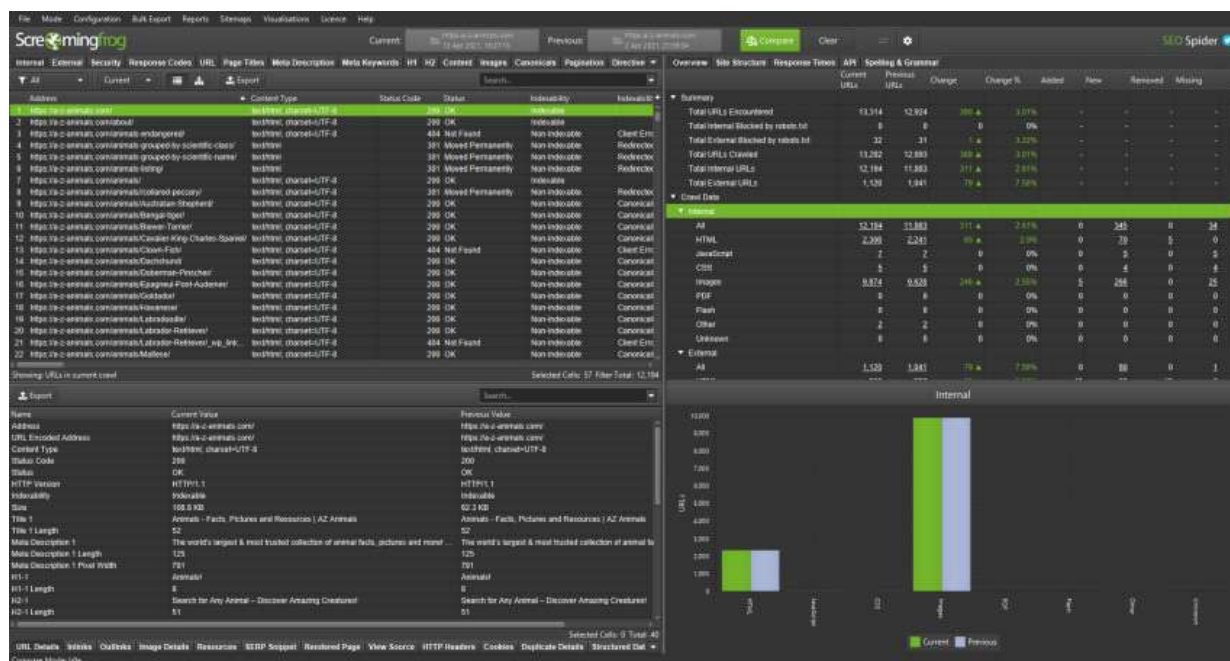


Seguendo gli step elencati precedentemente sarai in grado di confrontare i cookie scoperti con e senza accettazione.

COME COMPARARE I DATI DI DUE SCANSIONI

Questo tutorial ti spiega come confrontare due scansioni utilizzando Screaming Frog SEO Spider e analizzare come i dati di crawl siano cambiati nelle schede e nei filtri nel tempo.

Attraverso questa attività è possibile monitorare il progresso di eventuali ottimizzazioni o l'andamento delle criticità Seo avendo a disposizione utili dati granulari su ciò che è cambiato tra i diversi crawl.

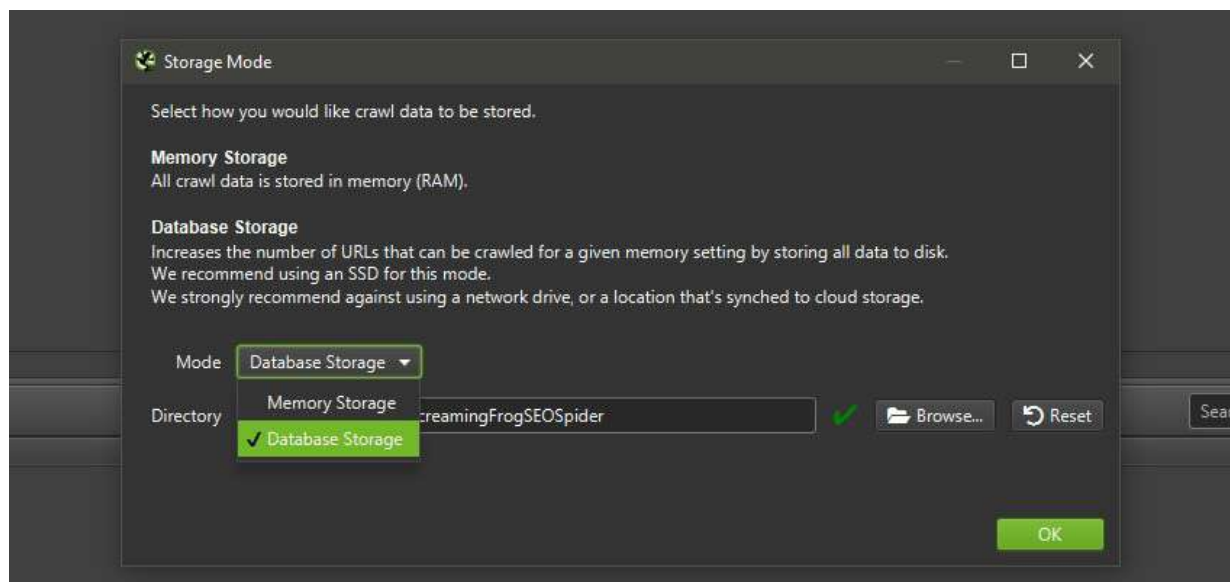


In sintesi questa analisi ti aiuta ad interpretare diversi scenari:

- confrontare i dati di crawl e rilevare nuovi problemi.
- Confrontare la struttura del sito e i cambiamenti nelle directory, nelle pagine e nella profondità di scansione tra i crawl.
- Rilevare i cambiamenti negli elementi e nelle metriche, come i titoli delle pagine, le meta descrizioni, la profondità di scansione, il conteggio delle parole, i collegamenti interni, il contenuto e altro ancora.

1. Utilizza della modalità di archiviazione su database

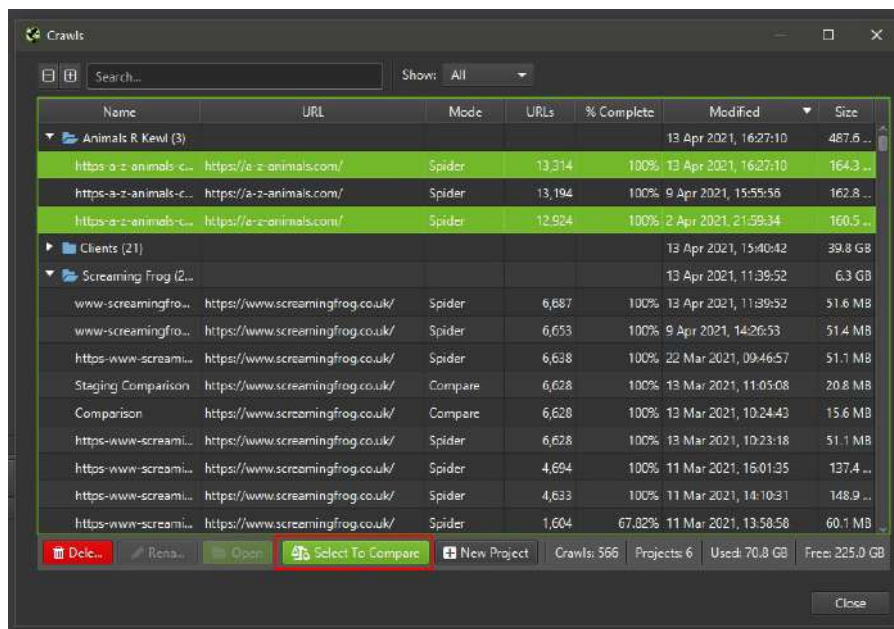
Per confrontare i crawl è necessario archiviare le diverse scansioni selezionando l'opzione "Database Storage".



L'archiviazione su database ti permette il salvataggio automatico delle scansioni, l'apertura rapida dei crawl e di avere sempre a disposizione più scansioni di un progetto direttamente in locale.

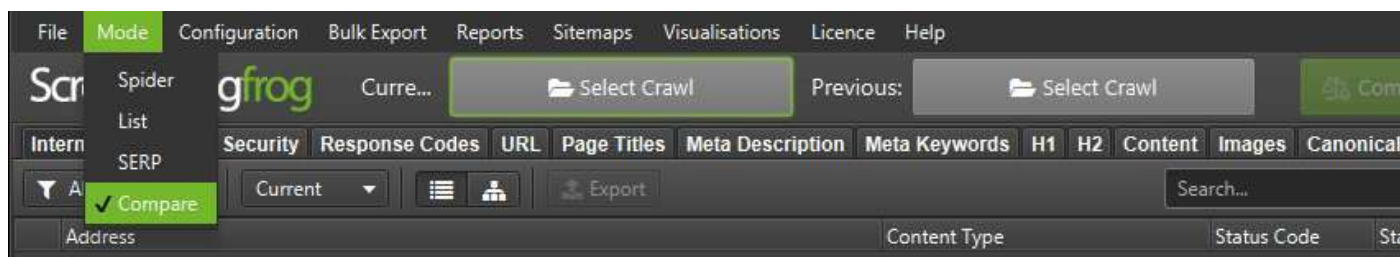
2. Seleziona i crawl da confrontare

Per iniziare il confronto è sufficiente selezionare le scansioni da “File > Crawls”, evidenziare i due crawl tenendo premuto “control” su Windows, o il tasto “command” su macOS e cliccare su “Select to Compare”, per passare dalla modalità “Spider” o “List” alla modalità “Compare”.



Alternativamente puoi scegliere la modalità “Compare” direttamente dal menu principale.

Mode > Compare



3. Change Detection

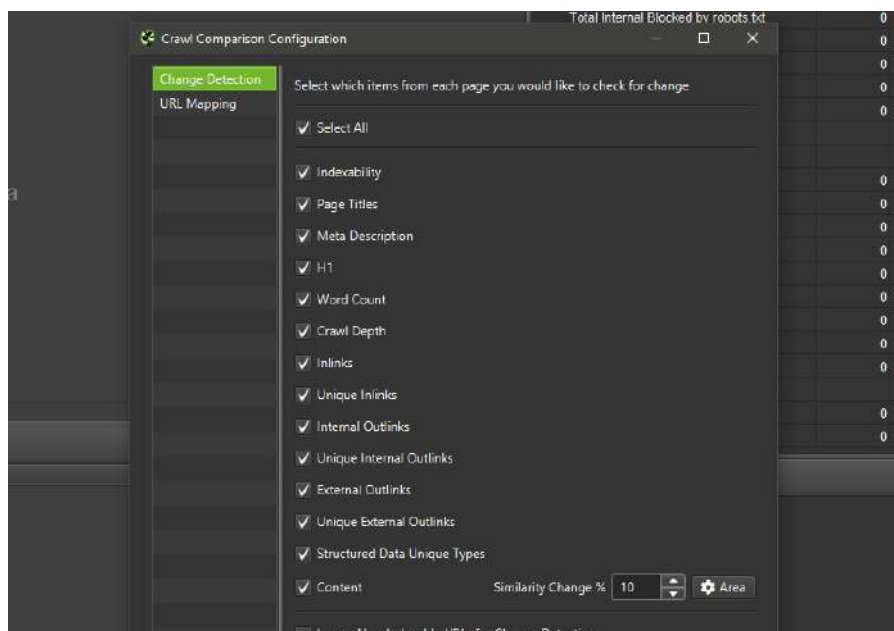
Il Seo Spider confronterà tutti i dati della scansione nelle schede e nei filtri di “Overviews” della sidebar destra.

Tuttavia, puoi scegliere di configurare lo strumento per eseguire un'analisi aggiuntiva sui cambiamenti più specifici agli elementi e alle metriche chiave come il titolo della pagina, la meta descrizione, l'h1, il conteggio delle parole, la profondità di scansione, i link interni, i dati strutturati, il contenuto e molto altro.

Questa opzione è utile quando un elemento è cambiato, ma non ha attivato un filtro. Per esempio, il titolo di una pagina potrebbe cambiare completamente, ma questo non verrebbe rilevato se non attivasse un filtro di cambiamento come “Missing”, “Over X Characters”, “Duplicates” ecc.

Il rilevamento delle modifiche è possibile con la funzione “Change Detection” che ti segnala ogni cambiamento ad un elemento o ad una metrica.

In modalità “Compare”, clicca sulla configurazione di confronto tramite “Config > Compare” (o l'icona “ingranaggio” in alto) e seleziona gli elementi e le metriche per i quali vuoi identificare i cambiamenti.

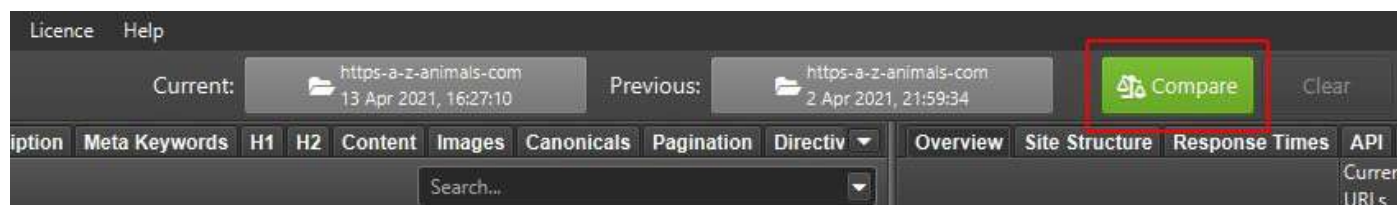


Abilitando queste opzioni i risultati del crawl attivano la nuova scheda “Change Detection” nella vista principale e nella scheda Panoramica, contenente tutti i filtri degli elementi e le metriche selezionati con i dettagli delle modifiche rilevate.

