

Lección 2. Posicionamiento Web

Índice

1. ¿Qué es el posicionamiento?
2. Factores que afectan al posicionamiento:
 - a. Factores Directos
 - b. Factores Indirectos
 - i. Directrices de diseño para factores indirectos
3. Penalizaciones
4. Herramientas SEO
5. Glosario

1. ¿Qué es el posicionamiento?

Por **posicionamiento** entendemos el orden en que un sistema de búsqueda devuelve los resultados según la **probabilidad de relevancia del documento a la consulta**.

Un sistema de recuperación como **Google** es extremadamente **ruidoso**. Por **ruido** entendemos la obtención de documentos relevantes a la consulta. Pero gracias a un buen **sistema de ordenación de los resultados** nuestra percepción es que el ruido es menor que en otros sistemas. Hay que tener en cuenta que muy pocos **usuarios** revisan más allá de la primera página de resultados de una de sus consultas. La consecuencia es que es poco relevante los resultados situados tras esa primera hoja de resultados.

Siempre va a existir un intento por “*convencer*” del valor de nuestra página al motor. Esta persuasión tiende a ser mayor cuanto más valor financiero tenga una página. Para “*convencer*” al motor podemos sacar partido de nuestro conocimiento del algoritmo de posicionamiento del motor de búsqueda. Esto obliga a estos sistemas de búsqueda a **variar e incorporar elementos** para evitar la penalización por un lado, y a mantener el algoritmo en **secreto** por otro. Este fenómeno se denomina **Efecto de la Reina Roja**. El conocimiento de los algoritmos suele ser analizado por herramientas informáticas y pruebas manuales.

La **optimización** se suele producir para una **consulta predeterminada**. Sin embargo existen factores que dan una optimización genérica independiente de la consulta.

SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

En esta lección nos ceñiremos al motor de recuperación **Google**, dada su importancia en número de consultas enviadas a través de la Web y por brevedad dentro del cometido de esta lección.

2. Factores que afectan al Posicionamiento

A continuación se verán algunos factores que afectan al posicionamiento. Estos factores pueden ser de dos tipos:

- **Directos:** Son los que influyen de forma directa en el posicionamiento en un motor determinado.
- **Indirectos:** Son aquellos que influyen en algún factor Directo. Y por tanto influyen en el posicionamiento indirectamente.

a. Factores Directos

Hay que volver a destacar que la optimización suele realizarse para unos términos de consulta predefinidos. Esto es así para evitar grandes manipulaciones en el posicionamiento. Los elementos dependientes de la consulta se han marcado como (Q). Google además da más valor a determinados elementos debido a su menor probabilidad de manipulación para una consulta determinada (en negrilla). Dentro de los factores directos hay que destacar los siguientes:

Ubicación Términos de Consulta

- **Aparición de los términos en el título de la página (Q)**
- **Aparición de los términos de consulta en la URL de la página (Q)**
- Aparición en las etiquetas META (keywords y description) (Q)
- Aparición en el body y sobre todo en encabezamientos. (Q)
- Nombres de ficheros con términos de consulta (Q)
- Aparición de los términos al comienzo del cuerpo del documento (Q)

Enlaces

- **Número de enlaces entrantes**
- Número de enlaces salientes (negativo)
- Prestigio de las páginas que contienen los enlaces entrantes
- **Texto de los enlaces entrantes (coincidencia con términos de consulta) (Q)**

Términos de consulta en el cuerpo del documento

- **Alta frecuencia de aparición de los términos de consulta (Q)**
- IDF, frecuencia de aparición de términos no frecuentes (Q)
- Proximidad de los términos en el texto (consultas con más de una palabra) (Q)

Tipografía

SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

- **Utilización en los términos de consulta de etiquetas HTML para resaltar la relevancia (mayúsculas, negrilla, encabezados (Hn), etc) (Q)**

URL

- Proximidad a la raíz de la URL del término de búsqueda (Q)
- Prestigio del dominio de la URL (org, edu, es, ...)
- Periodo para el que se hace el registro del sitio.

Criterios temporales

- Frecuencia de actualización de la página (incluido Google Site Map)
- **Antigüedad del sitio en el índice del buscador**

Contenidos

- Riqueza de documentos adjuntos en el sitio (pdf, ppt, doc, etc)

Visitas

- Visitas que tiene una página que ha sido recuperada por los términos de consulta

Ubicación

- Situación Geográfica del servidor o IP (cuanto más próxima mejor)

PageRank

El PageRank es el cálculo que hace Google a priori sobre la relevancia de una página Web. Depende de tres factores principalmente:

- Número de enlaces que apuntan a la página A.
- Número de enlaces salientes que tienen las páginas que apuntan.
- Importancia de las páginas que apuntan a la página A.

