

How To Use XML Sitemaps To Boost SEO via @sejournal, @jes_scholz

What was considered best practice yesterday does not hold true today and this is especially relevant when it comes to XML sitemaps, which are almost as old as SEO itself.

The problem is, it's time-consuming to sort valuable advice from all the misinformation on forums and social media about how to optimize XML sitemaps

So, while most of us recognize the importance of submitting sitemaps to Google Search Console and Bing Webmaster Tools, as well as in the robots.txt file – for faster content discovery and refresh, more efficient crawling of SEO-relevant pages, and valuable indexing reporting to identify SEO issues – the finer details of implementing sitemaps to improve SEO performance may be missed.

Let's clear up the confusion and dive into the current best practices for sitemap optimization.

In this article, we cover:

- What is an XML sitemap?
- How to create a sitemap.
- Valid XML sitemap format.
- Types of sitemaps.
- Optimization of XML sitemaps.
- XML sitemap best practice checklist.

What Is An XML Sitemap?

An XML sitemap is a file that lists all of your website's URLs.

It acts as a roadmap to tell the crawlers of indexing platforms (like search engines, but also large language models (LLMs)) what content is available and how to reach it.

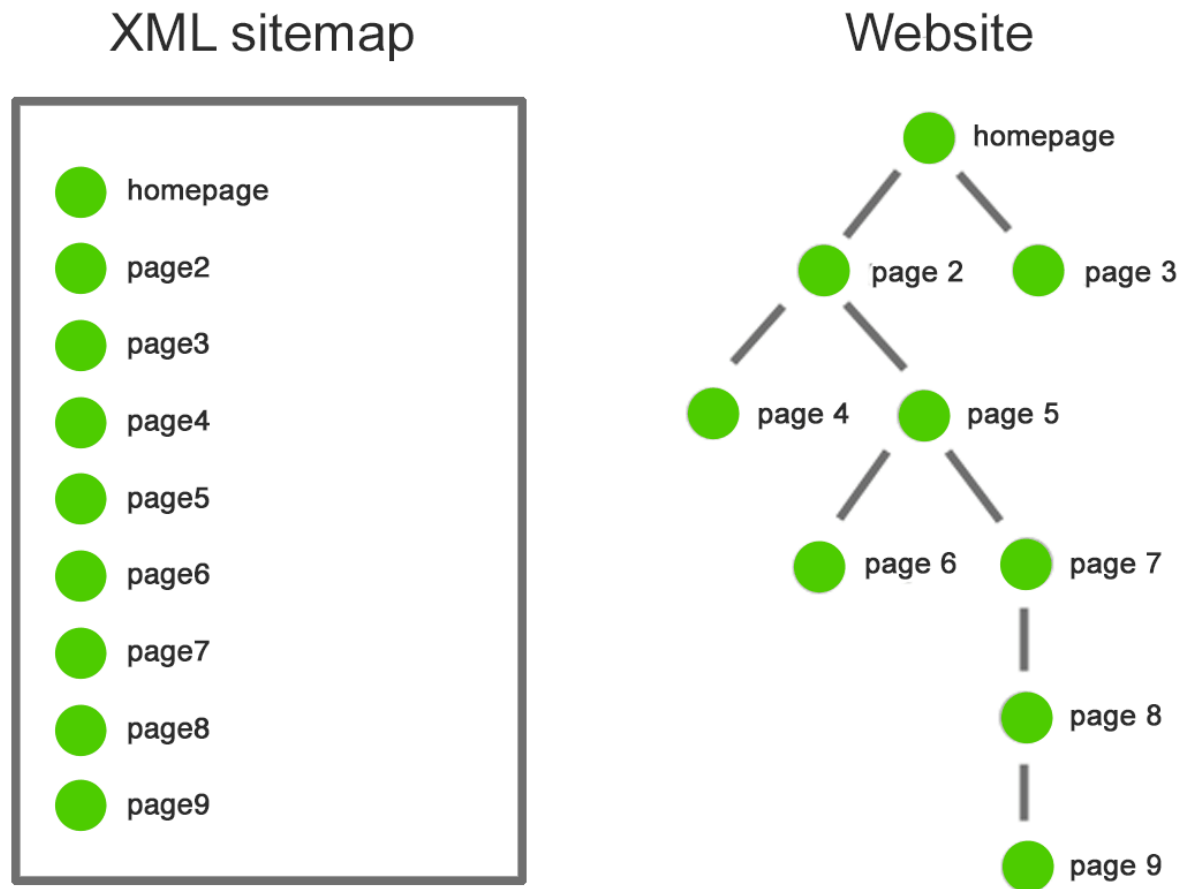


Image from author, February 2025

In the example above, a search engine will find all nine pages in a sitemap with one visit to the XML sitemap file.

On the website, it will have to jump through five internal links on five pages to find page 9.

This ability of XML sitemaps to assist crawlers in faster indexing is especially important for websites that:

- Have thousands of pages and/or a deep website architecture.
- Frequently add new pages.
- Frequently change the content of existing pages.
- Suffer from weak internal linking and orphan pages.
- Lack of a strong external link profile.

Even though indexing platforms could technically find your URLs without it, by including pages in an XML sitemap, you're indicating that you consider them to be quality landing pages.

And while there is no guarantee that an XML sitemap will get your pages crawled faster, let alone indexed or ranked, submitting one certainly increases your chances.

How To Create A Sitemap

There are two ways to create a sitemap: Static sitemaps that must be manually updated, or dynamic sitemaps that are updated in real-time or by a regular cron job.

Static sitemaps are simple to create using a tool such as Screaming Frog.

The problem is that as soon as you create or remove a page, your sitemap is outdated. If you modify the content of a page, the sitemap won't automatically update the lastmod tag.

So, unless you love manually creating and uploading sitemaps for every single change, it's best to avoid static sitemaps.

Dynamic XML sitemaps, on the other hand, are automatically updated by your server to reflect relevant website changes.

To create a dynamic XML sitemap you can do one of the following:

- Ask your developer to code a custom script, being sure to provide clear specifications.
- Use a dynamic sitemap generator tool.
- Install a plugin for your content management system (CMS), for example, Yoast plugin for WordPress.

Valid XML Sitemap Format

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>https://example.com/first-url</loc>
    <lastmod>2025-01-28T16:15:20+02:00</lastmod>
  </url>
  <url>
    <loc>https://example.com/second-url</loc>
    <lastmod>2025-01-26T13:15:24+02:00</lastmod>
