

# Google Index and Cache

Contributed by Suresh Mishra  
Last Updated Sunday, 11 April 2010

## What are Google's bots?

Google constantly seek out new pages and / or updated to add to your index and there is a charge of this program that is called Googlebot, the famous robots or spiders (spiders). So how Googlebots are calling the search bots whose sole mission in life is to collect web documents in order to build a database that is used by the search engine of its master.

The Googlebots employ a process based on algorithms that determine which sites to crawl, the frequency and number of pages to fetch from each site. These lists are comprehensive websites to identify links to other pages. What is indexing? Indexing is the processing of the pages scanned and is what creates the index that uses Google to give results when you search. In fact, the robots do not keep our pages but the analysis and make an index of all the words they see and their location. In addition, process information in the TITLE tag and the ALT attribute content of the images, nor do they do with all that he has a page, for example, do not process the content of most Flash files or dynamic pages . Just read HTML documents? No, also extract index information or other files: PDF, PS (Adobe PostScript), leaves of Lotus (wk1, wk2, wk3, wk4, WK5, WKI, wks, wku, lwp) and Excel (xls), documents MW text, DOC, WRI, RTF, ANS, TXT, PowerPoint presentations (ppt) files, Microsoft Works (wks, wps, wdb) and swf. This is done to give more results, in fact, can do a search indicating that we display only certain types of files, for example:

filetype: doc "search text" In most cases, even when we do not have the software necessary to interpret, we show the option of seeing them as HTML or plain text. Conversely, we can eliminate certain types of search results using a filter, for example:  
-filetype: pdf "search text" How often do we visit?

They say "regularly" but give no details, speak of many factors that can influence but, the truth is that often you access a site depends almost exclusively on PageRank you have. The higher, more will be visited regularly (wealth generates wealth). Then, they can do every day or take weeks. Google PageRank and is proud of us know that is the heart of his whole system: "The heart of our software is PageRank &trade;;, a system for ranking web pages developed by our founders Larry Page and Sergey Brin at Stanford University. And while we have dozens of engineers working to Improve every aspect of Google on a daily basis, PageRank continues to play a central role in many of our web search tools.