

SEO - XML Sitemap

A Sitemap: What Is It?

A document in which you list your website's web pages, videos, and other assets, along with their connections. Search engine spiders examine this file so they can crawl your website. Text, XML, and RSS are among the most common formats for writing sitemaps. The most popular form is XML. In this chapter, we will learn about XML.

Do You Require a Sitemap?

According to the size and design of the website, it depends. If your site has 100 or fewer URLs, and those URLs are all internally linked, then you generally don't require a sitemap. A sitemap, however, will be pretty helpful if you manage a sizable website with thousands, millions, or billions of URLs.

```
<url>
  <loc>https://example.com/blog/sample-blog-article/</loc>
  <xhtml:link rel="alternate" hreflang="en" href="https://example.com/blog/sample-blog-article/">
  <xhtml:link rel="alternate" hreflang="de" href="https://example.com/de/blog/sample-blog-article/">
  <xhtml:link rel="alternate" hreflang="fr" href="https://example.com/fr/blog/sample-blog-article/">
  <xhtml:link rel="alternate" hreflang="es" href="https://example.com/es/blog/sample-blog-article/">
  <xhtml:link rel="alternate" hreflang="it" href="https://example.com/it/blog/sample-blog-article/">
  <xhtml:link rel="alternate" hreflang="nl" href="https://example.com/nl/blog/sample-blog-article/">
</url>
```

Explore our [latest online courses](#) and learn new skills at your own pace. Enroll and become a certified expert to boost your career.

About XML

The most flexible sitemaps are those in XML format. It is readily expandable and may be used to provide extra details regarding the localized variations of your web pages, photographs, videos, and news items.

Pros



- It is adaptable and extensible.
- It can offer the most details regarding your URLs.
- Users of CMSs can locate plugins for creating sitemaps.

Cons

- High-skilled task.
- Keeping the mapping up to date on more prominent portals or websites where the URLs frequently change may be challenging.

XML tags make up the Sitemap protocol format. For any information or value contained in a Sitemap, Entity-escaping is required.

The Sitemap has a format

- `<urlset>` is used at the beginning and `</urlset>` is used at the end.
- The `<urlset>` tag's namespace must be defined.
- As a parent XML tag, add an `<url>` entry for every URL.
- Include a `<loc>` entry for every `<url>` tag.

The rest of the tags are configurable. Search engine spiders might or might not accommodate these additional tags. For details regarding the compatibility of each search engine, read its instruction manual. Furthermore, each URL in a Sitemap should be from the same host, such as www.tutorialspoint.com, guides.tutorialspoint.com, etc.

XML Sitemap

This represents a sample XML sitemap that shows where one URL is located –

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://www.example.com/foo.html</loc>
  </url>
</urlset>
```

Now, here is an example of an XML sitemap with multiple URLs –

```
<?xml version="1.0" encoding="UTF-8"?>

<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">

  <url>

    <loc>http://www.tutorialspoint.com/</loc>

    <lastmod>2023-06-01</lastmod>

    <changefreq>monthly</changefreq>

    <priority>0.8</priority>

  </url>

  <url>

    <loc>http://www.tutorialspoint.com/catalog?item=12&desc=guide_seo</loc>

    <changefreq>weekly</changefreq>

  </url>

  <url>

    <loc>http://www.tutorialspoint.com/catalog?item=73&desc=guide_xmlsitemaps</loc>

    <lastmod>2023-05-25</lastmod>

    <changefreq>weekly</changefreq>

  </url>

  <url>

    <loc>http://www.tutorialspoint.com/catalog?item=74&desc=guide_cybersecurity</loc>

