



Instituto Nacional  
de Tecnologías  
de la Comunicación

# **GUÍA DE RECOMENDACIONES “SEO” PARA POSICIONAMIENTO EN INTERNET**

Septiembre 2007

<b>1</b>	<b>Qué es SEO: Introducción</b>	<b>5</b>
1.1	Qué es SEO	5
1.1.1	Enlaces de interés	6
1.2	Por qué debería preocuparme el SEO	6
<b>2</b>	<b>Cómo funciona un buscador Web</b>	<b>7</b>
2.1	Qué es un buscador	7
2.1.1	Enlaces de interés	7
2.2	Cuáles son los principales buscadores de la red	7
2.2.1	Enlaces de interés	8
2.3	Cómo funciona a grandes rasgos un buscador	8
2.4	Qué procesos realiza un buscador	8
2.5	En qué consiste el rastreo de un buscador	9
2.6	En qué consiste la indexación de un buscador	9
2.6.1	Enlaces de interés	9
2.7	Cómo se resuelve una búsqueda en un buscador	10
<b>3</b>	<b>Criterios de ordenación de resultados en un buscador Web</b>	<b>11</b>
3.1	Qué criterios usa un buscador para ordenar los resultados	11
3.1.1	Enlaces de interés	11
3.2	Cuáles son los criterios de relevancia Extra-Site	12
3.3	Cuáles son los criterios de relevancia Intra-Site	13
3.4	Qué son los Meta-Tags y cuál es su importancia real	14
3.4.1	Ejemplos de uso	14
3.4.2	Enlaces de interés	15
3.5	Qué es el PageRank y cómo funciona	16
3.5.1	Enlaces de interés	16
3.6	Cuáles son los criterios de relevancia que más pesan en la actualidad	16
<b>4</b>	<b>Cómo optimizar un Web para mejorar su posicionamiento en buscadores</b>	<b>18</b>
4.1	Cómo consigo que los buscadores me indexen	18
4.1.1	Enlaces de interés	18
4.2	Cada cuánto me re-indexan los buscadores	18
4.3	Qué son barreras de rastreo	19
4.3.1	Enlaces de interés	20
4.4	Qué es la Web Oculta y por qué es un problema	20
4.4.1	Enlaces de interés	20

<b>4.5</b>	<b>Qué es un Sitemap y para qué sirve.....</b>	<b>20</b>
4.5.1	Ejemplo de Sitemap .....	21
4.5.2	Enlaces de interés .....	21
<b>4.6</b>	<b>Qué es el Robots.txt y para qué sirve .....</b>	<b>21</b>
4.6.1	Ejemplo de Robots.txt .....	22
4.6.2	Enlaces de interés .....	22
<b>4.7</b>	<b>Cómo debe ser una URL para que no plantee problemas a los buscadores .....</b>	<b>22</b>
<b>4.8</b>	<b>Qué es el Spam en buscadores .....</b>	<b>23</b>
4.8.1	Enlaces de interés .....	23
<b>4.9</b>	<b>¿Puedo ser penalizado por los buscadores?.....</b>	<b>23</b>
<b>4.10</b>	<b>Y qué ocurre con el contenido de las páginas, ¿es irrelevante? .....</b>	<b>24</b>
<b>4.11</b>	<b>Qué son las palabras clave y para qué sirven .....</b>	<b>24</b>
4.11.1	Enlaces relacionados .....	24
<b>4.12</b>	<b>Cómo puedo conseguir ranking para mis palabras clave .....</b>	<b>25</b>
<b>4.13</b>	<b>¿Tan Importantes son los Enlaces Externos? .....</b>	<b>25</b>
4.13.1	Enlaces de interés .....	25
<b>4.14</b>	<b>Las arañas cargan demasiado mi servidor, ¿qué puedo hacer? .....</b>	<b>26</b>
<b>5</b>	<b>Optimizando mi sitio Web paso a paso.....</b>	<b>27</b>
<b>5.1</b>	<b>Optimiza tu Sitio: Flujo de trabajo.....</b>	<b>27</b>
<b>5.2</b>	<b>Comprueba que las URLs de tus páginas son adecuadas .....</b>	<b>27</b>
<b>5.3</b>	<b>Revisa cómo están escritos los enlaces en tus páginas .....</b>	<b>27</b>
5.3.1	Enlaces de interés .....	28
<b>5.4</b>	<b>Optimiza los títulos .....</b>	<b>28</b>
<b>5.5</b>	<b>Optimiza el contenido de tus páginas.....</b>	<b>28</b>
5.5.1	Enlaces de interés .....	29
<b>5.6</b>	<b>Mide tu tráfico .....</b>	<b>29</b>
5.6.1	Enlaces de interés .....	29
<b>5.7</b>	<b>Analiza la actividad de las arañas de los buscadores.....</b>	<b>30</b>
5.7.1	Enlaces de interés .....	30
<b>5.8</b>	<b>Mide tu Ranking para tus principales palabras clave.....</b>	<b>30</b>
5.8.1	Enlaces de interés .....	30
<b>5.9</b>	<b>Mide el número de páginas que tienes indexadas.....</b>	<b>31</b>
5.9.1	Enlaces de interés .....	31
<b>5.10</b>	<b>Crea y optimiza un Robots.txt y un Sitemap .....</b>	<b>31</b>
5.10.1	Enlaces de interés .....	31
<b>5.11</b>	<b>Trabaja tus contenidos y consigue enlaces.....</b>	<b>31</b>

5.11.1	Enlaces de interés .....	32
5.12	¿Se puede sobre-optimizar? ¿Qué sucede entonces? .....	33
<b>6</b>	<b><i>Los principales problemas de rastreo en detalle .....</i></b>	<b>34</b>
6.1	URLs construidas con Javascript .....	34
6.2	URLs embebidas en Flash.....	34
6.3	Páginas con Frames.....	34
6.4	Complicaciones asociadas a ciertas páginas dinámicas .....	34
6.5	Formularios.....	35
6.6	Multitud de parámetros en URLs de páginas dinámicas .....	35
6.7	Estructura de enlaces del Sitio Web .....	36
<b>7</b>	<b><i>Más consejos para construir un sitio Web de éxito.....</i></b>	<b>37</b>
7.1	Consejos para la selección del proveedor de “hosting” .....	37
7.1.1	Requisitos mínimos recomendados .....	38
7.2	Consejos para la selección del software de publicación.....	38
<b>8</b>	<b><i>Enlaces y recursos de interés para SEO.....</i></b>	<b>40</b>
8.1	Herramientas SEO .....	40
8.2	Información SEO .....	42
8.3	Conceptos SEO .....	43
<b>9</b>	<b><i>Diccionario de Términos SEO.....</i></b>	<b>45</b>





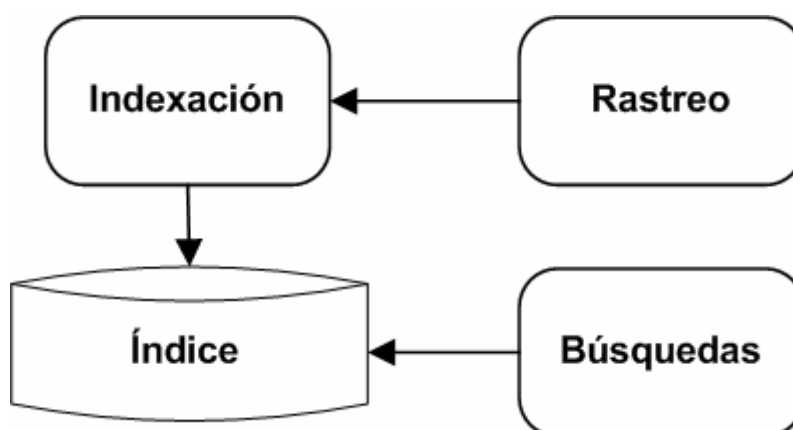


### 2.2.1 Enlaces de interés

- Top Buscadores en el Mundo según Nielsen//Netratings Junio 2007  
<http://searchenginewatch.com/showPage.html?page=3626726>
- Top Buscadores en España  
<http://www.ojobuscador.com/2006/12/09/el-dominio-de-google-en-europa/>
- Search Engine Watch  
<http://www.searchenginewatch.com/>  
Observatorio del universo de los buscadores: una de las principales referencias a nivel mundial en su clase.
- List of Search Engines – Wikipedia  
[http://en.wikipedia.org/wiki/List\\_of\\_search\\_engines](http://en.wikipedia.org/wiki/List_of_search_engines)  
Completo listado de buscadores de todo tipo.

### 2.3 Cómo funciona a grandes rasgos un buscador

Los grandes buscadores generalistas a muy alto nivel de abstracción, tienen un funcionamiento común. En ellos, todo gira alrededor del índice. El índice es la estructura de datos que el buscador consulta para satisfacer las búsquedas de los usuarios. El índice de un motor de búsqueda es análogo al índice de un libro: son referencias inversas que indican dónde se cubren determinados términos o dónde aparecen ciertas palabras. El proceso de construcción del índice se llama indexación.



### 2.4 Qué procesos realiza un buscador

Los procesos fundamentales de un buscador son dos: resolver las búsquedas propiamente dichas (búsqueda) y la construcción/actualización del índice (indexación). Para que una

página de Internet pueda ser ofrecida como resultado en un buscador, debe haber sido indexada previamente. Por tanto, para la construcción/actualización del índice, los buscadores deben recorrer (rastrear) la Web en busca de nuevas páginas.

## **2.5 En qué consiste el rastreo de un buscador**

Para que un buscador pueda devolver los resultados más recientes y relevantes ante las búsquedas de los usuarios es preciso que mantenga el índice lo más actualizado posible. Y para mantener el índice “fresco” debe estar continuamente recorriendo la Red en busca de nuevas páginas y páginas que hayan sido modificadas. ¿Cómo se lleva a cabo este proceso?.

