GOOGLEBOT

Googlebot es el robot de de búsqueda usado por Google (en ocasiones, también denominado "araña"). El rastreo es el proceso mediante el cual Googlebot descubre páginas nuevas y actualizadas y las añade al índice de Google.

Utilizamos una enorme cantidad de equipos informáticos para obtener (o "rastrear") miles de millones de páginas de la Web. Googlebot utiliza un proceso de rastreo algorítmico: a través de programas informáticos se determinan los sitios que hay que rastrear, la frecuencia y el número de páginas que hay que buscar en cada sitio.

El proceso de rastreo de Googlebot comienza con una lista de URL de páginas web generada a partir de procesos de rastreo anteriores y se amplía con los datos de los [sitemaps](http://www.google.com/support/webmasters/bin/answer.py?answer=156184) que ofrecen los webmasters. A medida que Googlebot visita cada uno de esos sitios web, detecta enlaces (SRC y HREF) en sus páginas y los añade a la lista de páginas que debe rastrear. Los sitios nuevos, los cambios en los existentes y los enlaces obsoletos se detectan y se utilizan para actualizar el índice de Google.

* 1. **VERSIONES:**
* Deepbot, como su nombre indica, es un rastreador en profundidad el cual trata de seguir cualquier enlace en la Web y descargar todas las páginas posibles  para la indexación. Suele realizar esta tarea una vez al mes.
* Freshbot rastrea la Web buscando contenido fresco. Visita las Web que cambian frecuentemente, ajustándose a los tiempos de cambio de dichas Web.

Para hacernos una idea, visitará diariamente páginas que suelen estar en continuo movimiento como pueden ser periódicos virtuales o Webs informativas. Otros sitios Web pueden ser visitados semanalmente ya que los cambios son menores.

La frecuencia con la que Googlebot accede a un sitio web depende del [**PageRank**](http://google.dirson.com/pagerank.php) de éste. Contra mayor sea este valor, el robot accederá más asiduamente a sus páginas.

* Por ejemplo, podemos comprobar que los sitios con PR10 (el valor más alto) como yahoo.com o han sido 'rastreados' por Googlebot ayer o incluso hoy mismo, mientras que otros han sido accedidos hace varias semanas. Esto se puede comprobar accediendo a la 'cache' de esta página.