  </url>
</urlset>
```

Image from author, February 2025

Your sitemap needs three items to introduce itself to indexing platforms:

- **XML Version Declaration:** Specifies the file type to inform indexing platforms what they can expect from the file.
- **UTF-8 Encoding:** Ensures all the characters used can be understood.
- **Specify The Namespace:** Communicates what rules the sitemap follows. Most sitemaps use the “http://www.sitemaps.org/schemas/sitemap/0.9” namespace to show that the file conforms to standards set by sitemaps.org.

This is followed by a URL container for each page. In a standard XML sitemap, there are only two tags that should be included for a URL:

1. **Loc (a.k.a. Location) Tag:** This compulsory tag contains the absolute, canonical version of the URL location. It should accurately reflect your site protocol (http or https) and if you have chosen to include or exclude www.
2. **Lastmod (a.k.a. Last Modified) Tag:** An optional but highly recommended tag to communicate the date and time

the page was published or the last meaningful change. This helps indexing platforms understand which pages have fresh content and prioritize them for crawling.

Google's documentation on sitemaps states:

"Google uses the <lastmod> value if it's consistently and verifiably (for example by comparing to the last modification of the page) accurate. The <lastmod> value should reflect the date and time of the last significant update to the page. For example, an update to the main content, the structured data, or links on the page is generally considered significant, however an update to the copyright date is not."

Bing's documentation agrees on the importance of the lastmod tag:

"The "lastmod" tag is used to indicate the last time the web pages linked by the sitemaps were modified. This information is used by search engines to determine how frequently to crawl your site, and to decide which pages to index and which to leave out."

Mistakes, such as updating the <lastmod> value when the sitemap is generated rather than when the individual page was last modified, or worse, trying to manipulate crawlers by updating the date without significantly altering the page, may result in this signal being ignored by search engines for your website. Damaging your ability to have your content efficiently crawled.

Do not include the Changelist (a.k.a. Change Frequency) Tag or priority tag. Once upon a time, these hinted at how often to crawl, but are now ignored by search engines.

Types Of Sitemaps

There are many different types of sitemaps. Let's look at the ones you actually need.

XML Sitemap Index

XML sitemaps have a couple of limitations:

- A maximum of 50,000 URLs.
- An uncompressed file size limit of 50 MB.

Sitemaps can be compressed using gzip to save bandwidth for your server. But once unzipped, the sitemap still can't exceed either limit.

Whenever you exceed either limit, you will need to split your URLs across multiple XML sitemaps.

Those sitemaps can then be combined into a single XML sitemap index file, often named `sitemap-index.xml`. Essentially, it is a sitemap for sitemaps.

You can create multiple sitemap index files. But be aware that you cannot nest sitemap index files.

For indexing platforms to easily find every one of your sitemap files, you will want to:

- Submit your sitemap index to Google Search Console and Bing Webmaster Tools.
- Specify your sitemap or sitemap index URL(s) in your `robots.txt` file. Pointing indexing platforms directly to your sitemap as you welcome them to crawl.