La Web está construida alrededor del concepto de hipervínculo: los enlaces. Tenemos páginas con referencias a otras páginas, y éstas que apuntan a otras a su vez. Los buscadores, para recorrer la Web, llevan a cabo procesos de rastreo para los cuales utilizan unos sistemas llamados arañas, robots o rastreadores. Las arañas avanzan, enlace a enlace, por los contenidos de la Red. Para que pueda realizarse un rastreo lo más eficaz posible, el recorrido de la Red no es lineal, sino que las páginas que más cambian y las más relevantes se rastrean con mayor frecuencia.

## **2.6 En qué consiste la indexación de un buscador**

Como resultado del proceso de rastreo el buscador obtiene una serie de páginas nuevas y/o modificadas que deben ser estudiadas e incorporadas al índice. Este proceso recibe el nombre de indexación. Como parte del proceso, el buscador deberá analizar no sólo las palabras y temáticas de las páginas, sino una serie de criterios que permiten medir su calidad e importancia. O dicho de otra manera, criterios de relevancia. El índice es una estructura de datos especialmente concebida para que las búsquedas se puedan resolver de modo inmediato.

### **2.6.1 Enlaces de interés**

- Wikilearning – Cómo funciona un buscador  
[http://www.wikilearning.com/como\\_funciona\\_un\\_buscador-wkccp-6079-2.htm](http://www.wikilearning.com/como_funciona_un_buscador-wkccp-6079-2.htm)
- Cómo funciona... un motor de búsqueda  
<http://www.pisitoenmadrid.com/blog/2007/04/como-funciona-un-motor-de-busqueda/>

## **2.7 Cómo se resuelve una búsqueda en un buscador**

El buscador recibe del usuario las palabras introducidas en la caja de búsquedas: las que describen su necesidad de información. A partir de ahí, realiza una serie de consultas en el índice para identificar las páginas en que están presentes dichas palabras. Recordemos que el índice ha sido especialmente construido para que la localización de páginas a partir de palabras sea extraordinariamente veloz. Al tiempo que descubre las páginas en cuestión, lee los criterios de relevancia que se insertaron en la fase de indexación y ordena las páginas de acuerdo a dicha relevancia. Por último, compone la página de respuesta al usuario en la que los resultados están ordenados de mayor a menor relevancia.

## 3 CRITERIOS DE ORDENACIÓN DE RESULTADOS EN UN BUSCADOR WEB

---

### 3.1 Qué criterios usa un buscador para ordenar los resultados

Por defecto, los buscadores generalistas de Internet ordenan los resultados por relevancia. Dicho de otra manera: el buscador calcula la probabilidad con que una determinada página va a satisfacer al usuario. A mayor probabilidad, mayor relevancia. Pero, ¿en qué se fija un buscador para decidir la relevancia de una página? ¿cuáles son los criterios de relevancia para ordenar los resultados?.

Esta es una de las preguntas más realizadas en los entornos en los que se discute sobre posicionamiento. Los propios buscadores no revelan esta información por un motivo doble:

- Por un lado, por cuestiones de propiedad industrial. La inferencia de relevancia es uno de los valores que diferencian a un buscador de su competencia.
- Por otro, para evitar que puedan crearse mecanismos artificiales para intentar “engañar” a los buscadores haciéndoles creer que una página es más relevante de lo que en realidad es y así conseguir un mayor ranking y más tráfico.

Cierto es que no existe información muy precisa al respecto de cuáles son estos criterios. Aunque, de la experiencia de la comunidad, se han identificado dos grandes grupos de criterios:

- Los Intra-Site, los que se refieren a información que se puede recolectar desde el propio sitio Web.
- Los Extra-Site, los que se infieren a partir de enlaces externos al propio sitio Web.

Las labores de optimización giran en torno a ambos tipos de criterios para conseguir una puntuación lo más elevada posible y así escalar posiciones en las páginas de resultados.

#### 3.1.1 Enlaces de interés

- Search Engine Ranking Factors  
<http://www.seomoz.org/article/search-ranking-factors>  
Uno de los listados más completos de factores de ranking
- Ranking Factors Simplified Quick Reference  
<http://www.searchenginejournal.com/ranking-factors-simplified-quick-reference/4764/>

- Most Important Search Engine Ranking Factors  
<http://www.quickonlinetips.com/archives/2007/04/most-important-search-engine-ranking-factors/>
- Ranking The SEO Ranking Factors  
<http://searchengineland.com/070403-145435.php>

### 3.2 Cuáles son los criterios de relevancia Extra-Site

Son los factores que los buscadores consideran para inferir la relevancia de una página y que se basan en datos que recogen FUERA de la propia página y/o sitio a que pertenece. Básicamente se analizan los enlaces externos que apuntan a las páginas del sitio Web y el texto de dicho enlace (texto ancla). En dicho análisis se tienen en cuenta factores como:

- El número de enlaces externos, si bien hay que tener en cuenta que no todos los enlaces tienen el mismo peso. No es lo mismo tener un enlace externo desde un sitio de alta relevancia que desde uno completamente anónimo.
- La relevancia de las páginas en la que se encuentran dichos enlaces. Y no solamente su relevancia en términos absolutos, sino también considerando la temática general de cada uno de los sitios. Por ejemplo, un enlace de un Web de alta relevancia sobre bricolaje a una ferretería cuenta mucho más que un enlace de dicho Web a una tienda de animales domésticos.
- Las palabras que se han usado en el enlace. Se tienen muy en cuenta. De hecho en ocasiones los buscadores muestran como resultado para una búsqueda dada, páginas que no contienen el término buscado. En su lugar hay un enlace externo que apunta a dicho resultado con los términos introducidos en la búsqueda. Pongamos un ejemplo: tenemos una página sobre alquiler de apartamentos en Almuñecar. Supongamos que dicha página no contiene la palabra 'reserva'. Y supongamos también que hay un enlace externo a dicha página con el texto ancla 'reserva de apartamentos'. Pues bien: por la importancia que conceden los buscadores a las palabras presentes en el texto ancla, podría darse el caso en que nuestra página de alquiler de apartamentos en Almuñecar apareciese como resultado para la búsqueda 'reserva de apartamentos'.
- La diversidad y dispersión de dichas palabras. Existen mecanismos no lícitos que permiten la creación automática de enlaces con el objeto de engañar a los buscadores y conseguir enlaces. Estos enlaces generados automáticamente suelen tener un texto ancla más o menos fijo. Sin embargo, los enlaces "normales" generados por humanos suelen tener textos ancla ligeramente diferentes. Los

buscadores han desarrollado mecanismos para diferenciar enlaces automáticos de enlaces naturales mediante el análisis de la dispersión de los textos ancla.

### 3.3 Cuáles son los criterios de relevancia Intra-Site

Son los factores que los buscadores consideran para inferir la relevancia de una página y que se basan en datos que recogen DENTRO de la propia página y/o sitio a que pertenece. Se trata, básicamente, de información de estructura del sitio y cómo están descritos los contenidos. Se tienen en cuenta:

- Localidad de la búsqueda. Puedes hacer la prueba si quieres. Busca algo en google.es y luego lo mismo en google.com. Verás que los resultados son diferentes: los de google.es están localizados. Es decir: en google.es se ha dado un plus de relevancia a los sitios con contenidos y/o servicios en español y para España.
- La antigüedad del dominio. Es una tendencia más o menos reciente. Es una de las medidas adoptadas para luchar contra el spam. En ocasiones supone un freno para un mejor posicionamiento de sitios de reciente creación.
- Las palabras clave que aparecen en el nombre del dominio. Tradicionalmente se le ha dado mucha importancia a este factor. Fundamentalmente el dominio, no tanto el subdominio. Ese es el motivo de que, en ocasiones, nos encontremos sitios Web con nombres como [www.maquinas-tragaperras-de-segunda-mano.com](http://www.maquinas-tragaperras-de-segunda-mano.com), cuyo principal objetivo es competir en posicionamiento por la búsqueda que coincide exactamente con su nombre de dominio.
- Las palabras clave que aparecen en el resto de la url. También tienen un peso significativo. Además, algunos estudios aseguran que cuando la url contiene los términos de búsqueda, los usuarios de los buscadores tienen una tendencia superior a hacer clic en ellos.
- Las palabras que aparecen en el título (<title>). Tiene un efecto doble beneficioso. Por un lado, los buscadores asocian gran parte de la carga descriptiva de un contenido al campo <title>. Por otro es el campo que aparece como título de un resultado en las páginas de resultados de los buscadores y, el hecho de contener los términos de búsqueda es un aliciente añadido para hacer clic.
- La utilización de secciones h1, h2, etc. Los buscadores interpretan estas etiquetas como texto relevante y descriptivo de las secciones de una página. Conviene tener en cuenta que, aunque lo valoran, penalizan utilizaciones abusivas. Como regla, no debería utilizarse más de un h1 por página.

- La frecuencia de aparición de las palabras. Se basa en el hecho de que, por ejemplo, una página muy relevante sobre quesos, utiliza muy a menudo en sus contenidos la palabra 'queso'. Igual que antes, conviene huir del uso excesivo y antinatural de un término para tratar de posicionarse para él.
- La anatomía de enlaces internos dentro del sitio Web. La estructura de enlaces internos también se usa. Las páginas de menor profundidad en un sitio tienden a considerarse más relevantes (sin tener en cuenta factores Extra-Site). También conviene tener en cuenta un par de detalles. Por un lado, a los buscadores les suelen gustar estructuras más bien aplanadas, sin muchos niveles de profundidad. Por otro, no conviene tener páginas con muchos enlaces. Algunos expertos sitúan la cifra de corte en torno a los no más de 100 enlaces por página.

