

5 June 2005

By:



[How does "Google Sitemaps" work?](#)

A short review by Softpedia

The purpose of the Google Sitemaps project is enabling webmasters to inform and direct the Google spiders through their website. Sitemaps provide crawlers with information about the website's structure as well as data about its pages, which leads to, according to Google, an improved indexing process. Another benefit brought by sitemaps is that webmasters can quicken the indexing of some pages by publishing them in the sitemap, without waiting for the usual „visit“ of the crawlers. This technique is called „content pushing“. However, by using sitemaps you will NOT achieve better placement in Google SERPS (search engine result pages). Sitemaps are text files in the XML format that contain information about the website, a list of webpages and some corresponding parameters. At most 50,000 entries can be placed into a sitemap, provided the sitemap's filesize is not more than 10MB uncompressed (sitemap access and transfer can be greatly improved by using gzip compression which yields smaller files due to its excellent handling of text data). Should you exceed these limits you will have to use multiple sitemaps and group them into a „sitemap-index“. Each webpage inserted into a sitemap is defined by the following characteristics:- Changefreq: how often the webpage's content is modified- Lastmod: the date of the last modification- Loc: the webpage URL- Priority: the priority of the webpage with respect to the other webpages in the site To ease the creation of sitemaps, Google has made available Sitemap Generator (sitemap_gen.py), an utility written in the Python language. It can create sitemaps through three methods, based on some parameters:- by reading a text file that contains all URLs to include in the sitemap- by reading directories on the server's filesystem- by parsing webserver log files Advanced users can take advantage from the XML format of sitemaps and choose to „automagically“ generate them (for example, a developer can integrate sitemap support into his/her content management system's (CMS)) Sitemaps can also be based on the „Open Archives Initiative (OAI) protocol for metadata harvesting“, a popular and standard protocol, so anyone that already has OAI-PMH 2.0 sitemaps can submit them unaltered to Google/For those that think XML/OAI sitemaps are too complex, there's a simpler solution: submitting a file that contains just an URL list. For more information about the Google Sitemaps protocol you may visit these pages: <http://www.google.com/webmasters/sitemaps/docs/en/protocol.html> <http://www.google.com/webmasters/sitemaps/docs/en/faq.html> One thing that should not be passed-by easily is that Google Sitemaps is still a Beta project, therefore Google does not guarantee it will crawl all URLs included in sitemaps. However, the benefits seem important enough to encourage all webmasters to use sitemaps.