

Bienvenido a la página de algoritmos de posicionamiento del grupo Wiki 2 de la asignatura Recuperación y Acceso a la Información de la [Universidad Carlos III de Madrid](#).

Introducción

Dentro de la recuperación y organización de la información se encuentran los algoritmos de posicionamiento. Éstos algoritmos son un conjunto de técnicas que nosotros nos encargamos de aplicar para que nuestro sitio aparezca en los primeros puestos de [Google](#), [Yahoo](#), [MSN Search](#) y muchos otros buscadores.

Posicionar un sitio no es una tarea rápida ni fácil, pero los beneficios son muy grandes como para no prestarles atención.

En esta página pueden encontrarse los algoritmos de posicionamiento utilizados por los principales buscadores.

Recuperación y organización de la información: Algoritmos

Por naturaleza, los buscadores se encargan de la recuperación de los sitios de acuerdo a un algoritmo. Mediante este algoritmo, sitúan en las primeras posiciones los sitios web que el buscador cree que serán más útiles al navegante debido a la información que contienen.

De este modo, los sitios no optimizados, no aparecen nunca en las primeras posiciones de las búsquedas, por lo que prácticamente no obtendremos por este medio ninguna visita. Así, es comprensible que nuestro objetivo sea que nuestro sitio se sitúe en las mejores posiciones en los principales buscadores como son Google, Yahoo, y MSN Search.

Existen diversas técnicas, totalmente limpias, para conseguir esto. Por ejemplo, debe optimizarse el sitio de acuerdo a palabras clave, realizar correctamente la organización de la información, optimizar el texto, los títulos, etiquetas meta, textos de enlaces, direcciones y decenas de factores más que los buscadores tienen en cuenta. También deben agregarse enlaces, aspecto muy importante en la mayoría de los buscadores, que influye a la hora de la recuperación de la información.

Algoritmos de posicionamiento en Google

Principalmente existen dos algoritmos utilizados por [Google](#) a la hora de indexar los resultados de las búsquedas:

- Algoritmo de Relevancia
- Algoritmo de PageRank

Algoritmo de Relevancia

La lógica de este algoritmo es la siguiente: Google desea saber si realmente tu página tiene la información que el usuario está buscando. Para ello comprueba si la palabra buscada aparece en tu página y en qué partes de ella lo hace. De este modo, no puedes desatender la organización de la información de tu página, porque sino el el buscador no procederá a su recuperación en la búsqueda.

Google considera que, por ejemplo, si la palabra está en el título principal de la página, seguramente tu página sí está dedicada al tema que está buscando el usuario. Si la palabra además aparece varias veces en el texto de la página en cuestión, y en muchas otras páginas de esa web, aun con más razón esa página estará dedicada al tema que busca el usuario, por lo que le otorgará un mejor posicionamiento.

Para representar todo esto de forma matemática Google utiliza el Algoritmo de Relevancia. Dicho algoritmo tiene en cuenta los siguientes factores:

- Relevancia de la palabra en el contexto general de páginas indexadas: en cuantas páginas de nuestra web, aparece esta palabra (en forma de densidad: $\text{núm}/\text{total pág.}$)
- La relevancia de las palabras que otras webs han utilizado para realizar los enlaces a la nuestra.
- Relevancia de la palabra en cada una de las páginas.

Orden de importancia de los factores

El orden de importancia(aproximado) de cada una de las localizaciones es el siguiente:

- Densidad en la URL
- Densidad en el Título de la página (tag "title")
- Densidad en la descripción (Google no lo tiene en cuenta, para buscadores que no son Google sí es importante)
- Densidad en cabeceras (H1,H2, etc...)
- Densidad en el nombre de enlaces
- Densidad en palabras en negrita
- Densidad en textos alternativos (ALT)

De esto se deduce que, para que nuestra página aparezca en las primeras posiciones de una búsqueda, debemos conseguir que la palabra que el usuario busca, aparezca en nuestra URL (en forma de subdominio, de directorio o de nombre de fichero), en el título de nuestra página, en la descripción y en las cabeceras.