### 3.4 Qué son los Meta-Tags y cuál es su importancia real

Los meta-tags son una serie de etiquetas que se escriben en la sección de cabecera de las páginas HTML, que no son visibles para el usuario final en el navegador, pero sí lo son para los buscadores. Se sitúan en la sección <head> de la página Web en la que se encuentran, típicamente tras el <title>. Algunas de las más importantes para los buscadores eran:

- Keywords
- Description
- Robots

```
<!DOCTYPE html PUBLIC
<html lang="es" xml:la
<head>
<title>Mi Sitio</title
<meta http-equiv="Cont
<meta name="descriptio
<meta name="keywords"
<meta name="robots" co
```

En la actualidad, y debido al abuso ilícito (spam) de los metas, los buscadores tienden a obviarlos. Únicamente se utilizan las etiquetas de robots como guía para el rastreo de las páginas y la descripción (description). El contenido de dicha etiqueta (description) se utiliza en ocasiones como snippet (pequeño texto que describe el contenido del resultado tras una búsqueda). En cualquiera de los casos, como no viene mal, la práctica recomendada es utilizarlas para describir el contenido real de las páginas.

#### 3.4.1 Ejemplos de uso

**Meta Keywords:** contiene palabras clave que describen la naturaleza del contenido de la página en la que se encuentra. Los principales buscadores no la tienen en cuenta hoy por hoy. Ejemplo:

```
<meta name="keywords" content="palabra1, palabra2, palabra3" />
```

**Meta Description:** el texto de este campo se suele utilizar como snippet en búsquedas genéricas. Merece la pena prestarle atención. Ejemplo:

```
<meta name="description" content="Aqui describo mi sitio" />
```

**Meta Robots:** contiene información sobre cómo deben rastrear los buscadores la página. Puede tomar diferentes valores:

- Index, Noindex. Indica a los buscadores si pueden indexar la página.
- Follow, Nofollow. Indica a los buscadores si pueden progresar por los enlaces que contiene la página.
- None. Indica a los buscadores que deberían ignorar la página.
- Noarchive. Indica a los buscadores que no deben almacenar copia de cache de la página. Por defecto, los buscadores ofrecen a sus usuarios la posibilidad de descargar cada una de las páginas resultado de un almacén temporal donde se guarda una copia de la página, llamada cache, tal y como estaba cuando fue rastreada.

Por ejemplo, si se quiere que una página sea indexada, que no avance por los enlaces a los que ésta apunta y que no se guarde copia de cache, habría que incluir:

```
<meta name="robots" content="index,nofollow,noarchive" />
```

En el más común de los casos, se querrá que los buscadores indexen y progresen por los enlaces de una página dada. Entonces, debería indicarse así.

```
<meta name="robots" content="index, follow" />
```

### 3.4.2 Enlaces de interés

- SearchEngineWatch: How to Use HTML Meta Tags  
<http://searchenginewatch.com/showPage.html?page=2167931>  
Artículo detallado con ejemplos sobre el correcto uso de los metas.
- Useful HTML Meta Tags  
<http://www.i18nguy.com/markup/metatags.html>  
Descripción detallada del modo de uso de los diferentes metas.

### 3.5 Qué es el PageRank y cómo funciona

El PageRank o PR es un algoritmo de análisis de enlaces cuyo objeto es el cálculo de la importancia de una página en función

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

del número de enlaces que contiene y recibe de otras fuentes. Fue desarrollado por Sergey Brin y Larry Page, fundadores de Google en el año 1995 y supuso una novedosa técnica de inferencia de relevancia. Durante varios años el posicionamiento en buscadores giraba alrededor del PR, hasta que el desarrollo de técnicas de spam de PR (granjas de enlaces, políticas de intercambio, etc) acabaron con su supremacía.

En Google todas las páginas indexadas tienen asociado un índice, el PR, que oscila de 0 a 10. No es una escala lineal, por lo que el salto entre PR7 y PR8 es mucho mayor que entre PR2 y PR3. En la actualidad, el PR se sigue utilizando como criterio de relevancia, pero su peso en el global es mucho menor de lo que fue en su tiempo.



#### 3.5.1 Enlaces de interés

- Page Rank Checker  
[http://www.prchecker.info/check\\_page\\_rank.php](http://www.prchecker.info/check_page_rank.php)
- Google Toolbar  
<http://toolbar.google.com/>  
Entre otras funcionalidades, te permite chequear el PageRank de la página que estás visitando.
- Como Funciona un Buscador  
[http://www.aulacli.com/internet/t\\_4\\_7.htm](http://www.aulacli.com/internet/t_4_7.htm)

### 3.6 Cuáles son los criterios de relevancia que más pesan en la actualidad

Debido al trabajo incansable de los spammers los criterios de relevancia están en continua evolución. Más que pensar en términos de cuáles son los criterios que más se consideran en la actualidad conviene pensar en términos de construir un sitio Web con contenido de calidad y sin barreras de rastreo. Los buscadores encaminan sus esfuerzos en el desarrollo de algoritmos que infieran de una manera precisa y sin errores, el valor real y la calidad de los contenidos. Es una buena política jugar a “convencer” a los buscadores más

que a “engañarlos”. No obstante, y para poder entender por qué el orden que muestran los buscadores es el que es, no debes perder de vista que pesa considerablemente:

- Que los términos buscados estén en el nombre del dominio
- La antigüedad del dominio
- El número y la relevancia de los enlaces externos
- El texto utilizado en los enlaces externos y la dispersión del mismo

## 4 CÓMO OPTIMIZAR UN WEB PARA MEJORAR SU POSICIONAMIENTO EN BUSCADORES

---

### 4.1 Cómo consigo que los buscadores me indexen

Algunos buscadores cuentan con formularios de solicitud de inclusión y otros no. Lo que funciona en cualquier caso es obtener enlaces desde sitios Web que ya están indexados para que los rastreadores recorran e indexen las páginas de tu sitio. En función de diferentes parámetros (antigüedad del dominio, relevancia, velocidad de actualización de los contenidos) el número de páginas indexadas y su frecuencia de rerrastreo será mayor. No vayas a pensar que la indexación es todo o nada. Normalmente empiezan indexando únicamente la página principal de tu sitio y conforme va pasando el tiempo y consigues más enlaces externos va aumentando el número de páginas indexadas.

#### 4.1.1 Enlaces de interés

- Formulario de Inserción de Google  
<http://www.google.com/addurl/>
- Formulario de Inserción de Yahoo  
<http://search.yahoo.com/info/submit.html>
- Formulario de Inserción de Live  
<http://search.live.com/docs/submit.aspx>

### 4.2 Cada cuánto me re-indexan los buscadores

Las arañas de los buscadores re-visitarán tu sitio Web en busca de nuevos contenidos y páginas en función de varios parámetros. Los factores que más afectan son los siguientes:

- La relevancia de tus páginas
- La velocidad de actualización de tus contenidos

Los buscadores suelen conceder un plus de relevancia a los contenidos más frescos, por lo que es interesante trabajar para conseguir que tus contenidos se reindexen a la mayor velocidad posible. Por un lado, tendrás tus últimos contenidos disponibles desde los buscadores. Por otro, conseguirás subir algún que otro puesto en el ranking de resultados.

### 4.3 Qué son barreras de rastreo

Para poder salir en las páginas de resultados de los buscadores, un sitio Web debe haber sido indexado previamente. Y para que pueda haber sido indexado, es necesario que una araña o robot lo haya rastreado; es decir, que haya descargado sus páginas y progresado por sus enlaces. Las barreras de rastreo son las dificultades técnicas que encuentran las arañas y que hacen imposible que los buscadores puedan rastrear un sitio Web. Dicho de otro modo, las barreras de rastreo impiden que un sitio Web sea visible para los buscadores.

Los robots han sido diseñados para poder realizar el rastreo de modo eficiente y muy rápido y no tienen la misma capacidad que un navegador para interpretar las páginas. En consecuencia, hay contenidos que un buscador no puede ver y enlaces que no puede seguir. En general, son problemáticos para los buscadores:

- El código Javascript. En especial los enlaces a páginas construidos con Javascript. Los principales buscadores no podrán seguir estos enlaces, así que conviene tener mecanismos alternativos a través de los cuales sí puedan llegar los buscadores.
- El Flash. Los buscadores comienzan a hacer tímidos movimientos en la indexación del contenido textual de los Flash. Desde luego, lo que no pueden indexar es el texto embutido en los Flash como imágenes. Y, hasta la fecha, no parecen progresar por los enlaces insertados en un componente Flash.
- Los Applets Java. Los buscadores no tienen capacidad para ejecutar los applets, por lo que no podrán hacer nada con estos elementos.
- El contenido textual en imágenes. Sea en imágenes en Flash o en otros componentes, cuando el texto está recogido como imagen, no puede ser reconocido ni indexado por los buscadores.
- Los formularios. Los formularios ofrecen páginas como resultado de haber introducido información en sus campos. Las arañas de los buscadores sólo siguen enlaces: no saben insertar información significativa en los campos de los formularios para obtener páginas.