Se opti per questo controllo degli elementi considera che il Seo Spider richiede maggiori risorse per elaborare, memorizzare l'HTML delle pagine, analizzare le relative somiglianze utilizzando il minhash e determinare la % di cambiamento tra il contenuto attuale e quello precedente.

4. Avvia la scansione di comparazione cliccando “Compare”

E' sufficiente cliccare su “Confronta” per avviare la scansione fra i due crawl in base alle configurazioni “Change Detection” settate in precedenza.

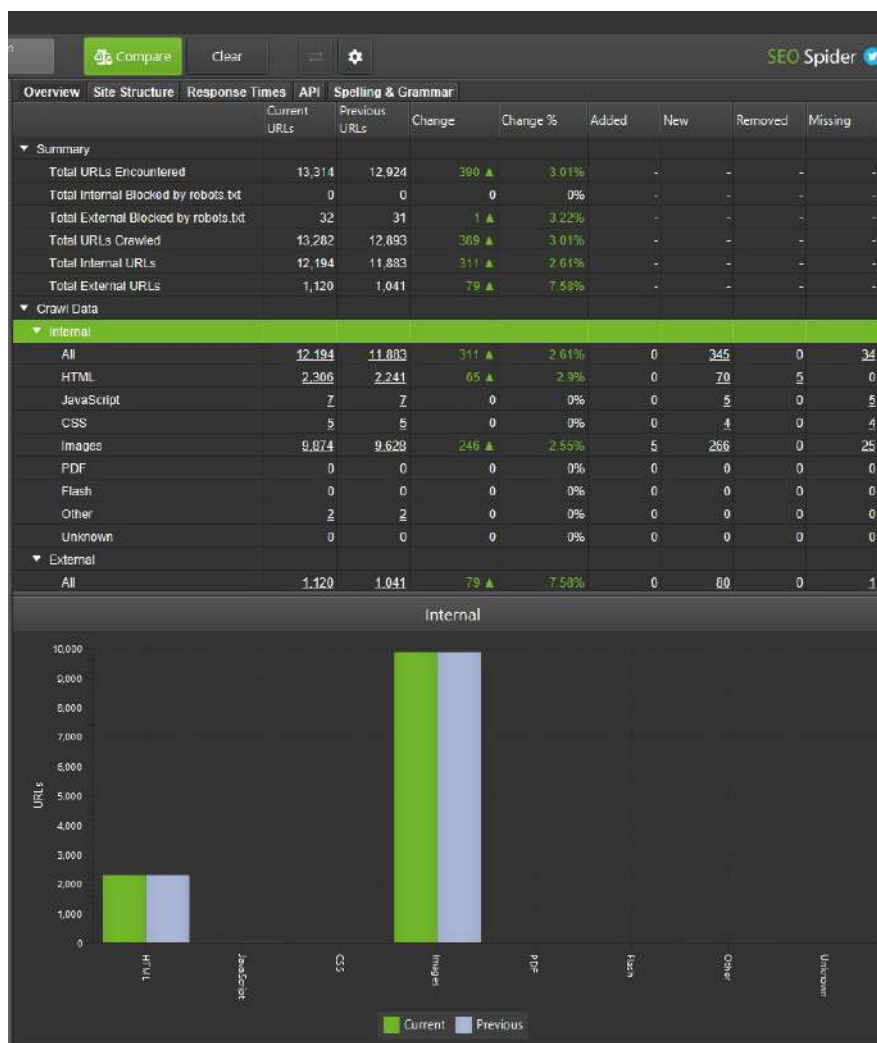


Nota Bene: il Seo Spider utilizza l'ultimo crawl in ordine di data come “Current” e la scansione più vecchia come “Previous” ma puoi invertire la selezione cliccando sulla freccia accanto all'icona dell'ingranaggio nella parte superiore dell'interfaccia.

5. Visualizza la scheda “Panoramica”

Quando la comparazione è stata completata, la scheda panoramica sulla sidebar di destra si popola per mostrare i dati e le modifiche rilevate.

Cliccando sulle schede puoi aggiornare i grafici sottostanti, così come cliccare sui filtri e sui numeri nelle colonne che aggiornano i dati della finestra principale.



La sidebar laterale presenta otto colonne mentre la finestra principale utilizza 4 filtri che ti permettono di segmentare gli URL che sono cambiati nelle schede e nei filtri.

- 1. Added:** identifica tutti gli URL che sono presenti in entrambe le scansioni ma che presentano degli elementi modificati che vengono inseriti in qualche filtro del crawl "Current".
Ad esempio un URL presentava un "meta_title" nella versione precedente mentre nella versione attuale questo elemento risulta mancante. In questo caso questa indicazione sarà presente nel filtro "Missing" dell'elemento "Meta Title".
- 2. New:** comprende tutti i nuovi URL che non erano presenti nella scansione precedente ma sono inclusi nella scansione attuale.
- 3. Removed:** URL presenti in entrambe le scansioni ma con elementi presenti nel filtro della scansione precedente, ma non nel filtro della scansione attuale.
Riprendendo l'esempio del "Meta Title" precedente, questo filtro identifica tutti gli URL che non presentavano il titolo della pagina ed erano inseriti nel filtro "Missing" ma nella scansione attuale non sono più presenti nel filtro perché ottimizzati.
- 4. Missing:** URL presenti nella scansione precedente ma non rilevati nel crawl "Current"

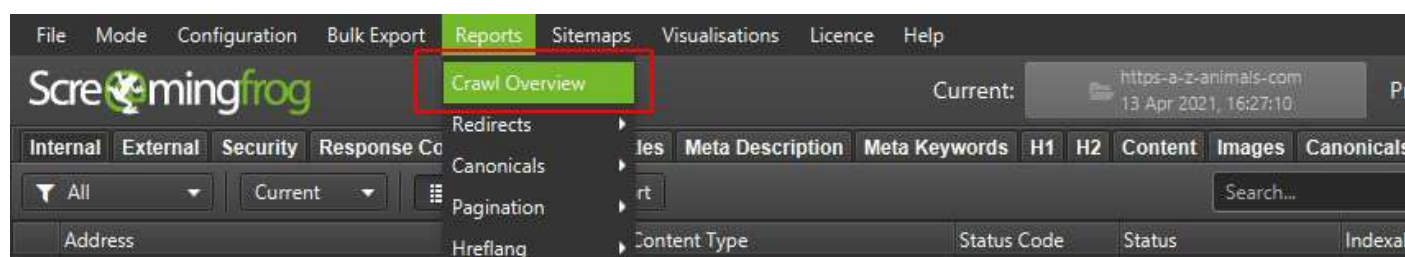
In sintesi, "aggiunti" e "rimossi" sono URL che esistono sia nella scansione attuale che in quella precedente ma differiscono nei filtri. "Nuovi" e "mancanti" sono URL che esistono solo in uno dei crawl. Questi filtri ti aiutano a determinare se i nuovi problemi o le correzioni riguardano URL esistenti o nuovi URL. Cliccando sui numeri nelle colonne sei in grado di vedere quali URL sono cambiati, e usare il filtro nella finestra principale per alternare le scansioni attuali e precedenti, così come aggiunte, nuove, rimosse o mancanti.

The screenshot shows the Screamingfrog software interface. The 'Internal' tab is selected, displaying a list of URLs and their occurrences. The 'Response Codes' column is highlighted with a green box. The 'Current' status is set to 'Added'.

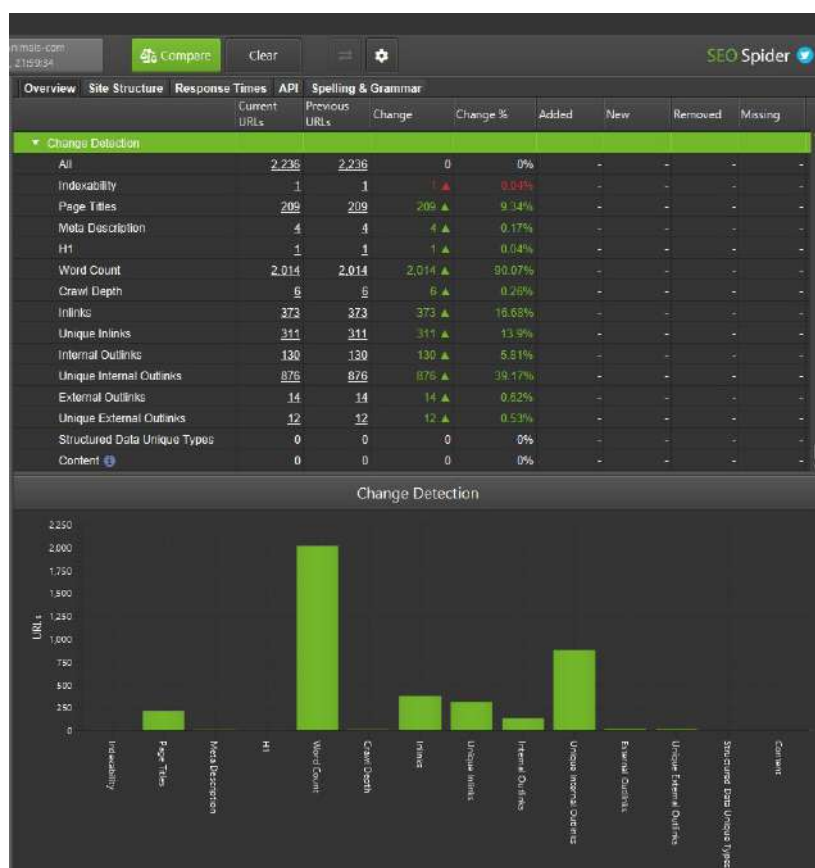
Address	Occurrences	Title 1
1 https://a-z-animals.com/animals/location/ocean/great-lakes/	1	Wildlife in Great Lakes Types of Great Lakes Animals AZ Animals
2 https://a-z-animals.com/animals/location/ocean/indian/	1	Wildlife in Indian Types of Indian Ocean Animals AZ Animals
3 https://a-z-animals.com/animals/location/ocean/australia/	1	Wildlife in Australia Types of Australian Animals AZ Animals
4 https://a-z-animals.com/animals/location/asia/vietnam/	1	Wildlife in Vietnam Types of Vietnamese Animals AZ Animals
5 https://a-z-animals.com/animals/location/africa/south-africa/	1	Wildlife in South Africa Types of South African Animals AZ Animals
6 https://a-z-animals.com/animals/location/asia/south-korea/	1	Wildlife in South Korea Types of Korean Animals AZ Animals
7 https://a-z-animals.com/animals/location/africa/madagascar/	1	Wildlife in Madagascar Types of Malagasy Animals AZ Animals
8 https://a-z-animals.com/animals/location/europe/hungary/	1	Wildlife in Hungary Types of Hungarian Animals AZ Animals
9 https://a-z-animals.com/animals/location/south-america/suriname/	1	Wildlife in Suriname Types of Surinamese Animals AZ Animals
10 https://a-z-animals.com/animals/location/africa/south-sudan/	1	Wildlife in South Sudan Types of South Sudanese Animals AZ Animals
11 https://a-z-animals.com/animals/location/asia/united-arab-emirates/	1	Wildlife in United Arab Emirates Types of Emirian Animals AZ Animals
12 https://a-z-animals.com/animals/location/africa/botswana/	1	Wildlife in Botswana Types of Motswana, Botswanan Animals AZ Animals
13 https://a-z-animals.com/animals/location/south-america/argentina/	1	Wildlife in Argentina Types of Argentine Animals AZ Animals
14 https://a-z-animals.com/animals/location/africa/ethiopia/	1	Wildlife in Ethiopia Types of Ethiopian Animals AZ Animals
15 https://a-z-animals.com/animals/location/north-america/united-states/	1	Wildlife in United States Types of USA Animals AZ Animals
16 https://a-z-animals.com/animals/location/europe/bulgaria/	1	Wildlife in Bulgaria Types of Bulgarian Animals AZ Animals
17 https://a-z-animals.com/animals/location/asia/afghanistan/	1	Wildlife in Afghanistan Types of Afghan Animals AZ Animals
18 https://a-z-animals.com/animals/location/central-america/nicaragua/	1	Wildlife in Nicaragua Types of Nicaraguan Animals AZ Animals
19 https://a-z-animals.com/animals/location/south-america/colombia/	1	Wildlife in Colombia Types of Colombian Animals AZ Animals
20 https://a-z-animals.com/animals/location/asia/bangladesh/	1	Wildlife in Bangladesh Types of Bangladeshi Animals AZ Animals
21 https://a-z-animals.com/animals/location/south-america/paraguay/	1	Wildlife in Paraguay Types of Paraguayan Animals AZ Animals
22 https://a-z-animals.com/animals/location/oceania/new-zealand/	1	Wildlife in New Zealand Types of New Zealand Animals AZ Animals
23 https://a-z-animals.com/animals/location/central-america/guatemala/	1	Wildlife in Guatemala Types of Guatemalan Animals AZ Animals
24 https://a-z-animals.com/animals/location/central-america/puerto-rico/	1	Wildlife in Puerto Rico Types of Puerto Rican Animals AZ Animals

Showing: URLs in previous crawl that moved to filter of current crawl. Selected Cells: 0 Filter Total: 74

I dati della sidebar laterale “Overview - Panoramica” possono essere esportati con il documento “Crawl Overview”.

