

Technical SEO for Coursology.io

Current Website Performance

The website sitemap contains 156 nested sitemaps with 5000 URLs in each.

It makes approximately $156 * 5000 = 780\ 000$ URLs

```
▼<sitemap>
  <loc>https://coursology.io/sitemaps/sitemap_152.xml</loc>
</sitemap>
▼<sitemap>
  <loc>https://coursology.io/sitemaps/sitemap_153.xml</loc>
</sitemap>
▼<sitemap>
  <loc>https://coursology.io/sitemaps/sitemap_154.xml</loc>
</sitemap>
▼<sitemap>
  <loc>https://coursology.io/sitemaps/sitemap_155.xml</loc>
</sitemap>
▼<sitemap>
  <loc>https://coursology.io/sitemaps/sitemap_156.xml</loc>
</sitemap>
</sitemapindex>
```

The Screaming Frog SEO Spider has spent more than 10 hours crawling the website. The final score was about 95 000 pages after crawling 32% of the pages that makes approximately 185 000 pages totally. This estimation is not precise, we can just expect the number of pages to be not less than 185K pages.

The screenshot shows the Screaming Frog SEO Spider interface. At the top, there are buttons for 'Resume', 'Clear', and a green button, with a progress indicator 'Crawl 32%'. The main interface is divided into several sections: 'Overview', 'Issues', 'Site Structure', and 'Response'. The 'Overview' section shows a table with the following data:

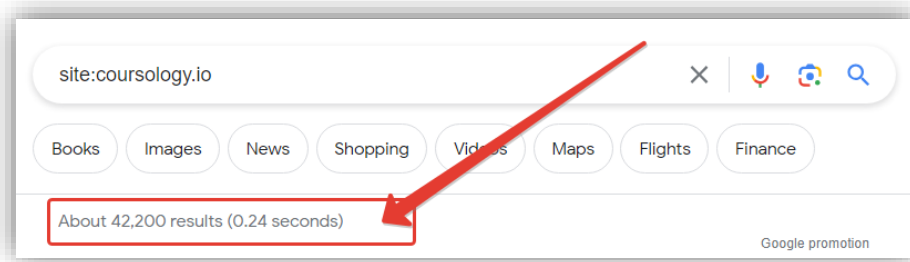
	URLs	% of T...
Total Internal URLs	94 901	99,99%
Total External URLs	8	0,01%
Total Internal Indexab...	94 893	99,99%
Total Internal Non-Ind...	8	0,01%

The 'Crawl Data' section is expanded to show 'Internal' data:

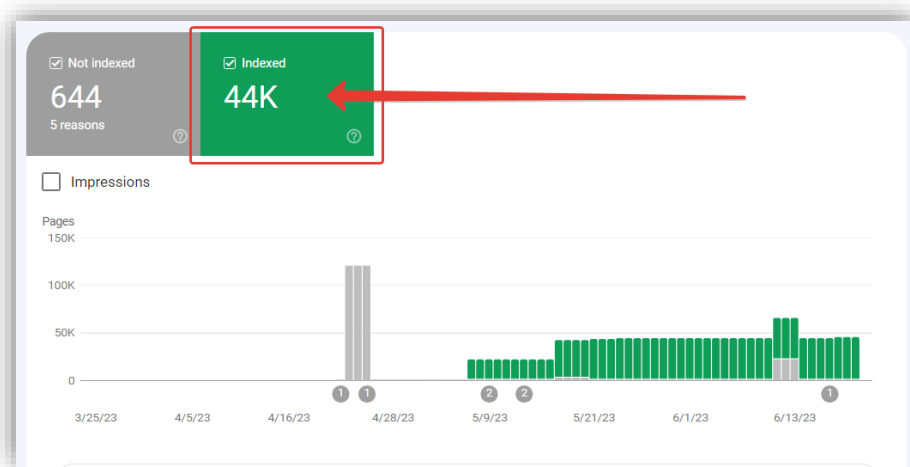
	URLs	% of T...
All	94 901	100%
HTML	94 823	99,92%
JavaScript	5	0,01%

A red arrow points from the green button at the top to the 'HTML' row in the 'Crawl Data' table, which is also highlighted with a red box.