La regla de oro en este caso es hacer posible que toda página pueda ser accesible (por medio de uno o más saltos) desde cualesquiera otra por medio de enlaces estáticos de tipo `<a href=`

#### 4.3.1 Enlaces de interés

- Directrices de Google para Webmasters  
<http://www.google.com/support/Webmasters/bin/answer.py?answer=35769>
- Yahoo Search Content Quality Guidelines  
<http://help.yahoo.com//us/yahoo/search/basics/basics-18.html>
- Guidelines for Successful Indexing  
[http://search.msn.com/docs/siteowner.aspx?t=SEARCH\\_WEBMASTER\\_REF\\_GuidelinesforOptimizingSite.htm](http://search.msn.com/docs/siteowner.aspx?t=SEARCH_WEBMASTER_REF_GuidelinesforOptimizingSite.htm)  
Guía de recomendaciones de Live para ser indexado.
- Poodle Predictor  
<http://www.gritechnologies.com/tools/spider.go>  
Herramienta que te ofrece información sobre cómo ven las arañas de los buscadores tu sitio Web.

#### 4.4 Qué es la Web Oculta y por qué es un problema

La Web Oculta es la parte de Web que queda “al otro lado” de las barreras de rastreo, y que los buscadores no pueden rastrear ni indexar. Puesto que los buscadores son los grandes canalizadores de tráfico, pertenecer a la Web Oculta es sinónimo de no poder disfrutar de tráfico directo procedente de los buscadores, lo cual significa, en la mayoría de los casos, de no poder disfrutar de tráfico.

##### 4.4.1 Enlaces de interés

- Deep Web en la Wikipedia  
[http://en.wikipedia.org/wiki/Hidden\\_Web](http://en.wikipedia.org/wiki/Hidden_Web)

#### 4.5 Qué es un Sitemap y para qué sirve

Sitemap o Mapa del Web es un concepto con una doble acepción. Tradicionalmente se trataba de una página HTML más de un sitio Web en la que se recogía la estructura principal del sitio con enlaces a las diferentes secciones. En tiempos más recientes, se usa el término sitemap para referirse a un fichero en formato XML generado para los buscadores. Contiene información sobre las páginas disponibles y frecuencia de actualización de sus contenidos.

Por un lado, los sitemaps XML son de inestimable ayuda para salvar problemas de barreras de rastreo y Web oculta. Por otro, los sitemaps HTML están concebidos más

como ayuda al usuario humano de un sitio Web, aunque, correctamente implementados, también pueden servir de ayuda a los buscadores para facilitar el rastreo e indexación de páginas.

#### 4.5.1 Ejemplo de Sitemap

Este es un ejemplo de un sitemap que contiene 2 URLs.

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://www.ejemplo.com/1</loc>
    <lastmod>2007-06-01</lastmod>
    <changefreq>weekly</changefreq>
    <priority>0.9</priority>
  </url>
  <url>
    <loc>http://www.ejemplo.com/2</loc>
    <lastmod>2007-01-01</lastmod>
    <changefreq>monthly</changefreq>
    <priority>0.7</priority>
  </url>
</urlset>
```

#### 4.5.2 Enlaces de interés

- Cómo se Construye un Sitemap  
<https://www.google.com/Webmasters/tools/docs/es/protocol.html>
- Sitemap Generator de Google  
<https://www.google.com/Webmasters/tools/docs/es/sitemap-generator.html>  
Se trata de un pequeño programa que genera automáticamente un sitemap a partir de tu sitio Web.
- Sitio Web de Referencia del Estándar Sitemap  
<http://www.sitemaps.org/>

#### 4.6 Qué es el Robots.txt y para qué sirve

El robots.txt es un fichero que suele estar disponible en la raíz de un sitio Web (es decir, <http://miWebsite/robots.txt>). Se trata de un fichero que las arañas de los buscadores piden antes de descargar ninguna página de un sitio Web. En el fichero robots.txt se pueden especificar qué páginas y/o directorios se pueden o no indexar en los buscadores. El fichero robots.txt suele utilizarse con diferentes propósitos:

- Para evitar que cierta información sensible esté indexada en los buscadores.

- Para evitar que información poco útil se indexe y concentrar la actividad del buscador sobre las páginas más importantes
- Para especificar la ruta del sitemap XML

Muchos sitios Web no disponen de robots.txt. En tal caso, los buscadores asumen que el propietario de dicho sitio no impone ningún tipo de reglas al rastreo e indexación.

#### 4.6.1 Ejemplo de Robots.txt

A continuación se incluye un ejemplo de robots.txt: incluye una referencia al sitemap del sitio Web e indica a los robots que no deben indexar el /admin.php ni nada del directorio /admin.