```
User-agent: *  
Disallow:  
  
Sitemap: https://www.example.com/sitemap-products-index.xml  
Sitemap: https://www.example.com/sitemap-articles-index.xml  
Sitemap: https://www.example.com/sitemap-categories.xml
```

Image from author, February 2025

Image Sitemap

Image sitemaps were designed to improve the indexing of image content, originally offering additional tags.

In modern-day SEO, however, it's best practice to utilize JSON-LD `schema.org/ImageObject` markup to call out image properties to indexing platforms, as it provides more attributes than an image XML sitemap.

Because of this, a dedicated XML image sitemap is unnecessary. Simply add the image XML namespace declaration and the image tags directly to the main XML sitemap within the associated URL.

```
<?xml version="1.0" encoding="UTF-8"?>  
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"  
  xmlns:image="http://www.google.com/schemas/sitemap-image/1.1">  
  <url>  
    <loc>https://example.com/first-url</loc>  
    <lastmod>2025-01-28T16:15:20+02:00</lastmod>  
    <image:image>  
      <image:loc>https://example.com/image.jpg</image:loc>  
    </image:image>  
  </url>  
</urlset>
```

Image from author, February 2025

Know that images don't have to be on the same domain as your website to be submitted in a sitemap. You can use a CDN as long as it's verified in Google Search Console.

Video Sitemap

Similar to images, Google says video sitemap tags can be added within an existing sitemap.

However, unlike images, video extensions in sitemaps offer a multitude of additional tags.

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:video="http://www.google.com/schemas/sitemap-video/1.1">
  <url>
    <loc>https://www.example.com/video-page-1</loc>
    <video:video>
      <video:thumbnail_loc>https://www.example.com/video-thumb.jpg</video:thumbnail_loc>
      <video:title>Example video title</video:title>
      <video:description>This is a description of the first video.</video:description>
      <video:content_loc>https://www.example.com/video1.mp4</video:content_loc>
      <video:player_loc>https://www.example.com/player/video-1</video:player_loc>
      <video:duration>600</video:duration>
      <video:expiration_date>2025-12-31T12:00:00+00:00</video:expiration_date>
      <video:rating>4.5</video:rating>
      <video:view_count>1500</video:view_count>
      <video:publication_date>2025-01-01T12:00:00+00:00</video:publication_date>
      <video:family_friendly>yes</video:family_friendly>
      <video:restriction relationship="allow">IE GB US CA</video:restriction>
      <video:requires_subscription>no</video:requires_subscription>
      <video:uploader
        info="https://www.example.com/users/profile-name">Profile Name
      </video:uploader>
      <video:live>no</video:live>
      <video:tag>tutorial</video:tag>
    </video:video>
  </url>
</urlset>
```

Image from author, February 2025

If you leverage these tags extensively, consider a dedicated video sitemap.

