

Search Engine Optimization-Web Crawlers

The terms web crawler, automatic indexers, bots, worms, web spiders, and web robots are programs or automated scripts with browse the World Wide Web in a methodical, automated manner. The term web crawler is the most commonly used term.

Web crawlers are a tool used for search engine optimization.

Search engines use web crawlers to provide up to date data and information. Web crawlers provide the requested information by creating copies of web pages that the search engine later processes. Once the information has been processed the search engines indexes the pages and are able to quickly download the pages during a search. The process of web crawling is a key factor in search engine optimization. Search engine optimization is the art and science of making web pages attractive to search engines. Computer people call the process of using a web crawler to rank a website spidering.

Some search engines use web crawlers for maintenance tasks. Web crawlers can also be used for harvesting e-mail addresses. The internet is a gaping ocean of information. In 2000, Lawrence and Giles manufactured a study that indicated the internet search engines have only indexed approximately sixteen percent of the Web. Web crawlers are designed to only download a tiny amount of the available pages. A miniscule sample of what the internet has to offer.

Search engines use web crawlers because they can fetch and sort data faster than a human could ever hope to. In an effort to maximize the download speed while decreasing the amount of times a webpage is repeated search engines use parallel web crawlers. Parallel web crawlers require a policy for reassigning new URLs. There are two ways to assign URLs. A dynamic assignment is what happens when a web crawler assigns a new URL dynamically. If there is a fixed rule stated from the beginning of the crawl that defines how to assign new URLs to the crawls it is called static assignment.

In order to operate at peak efficiency web crawlers have to have a highly optimized architecture.

URL nominalization is the process of modifying and standardizing a URL in a consistent manner. URL nomalization is sometimes called URL canonicalization. Web crawlers usually use URL nomilization to avoid multiple crawling of a source.

In an attempt to attract the attention of web crawlers, and subsequently highly ranked, webmasters are constantly redesigning their websites. Many webmasters

rely on key word searches. Web crawlers look for the location of keywords, the amount of keywords, and links.

If you are in the process of creating a website try to avoid frames. Some search engines have web crawlers that can not follow frames. Another thing some search engine are unable to read are pages via CGI or database -delivery, if possible try creating static pages and save the database for updates. Symbols in the URL can also confuse web crawlers. You can have the best website in the world and if a web crawler can't read it probably won't get the recognition and ranking it deserves.