```
User-agent: *  
Disallow: /admin.php  
Disallow: /admin/  
Sitemap: http://misitio.com/sitemap
```

#### 4.6.2 Enlaces de interés

- Robots.txt Generator  
<http://www.mcanerin.com/EN/search-engine/robots-txt.asp>  
Herramienta que te permite generar automáticamente un robots.txt a partir de la información que introduces en un formulario.
- Sitio Web del Estándar Robots.txt  
<http://www.robotstxt.org/>

### 4.7 Cómo debe ser una URL para que no plantee problemas a los buscadores

La URL es la dirección bajo la cual está accesible una determinada página. Piensa por un momento que las páginas de resultados de un buscador incluyen referencias a tus páginas y a otras páginas. Es por tanto imprescindible que una URL lleve siempre al mismo contenido. Es decir: que la correspondencia en URL y contenido sea biunívoca. En ocasiones, cuando se utiliza javascript o se manejan cookies y variables de sesión de una manera no adecuada, la página que encontró la araña del buscador cuando estaba rastreando tu sitio es distinta de la que se va a encontrar el usuario cuando acceda desde la página de resultados. Asegúrate de que la página que se obtiene cuando se mete cualquier URL en el cajetín de tu navegador es la misma en todo momento.

Por otro lado, y de cara al posicionamiento, a los buscadores les suele gustar que:

- La URL contenga las palabras que el usuario ha buscado. Si es en el nombre del dominio de tu sitio Web, mejor que mejor.
- La URL no contenga más de 2 o 3 parámetros. Los parámetros de una URL <http://midominio/dir1/fich1?par1=val1&par2=val2> son par1 y par2. Existen técnicas de configuración avanzada de servidores para la ocultación de parámetros (reescritura de urls) pero su descripción queda fuera del alcance de la presente guía.

## 4.8 Qué es el Spam en buscadores

Existe un valor económico claro asociado a un alto ranking en las búsquedas. No en vano, los buscadores cuentan con programas publicitarios que permiten situar pequeños anuncios en las partes de mayor visibilidad de las páginas de resultados. Como consecuencia, muchos miembros de los sectores más oscuros de la Red están continuamente ideando técnicas para “engañar” a los buscadores y hacerles creer que ciertas páginas tienen más relevancia de la que en realidad tienen. Estas páginas reciben el nombre de spam. A lo largo de la historia de los buscadores se han sucedido diferentes episodios de spam. En la actualidad, el spam es una de las principales amenazas para la calidad de los resultados de búsqueda y el negocio de los buscadores. Es por ello que los grandes de la búsqueda dedican gran parte de su esfuerzo y talento a desarrollar algoritmos y técnicas que no sean “spammeables”.

### 4.8.1 Enlaces de interés

- Spamdexing en la Wikipedia  
<http://en.wikipedia.org/wiki/Spamdexing>

## 4.9 ¿Puedo ser penalizado por los buscadores?

Los buscadores disponen de mecanismos automáticos para la detección de prácticas ilícitas para mejorar artificialmente el ranking en las páginas de resultados. Del mismo modo, cuentan también con páginas que permiten a los usuarios finales denunciar el spam. Generalmente, la reacción de los buscadores frente a estos casos es la penalización que puede variar desde reducir el ranking en x posiciones a todas las páginas del sitio Web infractor hasta su eliminación completa del índice.

#### 4.10 Y qué ocurre con el contenido de las páginas, ¿es irrelevante?

En absoluto es irrelevante. Piensa que los buscadores lo que hacen es tratar de desarrollar técnicas para decidir qué contenidos son realmente relevantes y cuáles no. Hay que trabajar en la línea de tratar de facilitar al máximo el rastreo por parte de los buscadores y describir correctamente la información. Pero, en absoluto es menos importante trabajar en la línea de generar contenidos realmente relevantes y útiles. Los enlaces externos son uno de los criterios de relevancia que más pesan en los buscadores porque precisamente hablan muy claro al respecto de la calidad y utilidad de las páginas: las más enlazadas son las más relevantes. Una de las frases célebres en la Red es *Content is King* (“el contenido es el rey”).

En consecuencia, debes pensar a quién quieres dirigirte con tu sitio Web y pensar en cuáles son sus necesidades. A continuación generar contenido de valor para ellos, pequeñas o grandes muestras de contenido, pero algo que le aporte valor. Poco a poco, ese valor se traduce en enlaces externos y estos enlaces en ranking en los buscadores.

#### 4.11 Qué son las palabras clave y para qué sirven

Las palabras clave son los términos más representativos de un contenido. Las necesidades de la audiencia potencial de tu sitio giran en torno a las palabras clave. Típicamente son estas palabras las que se usan para realizar búsquedas en Internet. Por ejemplo, las palabras clave de un sitio Web de una tienda de televisores pueden ser, entre otras: televisiones, plasma, tv samsung, televisores sony, televisiones baratas, etc. Las palabras clave de tu negocio que más se busquen y más se traduzcan en compras finales serán las que más te interesen, las que más valor tenga para ti, y para las que más te interese estar bien posicionado. Una de las tareas clave del posicionamiento es identificar tus palabras clave y competir por el ranking en los buscadores para las mismas.

##### 4.11.1 Enlaces relacionados

- SEO Book Keyword Suggestion Tool  
<http://tools.seobook.com/general/keyword/>  
Herramienta que te ayuda a identificar palabras clave para tu negocio.
- Yahoo Search Keyword Terms Extraction Tool  
<http://tools.seobook.com/yahoo-keywords/>  
Herramienta que te permite extraer las palabras clave contenidas en un párrafo.
- Keyword Density Analyzer  
<http://tool.motoricerca.info/keyword-density.phtml>  
Herramienta que te calcula la frecuencia de palabras en una determinada página.

## 4.12 Cómo puedo conseguir ranking para mis palabras clave

Esa es la clave del posicionamiento. La respuesta es clara: consiguiendo que los buscadores te consideren relevante para esos términos. ¿Y cómo lo hago? Eliminando las barreras de rastreo, generando contenido de calidad y prestando un buen servicio. En otras secciones de esta guía obtendrás información más detallada sobre cómo conseguirlo.

## 4.13 ¿Tan importantes son los Enlaces Externos?

Son ciertamente importantes, no sólo los enlaces sino también el texto ancla (el texto que se utiliza en el enlace). Pero ojo: no vale cualquier enlace. Piensa que, en los últimos tiempos, se ha abusado de prácticas no lícitas para conseguir enlaces externos y los buscadores han diseñado técnicas para distinguir enlaces válidos de los que no lo son. En concreto, consideran lo siguiente:

- La diversidad de los textos ancla. En los planes de intercambio de enlaces generalmente se utiliza siempre el mismo texto ancla. Cuando los enlaces se generan espontáneamente, el texto ancla no suele ser idéntico. En su lucha contra el spam, los buscadores han empezado a analizar factores como éste para poder distinguir entre un enlace realmente relevante y uno que no lo es.
- El origen de los enlaces. Los enlaces procedentes de tu vecindario temático pesan mucho más. Raramente un enlace en un sitio Web sobre animales domésticos apuntando a un sitio Web sobre pastelería reflejará relevancia.
- La distribución de los enlaces en el tiempo. La evolución de los nuevos enlaces externos que se generan a una página es otro de los patrones que valoran los buscadores.

Por la gran importancia que tienen los enlaces externos a tu sitio, es una práctica interesante monitorizar regularmente los enlaces que van apareciendo en la Red y apuntan a tus páginas. Es una de las principales referencias que tienes para poder entender si tus esfuerzos tienen éxito o no.

### 4.13.1 Enlaces de interés

- Yahoo Site Explorer  
<http://siteexplorer.search.yahoo.com/>  
Herramienta que te permite descubrir los enlaces entrantes que tiene tu sitio Web.

- Link Popularity Check

<http://www.marketleap.com/publinkpop/>

Completa herramienta que te informa acerca de la densidad de enlaces que tu sitio recibe.

#### **4.14 Las arañas cargan demasiado mi servidor, ¿qué puedo hacer?**

Antes de nada, si se trata de las arañas de los grandes buscadores de Internet ¡enhorabuena!. Eso significará que consideran tu sitio relevante y se traducirá luego en tráfico procedente de las páginas de resultados. Merece la pena que dediques esfuerzo a optimizar tus aplicativos web para que sean más eficientes y/o contratar un paquete de hosting de prestaciones superiores.

En el caso de que se trate de arañas desconocidas, si quieres puedes identificar la IP de la que proceden y prohibir su acceso en la configuración del firewall de tu sitio. Si no sabes como hacerlo, ponte en contacto con tu proveedor de hosting.

## 5 OPTIMIZANDO MI SITIO WEB PASO A PASO

---

### 5.1 Optimiza tu Sitio: Flujo de trabajo

Para ayudarte en el proceso de optimización de tu sitio Web para buscadores, te proponemos un flujo de trabajo y los elementos más importantes que deberás tener en cuenta en cada paso.

1. Comprueba que las URLs de tus Páginas son Adecuadas.
2. Revisa Cómo están Escritos los Enlaces en tus Páginas.
3. Optimiza los Títulos.
4. Optimiza el Contenido de Tus Páginas.
5. Mide tu Tráfico.
6. Analiza la Actividad de las Arañas de los Buscadores.
7. Mide tu Ranking para tus Principales Palabras Clave.
8. Mide el Número de Páginas que Tienes Indexadas.
9. Crea y Optimiza un Robots.txt y un Sitemap.
10. Trabaja Tus Contenidos y Consigue Enlaces.

### 5.2 Comprueba que las URLs de tus páginas son adecuadas

Para que un buscador pueda rastrear tu sitio correctamente es imprescindible que las URLs de las páginas de tu sitio sean únicas y que devuelvan siempre el mismo resultado independientemente de la sesión y otros factores. Procura que no contengan muchos parámetros y que sean lo más cortas posibles. Para más información, puedes echar un vistazo al Apartado 6. Si quieres ir un paso más allá en la optimización considera la labor técnica de reescritura de urls.

### 5.3 Revisa cómo están escritos los enlaces en tus páginas

Lo más importante que debes tener en cuenta es que los buscadores no saben leer ni Javascript ni Flash como lo hace tu navegador. En consecuencia, todos los enlaces de tus

páginas deberían estar descritos en HTML. Estos son algunos ejemplos de enlaces válidos:

- `<a href="http://dominio/dir/fich.html">texto ancla</a>`
- `<a href="/dir/fich.html">texto ancla</a>`
- `<a href="../fich.html">texto ancla</a>`

Debes asegurarte de que desde la página principal de tu sitio se puede llegar a cualquier otra página siguiendo enlaces de este tipo. Si no fuera posible, crea enlaces adicionales desde tus páginas con enlaces de este tipo para favorecer que los rastreadores puedan progresar por los contenidos de tu sitio.

### 5.3.1 Enlaces de interés

- Xenu's Link Sleuth  
<http://home.snafu.de/tilman/xenulink.html>  
Software que te ayuda a descubrir los enlaces rotos que hay en tu sitio.