Algoritmo de PageRank

El principal algoritmo de recuperación de Google se llama PageRank y consiste en la creación de un sistema de clasificación de páginas webs en función de:

- El número de páginas relacionadas con lo que busca el usuario que linkean o enlazan con la nuestra. Cada página que nos enlaza nos está dando 1 voto desde el punto de vista de Google.
- El contenido e información de nuestra página. Para evaluarlo mira la densidad de palabras clave, título de la página, etc. y la compara con la búsqueda del usuario.
- Qué es el PageRank

El PageRank es un valor numérico que representa la importancia que una página web tiene en Internet. Google se hace la idea de que cuando una página coloca un enlace (link) a otra, es de hecho un voto para esta última. Cuantos más votos tenga una página, será considerada más importante por Google. Además, la importancia de la página que emite su voto también determina el peso de este voto. De esta manera, Google calcula la importancia de una página gracias a todos los votos que reciba, teniendo en cuenta también la importancia de cada página que lo emite.

De este modo, el PageRank es la manera que tiene Google de decidir la importancia de una página. Es un dato valioso, porque es uno de los factores que determinan la posición que va a tener una página dentro de los resultados de la búsqueda. No es el único factor que Google utiliza para clasificar las páginas y realizar su recuperación, pero sí es uno de los más importantes.

Cómo ganar PageRank

El PageRank se transmite de unas páginas web a otras. Una página web que nos enlace desde una subpágina de la misma, con Pagerank 5, dependiendo de como tenga los enlaces, nos podría pasar hasta un Pagerank 4. Por tanto, la forma más fácil de conseguir más PageRank es, ser enlazado por webs con un valor más alto que el nuestro. Uno de los métodos más utilizados para subirlo es el intercambio de enlaces, hay que tener en cuenta que últimamente este tipo de enlaces recíprocos, no están tan valorados como los enlaces no recíprocos. Cuantas más webs nos enlacen más posibilidades tendremos de obtener un Pagerank más alto, pero pasar de un valor a otro

conlleva un cálculo exponencial, esto quiere decir, que podemos tener 1000 webs de PageRank 2 enlazándonos y, seguramente, nunca llegaremos a un valor de 5, y con un único enlace que tuviésemos de una web de PageRank 6, podríamos obtener un 5.

Cómo conocer tu PageRank

Una manera de conocer el PageRank de una página es descargándose la [barra de búsqueda de Google](#). Aparece una barra en la que se muestra en color verde el valor de PageRank en una escala de 0 a 10. Sitios web con PR10 son Yahoo!, Microsoft, Adobe, Macromedia, o el propio Google.

La fórmula del PageRank

La fórmula del Page Rank es la siguiente:

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

Siendo:

- PR(A) es el PageRank de la página que tomamos como referencia.
- d es el factor de debilitación.
- (1-d) asegura que cualquier página indexada por Google, aunque no reciba ningún enlace, tendrá un PR mínimo de 0'15.
- PR(T1)/C(T1) será el PageRank (PR) de una de las páginas que nos enlaza, (T1), dividido por todos los enlaces (C) que también salen de esa página T1, es decir, el PR que nos transmite.
- ... + PR(Tn)/C(Tn) lo mismo que el punto anterior, repetido por cada página que nos enlace.

Algoritmos de posicionamiento en Yahoo

Existen diversos factores que influyen a la hora de aparecer como primer resultado en Yahoo. Algunos de estos factores son:

- Asignación de pesos denominada WebRank.
- Interfaz.
- Directorio Yahoo.

- Otros pesos de posicionamiento copiados de Google.

Algoritmo de Yahoo: WebRank

El WebRank sería análogo al PageRank de Google, y es la escala (de 1 a 10) que [Yahoo](#) utiliza para medir la popularidad de una página web. Sin embargo, el WebRank no se calcula en función del número de enlaces hacia dicha página (como lo hace Google), sino que depende de la información obtenida a través de la barra instalada en los navegadores de los usuarios.

Si instalas la versión "beta" de esta barra, la "[Companion Toolbar](#)", desde tu navegador se enviará la información de la URL visitada a los servidores de Yahoo, los cuales realizarán los cálculos del WebRank. Además, podrás ver el valor del WebRank en color lila.

Esta misma información sirve al robot de Yahoo, Slurp, a rastrear (si no lo ha hecho ya) los sitios web visitados por los usuarios para, de esta manera, incorporarlos a los resultados de las búsquedas.