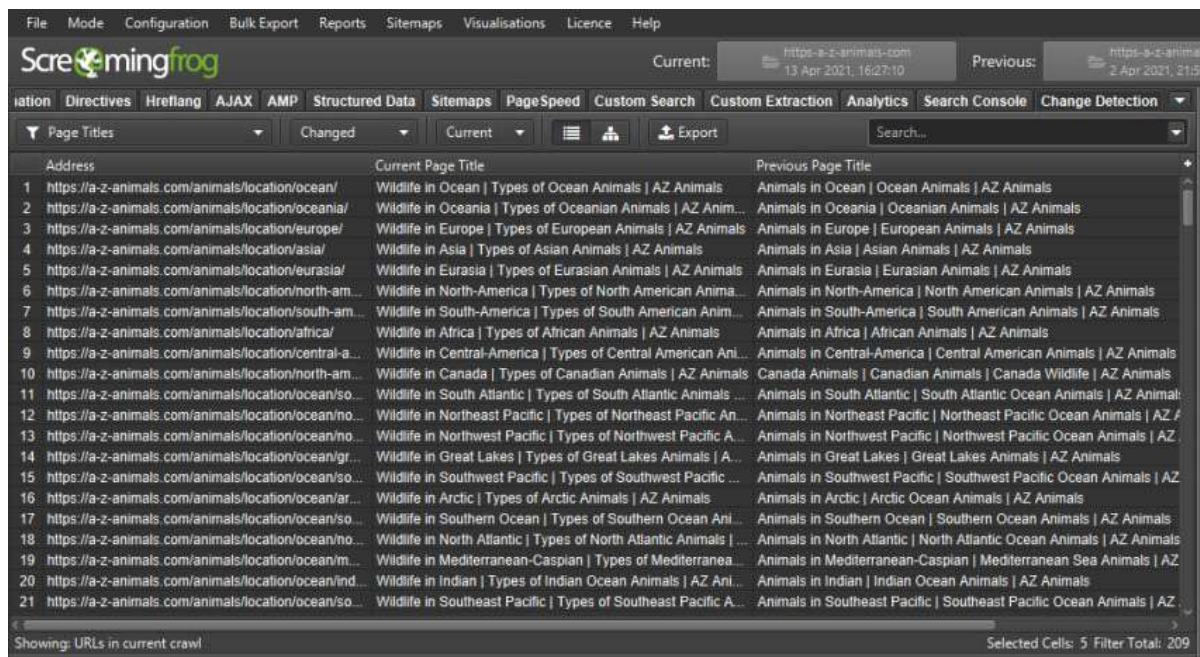


6. Analizza i cambiamenti nella scheda “Change Detection”.



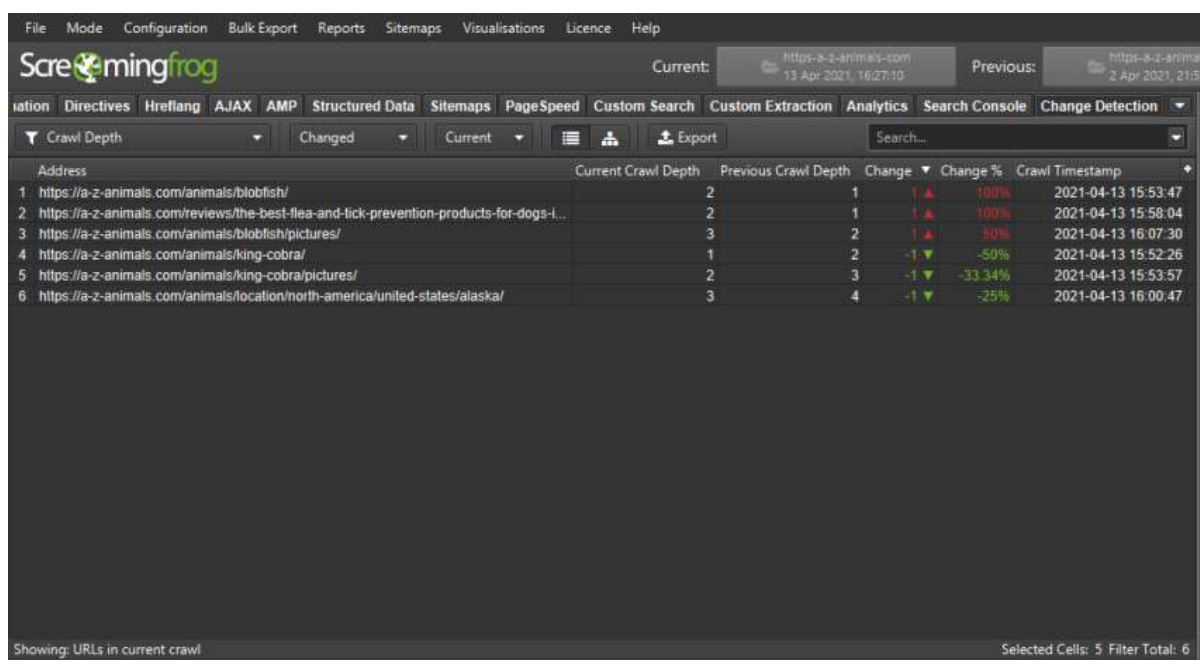
Nell'esempio precedente, ci sono un totale di 2.236 URL esistenti in entrambe le scansioni, 1 URL ha modificato l'H1, 209 URL hanno cambiato il titolo della pagina, 6 hanno cambiato la profondità di scansione, ecc.

Per interagire con questi dati è sufficiente cliccare su un elemento e visualizzare i cambiamenti nella finestra principale. Per esempio, se vuoi visualizzare i cambiamenti dei titoli di pagina, puoi cliccare sull'elemento e confrontarli fianco a fianco nella scheda principale.



Address	Current Page Title	Previous Page Title
1 https://a-z-animals.com/animals/location/ocean/	Wildlife in Ocean Types of Ocean Animals AZ Animals	Animals in Ocean Ocean Animals AZ Animals
2 https://a-z-animals.com/animals/location/oceania/	Wildlife in Oceania Types of Oceanian Animals AZ Animals	Animals in Oceania Oceanian Animals AZ Animals
3 https://a-z-animals.com/animals/location/europe/	Wildlife in Europe Types of European Animals AZ Animals	Animals in Europe European Animals AZ Animals
4 https://a-z-animals.com/animals/location/asia/	Wildlife in Asia Types of Asian Animals AZ Animals	Animals in Asia Asian Animals AZ Animals
5 https://a-z-animals.com/animals/location/eurasia/	Wildlife in Eurasia Types of Eurasian Animals AZ Animals	Animals in Eurasia Eurasian Animals AZ Animals
6 https://a-z-animals.com/animals/location/north-am...	Wildlife in North-America Types of North American Anima...	Animals in North-America North American Animals AZ Animals
7 https://a-z-animals.com/animals/location/south-am...	Wildlife in South-America Types of South American Anim...	Animals in South-America South American Animals AZ Animals
8 https://a-z-animals.com/animals/location/afrika/	Wildlife in Africa Types of African Animals AZ Animals	Animals in Africa African Animals AZ Animals
9 https://a-z-animals.com/animals/location/central-a...	Wildlife in Central-America Types of Central American Ani...	Animals in Central-America Central American Animals AZ Animals
10 https://a-z-animals.com/animals/location/north-am...	Wildlife in Canada Types of Canadian Animals AZ Animals	Canada Animals Canadian Animals Canada Wildlife AZ Animals
11 https://a-z-animals.com/animals/location/ocean/so...	Wildlife in South Atlantic Types of South Atlantic Animals ...	Animals in South Atlantic South Atlantic Ocean Animals AZ Animal...
12 https://a-z-animals.com/animals/location/ocean/no...	Wildlife in Northeast Pacific Types of Northeast Pacific An...	Animals in Northeast Pacific Northeast Pacific Ocean Animals AZ /
13 https://a-z-animals.com/animals/location/ocean/no...	Wildlife in Northwest Pacific Types of Northwest Pacific A...	Animals in Northwest Pacific Northwest Pacific Ocean Animals AZ
14 https://a-z-animals.com/animals/location/ocean/gr...	Wildlife in Great Lakes Types of Great Lakes Animals A...	Animals in Great Lakes Great Lakes Animals AZ Animals
15 https://a-z-animals.com/animals/location/ocean/so...	Wildlife in Southwest Pacific Types of Southwest Pacific ...	Animals in Southwest Pacific Southwest Pacific Ocean Animals AZ
16 https://a-z-animals.com/animals/location/ocean/ar...	Wildlife in Arctic Types of Arctic Animals AZ Animals	Animals in Arctic Arctic Ocean Animals AZ Animals
17 https://a-z-animals.com/animals/location/ocean/so...	Wildlife in Southern Ocean Types of Southern Ocean Ani...	Animals in Southern Ocean Southern Ocean Animals AZ Animals
18 https://a-z-animals.com/animals/location/ocean/no...	Wildlife in North Atlantic Types of North Atlantic Animals ...	Animals in North Atlantic North Atlantic Ocean Animals AZ Animals
19 https://a-z-animals.com/animals/location/ocean/m...	Wildlife in Mediterranean-Caspien Types of Mediterranean...	Animals in Mediterranean-Caspien Mediterranean Sea Animals AZ
20 https://a-z-animals.com/animals/location/ocean/ind...	Wildlife in Indian Types of Indian Ocean Animals AZ Ani...	Animals in Indian Indian Ocean Animals AZ Animals
21 https://a-z-animals.com/animals/location/ocean/so...	Wildlife in Southeast Pacific Types of Southeast Pacific A...	Animals in Southeast Pacific Southeast Pacific Ocean Animals AZ

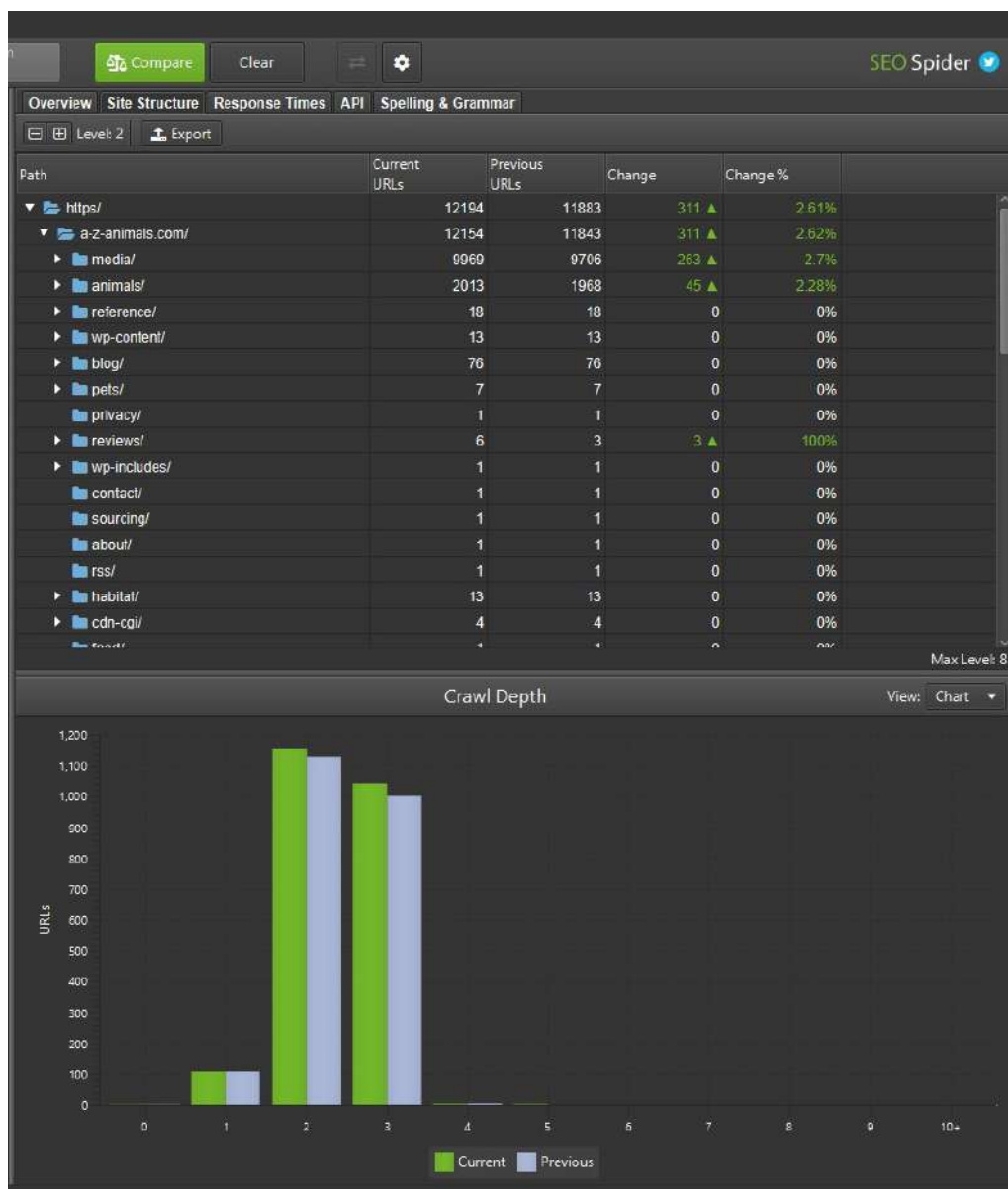
Un altro esempio di applicazione potrebbe essere quello di valutare se è cambiata l'architettura del sito relativamente al "Crawl Depth".