## 5.4 Optimiza los títulos

El criterio intra-site de mayor importancia es el campo `<title>` de una página. Busca títulos lo más descriptivos posible y asegúrate de que todas las páginas de tu sitio tienen títulos diferentes. Trata de que los títulos contengan palabras clave de tu negocio. Cada página debería tratar de estar optimizada para 2 o 3 palabras clave como mucho. Procura no utilizar más, pues entonces la optimización no tendrá el mismo efecto.

## 5.5 Optimiza el contenido de tus páginas

Ofrece contenido de calidad en tus páginas. En la Web, eres lo que publicas. A la hora de escribir el contenido de cada página, piensa en las palabras clave para las que quieres posicionar esa página y procura usarlas con frecuencia. Aquí hay un puñado de consejos que te pueden ayudar:

- Utiliza etiquetas `<h1>`, `<h2>`, etc. para definir títulos de secciones etc.
- Trata de escribir en el lenguaje que utiliza el público al que te diriges.
- Rellena los Meta-Tags con información significativa.
- Procura que tu código HTML sea correcto (que cumpla el estándar).

Existen algunas herramientas que realizan un análisis del contenido de las páginas e informan de la frecuencia de las diferentes palabras. Procura que las palabras clave de dicha página estén entre las más frecuentes y aparezcan en los títulos de las secciones. De todos modos, no fuerces artificialmente el texto: piensa que también lo van a leer personas.

### 5.5.1 Enlaces de interés

- Keyword Density Analyzer  
<http://tool.motoricerca.info/keyword-density.phtml>  
Herramienta que te calcula la frecuencia de palabras en una determinada página.
- W3C HTML Validation Service  
<http://validator.w3.org/>  
Herramienta que te permitirá descubrir los errores HTML de tus páginas
- Search Engine Spider Test  
<http://tools.seobook.com/general/spider-test/index.php>  
Herramienta que te indica datos como el número de palabras de la página, la frecuencia de las mismas, el tamaño de la página, los enlaces que hay en ella, etc.

## 5.6 Mide tu tráfico

Si no tienes una herramienta de analítica Web, instala una ya. Desde aquí te recomendamos Google Analytics: es una herramienta de muy elevadas prestaciones y además gratuita. Te será muy sencilla de instalar puesto que solamente deberás incorporar unas líneas de código HTML en cada página que quieras monitorizar. Con cierta regularidad echa un vistazo a las estadísticas de acceso a tu sitio Web, cuáles son los contenidos que mejor están funcionando, las palabras clave que más tráfico te están dando, de dónde procede tu tráfico, etc. Identifica dónde estás y dónde no estás cara a tenerlo en cuenta en tus labores de optimización.

### 5.6.1 Enlaces de interés

- Google Analytics  
<http://www.google.com/analytics/>  
Probablemente la mejor herramienta gratuita para analizar el tráfico de tu sitio Web.
- Crazy Egg  
<http://crazyegg.com/>

Servicio que genera un heatmap de tus páginas. Te permite conocer cuáles son las áreas en que más hace clic tu audiencia.

## 5.7 Analiza la actividad de las arañas de los buscadores

Te interesa saber qué páginas se han descargado las arañas, cuándo lo han hecho, y, sobre todo, si han tenido problemas en ese proceso. Podrás hacerlo a través de los logs de acceso de tu servidor quedándote con aquéllos que corresponden a user-agents de los robots. En el caso de Google, podrás realizar un seguimiento más cómodo con una herramienta que proporciona el principal buscador en España.

### 5.7.1 Enlaces de interés

- Google Webmaster Tools

<https://www.google.com/Webmasters/tools/>

Herramienta de Google que te permitirá conocer cada cuánto tiempo la araña de Google (el Googlebot) rastrea tu sitio, los problemas que ha encontrado, etc.

## 5.8 Mide tu Ranking para tus principales palabras clave

Realiza chequeos periódicos de la posición que ocupas en los buscadores frente a tus principales palabras clave. Realiza un seguimiento y anota la evolución para saber si tus últimas acciones han tenido un efecto favorable. Para automatizar el proceso podrás utilizar una serie de herramientas, pero deberás tomar sus datos con precaución. Los buscadores modernos tienen diferentes rankings en función del país desde donde se busque. Las herramientas de chequeos de ranking te van a mostrar la posición de tus resultados que ellos ven, desde el país en que se encuentran.

### 5.8.1 Enlaces de interés

- Search Engine Keyword Tracker

<http://www.digitalpoint.com/tools/keywords/>

Herramienta gratuita de Digital Point que te permite monitorizar el ranking de tu sitio para tus principales palabras clave.

- SEO for Firefox

<http://tools.seobook.com/firefox/seo-for-firefox.html>

Add-on para Firefox que te permitirá ver los principales criterios de relevancia para cada una de las URL's que aparecen en las páginas de resultados.

## 5.9 Mide el número de páginas que tienes indexadas

Es un índice muy interesante a manejar. Si es más bajo de lo que debería, trata de ver si tu sitio tiene barreras de rastreo. Trabaja para conseguir enlaces externos. Cuantos más enlaces externos tengas, los buscadores estarán más interesados en rastrear tus contenidos.

### 5.9.1 Enlaces de interés

- Yahoo Site Explorer  
<http://siteexplorer.search.yahoo.com/>  
Herramienta de Yahoo que te permite ver el número de páginas que tiene indexadas de tu sitio y los enlaces externos que tienes.
- Google Webmasters Tools  
<http://www.google.com/Webmasters/>  
Herramientas de Google que te dan multitud de información sobre el modo en que Google rastrea tu sitio.

## 5.10 Crea y optimiza un Robots.txt y un Sitemap

Crea un robots.txt para indicarle a los buscadores dónde deben rastrear y dónde no. Si hay páginas que no tiene sentido indexar, exclúyelas en el robots. Muchos expertos aseguran que es preferible centrar la actividad de las arañas en las páginas realmente interesantes. Crea un sitemap y referéncialo desde el robots.txt para facilitar la tarea a las arañas de los buscadores.

### 5.10.1 Enlaces de interés

- Google Webmasters Tools  
<http://www.google.com/Webmasters/>  
Contiene herramientas que te ayudan a escribir tu robots.txt y ver qué páginas bloquea. También ofrece información de la última vez que Google analizó tu sitemap y si encontró errores en el proceso.

## 5.11 Trabaja tus contenidos y consigue enlaces

Una vez que has comprobado que tu sitio no tiene barreras de rastreo, este puede ser uno de los puntos más interesantes. Y es que, uno de los criterios que más consideran los

buscadores para darte un alto ranking son tus enlaces externos. Pero, ¿cómo se consiguen enlaces externos?. Hay varias aproximaciones para hacerlo:

- Dando de alta tu sitio en directorios. Los directorios son recursos de la Red estructurados en categorías con enlaces a diferentes sitios Web. Lo que tienes que hacer para aparecer en ellos es rellenar un formulario y, a menudo, pagar una cuota que puede ser fija o periódica. Tradicionalmente, los directorios eran un mecanismo que funcionaba muy bien, pero en la actualidad, los buscadores tienden a valorar cada vez menos la aparición en los directorios.
- Intercambiando enlaces. Se basa en acuerdos a los que puedes llegar con otros sitios Web: tu me pones un enlace, y yo te pongo un enlace. Esta técnica también ha tenido su momento. En la actualidad solo funciona bien si los sitios que se enlazan recíprocamente son de la misma temática. Por otro lado, también tiene un riesgo y es quedar vinculado a lo que se llaman malos vecindarios: se trata de redes de sitios que llevan a cabo prácticas no lícitas de posicionamiento y a los que los buscadores pueden penalizar. Si tú intercambias enlaces con sitios de malos vecindarios, te conviertes en un miembro más. Así que conviene que cuides mucho cómo y a quién enlazas.
- Comprando enlaces. Para evitar riesgos, puedes comprar directamente enlaces. Deberás seleccionar bien quién y cómo, y si te merece realmente la pena.
- Con Contenidos de Calidad. Es la técnica más segura de conseguir enlaces. Si tus contenidos son realmente útiles y valiosos empezarán a lloverte enlaces de modo espontáneo. Estos son los enlaces que más gustan a los buscadores, porque son los que mejor reflejan la relevancia.

Existen otras técnicas para conseguir enlaces pero la que garantiza resultados a medio y largo plazo es la que propone trabajar en la línea de los contenidos. Escribe contenidos útiles, de valor, bien descritos, y los enlaces llegarán solos.

#### **5.11.1 Enlaces de interés**

- Directory Archives  
<http://www.directoryarchives.com/>  
Es un directorio de directorios. Interesante si estás buscando directorios en los que darte de alta.
- Yahoo Site Explorer  
<http://siteexplorer.search.yahoo.com/>  
Te permite descubrir los enlaces externos que apuntan a tu sitio Web.

- CopyScape

<http://www.copyscape.com/>

Herramienta que te permite descubrir si alguien está copiando tus contenidos.

## 5.12 ¿Se puede sobre-optimizar? ¿Qué sucede entonces?

Evidentemente, en lo que se refiere a facilitar el rastreo eliminando barreras, toda optimización es poca. Ahora bien, en lo que se refiere a los contenidos, la sobre-optimización es peligrosa. Pongamos por ejemplo: sabemos que a los buscadores les gustan las palabras clave en el título, por lo tanto genero una página en la que mi palabra clave más importante aparece 5 veces en el título. Notar que acciones de este tipo están penalizadas por los buscadores de modo que, en el mejor de los casos, pueden hacerte perder ranking, y en el peor, te pueden excluir del índice. Todo aquello que escape a un contenido naturalmente escrito puede ser tachado de sobre-optimización no lícita por parte de los buscadores y ser penalizado.

## 6 LOS PRINCIPALES PROBLEMAS DE RASTREO EN DETALLE

---

### 6.1 URLs construidas con Javascript

Funciones Javascript que construyen las URL's destino al vuelo a partir de eventos generados por el usuario (como hacer clic en una entrada de menú, pasear el ratón por una zona, etc.) o que incluyen dentro del tag <A HREF> una llamada a una función Javascript. Los rastreadores de los buscadores no cuentan con un intérprete Javascript; no por dificultad de implementación, sino por ralentización del rastreo debido al tiempo extra de interpretación.

### 6.2 URLs embebidas en Flash

URL's que se muestran interaccionando con una animación en formato 'swf' (formato "Macromedia Flash") o cualquier otro formato multimedia interactivo. El formato es pesado y complicado de analizar, así como dificulta la interacción del usuario con la representación visual. Los robots de los buscadores no consiguen progresar por estas URL's.

### 6.3 Páginas con Frames

El uso de 'marcos' ("*frames*") complica la progresión lineal de los robots por el grafo de enlaces del sitio Web, apareciendo dificultades como:

- Problemas de recursividad en la progresión.
- Necesidad de descargas 'extra' por residir el código de los 'marcos' fuera de la página descargada.
- Ambigüedades en la titulación de las páginas.