Adding these extensions increases the file size of your sitemap significantly and may lead to you exceeding the file size limits.

Either method will help Google discover, crawl, and index your video content as long as the video is related to the content of the host page and is accessible to Googlebot.

While Bing does support video extensions in XML sitemaps, Fabrice Canel confirmed to me that they prefer submission via IndexNow. Although Bing's documentation still mentions the

mRSS format.

Google News Sitemap

Google News sitemaps can only be used for article content that was created in the last two days. Once the articles are older than 48 hours, remove the URLs from the sitemap.

Again, while Google News sitemap tags can be included in your regular sitemap, this is not recommended.

Unlike for image and video, only Google leverages the news sitemap extension, not Bing or other indexers.

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:news="http://www.google.com/schemas/sitemap-news/0.9">
  <url>
    <loc>http://www.example.com/articles/article-url</loc>
    <news:news>
      <news:publication>
        <news:name>The Example Times</news:name>
        <news:language>en</news:language>
      </news:publication>
      <news:publication_date>2025-01-28T16:15:20+02:00</news:publication_date>
      <news:title>Companies A, B in Merger Talks</news:title>
    </news:news>
  </url>
</urlset>
```

Image from author, February 2025

Contrary to some online advice, Google News sitemaps don't support image URLs.

HTML Sitemap

XML sitemaps take care of indexing platform needs. HTML sitemaps were designed to assist human users in finding content.

The question becomes: If you have a good user experience and well-crafted internal links, do you need an HTML sitemap?

Check the page views of your HTML sitemap in Google Analytics.

Chances are, it's very low. If not, it's a good indication that you need to improve your website navigation.

HTML sitemaps are generally linked in website footers. Taking link equity from every single page of your website.

Ask yourself. Is that the best use of that link equity? Or are you including an HTML sitemap as a nod to legacy website best practices?

