

If you're having problems getting web pages indexed by Google, you may need a sitemap. A sitemap is a specially formatted file used by Google to spider your website and find web pages. The file is useful for people who have pages that can't be crawled by Google including AJAX, dynamic product pages, and archived files with no internal links. If you search Google and can't find your website pages indexed, a sitemap increases the chance of more search engine exposure.

Sitemap Format

A sitemap uses XML tags to list the URLs located on the website. Google guidelines indicate that sitemaps should have no more than 50,000 links and shouldn't be larger than 10MB in size. For sites with more than 50,000 links, the webmaster can separate the sitemap into a series of smaller files. Webmasters can also use gzip on files that are too large. Gzip software compresses files, so Googlebot downloads a smaller file and decompresses it after retrieval.

A properly formatted sitemap requires a few XML tags, so Google can read the information and index URLs without errors. First, the sitemap must open with the "<urlset>" tag. This tag indicates the schema for the file. The next two requirements are the "<url>" and "<loc>" tags. Each set of URLs should be encapsulated with the "<url>" tag. Within this tag is the list of links on the website. Each link is identified using the "<loc>" tag. The links located in these tags can be dynamic product links with query string variables, or it can contain static HTML files.

An example of a simple sitemap is below:

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

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

```
<url>
```

`<loc>http://mysite.com</loc>`

`<lastmod>2009-12-20</lastmod>`

`<changefreq>monthly</changefreq>`

`<priority>0.8</priority>`

`</url>`

`</urlset>`

Sitemap Options

Three other XML tags are optional tweaks to help control the frequency Google crawls and indexes pages. These tags are optional, and leaving them out of the sitemap will not cause errors with Googlebot.

The first optional XML unit is the "`<lastmod>`" tag. This tag tells the search engines the time and date of the last file modification. If you update your web page daily, this date and time changes frequently. However, you can also specify the amount of times a web page changes using the "`<changefreq>`" tag. This tag tells Googlebot the frequency of web page modifications such as daily, monthly, yearly or weekly. This tag affects the frequency Googlebot crawls pages. However, abuse of the "`<changefreq>`" tag can lower the amount of times Google crawls your site, so use it properly.

Finally, the "<priority>" tag is the last option for a sitemap. Many webmasters leave this value as 1.0, but this degrades the value of the sitemap. The priority tag is used to indicate the importance of each page, but it does not change the website's ranking. The priority tag influences the URL chosen for display by the search engines. For instance, the home page should have a priority of 1.0 while a page deeper within the site structure would have a value of 0.5.

Once the file is created, upload the sitemap to your website's host directory. Login to your webmaster tools at Google, and register the sitemap's location. Within a few hours, Google will crawl your sitemap. Although Google doesn't guarantee all available pages are indexed, it improves the chance of having more web pages included in the search engine results.