Address	Current Crawl Depth	Previous Crawl Depth	Change	Change %	Crawl Timestamp
1 https://a-z-animals.com/animals/blobfish/	2	1	1 ▲	100%	2021-04-13 15:53:47
2 https://a-z-animals.com/reviews/the-best-flea-and-tick-prevention-products-for-dogs-i...	2	1	1 ▲	100%	2021-04-13 15:58:04
3 https://a-z-animals.com/animals/blobfish/pictures/	3	2	1 ▲	50%	2021-04-13 16:07:30
4 https://a-z-animals.com/animals/king-cobra/	1	2	-1 ▼	-50%	2021-04-13 15:52:26
5 https://a-z-animals.com/animals/king-cobra/pictures/	2	3	-1 ▼	-33.34%	2021-04-13 15:53:57
6 https://a-z-animals.com/animals/location/north-america/united-states/alaska/	3	4	-1 ▼	-25%	2021-04-13 16:00:47

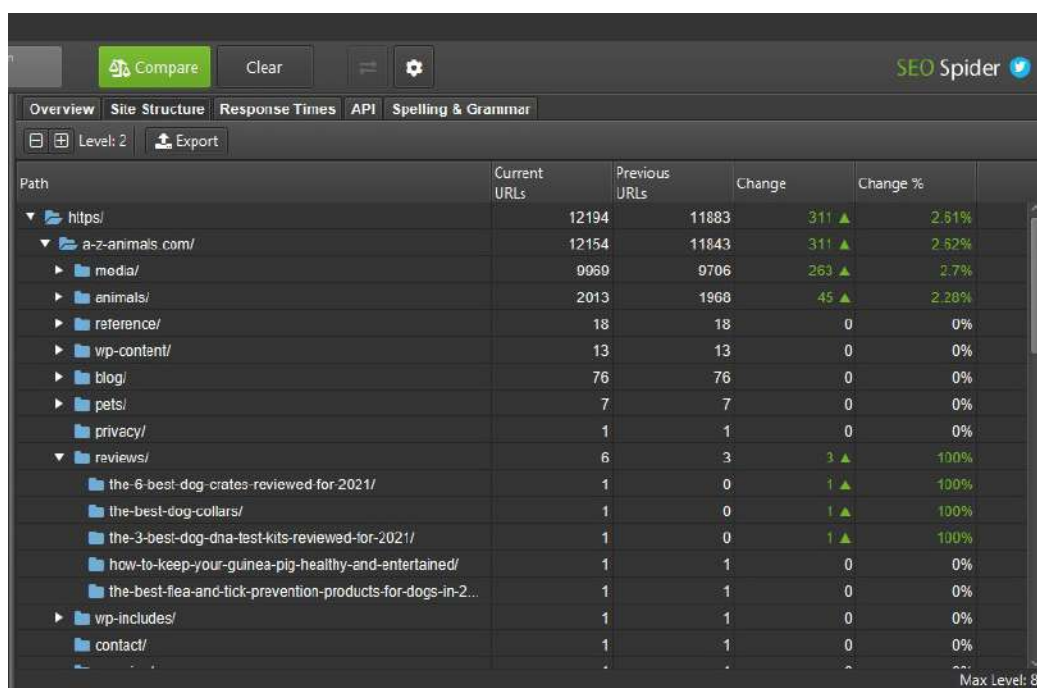
7. Visualizza la scheda "Struttura del sito" per esaminare le directory

La scheda "Site Structured" ti mostra una panoramica ad albero delle directory del sito web e può essere utile per identificare come l'architettura del sito si sia evoluta tra le scansioni identificando le eventuali modifiche alle directory.



Nell'esempio precedente puoi notare che sono stati trovati nuovi file nelle directory media, animali e recensioni.

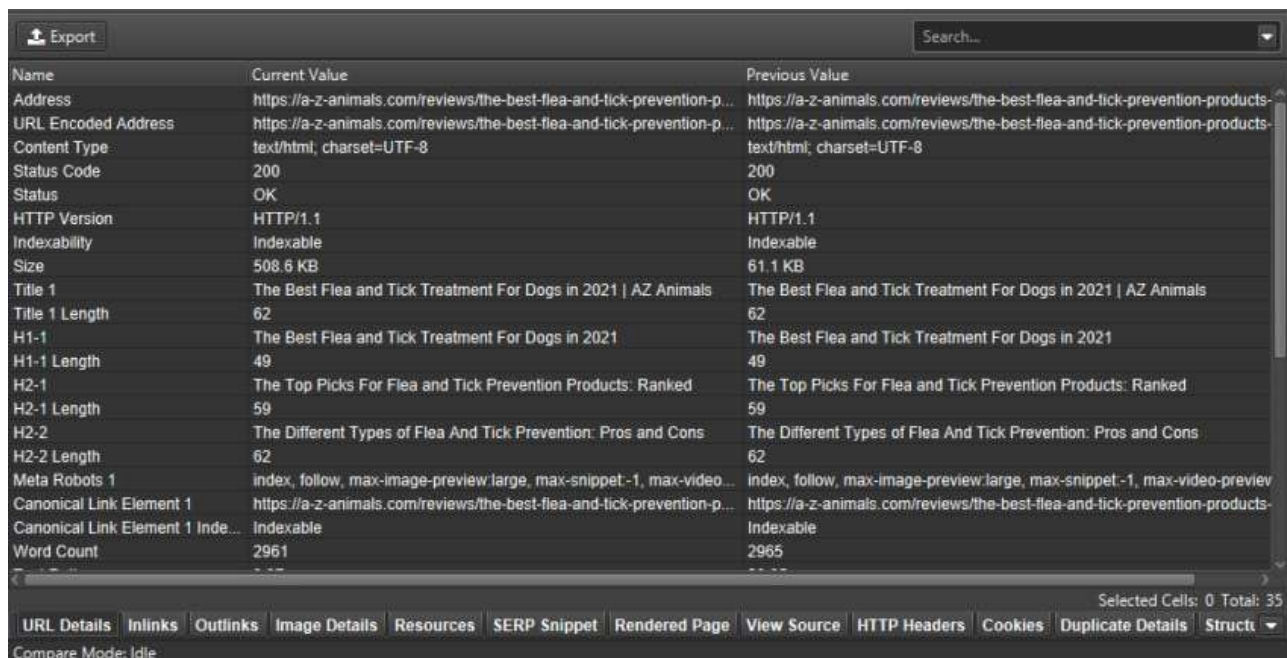
Anche con la visualizzazione del "Site Structured" sei in grado di visualizzare come è cambiata la profondità di scansione, molto utile per valutare i cambiamenti nel linking interno.



È possibile anche cliccare ed espandere le directory per vedere quali URL specifici sono cambiati, come nell'esempio nella directory delle recensioni. Le frecce verdi verso l'alto e i numeri indicano gli URL aggiunti, mentre le frecce rosse verso il basso e i numeri indicano gli URL che ora mancano. Nel nostro esempio la directory recensioni presenta 3 nuovi URL.

8. Confronta i dati nelle schede della finestra inferiore

Un'altra opportunità di confronto è data dalla consultazione della finestra inferiore del Seo Spider dove è possibile visualizzare i dati delle scansioni confrontandoli fianco a fianco.

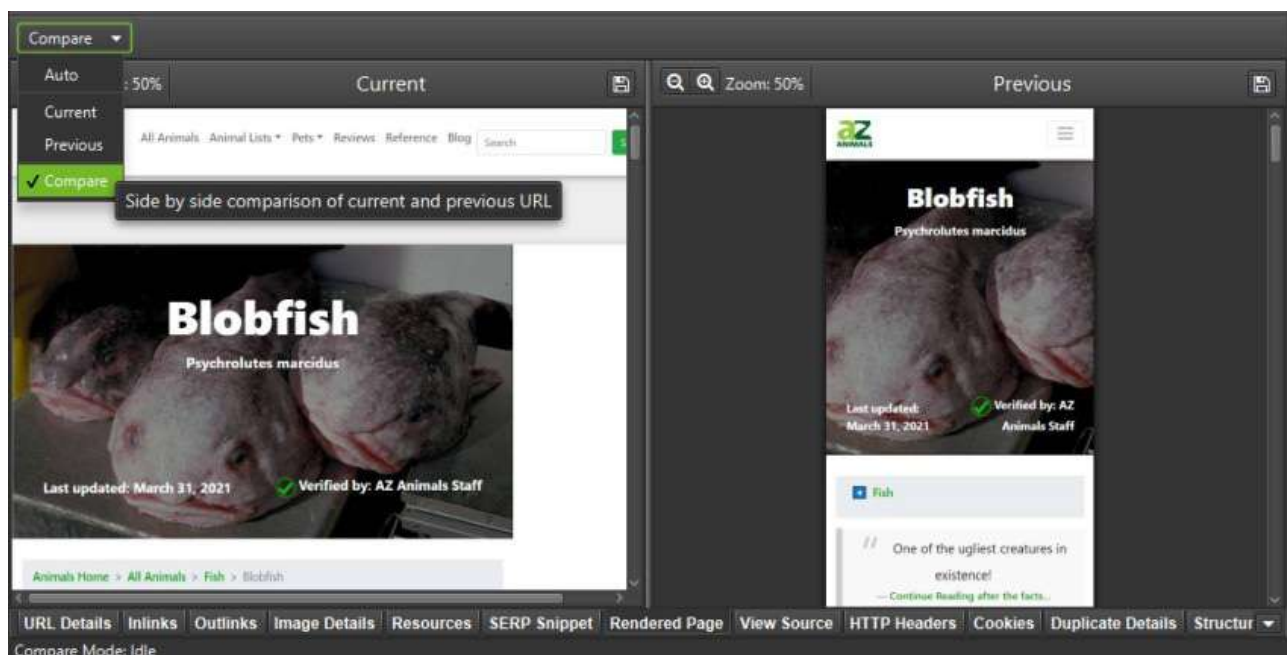


Name	Current Value	Previous Value
Address	https://a-z-animals.com/reviews/the-best-flea-and-tick-prevention-p...	https://a-z-animals.com/reviews/the-best-flea-and-tick-prevention-products-
URL Encoded Address	https://a-z-animals.com/reviews/the-best-flea-and-tick-prevention-p...	https://a-z-animals.com/reviews/the-best-flea-and-tick-prevention-products-
Content Type	text/html; charset=UTF-8	text/html; charset=UTF-8
Status Code	200	200
Status	OK	OK
HTTP Version	HTTP/1.1	HTTP/1.1
Indexability	Indexable	Indexable
Size	508.6 KB	61.1 KB
Title 1	The Best Flea and Tick Treatment For Dogs in 2021 AZ Animals	The Best Flea and Tick Treatment For Dogs in 2021 AZ Animals
Title 1 Length	62	62
H1-1	The Best Flea and Tick Treatment For Dogs in 2021	The Best Flea and Tick Treatment For Dogs in 2021
H1-1 Length	49	49
H2-1	The Top Picks For Flea and Tick Prevention Products: Ranked	The Top Picks For Flea and Tick Prevention Products: Ranked
H2-1 Length	59	59
H2-2	The Different Types of Flea And Tick Prevention: Pros and Cons	The Different Types of Flea And Tick Prevention: Pros and Cons
H2-2 Length	62	62
Meta Robots 1	index, follow, max-image-preview:large, max-snippet:-1, max-video-...	index, follow, max-image-preview:large, max-snippet:-1, max-video-...
Canonical Link Element 1	https://a-z-animals.com/reviews/the-best-flea-and-tick-prevention-p...	https://a-z-animals.com/reviews/the-best-flea-and-tick-prevention-products-
Canonical Link Element 1 Inde...	Indexable	Indexable
Word Count	2961	2965

Selected Cells: 0 Total: 35

Compare Mode: Idle

Molto utile ad esempio per confrontare un crawl di un sito in JavaScript e il suo comportamento su desktop e mobile per visualizzare il rendering mobile e desktop fianco a fianco.



Sempre nella finestra inferiore puoi beneficiare (Tab Inlink e Outlink) del filtro aggiuntivo “Auto” che per default ti mostra i dati dei link in base alla vista della finestra superiore.

File Mode Configuration Bulk Export Reports Sitemaps Visualisations Licence Help

Screamingfrog

Current: https://a-z-animals.com 13 Apr 2021, 16:27:10 Previous: https://a-z-animals.com 2 Apr 2021, 21:59:34

Pagination Directives Hreflang AJAX AMP Structured Data Sitemaps PageSpeed Custom Search Custom Extraction Analytics Search Console Change Detection

▼ Crawl Depth ▼ Changed ▼ Current ▼ Export

Address	Current Crawl Depth	Previous Crawl Depth	Change	Change %	Crawl Timestamp
1 https://a-z-animals.com/animals/blobfish/	2	1	1 ▲	100%	2021-04-13 15:53:47
2 https://a-z-animals.com/reviews/the-best-flea-and-tick-prevention-products-for-dogs-l...	2	1	1 ▲	100%	2021-04-13 15:58:04
3 https://a-z-animals.com/animals/blobfish/pictures/	3	2	1 ▲	50%	2021-04-13 16:07:30
4 https://a-z-animals.com/animals/king-cobra/	1	2	-1 ▼	-50%	2021-04-13 15:52:26
5 https://a-z-animals.com/animals/king-cobra/pictures/	2	3	-1 ▼	-33.34%	2021-04-13 15:53:57
6 https://a-z-animals.com/animals/location/north-america/united-states/alaska/	3	4	-1 ▼	-25%	2021-04-13 16:00:47

Showing: URLs in current crawl Selected Cells: 1 Filter Total: 6

Auto ▼ All ▼ Export

Auto

Current

Previous

Added

Removed

Links for the selected URL from either the previous or current crawl depending on the comparison mode off the master view

From	To
https://a-z-animals.com/animals/scientific/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/scientific/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/fish/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/fish/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/endangered/critically-endangered/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/endangered/critically-endangered/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/endangered/critically-endangered/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/class/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/toadfish/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/location/ocean/northeast-pacific/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/location/ocean/northeast-pacific/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/location/ocean/northeast-pacific/	https://a-z-animals.com/animals/blobfish/
https://a-z-animals.com/animals/location/ocean/northwest-pacific/	https://a-z-animals.com/animals/blobfish/

Selected Cells: 0 Total: 34

URL Details Inlinks Outlinks Image Details Resources SERP Snippet Rendered Page View Source HTTP Headers Cookies Duplicate Details Structured Data Details

Compare Mode: Idle

Nel nostro esempio il Seo Spider mostra i dati dei collegamenti per la scansione “corrente”, come selezionato dalla finestra principale superiore. Tuttavia la finestra inferiore può essere impostata in modo indipendente per mostrare i dati della scansione “precedente”, anche se la finestra superiore è selezionata per la scansione corrente.