Interfaz

A diferencia de Google, Yahoo tiene en cuenta diversos aspectos de la interfaz a la hora de realizar la recuperación de la información:

- Etiquetas META: Yahoo sigue teniendo en cuenta la Meta Tag Keywords. Esta etiqueta Meta va colocada en la cabecera del código HTML. La Meta Tag Keywords, en desuso por los buscadores desde hace tiempo, cobra una importancia vital en Yahoo. Su peso es similar a la utilización de las keywords en el body de la página. Yahoo está valorando las 27 primeras palabras que van en esta Meta Tag o los 160 primeros caracteres.
- Título: Yahoo sigue teniendo en cuenta el título de la página. Para conseguir buenas posiciones en Yahoo es recomendable utilizar un título profesional.

Directorio Yahoo

El único directorio actualmente comparable a [Dmoz](#) es el directorio de Yahoo, el gran dinosaurio de Internet que ha logrado adaptarse a cada nueva época de la red.

Si se quiere conseguir una buena posición en una búsqueda de información en Yahoo, no basta con tener una buena organización de la información. Es recomendable que te des de alta en su directorio en el apartado que mejor se adapte a la temática de tu web. Esto es debido a que Yahoo prevalece la búsqueda de Webs en su directorio.

Algoritmos de posicionamiento en MSN

El más reciente de los grandes buscadores es [MSN Search](#), de Microsoft, que previamente dependía de otros para realizar la recuperación de la información y listar sus búsquedas. En 2004 debutó una versión beta con sus propios resultados, impulsada por su propio robot (llamado msnbot). Tras tres meses de pruebas, el 1 de Febrero de 2005 MSN Search dejó de ser una Versión Beta y fue lanzado en más de 25 mercados simultáneamente. Supuso el estreno de un proyecto que llevaba más de dos años gestándose y que ha supuesto una inversión hasta el momento de más de 80 millones de dólares.

Google VS MSN

Definitivamente Microsoft ha decidido dejar de depender en sus búsquedas de Inktomi (propiedad de Yahoo) y nos presenta un nuevo buscador programado desde cero por su equipo y muy similar a Google en cuanto a forma. Analicemos las similitudes y diferencias:

MSN Search tiene una página de inicio bastante sencilla, que va al grano y uno resultados dispuestos de una manera muy similar a los de Google:

Mismos colores, mismas opciones, los anuncios se muestran de manera similar:

Google [Arturo Soria \(Madrid\) - Sólo lo mejor](#)
La zona **Arturo Soria** de Madrid ya tiene un web. Con bolsa inmobiliaria y artículos sobre medicina, animales, fisioterapia, psicología ...
www.arturosoria.com/ - 7k - 5 Feb 2005 - [En caché](#) - [Páginas similares](#)

Search [Arturo Soria \(Madrid\) - Sólo lo mejor](#)
MSN La zona **Arturo Soria** de Madrid ya tiene un web. Con bolsa inmobiliaria y artículos sobre medicina, animales, fisioterapia, psicología, inmobiliaria, historia, cine... ¡ Participa! Madrid ...
www.arturosoria.com [Página guardada en caché](#) 05/02/2005

La recuperación de la información parece tardar el mismo tiempo en los dos buscadores. En ocasiones MSN Search incluso supera la rapidez de Google, no obstante no existen grandes diferencias.

En cuanto a los resultados, en ocasiones el primer resultado suele coincidir en ambos buscadores. Determinar qué buscador busca mejor requeriría un análisis más profundo. En principio MSN Search ha indexado 5.000 millones de páginas para esta primera fase y esperará un tiempo antes de ampliar el número de páginas indexadas. Por su parte Google, recientemente duplicó el número de páginas indexadas superando en la actualidad los 8.000 millones. De todos modos, lo que importa no es tanto la cantidad de páginas indexadas como la calidad del algoritmo de búsqueda.

No obstante, el que MSN Search esté considerado uno de los buscadores más utilizados está relacionado directamente con el hecho de que muchos de los usuarios de Windows no saben como cambiar la página de inicio de sus navegadores de Internet, ni tampoco como cambiar el motor de búsqueda instalado por defecto.

A diferencia del resto de buscadores analizados en este sitio, la propia MSN ha publicado una página acerca de su algoritmo:

[Algoritmo de posicionamiento MSN](#)