### 6.4 Complicaciones asociadas a ciertas páginas dinámicas

Existen problemas de rastreo asociados a páginas que redirigen a otra (o bien sirven un fichero estático) en función de los parámetros que le llegan vía "POST", "GET" o variables de sesión (se entiende por variables de sesión tanto el uso de 'cookies' como el uso de variables de servidor y que se utilizan para modificar el contenido de una misma URL o para generar enlaces). Los mismos parámetros pueden generar diferentes páginas en

función de variables de sesión y parámetros diferentes pueden dar lugar a las mismas páginas rompiendo la regla de la unicidad en la indexación, puesto que una misma url pasa a representar diferentes contenidos y viceversa. Si a pesar de todo los robots progresan por un sitio Web con páginas de este tipo, su presencia en los buscadores implicará:

- ausencia de completitud: no todos los contenidos se han podido alcanzar mediante descubrimiento de enlaces, pues hay implicadas variables de formularios y/o de sesión que nos restringen el dominio de los documentos.
- ausencia de unicidad: al no haber correspondencia biunívoca entre contenidos y urls, la indexación, y por tanto la búsqueda, carece de sentido. Según las diferencias entre el estado de sesión del usuario y el del rastreador original, la url resultante en la búsqueda se resolverá a documentos diferentes.

No obstante, no todas las páginas dinámicas producen estos impedimentos. El uso de cookies no es perjudicial siempre y cuando no se utilicen para variar el contenido asociado con una misma URL. De igual forma, la utilización de variables de servidor no es problemática excepto cuando dichos valores se utilizan para generar enlaces o contenidos distintos. En estos casos donde el uso de estos parámetros no es perjudicial, se debe configurar un tiempo de expiración suficientemente grande para que las arañas sean capaces de progresar por todo el sitio Web.

## 6.5 Formularios

La inclusión de formularios en las páginas Web presupone una interacción humana, en la cual se rellenan una serie de campos y se seleccionan unos determinados valores. La automatización de este comportamiento no es recomendable. En el caso de la utilización del método de envío de formularios GET, se puede mantener la asociación biunívoca entre URL y contenido puesto que los parámetros del formulario pasan a formar parte de la URL. Esto no es posible si se utiliza el método POST. La utilización de formularios Web que utilizan el método POST es la forma más común de crear lo que se conoce como 'Web oculta'.

## 6.6 Multitud de parámetros en URLs de páginas dinámicas

Para evitar bucles en el recorrido de la estructura de páginas de los sitios Web, los robots de los buscadores no suelen descargar URLs con más de tres o cuatro parámetros.

## 6.7 Estructura de enlaces del Sitio Web

Los robots de rastreo de los buscadores no progresan indefinidamente por la estructura de enlaces de los sitios Web. Se suelen producir cortes al 5 o 6 nivel de profundidad. Dicho nivel disminuye si hablamos de sitios de antigüedad reducida y/o considerados poco relevantes. En general, es recomendable tener una estructura lo más aplanada posible, lo cual permite alcanzar cualquier página en el menor número de saltos; aunque tampoco se recomienda llevar este punto hasta el extremo. Conviene además tener en cuenta que a los buscadores no les suelen gustar páginas con muchos enlaces (por debajo de 100 suele ser un valor seguro).

## 7 MÁS CONSEJOS PARA CONSTRUIR UN SITIO WEB DE ÉXITO

---

### 7.1 Consejos para la selección del proveedor de “hosting”

500GB de transferencia mensual, 30GB de espacio, 50 buzones de correo... A menudo la publicidad nos llena los ojos con cifras que suelen significar muy poco para clarificar cuáles son los criterios fundamentales en los que nos hemos de fijar. Estos son algunas de las recomendaciones más interesantes para la selección de proveedor de hosting si buscas un posicionamiento adecuado:

- Antes de nada debes saber cual es el software de gestión de contenidos que vas a usar. Si tu sitio Web va a estar formado por páginas estáticas, puedes saltarte este punto. Si no, deberás informarte cuáles son los requisitos técnicos de plataforma que tiene dicho software. Una de las combinaciones más frecuentes es PHP y MySQL. Sean cuales sean, deberás buscar un proveedor que lo soporte y un plan que satisfaga tus necesidades.
- Un factor importante es la estabilidad del servicio (uptime). Internet es un canal 24x7. De poco sirve un plan espectacular si tu sitio Web se cae cada dos por tres. Piensa que, cuando tu sitio está no disponible, no sólo estás perdiendo la oportunidad de hacer negocio, sino que también estás dañando la imagen de tu organización. Además, los buscadores suelen penalizar páginas que no están disponibles cuando se visitan. Puedes utilizar servicios como Netcraft (<http://uptime.netcraft.com/perf/reports/Hosters>) para comprobar los ratios de uptime de los diferentes proveedores.
- Las ofertas suelen dar unas cifras de máxima transferencia mensual muy elevadas. Tanto, que casi nunca se alcanzan dichos límites. En cambio, lo que interesa es que los servidores en los que se va a ejecutar tu sitio Web no estén muy cargados y dispongan de líneas de conexión lo suficientemente veloces como para que la carga de tus páginas sea lo más ágil posible.
- Desconfía generalmente de los proveedores y planes que te prometen espacio ilimitado, número de cuentas de correo ilimitadas, etc. Es prácticamente imposible garantizar un servicio de calidad sin limitaciones. Es preferible tener cifras limitadas aunque suficientes.
- Otro factor importante de cara al posicionamiento es la ubicación geográfica o región en la que se encuentra la IP. Los buscadores actuales saben desde dónde (ubicación geográfica) buscan los usuarios, y uno de los criterios de relevancia que suelen utilizar es la proximidad geográfica. Contar con tu sitio Web disponible

desde una IP cercana a tu usuario puede mejorar tu ranking, especialmente si tu dominio no es “.es”.

### **7.1.1 Requisitos mínimos recomendados**

A la hora de realizar la elección de un proveedor de hosting debes considerar los siguientes requisitos como mínimos recomendables:

- El proveedor debe garantizarte el registro y mantenimiento de un nombre de dominio bajo el código “.es”, por los motivos de proximidad geográfica ya comentados.
- Al menos 1 GB de transferencia mensual, que garantice la agilidad de acceso a tu Web así como la disponibilidad de la misma, ya sea para los usuarios que la visitan o para las arañas que la rastrean.
- Acceso a estadísticas de uso, que te permitan monitorizar el tráfico de la Web.

### **7.2 Consejos para la selección del software de publicación**

Debes considerar si necesitas un software de publicación, especialmente en el sector PYME. En muchos casos, una presencia efectiva en Internet no requiere más de unas 20 o 30 páginas. En tal caso casi nunca se hace necesario contar con ningún software: así es más sencillo el manejo y requieres menos participación de mano de obra experta. Si no es tu caso, lo primero que debes hacer es hablar con la persona que te va a construir tu sitio Web para que te oriente sobre las diferentes opciones. Además, asegúrate de tener en cuenta que:

- Es preferible usar un software que ya esté hecho frente a desarrollar uno a medida.
- Existen excelentes gestores de contenido de uso gratuito.
- El entorno de tecnologías PHP y MySQL es uno de los más utilizados y versátiles.
- Desde la óptica del posicionamiento, asegúrate de que:
  - Es muy sencillo y cómodo editar el contenido de las páginas.
  - Puedes modificar de modo muy sencillo el título (<title>) de cada una de las páginas.
  - Los enlaces de navegación que genera la estructura son de tipo <a href= y no incluyen javascript.
  - Puedes modificar a tu voluntad el etiquetado de los subtítulos (h1, h2, etc)

- Que las urls que genera no contienen un elevado número de parámetros ni contienen un código de sesión.
- Que las urls sean únicas y que siempre devuelvan el mismo contenido independientemente de la sesión y la historia de navegación.
- Que su estructura de presentación no esté basada ni en frames ni en iframes.

## 8 ENLACES Y RECURSOS DE INTERÉS PARA SEO

---

### 8.1 Herramientas SEO

- Page Rank Checker  
[http://www.prchecker.info/check\\_page\\_rank.php](http://www.prchecker.info/check_page_rank.php)  
Herramienta que permite obtener el PageRank de una página dada.
- Google Toolbar  
<http://toolbar.google.com/>  
Entre otras funcionalidades, te permite chequear el PageRank de la página que estás visitando.
- Sitemap Generator de Google  
<https://www.google.com/Webmasters/tools/docs/es/sitemap-generator.html>  
Se trata de un pequeño programa que genera automáticamente un sitemap a partir de tu sitio Web.
- Robots.txt Generator  
<http://www.mcanerin.com/EN/search-engine/robots-txt.asp>  
Herramienta que te permite generar automáticamente un robots.txt a partir de la información que introduces en un formulario.
- SEO Book Keyword Suggestion Tool  
<http://tools.seobook.com/general/keyword/>  
Herramienta que te ayuda a identificar palabras clave para tu negocio.
- Yahoo Search Keyword Terms Extraction Tool  
<http://tools.seobook.com/yahoo-keywords/>  
Herramienta que te permite extraer las palabras clave contenidas en un párrafo.
- Keyword Density Analyzer  
<http://tool.motoricerca.info/keyword-density.phtml>  
Herramienta que te calcula la frecuencia de palabras en una determinada página.
- Poodle Predictor  
<http://www.gritechnologies.com/tools/spider.go>  
Herramienta que te ofrece información sobre cómo ven las arañas de los buscadores tu sitio Web.
- Link Popularity Check  
<http://www.marketleap.com/publinkpop/>  
Herramienta que te informa acerca de la densidad de enlaces que tu sitio recibe.

- Netcraft  
<http://uptime.netcraft.com/perf/reports/Hosters>  
Servicio que permite comprobar los ratios de uptime o disponibilidad de servicio de los diferentes proveedores de hosting.
- Xenu's Link Sleuth  
<http://home.snafu.de/tilman/xenulink.html>  
Software que te ayuda a descubrir los enlaces rotos que hay en tu sitio.
- W3C HTML Validation Service  
<http://validator.w3.org/>  
Herramienta que te permitirá descubrir los errores HTML de tus páginas.
- Search Engine Spider Test  
<http://tools.seobook.com/general/spider-test/index.php>  
Herramienta que te indica datos como el número de palabras de la página, la frecuencia de las mismas, el tamaño de la página, los enlaces que hay en ella, etc.
- Google Analytics  
<http://www.google.com/analytics/>  
Probablemente la mejor herramienta gratuita para analizar el tráfico de tu sitio Web.
- Crazy Egg  
<http://crazyegg.com/>  
Servicio que genera un heatmap de tus páginas. Te permite conocer cuáles son las áreas en las que más hace "clic" tu audiencia.
- Google Webmaster Tools  
<https://www.google.com/Webmasters/tools/>  