Potrai anche utilizzare i filtri “Added” e “Removed” ed evidenziare i relativi URL per evincere dove i link interni sono cambiati.

Auto ▼ All ▼ Export

Auto

Current

Previous

Added

Removed

From	To	Anchor Text
https://a-z-animals.com/animals/location/ocean/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/location/ocean/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/scientific/	https://a-z-animals.com/animals/blobfish/	Psychrolutes marcidus
https://a-z-animals.com/animals/scientific/	https://a-z-animals.com/animals/blobfish/	Psychrolutes marcidus
https://a-z-animals.com/animals/fish/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/fish/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/endangered/critically-endangered/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/endangered/critically-endangered/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/endangered/critically-endangered/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/class/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/toadfish/	https://a-z-animals.com/animals/blobfish/	blobfish
https://a-z-animals.com/animals/location/ocean/northeast-pacific/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/location/ocean/northeast-pacific/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/location/ocean/northeast-pacific/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/location/ocean/northwest-pacific/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/location/ocean/northwest-pacific/	https://a-z-animals.com/animals/blobfish/	Blobfish
https://a-z-animals.com/animals/location/ocean/northwest-pacific/	https://a-z-animals.com/animals/blobfish/	blobfish

Selected Cells: 0 Total: 34

URL Details Inlinks Outlinks Image Details Resources SERP Snippet Rendered Page View Source HTTP Headers Cookies Duplicate Details Structured Data Details

Compare Mode: Idle

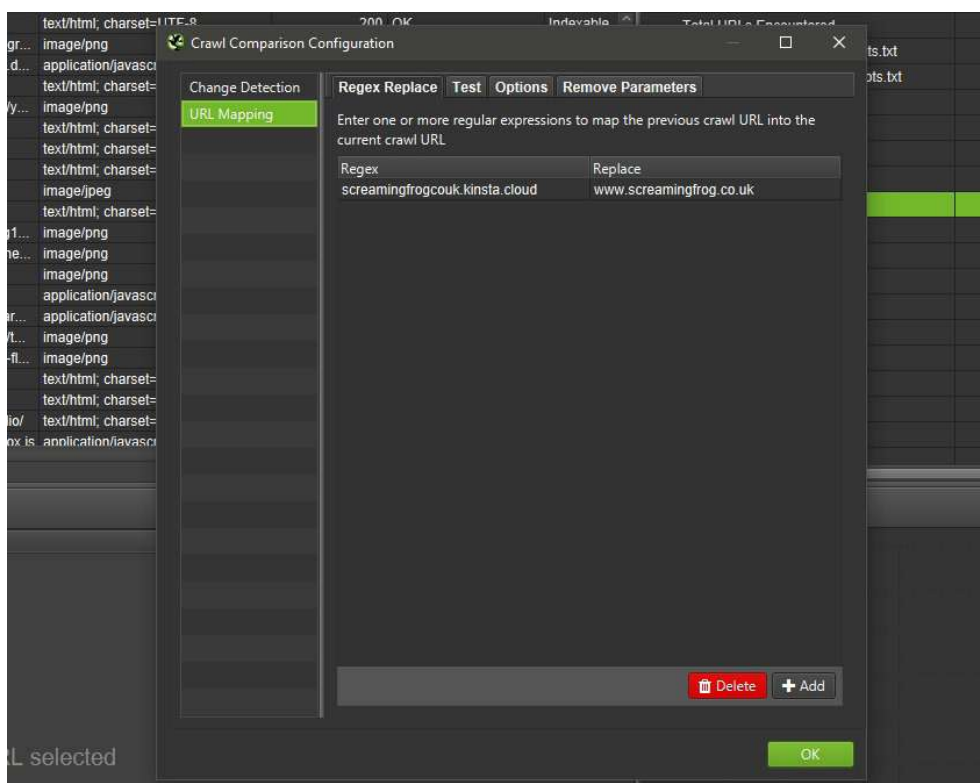
COME CONFRONTARE STAGING E ALTRE STRUTTURE DI URL CON LA MAPPATURA DEGLI URL

Finora abbiamo sempre confrontato due sottodomini con lo stesso URL ma utilizzando la funzione “URL Mapping” è possibile anche mettere a confronto scansioni fra sottodomini differenti. Un’applicazione molto comune può coinvolgere il confronto tra un sito live e uno staging, o un sito mobile e il suo equivalente desktop. Attraverso gli ultimi aggiornamenti di Screaming Frog è possibile confrontare nomi host completamente diversi, directory o modifiche più marginali.

Per confrontare due diverse strutture di URL, clicca sulla configurazione di confronto (“Config > Compare”) e su “URL Mapping” e utilizza delle regole RegEx per la corrispondenza.

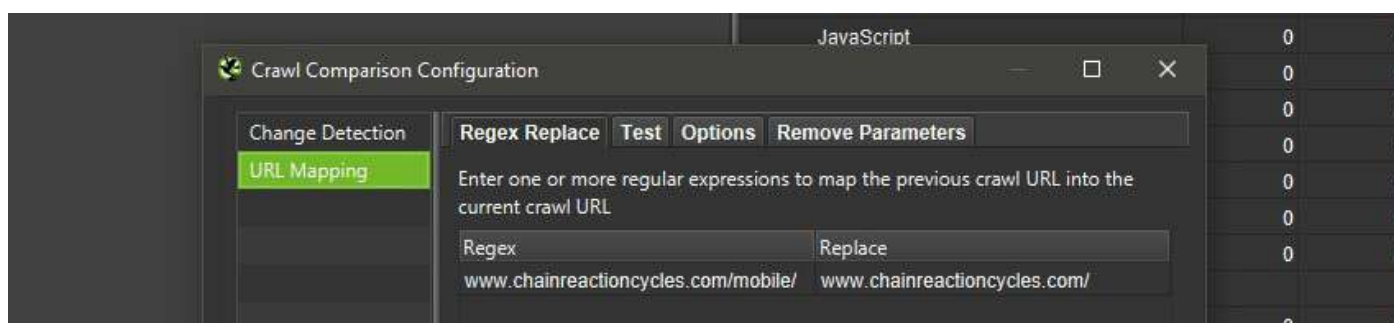
ESEMPLI:

Vediamo come Screaming Frog abbia cambiato host in Kinsta e testato il sito di staging prima della migrazione con URL Mapping.



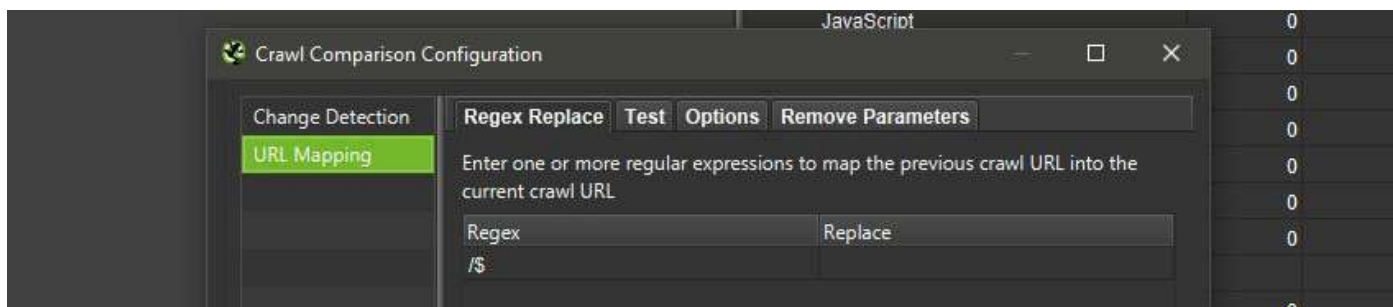
In questo caso gli URL del sito di staging su Kinsta presentano un hostname diverso e attraverso la mappatura sono stati fatti corrispondere al sito live esistente per permettere il confronto dei dati.

Un altro scenario per la mappatura degli URL potrebbe essere il confronto tra siti web mobile e desktop se non è prevista la versione responsive.



Nell'esempio precedente il sito web di Chain Reactions Cycles ha una sottocartella **/mobile/** dedicata alla versione per smartphone, quindi per confrontare il sito web mobile con quello desktop, gli URL mobili sono stati mappati all'equivalente desktop con il seguente URL Mapping.

Infine, se le due versioni da comparare differiscono solamente per lo slash finale nel sottodominio puoi usare **"/\$"** nella mappatura degli URL.

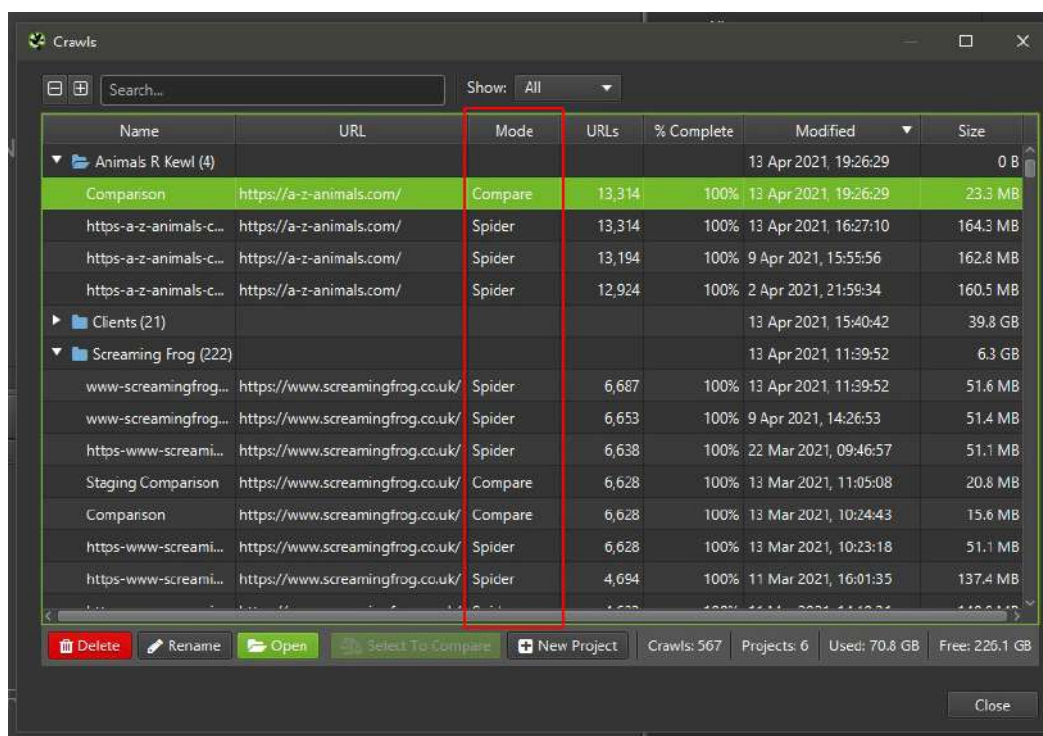


Prima di applicare l'Url Mapping puoi anche fare dei test usando la scheda dedicata per evitare di fare errori e perdere tempo.

COME VISUALIZZARE I VECCHI CONFRONTI DI CRAWL

Una volta completata la comparazione i dati saranno salvati automaticamente e potrai utilizzarli nuovamente andando su **"File > Crawls"**.

Le scansioni sono identificate con la voce **"Compare"**.



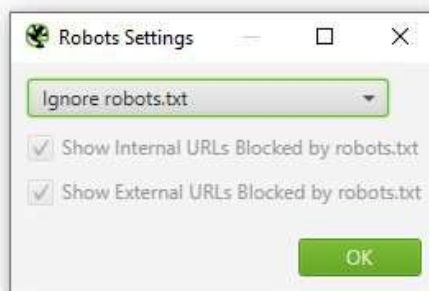
I file di confronto utilizzano i crawl attuali e precedenti memorizzati e confrontati, se uno o l'altro viene rimosso, anche il confronto non sarà più accessibile.

SITO IN STAGING: PROBLEMI E SOLUZIONI DI SCANSIONE

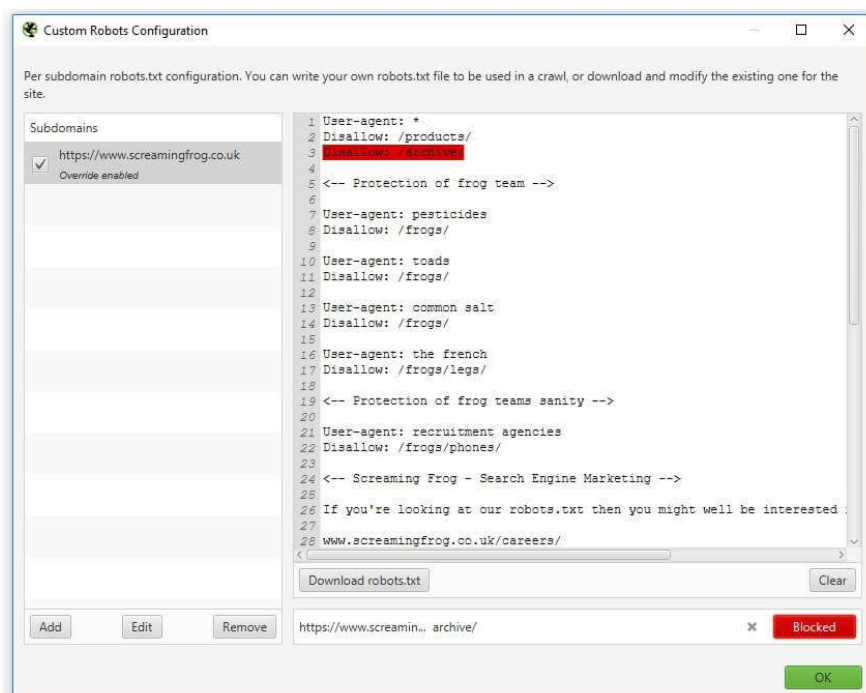
I siti web in fase di staging di solito non possono essere scansionati dai Motori di Ricerca e dai crawler. Esistendo vari metodi per impedire il crawling ci sono diversi approcci o configurazioni di Screaming Frog da attuare per ottenere una scansione ed aggirare questi limiti.