Google has as little as 42K pages in the index.



Google Search Console has 44K indexed pages that correlates with the search results.



Conclusions:

- The website coursology.io has certain issues with indexing
- Total number of pages on the website is about 780K, the indexed part is 42K that makes about 5.3%
- Almost all pages that Google has crawled are indexed well.
- The main efforts on the current stage should be aimed to make Google crawl the rest of pages. The ratio of indexing the crawled pages according to the GSC is about 99.98%

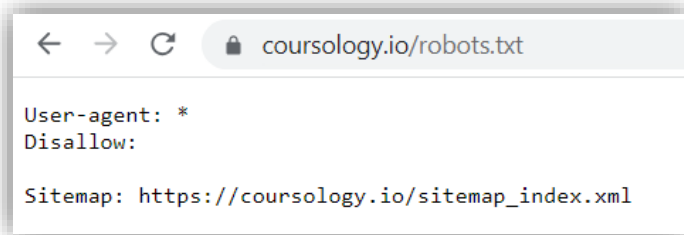
The Current Website Performance

The main mirror of the website is [https non-www](https://coursology.io).

Redirect rules from the [http-](http://coursology.io) and [www-](http://www.coursology.io)alternatives to the main mirror are set up correctly.

The robots.txt file is set up correctly and is available on the default address, [/robots.txt](https://coursology.io/robots.txt).

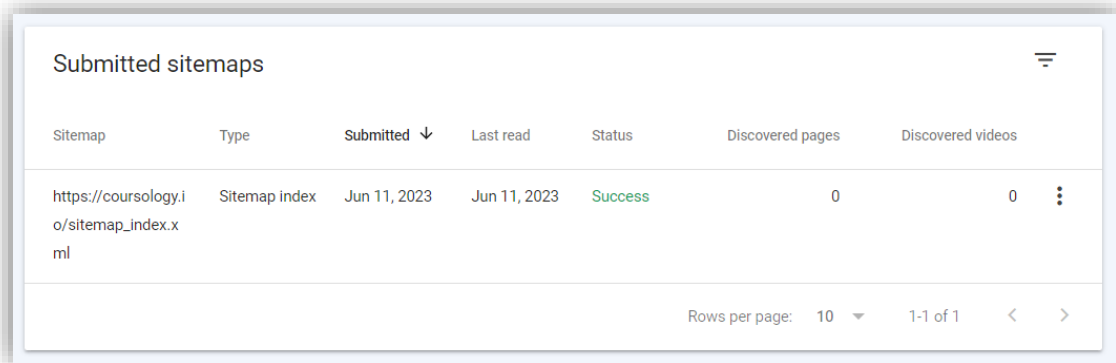
It has a Sitemap directive with the link following to the valid sitemap address.



```
← → ↻ coursology.io/robots.txt  
User-agent: *  
Disallow:  
Sitemap: https://coursology.io/sitemap\_index.xml
```

Default sitemap address <https://coursology.io/sitemap.xml> doesn't work.

The best SEO practice is to redirect the default sitemap address to the custom address, the [/sitemap.xml](https://coursology.io/sitemap.xml) to the [/sitemap_index.xml](https://coursology.io/sitemap_index.xml)



Sitemap	Type	Submitted ↓	Last read	Status	Discovered pages	Discovered videos	
https://coursology.io/sitemap_index.xml	Sitemap index	Jun 11, 2023	Jun 11, 2023	Success	0	0	⋮

Rows per page: 10 1-1 of 1 < >

Google Search Console has indexed the main [sitemap.xml](https://coursology.io/sitemap.xml) file but it doesn't display any discovered pages in it.