If few humans use it, and indexing platforms don't need it as you have strong internal linking and an XML sitemap, does that HTML sitemap have a reason to exist? I would argue no.

XML Sitemap Optimization

XML sitemap optimization involves how you structure your sitemaps and what URLs are included.

How you choose to do this impacts how efficiently indexing platforms crawl your website and, thus, your content visibility.

Here are four ways to optimize XML sitemaps:

1. Only Include SEO Relevant Pages In XML Sitemaps

An XML sitemap is a list of pages you want to be crawled (and subsequently given visibility to by indexing platforms), which isn't necessarily every page of your website.

A bot arrives at your website with an "allowance" for how many pages it will crawl.

The XML sitemap indicates that you consider the included URLs more important than those that aren't blocked but not in the sitemap.

You're using it to tell indexing platforms, "I'd really appreciate it if you'd focus on these URLs in particular."

To help them crawl your site more intelligently and reap the benefits of faster (re)indexing, do not include:

- 301 redirect URLs.
- 404 or 410 URLs.
- Non-canonical URLs.
- Pages with noindex tags.
- Pages blocked by robots.txt.
- Paginated pages.
- Parameter URLs that aren't SEO-relevant.
- Resource pages accessible by a lead gen form (e.g., white paper PDFs).
- Utility pages that are useful to users, but not intended to be landing pages (login page, contact us, privacy policy, account pages, etc.).

I've seen recommendations to add 3xx, 4xx, or non-indexable pages to sitemaps in the hope it will speed up deindexing.

But similar to manipulation of the last mod date, such attempts to get these pages processed faster may result in the sitemaps being ignored by search engines as a signal, damaging your ability to have your valuable content efficiently crawled.

But remember, Google is going to use your XML submission only as a hint about what's important on your site.

Just because it's not in your XML sitemap doesn't necessarily mean that Google won't index those pages.

2. Ensure Your XML Sitemap Is Valid

XML sitemap validators can tell you if the XML code is valid. But this alone is not enough.

There might be another reason why Google or Bing can't fetch your sitemap, such as robots directives. Third-party tools won't be able to identify this.