1. Robots.txt

Il caso più comune riguarda il file robots.txt che blocca la scansione. In questo caso il Seo Spider colleziona un solo URL, il record mostra il messaggio "Blocked by robots.txt" e lo stato è "Indexable". Per oltrepassare questa barriera è sufficiente impostare "Ignore Robots.txt" da Config. > Robots.txt



Un'alternativa è quella di configurare un robots.txt personalizzato tramite "Config > robots.txt > Custom" per rimuovere eventuali direttive "disallow" e aggiungere se necessario altre personalizzazioni che il Seo Spider deve seguire.



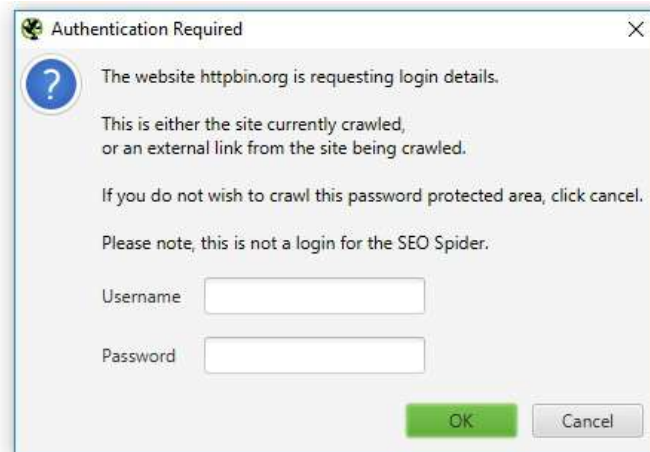
2. Autenticazione

Nel caso in cui il server richieda l'autenticazione è possibile fornire username e password al Seo Spider per scansionare il sito.

AUTENTICAZIONE DI BASE E DIGEST

L'autenticazione di base e digest viene rilevata automaticamente dal Seo Spider quando si esegue la scansione del sito web.

Eseguendo la scansione del sito in fase di sviluppo appare una finestra pop-up di autenticazione, proprio come avviene in un browser web, chiedendo un nome utente e una password.



Inserendo le corrette credenziali il Seo Spider inizia il normale crawling delle pagine del sito web in staging.

AUTENTICAZIONE WEB FORM

Se ci sono altri moduli web o aree che richiedono il login con i cookie per l'autenticazione, il Seo Spider permette agli utenti di accedere a questi moduli web all'interno del browser Chrome integrato nel tool. Per effettuare il login utilizzando l'autenticazione dei moduli web, clicca su "Configurazione > Autenticazione > Basato su moduli". Quindi clicca sul pulsante "Aggiungi", inserisci l'URL del sito su cui desideri eseguire il crawling e effettua il login dal browser.



IP ADDRESS

Alcune piattaforme di staging possono limitare l'accesso in base all'indirizzo IP.

Poiché il SEO Spider effettua il crawling locale dalla macchina da cui viene eseguito, è necessario fornire questo indirizzo IP per essere inclusi nella "whitelist".

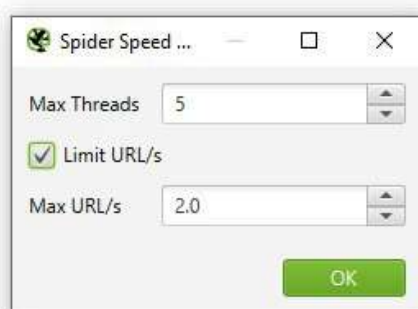
COME CONFIGURARE LE IMPOSTAZIONI PER SCANSIONARE I SITI IN FASE DI SVILUPPO

I siti in sviluppo possono rispondere alle richieste HTTP in modo diverso da quelli in un ambiente live, e spesso possono avere direttive del robot che richiedono ulteriori configurazioni nel SEO Spider.

CONFIGURARE LA VELOCITÀ DI SCANSIONE

I siti web in staging sono generalmente più lenti e non possono sopportare lo stesso carico di un server di produzione. Se noti instabilità del sito ed errori di timeout o errori del server devi ridurre la velocità di scansione dello Spider.

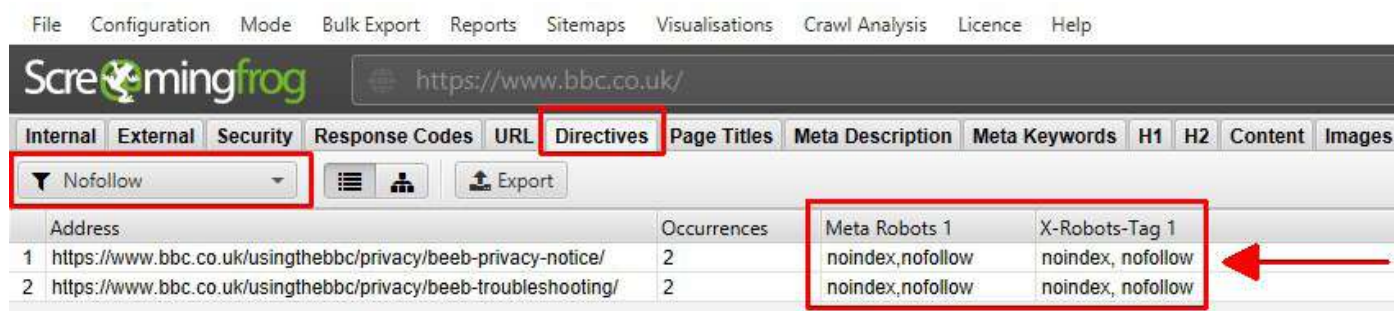
Config > Speed



Oppure gestire il tempo di timeout e il tempo di AJAX timeout nel caso di scansioni con il rendering Javascript.

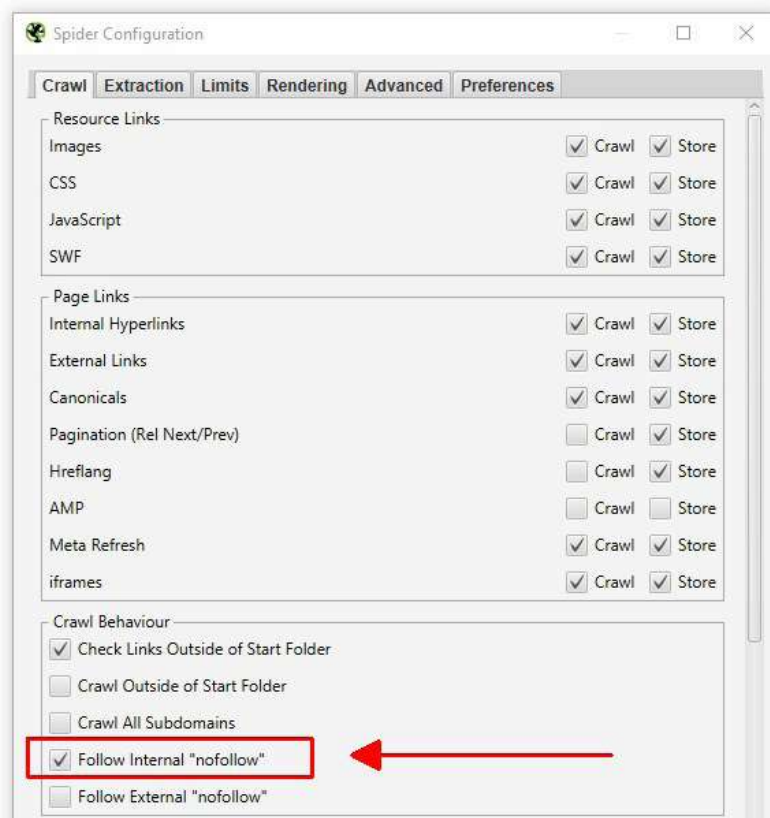
NOFOLLOW

Per proteggere eventuali intromissioni dei BOT molti siti in sviluppo potrebbero presentare un meta robots tag "nofollow" in tutto il sito, o X-Robots-Tag nell'intestazione HTTP. Un "nofollow" è una direttiva molto diversa da un noindex, e indica a un crawler di non seguire alcun outlink da una pagina.



Address	Occurrences	Meta Robots 1	X-Robots-Tag 1
1 https://www.bbc.co.uk/usingthebbc/privacy/beeb-privacy-notice/	2	noindex, nofollow	noindex, nofollow
2 https://www.bbc.co.uk/usingthebbc/privacy/beeb-troubleshooting/	2	noindex, nofollow	noindex, nofollow

Nel caso presenti la condizione dell'esempio vai a "Config > Spider" e abilita "Follow Internal Nofollow" per scansionare gli outlink da queste pagine ed ottenere un Audit completo.



NO INDEX

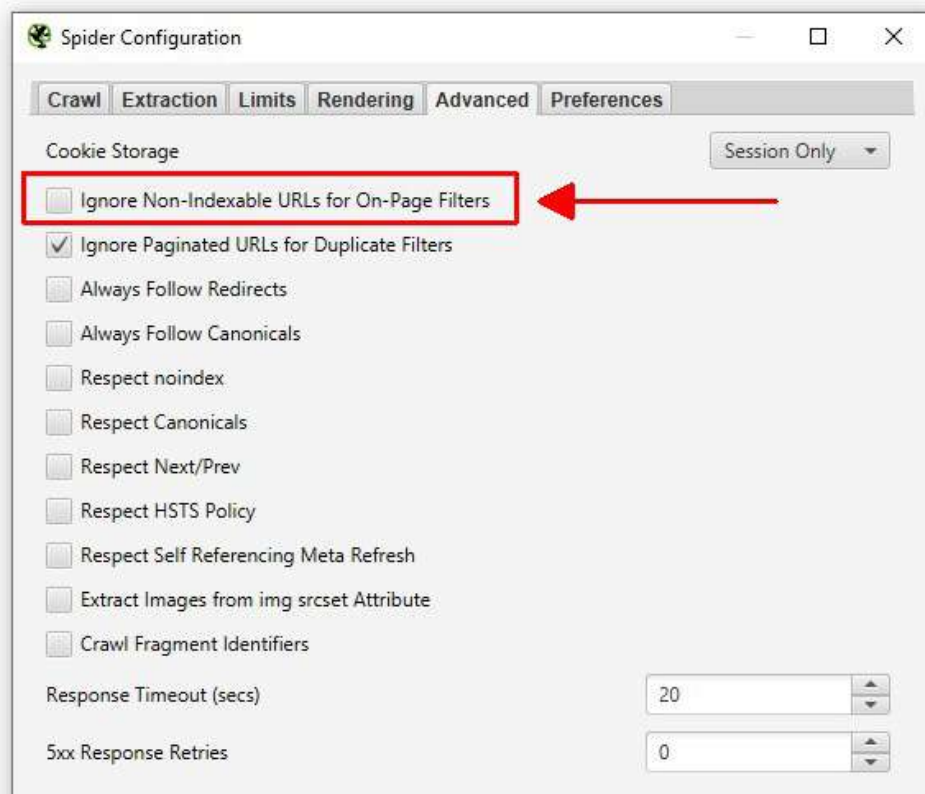
Un altro tag che puoi incontrare in un sito in staging è il noindex, che permette la scansione del crawler ma non la sua indicizzazione. L'uso di noindex può essere visto sotto la scheda "Directives" e il filtro "noindex". Quello che ti consiglio è di controllare che questo tag venga rimosso immediatamente alla messa online del sito web, troppe volte si vedono siti con il file robots.txt o il tag noindex ancora attivi anche se già pubblicati online.



Oltre agli aspetti legati alla messa online dovresti considerare che il Seo Spider scansiona le pagine con un noindex (come un normale BOT), ma vede queste pagine come "non indicizzabili". Questo significa che per impostazione prefinita lo Spider non considera le relative metriche nei filtri e non ti mette a disposizione elementi come i contenuti duplicati, o titoli di pagina mancanti, meta descrizioni ecc. quindi l'analisi potrebbe essere molto parziale.

La soluzione è quella di disabilitare "Ignore Non-Indexable URLs for On-Page Filters " per una scansione completa del sito web di staging.

Config > Spider > Advanced

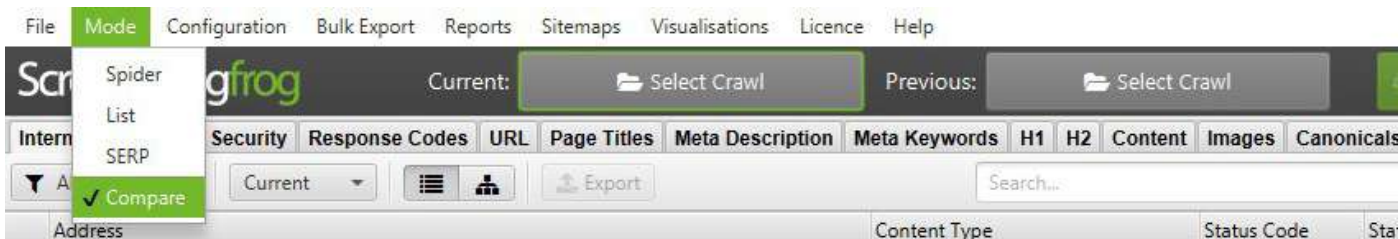


NONE

Troppe volte la direttiva “None” viene sottovalutata e male interpretata pensando che significhi la mancanza di direttive invece la stessa è equivalente a “noindex, nofollow”! Se fosse questo il caso devi seguire gli stessi passi descritti precedentemente.