Herramienta de Google que te permitirá conocer cada cuánto tiempo la araña de Google (Googlebot) rastrea tu sitio, los problemas que ha encontrado, etc.
- Search Engine Keyword Tracker  
<http://www.digitalpoint.com/tools/keywords/>  
Herramienta gratuita de Digital Point que te permite monitorizar el ranking de tu sitio para tus principales palabras clave.
- SEO for Firefox  
<http://tools.seobook.com/firefox/seo-for-firefox.html>  
Add-on para Firefox que te permitirá ver los principales criterios de relevancia para cada una de las URL's que aparecen en las páginas de resultados.

- Yahoo Site Explorer  
<http://siteexplorer.search.yahoo.com/>  
Herramienta de Yahoo que te permite ver el número de páginas que tiene indexadas de tu sitio y los enlaces externos que tienes.
- Google Webmasters Tools  
<http://www.google.com/webmasters/>  
Contiene herramientas que te ayudan a escribir tu robots.txt y ver qué páginas bloquea. También ofrece información de la última vez que Google analizó tu sitemap y si encontró errores en el proceso. Información general acerca del modo en que Googlebot rastrea tu sitio Web.
- Directory Archives  
<http://www.directoryarchives.com/>  
Es un directorio de directorios. Interesante si estás buscando directorios en los que darte de alta.
- CopyScape  
<http://www.copyscape.com/>  
Herramienta que te permite descubrir si alguien está copiando tus contenidos.
- Formulario de Inserción en Google  
<http://www.google.com/addurl/>
- Formulario de Inserción en Yahoo  
<http://search.yahoo.com/info/submit.html>
- Formulario de Inserción en Live  
<http://search.live.com/docs/submit.aspx>

## 8.2 Información SEO

- Search Engine Ranking Factors  
<http://www.seomoz.org/article/search-ranking-factors>  
Uno de los listados más completos de factores de ranking
- Ranking Factors Simplified Quick Reference  
<http://www.searchenginejournal.com/ranking-factors-simplified-quick-reference/4764/>

- Most Important Search Engine Ranking Factors  
<http://www.quickonlinetips.com/archives/2007/04/most-important-search-engine-ranking-factors/>
- Ranking The SEO Ranking Factors  
<http://searchengineland.com/070403-145435.php>
- Search Engine Watch  
<http://www.searchenginewatch.com/>  
Observatorio del universo de los buscadores: una de las principales referencias a nivel mundial en su clase.
- SearchEngineWatch: How to Use HTML Meta Tags  
<http://searchenginewatch.com/showPage.html?page=2167931>
- Directrices de Google para Webmasters  
<http://www.google.com/support/webmasters/bin/answer.py?answer=35769>
- Yahoo Search Content Quality Guidelines  
<http://help.yahoo.com/l/us/yahoo/search/basics/basics-18.html>  
Guía de recomendaciones de Yahoo para ser indexado.
- Guidelines for Successful Indexing  
[http://search.msn.com/docs/siteowner.aspx?t=SEARCH\\_WEBMASTER\\_REF\\_GuidelinesforOptimizingSite.htm](http://search.msn.com/docs/siteowner.aspx?t=SEARCH_WEBMASTER_REF_GuidelinesforOptimizingSite.htm)  
Guía de recomendaciones de Live para ser indexado.
- Cómo se Construye un Sitemap  
<http://www.google.com/webmasters/tools/docs/es/protocol.html>

### 8.3 Conceptos SEO

- Qué es SEO – Wikipedia  
[http://en.wikipedia.org/wiki/Search\\_engine\\_optimisation](http://en.wikipedia.org/wiki/Search_engine_optimisation)
- Motor de Búsqueda en la Wikipedia  
[http://es.wikipedia.org/wiki/Motor\\_de\\_b%C3%BAsqueda](http://es.wikipedia.org/wiki/Motor_de_b%C3%BAsqueda)
- Top Buscadores en el Mundo según Nielsen//Netratings Junio 2007  
<http://searchenginewatch.com/showPage.html?page=3626726>
- Top Buscadores en España  
<http://www.ojobuscador.com/2006/12/09/el-dominio-de-google-en-europa/>

- Wikilearning – Cómo funciona un buscador  
[http://www.wikilearning.com/como\\_funciona\\_un\\_buscador-wkccp-6079-2.htm](http://www.wikilearning.com/como_funciona_un_buscador-wkccp-6079-2.htm)
- Como Funciona un Buscador  
[http://www.aulaclie.es/internet/t\\_4\\_7.htm](http://www.aulaclie.es/internet/t_4_7.htm)
- Cómo funciona... un motor de búsqueda  
<http://www.pisitoenmadrid.com/blog/2007/04/como-funciona-un-motor-de-busqueda/>
- Deep Web en la Wikipedia  
[http://en.wikipedia.org/wiki/Hidden\\_web](http://en.wikipedia.org/wiki/Hidden_web)
- Sitio Web de Referencia del Estándar Sitemap  
<http://www.sitemaps.org/>
- Sitio Web del Estándar Robots.txt  
<http://www.robotstxt.org/>
- Spamdexing en la Wikipedia  
<http://en.wikipedia.org/wiki/Spamdexing>

## 9 DICcionario DE TÉRMINOS SEO

---

- **Araña, Robot, Rastreador, Crawler o Spider:** se refiere a la parte de un buscador que recorre la Web enlace a enlace para descubrir sus páginas. Para que cualquier página pueda formar parte de las páginas de resultados de un buscador, su araña debe haberla descargado previamente.
- **Enlace Externo, Backlink o Inbound Link:** se trata de un enlace que apunta a una página de un sitio desde otro sitio diferente. En la actualidad, los buscadores los utilizan como criterio de relevancia.
- **Enlace Interno:** se trata de un enlace que apunta a una página de un sitio desde otra página de ese mismo sitio. La estructura de enlaces internos ayudan a los buscadores a reconocer la estructura del sitio.
- **Enlace Roto:** se trata de un hipervínculo en el que la página apuntada no existe o es errónea.
- **Frecuencia de Rastreo:** viene determinada por el tiempo que transcurre desde que la araña de un buscador rastrea una página, hasta que la vuelve a rastrear de nuevo. Los sitios Web más relevantes y los que más actualizan sus contenidos suelen ser rastreados con mayor frecuencia por los buscadores.
- **Googlebot:** es la Araña de Google.
- **Índice:** en buscadores, se trata de la estructura de datos que contiene referencias a las páginas que contienen una determinada palabra. Para que un buscador pueda ofrecer resultados de búsqueda es necesario que cree previamente un índice. Para que una URL pueda aparecer en las páginas de resultados de un buscador, debe haber sido previamente rastreada e indexada.
- **Indexación:** es el proceso que tiene lugar en los buscadores por el cual se actualiza el índice a partir de la información obtenida en el rastreo.
- **Logs:** es un concepto amplio. En esta guía se refiere a los registros de actividad de un sitio Web. Por cada página que solicita un internauta, se generan entradas en los logs que registran datos como: página solicitada, fecha y hora, navegador desde el que se ha realizado la solicitud, etc.
- **Metas:** son etiquetas que forman parte de las páginas HTML, permanecen invisibles para los internautas, pero no así para los buscadores. Aportan

información que tiene que ver con palabras clave, contenido e información de ayuda a las arañas para su rastreo. Hoy día no tienen tanta importancia como un tiempo atrás.

- **PageRank:** es un algoritmo de análisis de enlaces cuyo objeto es el cálculo de la importancia de una página en función del número de enlaces que contiene y recibe de otras fuentes. Fue desarrollado por Google y muchos expertos aseguran que fue la clave de su éxito a finales de los 90. En la actualidad no es tan considerado como criterio de relevancia como lo fue en su tiempo.
- **Palabra Clave o Keyword:** se trata de una palabra con fuerte contenido semántico para tu sitio. Las palabras clave de un negocio lo definen y son utilizadas por los internautas para localizar dicho negocio a través de los buscadores. Conviene, por tanto, prestar especial atención en la selección de dichas palabras así como ubicarlas correctamente en el contenido de las páginas de tu sitio Web.
- **Profundidad de Rastreo:** la profundidad de una página se mide, , partiendo de la página principal del sitio Web al que pertenece, según el número de enlaces en los que hay que hacer clic hasta llegar a ella. La profundidad de rastreo de un sitio Web coincide con el nivel de profundidad de la página más profunda de dicho sitio que la araña de un buscador ha rastreado. No todos los sitios Web son rastreados por los buscadores al mismo nivel de profundidad. Cuanto más relevante sea un sitio, mayor profundidad tiene su rastreo (normalmente).
- **Rastreo:** es el proceso que tiene lugar en los buscadores por el cual estos recorren la Red enlace a enlace. Es la etapa previa a la indexación.
- **Relevancia:** es el criterio bajo el cual ordenan típicamente los resultados los buscadores. El resultado de mayor relevancia ocupa la primera posición. La relevancia es una propiedad de la relación búsqueda-página y resultado. Los buscadores miden cientos de parámetros para calcular la relevancia de una página dada para una búsqueda en concreto.
- **Robots.txt:** es un fichero que suele estar disponible en la raíz de un sitio Web (es decir, <http://mywebsite/robots.txt>). En el fichero robots.txt se pueden especificar qué páginas y/o directorios pueden ser indexados por los buscadores.
- **SEO o Search Engine Optimization:** es un proceso que tiene por objeto la disposición óptima de los contenidos de un sitio Web para mejorar su presencia en los buscadores.
- **Snippet:** es el fragmento de texto que acompaña a cada uno de los resultados que ofrece un buscador en sus páginas de resultados. Típicamente se trata del

fragmento del contenido de la página en la que se encuentran los términos introducidos en la búsqueda.

- **Sitemap o Mapa del Web:** es un concepto con una doble acepción. Tradicionalmente se trataba de una página HTML más de un sitio Web en la que se recogía la estructura principal del site con enlaces a las diferentes secciones. En tiempos más recientes, se usa el término sitemap para referirse a un fichero en formato XML generado para los buscadores. Contiene información sobre las páginas disponibles y frecuencia de actualización de sus contenidos. Como se ha indicado a lo largo del documento, desde el fichero robots.txt se puede referenciar la ubicación del sitemap para ayudar a los robots de los buscadores en el proceso de rastreo y posterior indexación del sitio.
- **Spam de Buscadores o Spamdexing:** describe a aquellas páginas, generalmente de muy baja calidad, que han sido creadas con el objeto de conseguir altos rankings en los buscadores. Tratan de explotar debilidades de los algoritmos de ordenación.
- **Texto Ancla:** se trata del texto constitutivo de un enlace. Dicho de otro modo, es el texto que aparece subrayado. Los buscadores utilizan el texto ancla como un elemento que determina la temática de la página apuntada.
- **User-Agent o Agente de Usuario:** Los diferentes programas que descargan páginas Web (navegadores, arañas de buscadores, etc.) se identifican con un nombre. Éste es el agente de usuario. Normalmente el agente de usuario queda registrado en los logs. De este modo, se puede realizar un seguimiento a la actividad de las arañas de los buscadores.
- **Web Oculta, Web Invisible o Web Profunda:** se trata de las partes de la Red que los buscadores no pueden rastrear por encontrarse del otro lado de las barreras de rastreo. Dicho de otro modo, existen ciertos elementos tecnológicos hacen que los contenidos de la Web Oculta sean invisibles a los buscadores.