As such, the most efficient way to ensure your sitemap is valid is to submit it directly to Google Search Console and Bing Webmaster Tools.

Submitted sitemaps							
Sitemap	Type	Submitted	Last read	Status	↓ Discovered pages	Discovered videos	
/sitemap-articles-in-dex.xml	Sitemap index	27 Jan 2025	27 Jan 2025	Success	782,374	36,218	⋮

Image from author, February 2025

When valid in GSC and BWT, you will see the green “Success” status.

!

Sitemap index can be read, but has errors

Invalid date

2 instances ^

An invalid date was found. Please fix the date or formatting before resubmitting.

Examples

Line 2

Parent tag:

sitemap

Tag:

lastmod

Line 2

Parent tag:

sitemap

Tag:

lastmod

Image from author, February 2025

If you get a red message instead, click on the error to find out why, fix it, and resubmit.

But in Google Search Console and Bing Webmaster Tools, you can do so much more than simple validation.

3. Leverage Sitemap Reporting For Indexing Analysis

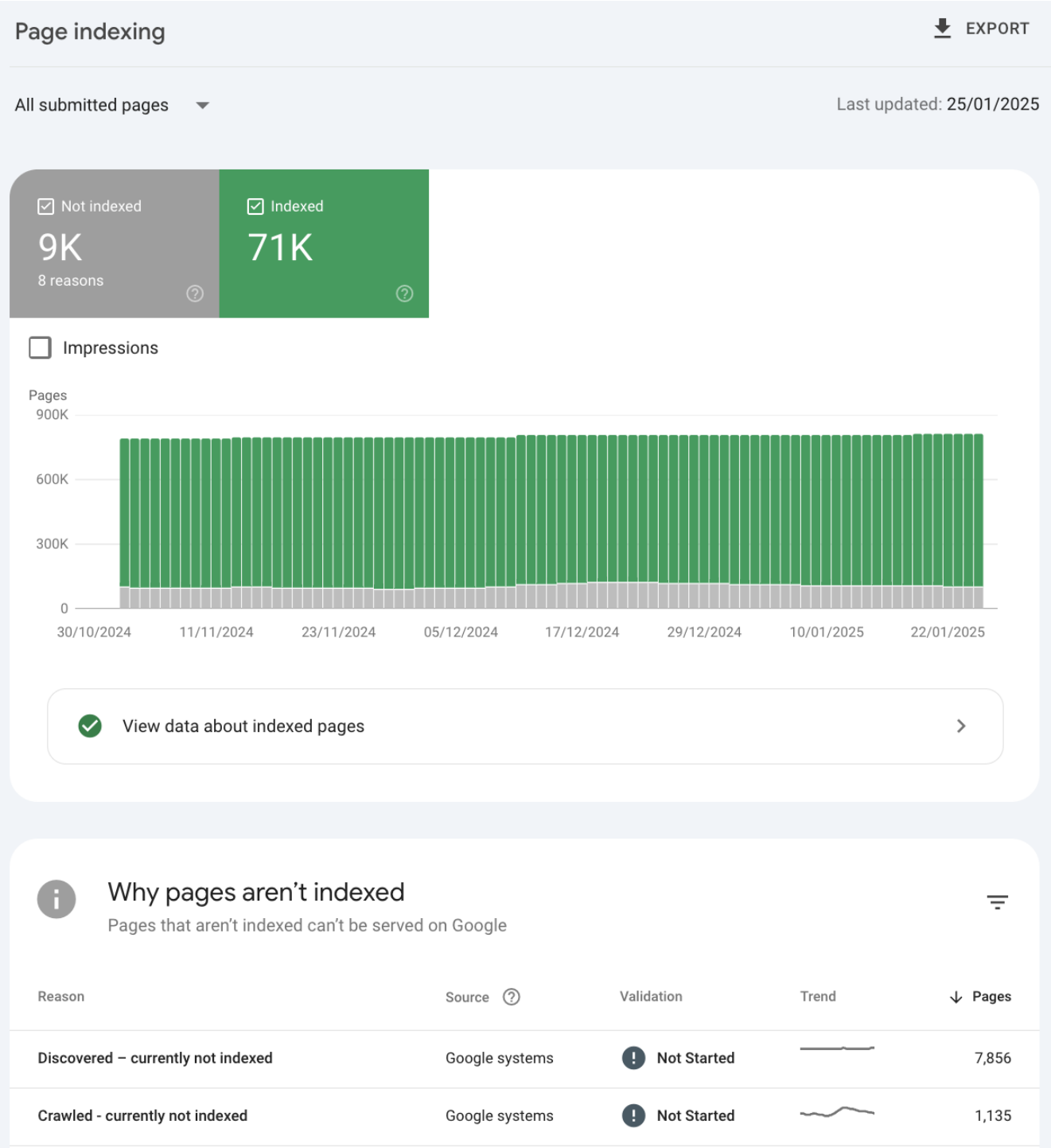


Image from author, February 2025

Say you submit 80,000 pages all in one sitemap index, and 9,000 are excluded by both Google and Bing.

Sitemap reporting will help you to understand overarching why, but provides limited reporting on which URLs are problematic.

So, while it's valuable information, it's not easily actionable. You need to discover which types of pages were left out.

What if you use descriptive sitemap names that reflect the sections of your website – one for categories, products, articles, etc.?

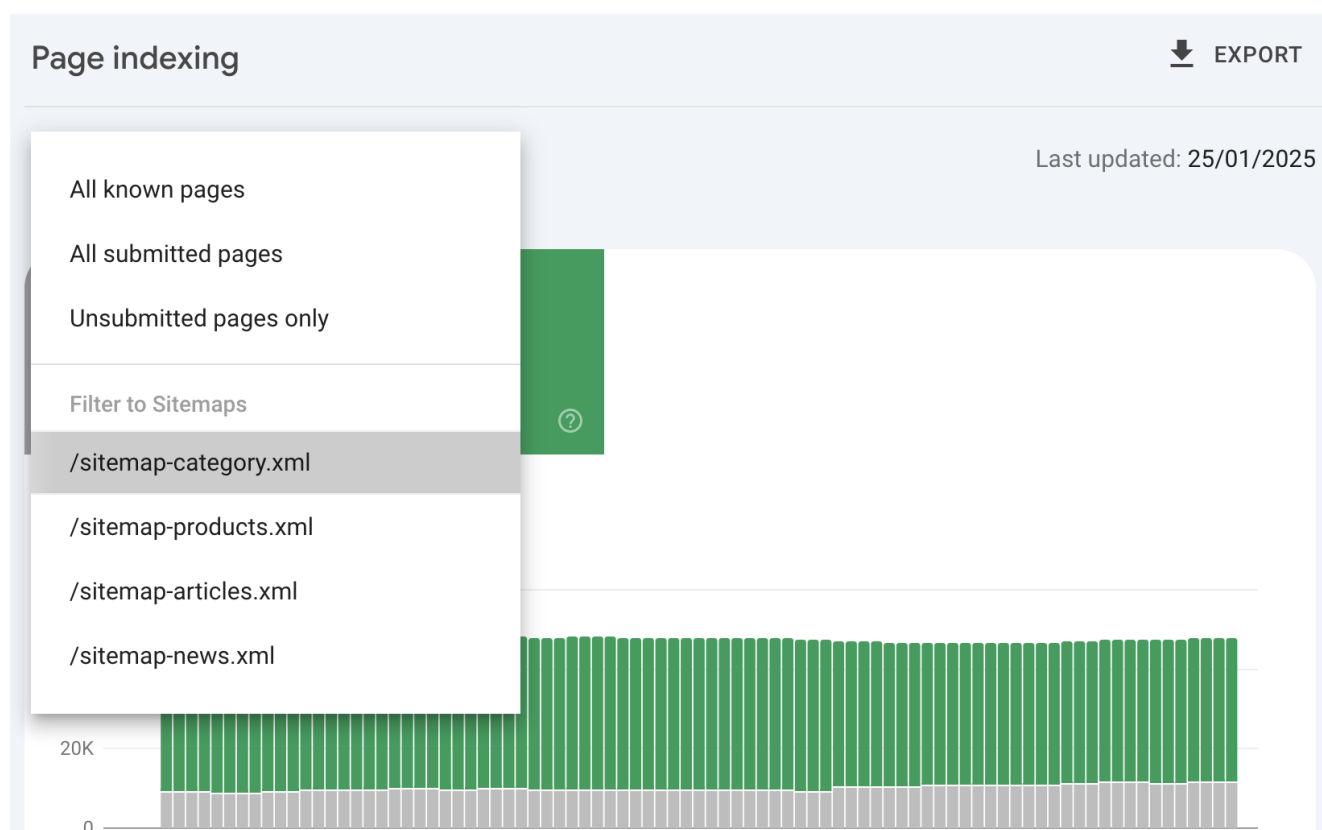


Image from author, February 2025

Then, we can drill down to see that 7,000 of the 9,000 non-indexed URLs are category pages – and clearly know where to focus attention.

This can also be done within a sitemap index file.

Now, I know both Google and Schema.org show examples encouraging numbered naming. So, you may have ended up in a /sitemap-products-index.xml file with something like this:

- /products-1.xml
- /products-2.xml

Which is not the most insightful naming convention. What if we

break it down into parent categories? For example:

- /products-mens.xml
- /products-womens.xml
- /products-kids.xml