CONFRONTA SITO IN STAGING VS LIVE

Grazie alle ultime release di Screaming Frog il Seo Spider puoi confrontare due crawls per vedere le differenze attraverso la funzione “URL Mapping”, che permette di confrontare due diverse strutture di URL, come un sito web di staging contro un ambiente di produzione o live.



Per confrontare lo staging con il sito web live, ti basta cliccare su “Mode > Compare” e selezionare i due crawls.



Per maggiori dettagli riguardanti l'URL Mapping ti consiglio di seguire l'articolo dedicato nella guida.

WEB SCRAPING & CUSTOM EXTRACTION

Vediamo come poter utilizzare Screaming Frog per fare Web Scraping utilizzando la funzione Custom Extraction.

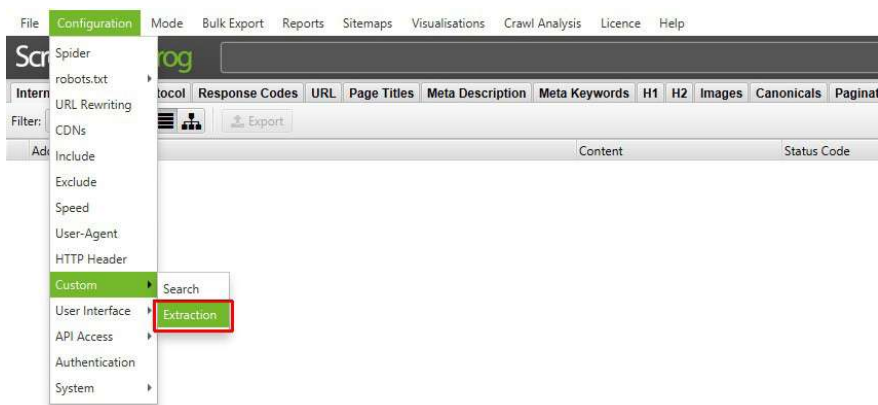
Attraverso questa funzionalità sei in grado di recuperare qualsiasi dato HTML da una pagina web mediante CSSPath, XPath e RegEX.

L'estrazione viene eseguita sull'HTML statico degli URL scansionati dal SEO Spider che rispondono con uno status code 200 'OK'.

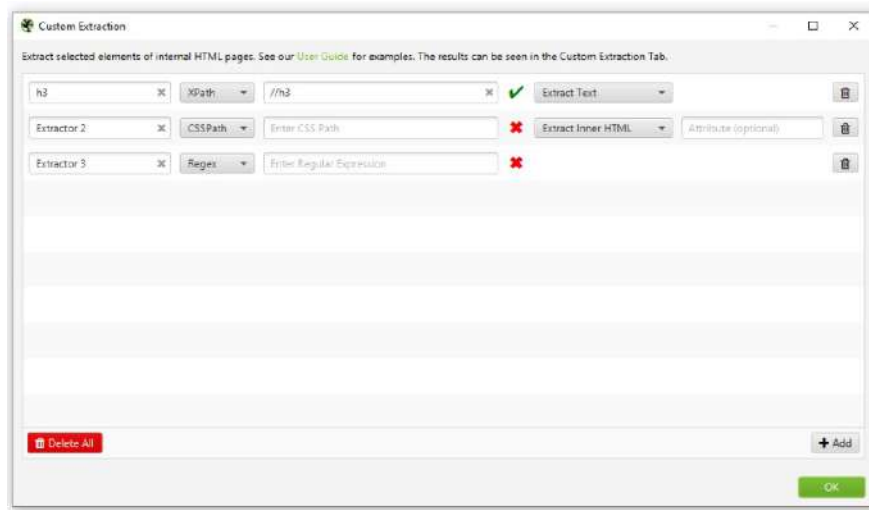
Se vuoi fare delle estrazioni da dati renderizzati è possibile abilitare la modalità "Javascript Rendered".

1. Configurazione della Custom Extraction

Per impostare la tua ricerca personalizzata vai su Configuration > Custom > Extraction.



Attraverso questa funzione sei in grado di impostare fino a 100 richieste di estrazione dati personalizzate.



2. Seleziona il percorso CSS, XPath o Regex da utilizzare per lo scraping

Il Seo Spider offre tre opportunità di scraping dei dati nei siti web:

- **XPath:** attraverso XPath sei in grado di selezionare i nodi da un documento dove eseguire un'interrogazione usando i selettori XPath, compresi gli attributi.
- **CSS Path:** questa opzione è la più veloce fra i metodi menzionati e permette di fare scraping usando i selettori CSS Path.
- **Regex:** questa interrogazione di dati utilizza le espressioni regolari RegEx e viene consigliata per usi avanzati come lo scraping di commenti HTML o JavaScript in linea.

Optando per XPath o CSS Path per interrogare l'HTML, puoi scegliere tra diversi filtri del Seo Spider:

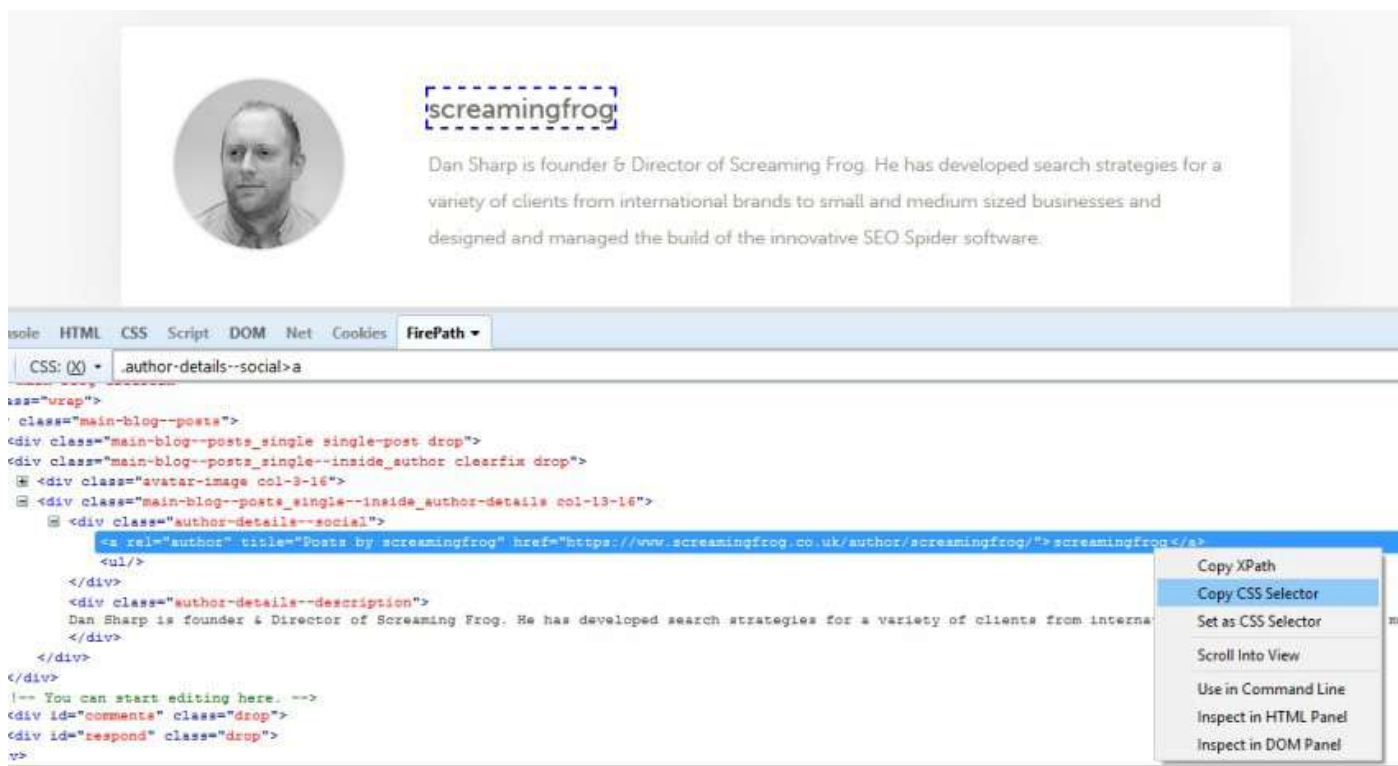
- **Extract HTML Elements:** raccoglie le informazioni dell'elemento selezionato e tutto il suo contenuto HTML interno.
- **Extract Inner HTML:** raccoglie il contenuto HTML interno dell'elemento selezionato. Se, ad esempio l'elemento selezionato contiene altri elementi HTML, anche questi saranno inclusi.
- **Extract Text:** raccoglie il contenuto testuale dell'elemento selezionato e del suo sotto elemento.
- **Function Value:** restituisce il numero complessivo dell'elemento richiesto, ad esempio se cerchi quanti h3 sono presenti in una pagina puoi utilizzare "count(//h3)".

3. Inserisci la tua sintassi

Una volta scelta la modalità di scraping non ti resta che definire la sintassi di estrazione. Per trovare il relativo percorso CSS o Xpath puoi aprire semplicemente la pagina web in Chrome e ispezionare l'elemento desiderato, quindi fare clic con il pulsante destro del mouse e copiare il relativo percorso di selezione fornito.

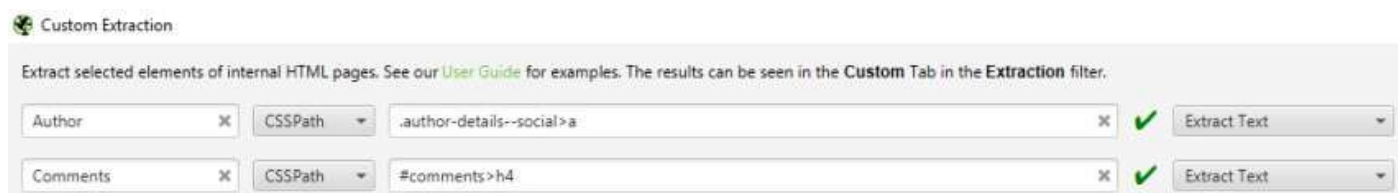
ESEMPIO:

Prendiamo in esame il blog di Screaming Frog.



Apri qualsiasi post del blog in Chrome, clicca con il tasto destro e "ispeziona l'elemento" sul nome dell'autore.

Clicca con il tasto destro del mouse sulla linea HTML pertinente (con il nome dell'autore), copia il percorso CSS o XPath pertinente e incollalo nel rispettivo campo del Seo Spider.



Se la sintassi inserita è valida (.author-details-social>a) vedrai una spunta verde accanto al tuo input, altrimenti sarà presente un warning con una croce rossa che identifica che la sintassi non è considerata corretta.

Completata questa operazione ti basta cliccare il tasto “ok” ed avviare il crawl.

Per approfondire sui selettori CSS e su XPath ti consiglio di seguire [w3schools](https://www.w3schools.com/).

4. Scansiona il sito web

Inserita la sintassi e convalidata non ti resta che eseguire la scansione del sito web per iniziare lo scraping.

5. Visualizza i dati dello scraping nella scheda “Custom Extraction”

I dati del web scraping sono a tua disposizione in tempo reale durante la scansione, nella scheda ‘Custom Extraction’, e nella tab “Internal”.

File Configuration Mode Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help

Screamingfrog

https://www.screamingfrog.co.uk/blog/

Internal External Protocol Response Codes URL Page Titles Meta Description Meta Keywords H1 H2 Custom Extraction Images Canonical

Filter: All

Export

	Address	Status Code	Status	Author 1	Comments 1
4	https://www.screamingfrog.co.uk/seo-spider-7-0/	200	OK	screamingfrog	85 Comments
5	https://www.screamingfrog.co.uk/seo-spider-companion-tools/	200	OK	Richard Lawther	7 Comments
6	https://www.screamingfrog.co.uk/log-file-analyser-2-0/	200	OK	screamingfrog	7 Comments
7	https://www.screamingfrog.co.uk/find-killer-content/	200	OK	Daniel Cartland	7 Comments
8	https://www.screamingfrog.co.uk/seo-spider-10/	200	OK	screamingfrog	72 Comments
9	https://www.screamingfrog.co.uk/the-beginners-guide-to-seo-competitor...	200	OK	Freddy Barnes	6 Comments
10	https://www.screamingfrog.co.uk/learn-to-crawl/	200	OK	Oliver Brett	6 Comments
11	https://www.screamingfrog.co.uk/seo-spider-9-0/	200	OK	screamingfrog	62 Comments
12	https://www.screamingfrog.co.uk/seo-spider-12/	200	OK	screamingfrog	60 Comments
13	https://www.screamingfrog.co.uk/chasing-for-a-link/	200	OK	Mark Porter	5 Comments
14	https://www.screamingfrog.co.uk/brightonseo-crawling-clinic/	200	OK		5 Comments
15	https://www.screamingfrog.co.uk/seo-spider-6-0/	200	OK	screamingfrog	57 Comments

Nel nostro esempio è stata avviata una scansione completa di un sito web ma se vuoi fare scraping da un elenco specifico di URL puoi decidere di utilizzare la modalità di scansione “List”.