All nested sitemap files are unavailable

Sitemaps read				
Sitemap	Last read	Status	Discovered URLs	Discovered videos
https://coursology.io/sitemaps/sitemap_0.xml		Couldn't fetch	0	0
https://coursology.io/sitemaps/sitemap_1.xml		Couldn't fetch	0	0
https://coursology.io/sitemaps/sitemap_10.xml		Couldn't fetch	0	0
https://coursology.io/sitemaps/sitemap_14.xml		Couldn't fetch	0	0
https://coursology.io/sitemaps/sitemap_16.xml		Couldn't fetch	0	0
https://coursology.io/sitemaps/sitemap_17.xml		Couldn't fetch	0	0
https://coursology.io/sitemaps/sitemap_2.xml		Couldn't fetch	0	0
https://coursology.io/sitemaps/sitemap_23.xml		Couldn't fetch	0	0

The scraped file by Screaming Frog with the internal links has a length of about 1.3M rows (it contains only about 32% of the total internal links). Its size is about 400M, I won't share it.

Its analysis has shown that internal links on the website pages have the following nature:

- Links like <https://pagead2.googlesyndication.com/pagead/js/adsbygoogle.js?client=ca-pub-9876122130467806> that are related to the Google Ads service (5 links on every page)
- Links like [/_next/static/media](#), [/_next/static/chunks](#), [/_next/static/css](#) (totally 11 on every page). I am not 100% sure in their role on the website but they are surely not used for the internal linking
- And links from the sitemap files following to the current html pages.

This must be the main reason of the poor website indexing. Internal pages don't have internal links following from the homepage to them and following across them.

Thus, the Google crawlers don't have a chance to reach every internal page except with use of the sitemap links. As long as we now have issues with indexing the sitemap files, the entire website indexing is at risk.

The other reasons that support this assumption are the following:

- The Screaming Frog couldn't crawl the website until reading the sitemap files was allowed.
- The orphaned pages report (pages without internal links followed to them) has the same length as the list of crawled pages and includes the sitemap pages.

How to solve problems with indexing

Step 1. To create a html sitemap

1. Create a simple html page containing a list of titles of internal /search/ pages with links following to them. As soon as this list will have a huge length, it can be fractured into parts using pagination. The length of every page is up to developer.
2. Set up a link from the homepage following to this html sitemap, footer looks like a good place for this link.

After implementing this simple step Google will shortly index this html sitemap and all links on it.

Step 2. Update the current xml sitemaps

1. The number of URLs in any sitemap could be increased up to 50 000 rows (this is a limit for the number of URLs in each sitemap). Though Google does not have any limits on the nested sitemap files, number of sitemaps will get 10 times less that may make their indexing easier for Google.
2. Set up the link from the /sitemap.xml to the sitemap_index.xml
3. Add these sitemaps to the Google Search Console (I am short of credentials)

Step 3. Update the internal /search/ pages

Every internal /search/ page may contain links to the other /search/ pages at the bottom of the page just above the footer. This will surely help website indexing, even 5 – 10 random links will make crawlers index the website every time they visit the website.

It shouldn't break the page template because there is a module for user comments on every page, that will take significant place on the page then users start to write comments. So extending the length of this page with some additional rows with relevant/random questions shouldn't be crucial.

Some other ideas:

The image shows a screenshot of a quiz interface with three questions. Each question is followed by a yellow 'View Answer >' button and a red arrow pointing to the button. The questions are:

- What** is the process of analyzing all inputs and outputs through the entire life of a product to determine the total environmental impact? Multiple choice question. Ecocentric management Sustainable analysis Environmental analysis Life-cycle analysis
- What** two policies do marketing managers typically choose between when deciding how to price their goods? (Check all that apply.) Multiple select question. One-price policy Flexible-price policy Status quo policy Zero-price policy
- What** was the nineteenth-century musical movement that emphasized the distinctive cultural and political themes of the composer's country? Multiple choice question. nationalism federalism republicanism regionalism

- The “View answer” buttons can be removed; links can be applied directly to the question texts. It would be good from the SEO point of view (relevant link anchors) and from the usability point of view (the length of the search results page will be less).
- A “Quick learn” text could be also displayed on the answer page. This text will be also indexed and will increase the page word count and relevancy.

The latest Result