And if your website is multilingual, be sure to leverage language as an additional separation layer.

Such smart structuring of sitemaps to group by page type allows you to dive into the data more efficiently and isolate indexing issues.

Just remember, for this to effectively work, sitemaps need to be mutually exclusive, with each URL existing in only one sitemap. The exception is the Google News sitemap.

4. Strategize Sitemap Size

As mentioned before, search engines impose a limit of 50,000 URLs per sitemap file.

Some SEO specialists intentionally reduce this number, say to 10,000. This can be helpful to speed up indexing.

However, you can only download 1,000 URLs in GSC. So, if 2,000 URLs in a certain sitemap are not indexed, you can only access half of them. If you are trying to do content cleanup, this will not be enough.

Page indexing > Discovered – currently not indexed

EXPORT [SHARE](#)

Examples ? 

URL	↓ Last crawled
https://www.examples.com	N/A

Rows per page: 10 ▼ 1-10 of 1000 < >

Image from author, February 2025

To gain full visibility on all URLs causing issues, break sitemaps down into groups of 1,000.

The obvious downside is that this has a higher setup time as all URLs need to be submitted in Google Search Console and Bing Webmaster Tools. This may also require high levels of ongoing management.

XML Sitemap Best Practice Checklist

Do invest time to:

- ✓ Dynamically generate XML sitemaps.
- ✓ Compress sitemap files.
- ✓ Use a sitemap index file.
- ✓ Include the <loc> & <lastmod> tags.
- ✓ Use image tags in existing sitemaps.
- ✓ Use video and Google News sitemaps if relevant.
- ✓ Reference sitemap URLs in robots.txt.
- ✓ Submit sitemaps to both Google Search Console and Bing Webmaster Tools.

- ✓ Include only SEO-relevant pages in XML sitemaps.
- ✓ Ensure URLs are included only in a single sitemap.
- ✓ Ensure the sitemap code is error-free.
- ✓ Group URLs in descriptively named sitemaps based on page type.
- ✓ Strategize how to break down large sitemap files.
- ✓ Use Google Search Console and Bing Webmaster Tools to analyze indexing rates.

Now, go check your own sitemaps and make sure you're doing it right.

More Resources:

Featured Image: BEST-BACKGROUNDS/Shutterstock

<https://www.searchenginejournal.com/technical-seo/how-to-use-xml-sitemaps/>