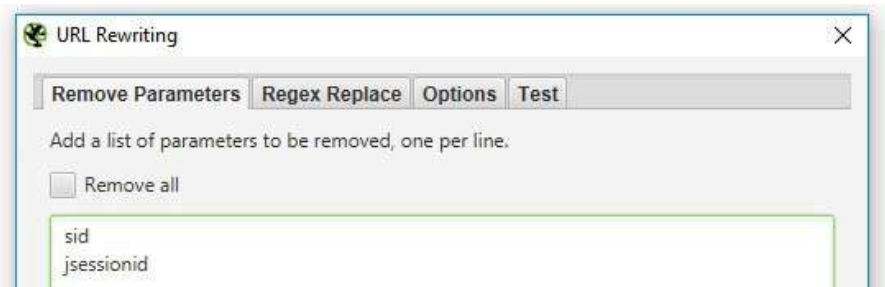
I campi di applicazione sono infiniti e dipendono dalla tipologia di analisi in essere, questa funzionalità può rivelarsi molto utile ad esempio per collezionare l’ID di Analytics o GTM, i social meta tags, i valori degli attributi Hreflang, i prezzi dei prodotti di un ecommerce, alcuni prezzi scontati etc.

REWRITE URL

La funzione “Rewrite URL” ti permette di riscrivere gli URL in pochi secondi ed è estremamente utile per la “rimozione dei parametri” e ID di sessione.

Questa funzione viene applicata a tutti gli URL scoperti durante il crawl tranne l'URL inserito come pagina di “inizio scansione” nella modalità “Spider” o all'elenco di URL aggiunti dall'utente nella modalità “List”.

Ad esempio se un sito web presenta ID di sessione che fanno apparire gli URL come questo **“example.com/?sid=6578”** è sufficiente aggiungere “sid” (senza gli apostrofi) all'interno del campo “Remove Parameters” della funzione “URL Rewriting”.

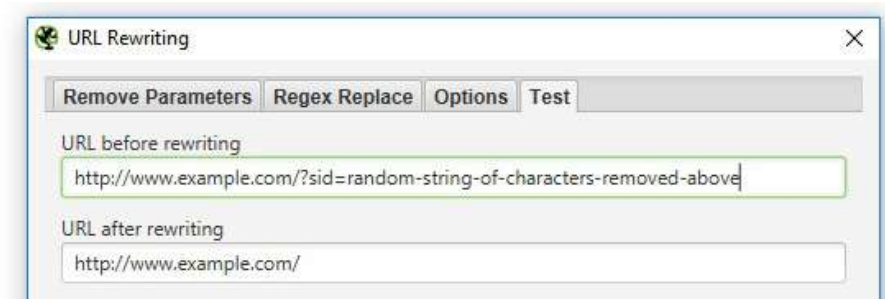


Un altro caso molto comune per siti che utilizzano campagne a pagamento PPC, newsletter o altro è rappresentato dall'utilizzo di parametri per il tracciamento come:

```
utm_source
utm_medium
utm_campaign
```

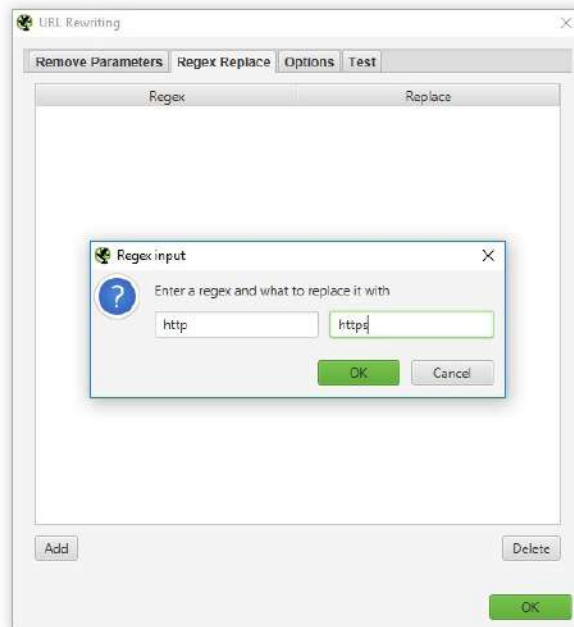
Anche in questo caso è sufficiente inserirli nel popup “URL Rewriting” per ottenere gli URL puliti dai parametri.

Identificati i parametri da rimuovere ti consiglio, prima di avviare la scansione, di utilizzare l'opzione “test” per verificare che tutto funzioni correttamente in base all'input inserito.



Oltre alla funzione sopra citata puoi optare per la funzione avanzata “Regex Replace”.

Questa funzione avanzata viene eseguita su ogni URL trovato durante un crawl o in modalità elenco. In modo molto semplice attraverso la sintassi RegEX sei in grado di sostituire ogni sottostringa di un URL che corrisponde all'input con la stringa prescelta. Anche questa opzione può essere testata nella scheda “Test” della finestra di configurazione “URL Rewriting”.



Attraverso la funzione Regex i campi di Applicazione sono infiniti, vediamo alcuni:

1. Cambiare il protocollo degli URL da HTTP a HTTPS

Regex: http
Replace: https

2. Modificare tutti gli URL da dominio.com in dominio.it

Regex: .com
Replace: .it

3. Uniformare tutti i link contenenti page=number ad un numero fisso predefinito, ad esempio

www.example.com/page.php?page=1
www.example.com/page.php?page=2
www.example.com/page.php?page=3
www.example.com/page.php?page=4

Regex: page=\d+
Replace: page=1

4. Rimuovere il dominio www. da qualsiasi URL usando un 'Replace' vuoto.

Regex: www.
Replace:

5. Eliminare tutti i parametri

Regex: \?.*
Replace:

6. Cambiare i link solo per i sottodomini di example.com da HTTP a HTTPS

Regex: http://(.example.com)
Replace: https://\$1

7. Rimuovere qualsiasi elemento presente in modalità di rendering JavaScript

```
Regex: #.*  
Replace:
```

8. Aggiungere alcuni parametri agli URL

```
Regex: $  
Replace: ?parameter=value
```

Questo aggiungerà '?parameter=value' alla fine di qualsiasi URL incontrato

Nelle situazioni in cui il sito ha già dei parametri, questo richiede espressioni più complicate perché il parametro sia aggiunto correttamente:

```
Regex: (.*?\?.*)  
Replace: $1&parametro=valore  
Regex: (^((?!?).)*$)  
Replace: $1?parametro=valore
```

Un'ultima opportunità di "Url Rewriting" è la possibilità di rimuovere dalla scansione tutti gli URL duplicati da caratteri maiuscoli utilizzando "Lowercase discovered URLs".

COME TROVARE LE PAGINE WEB CON POCO CONTENUTO

Partendo dalla considerazione che il contenuto è uno dei principali elementi di Ranking potrebbe essere molto utile definire quali pagine presentano un contenuto molto ridotto.

1. Avvia e concludi il crawl.
2. Utilizza il filtro HTML della schermata principale.
3. Clicca sulla scheda "Internal".
4. Spostati sulla colonna "Word Count".
5. Ordina la colonna.
6. Clicca su Export ed esporta il tuo file .CSV.

COME TROVARE LE PAGINE CON TEMPI LUNGHI DI CARICAMENTO

Per un controllo rapido delle performance delle singole pagine HTML:

4. Esegui una scansione del sito web.
5. Clicca sulla scheda "Response Code".
6. Ordina i dati della colonna Response Time.
7. Esporta i dati.

COME TROVARE SPAM E MALWARE SUL SITO WEB

Con la crescita esponenziale nell'utilizzo di Wordpress, Joomla e Drupal sono aumentati i casi di Malware e Spam che colpiscono sempre più frequentemente le vulnerabilità dei CMS. Per controllare se anche il tuo sito web è stato vittima di hackeraggio:

1. Identifica lo snippet di codice del malware o spam o semplicemente definisci delle query che potrebbero essere state usate per lo spam (es. viagra, sex, gambling, casino etc.).
2. Clicca Custom dal menu Configurazioni.
3. Utilizza i filtri personalizzati del Seo Spider e avvia la scansione.
4. Visualizza i risultati nella tab "Custom" della finestra principale.
5. Analizza gli URL che contengono questi elementi.

COME VERIFICARE LA PRESENZA DI SCRIPT E ID

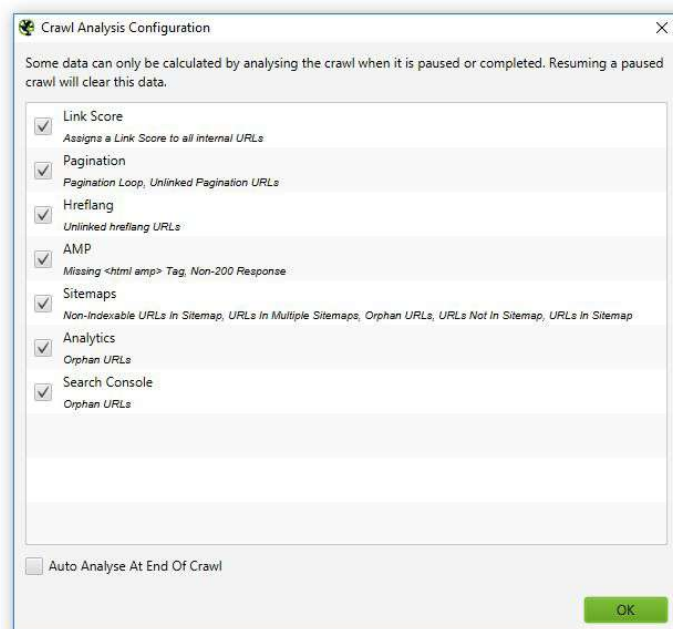
Vediamo come controllare in modo semplice se l'installazione di Google Analytics, GTM o altro è stata implementata correttamente ed è presente in ogni pagina del sito web.

1. Apri i filtri personalizzati dalle configurazioni di Screaming Frog.
2. Inserisci gli ID o lo script (Source Code) dell'elemento da tracciare utilizzando l'opzione "Contains" o "Does Not Contains".
3. Esegui la scansione.
4. Visualizza i risultati nella scheda "Custom".

ANALISI DEI COMPETITORS USANDO IL LINK SCORE

Per conoscere quali siano le pagine più importanti dei tuoi competitors puoi utilizzare la funzione "Link Metrics" che, identificando il numero di link interni ricevuti dalle altre pagine del sito web ne determina una graduatoria.

La funzionalità Link Metrics non viene popolata con i dati alla fine della scansione in automatico ma richiede la "Crawl Analysis".



ANALISI DEI COMPETITORS E DEGLI “ANCHOR TEXT” DEI LINK INTERNI

Un aspetto molto interessante da mettere sempre nelle tue analisi dei competitors è quello di considerare le parole chiave utilizzate come “Anchor Text” per i link interni. In questo modo riesci ad avere una visione d’insieme di quali siano le query su cui i competitors sta puntando.

1. Esegui il crawling del sito web del competitors.
2. Esporta gli Anchor Text da Bulk Export > Links > All Anchor Text

SCRAPING DEI RISULTATI SU GOOGLE PER IDENTIFICARE I TOP 10 COMPETITORS

Per un’analisi veloce dei competitors su delle specifiche query puoi decidere di interrogare direttamente Google per capire quali siano i top 10 per una specifica query.

ESEMPIO

Se ad esempio vuoi cercare quali siano i concorrenti per la parola chiave “**negozio per animali**” puoi fare una ricerca sul Motore di Ricerca. L’interrogazione ti restituisce un URL come il seguente:

“q=negozio+per+animali”

<https://www.google.com/search?q=negozio+per+animali&oq=negozio+per+animali&aqs=chrome..69i57j0i402l2j0i512l7.6084j0j15&sourceid=chrome&ie=UTF-8>

Questa informazione la puoi utilizzare con Screaming Frog per fare attività di scraping:

- Configura lo user Agent e impostalo su Chrome.
- Configura il file Robots.txt scegliendo l’opzione “Ignore Robots.txt”.
- Inserisci su “Include” l’URL da scansionare (azione che dovrebbe permetterti di non subire penalizzazioni dell’IP).
- Inserisci l’URL nel Seo Spider e avvia la scansione.
- Clicca sull’unico URL scansionato e spostati nella finestra in basso. Nella tab Outlinks troverai, oltre a Google, i primi 10 concorrenti nella colonna “To” e relativo “Meta Title” nella colonna “Anchor Text” dello Spider.

CONCLUSIONI

Anche questo viaggio sta volgendo al termine.

Il viaggio ci cambia, perfeziona e arricchisce e spero che questa lettura ti abbia dato qualche spunto per riflettere, metterti in gioco e sperimentare la tua nuova conoscenza di questo mondo. Ti voglio lasciare con una metafora:

“
Il vero viaggio di scoperta non consiste nel cercare nuove terre, ma nell'avere nuovi occhi.

Marcel Proust

Testa, sperimenta, osserva e ricomincia, sii curioso e indiscreto.

Il settore della Seo è in continua evoluzione e solo un vero esploratore riesce a domare il mare in tempesta e giungere sano e salvo sull'isola del tesoro.

Finalmente posso darti il benvenuto nel **Club di Screaming Frog!**

.

CREDITS:

Tutte le immagini e relativo copyright sono di proprietà di Screaming Frog Srl

Logo: Ally-Strategic Design

Supporto Grafico: Federica Scinetti

Supporto Sito Web: Matteo Tognolini

Pubblicazione: Novembre 2021

Autore: Raffaele Visintin

LICENZA DELLA GUIDA:

@ 2021 Questa guida è coperta da copyright e la distribuzione o adattamento intero o parziale del contenuto richiede la dichiarazione della fonte secondo le seguenti regole:.

Credits: **Guida KILL SEO - Volume 1**

con link "dofollow" al sito www.screamingfrog.club