    <lastmod>2023-06-10T18:00:15+00:00</lastmod>

    <priority>0.3</priority>

  </url>

  <url>

    <loc>http://www.tutorialspoint.com/catalog?item=83&desc=guide_ecommerce</loc>

    <lastmod>2023-05-30</lastmod>

  </url>

</urlset>
```

Tag specifications for XML

The XML tags that are accessible are as follows –

XML Tag	Status	Definition	^
---------	--------	------------	---

<urlset>	Necessary	It references the current protocol standard and encodes the file's contents.
<url>	Necessary	It refers to the parent tags for every URL present. This tag's children are the rest of the titles.
<loc>	Necessary	The website's URL. If the server hosting the website demands it, the protocol name (HTTP) must come first in this URL and have a trailing slash at the end. A smaller amount than 2,048 bits must make up this value.
<lastmod>	Optional	This represents the website's most recent update date. W3C Datetime format must be used for this date. If you like, you can use YYYY-MM-DD instead of the time component in this format. Remember that the date must be configured to reflect when the linked website's last content was edited, not when the developer created the sitemap.
<changefreq>	Optional	<p>The likelihood that the page will update regularly. Although it may not accurately represent how frequently search engines crawl the website, this value gives them a broad overview. Valid codes include –</p> <ul style="list-style-type: none"> ● always. ● hourly. ● daily. ● weekly. ● monthly. ● yearly. ● never. <p>always - Document updates whenever accessed. never - Archive URLs need to be assigned this value.</p>
<priority>	Optional	This tag represents the importance of an URL in comparison with additional URLs on your website. Between 0.0 to 1.0 are valid values. This number informs search engines about the pages you think are most crucial for the crawling mechanisms. Still, it has no impact on how your web pages are evaluated compared to content on other web pages. Default Value- 0.5

Note

- All tag values need to be entity escaped, as for all XML files.
- The values for <priority> and <changefreq> are ignored by Google.
- If the <lastmod> value can be regularly and independently verified as accurate, Google will utilise it.

Entity Eluding Capture

UTF-8 encoding is required for your sitemap file. All XML files must employ entity escape values for any characters indicated below, including URLs and any information contents and values and URLs.

Character	Symbol	Escape Code
Ampersand	&	&
Single Quote	'	'
Double Quotes	"	"
Greater Than	>	>
Lesser Than	<	<

Sitemap Index Files

The total number of URLs in each Sitemap file you provide is restricted to 50,000, and their combined size must not exceed 50MB. If you want to minimize the bandwidth needed, you can compress the contents of your sitemap documents using gzip; nevertheless, the sitemap file's uncompressed size cannot exceed 50MB. You need to make multiple Sitemap files to include over fifty thousand URLs.

You should create a directory file called "Sitemap" with a list of all the Sitemap files if your website(s) have a lot of distinct Sitemaps. 50,000 Sitemaps or less, a 50MB maximum file size, and compression, are all criteria for sitemap index files. Multiple Sitemap index files may exist.

A Sitemap file's XML format and its index file's XML format are identical. The Sitemap index template needs to –

- An opening tag `<sitemapindex>` at the start and a closing tag `</sitemapindex>` at the conclusion.
- Every sitemap should have its entry `<sitemap>` in the parent XML tag.
- For every `<sitemap>` parent tag, provide a child entry `<loc>`.
- Similarly, Sitemap index files support the optional tag `<lastmod>`.

```
<?xml version="1.0" encoding="UTF-8"?>
<sitemapindex xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <sitemap>
    <loc>http://www.tutorialspoint.com/sitemap01.xml.gz</loc>
    <lastmod>2023-06-18T18:21:00+00:00</lastmod>
  </sitemap>
  <sitemap>
    <loc>http://www.tutorialspoint.com/sitemap02.xml.gz</loc>
    <lastmod>2023-06-18</lastmod>
  </sitemap>
</sitemapindex>
```

Uploading XML Sitemap

Upload your XML sitemap to Google using Search Console; follow these steps –

- Login to Google Search Console.
- Select "Sitemaps".
- Add your sitemap's URL to the "Add a new sitemap" section at the top of the webpage.
- Press 'Submit,' and Google will go through your freshly constructed XML sitemap.

Conclusion

Your website's key pages are all directed by a good XML sitemap, which serves as a route map for Google. Regardless of whether the internal structure of your website could be better, XML sitemaps may prove beneficial for SEO because they help Google identify your key pages efficiently.