Su cálculo se hace según la siguiente fórmula:

$$PR(A)=(1-d)+d(PR(t1)/C(t1)+PR(t2)/C(t2)....)$$

PR(A), es el PageRank de la página A

d, factor de amortiguación, probabilidad de que el usuario abandone la página. Por defecto 0.85

PR (t1)...t1, t2 son los Pr de las páginas que apuntan a A

C(t1), enlaces que salen de la página t1, para evitar división por cero, la página se considera autolenlace

La elección de los términos de consulta es quizás el factor más crítico a la hora de realizar de posicionar un site. Hay que tener en cuenta que:

- Es más difícil posicionar una página cuyas palabras de consulta (keywords) están presentes en un gran número de páginas web.

SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

- Se deben seleccionar términos que realmente empleen los usuarios y que sean empleados para nombrar el concepto que deseamos en ese dominio.
- Hay que incluir en sitios clave, como etiquetas ALT, las variaciones ortográficas. Sobre todo al utilizar términos con ñ o acentos.

b. Factores Indirectos

Como se ha comentado estos factores son los que influyen en los factores directos. Si una página por ejemplo sube en PageRank gracias al número de enlaces entrantes, aparecerá antes en las consultas. La consecuencia es que al aparecer en las primeras páginas de resultados será más conocido y por tanto más enlazado. Y al recibir más enlaces mejorará su posicionamiento.

Estos factores son de varios tipos. Principalmente:

- Contenido de calidad
- Estructura del Site
- Interoperabilidad
- Credibilidad
- Accesibilidad
- Usabilidad

Contenido de calidad

No hay que olvidar que los usuarios acceden a nuestro site buscando respuesta a determinado contenido, no buscando determinado diseño. Así un contenido sobre un tema determinado de calidad, novedoso y claro tiene influencia en el posicionamiento. Por otro lado una página con un porcentaje alto de código informático frente a contenido perjudica al posicionamiento.

Algunas pautas son:

- Dado que el usuario a accedido a nuestra página buscando una información o un dato el contenido más relevante debe ser ubicado al principio de la página. Crear expectativa, en términos de posicionamiento Web, es contraproducente.
- Tratar en la misma página temas accesorios o irrelevantes defrauda a los usuarios.
- Por ejemplo, una redacción que intente bromear con el lector utilizando dobles significados no es útil. Hay que tener en cuenta que los dobles sentidos tienen una interpretación en contextos geográficos, sociales o temáticos locales. Fuera de ese área, el doble significado, puede ser incluso malinterpretado.

SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

- Debido a la menor legibilidad del texto digital es recomendable párrafos de menos de tres líneas y utilización de enumeraciones mediante viñetas.

Estructura del Site

La estructura del sitio, incluido su jerarquía de páginas y temas, es de crucial importancia para la usabilidad y asimilación de contenidos. Así algunas pautas pueden ser establecer el contenido más genérico en la página de inicio del Site, ubicando el contenido más específico en las subpáginas. También es aconsejable asignar cada tema a una página específica con una URL única y fácil de recordar.

Interoperabilidad

Es importante utilizar una codificación estandarizada y formalizada. Una forma de conseguir este fin es validar el HTML y la CSS, o utilizar vocabularios estándar como Dublin Core.

Los lenguajes de la Web semántica orientados a mejorar el acceso de aplicaciones software al contenido deben complementar la página diseñada para el usuario. Una buena práctica en este sentido es incorporar documentos RSS.

Credibilidad

Un aspecto frecuentemente minusvalorado es la credibilidad. La Web tiene multitud de contenidos e infinidad de autores. No existe ningún filtro para que los autores publiquen sus contenidos. Así el usuario debe confiar en ciertos aspectos de la página para decidir si un contenido es o no creíble.

Los aspectos que valora el usuario para valorar esta credibilidad han sido analizados. De las decenas de factores que entran en juego hay algunos que se pueden indicar por su importancia. Estos son: Evitar errores ortográficos, hacer un diseño que muestre profesionalidad, poner la autoría y una dirección vinculada al autor, y poner contenidos reales y claros.

Accesibilidad y Usabilidad

Resulta conveniente diseñar para versiones no demasiado recientes de los navegadores ni incorporar tecnología que no esté ampliamente difundida. Así no utilizar Flash o frames, que puedan dificultar la indexación podría mejorar el posicionamiento. Una de las formas más sencillas de comprobar la accesibilidad es ejecutar el test TAW.

La usabilidad es quizás otro factor crucial en el posicionamiento. Si el usuario permanece y recuerda nuestro site podremos convencerle sobre su utilidad. Algunas de las directrices de diseño asociadas a este aspecto son: utilizar una buena estructura, intuitiva y clara, utilizar un reducido número de fuentes y colores o usar un buen contraste entre fondo y color de fuente.

SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

3. Penalizaciones

Existe sin embargo el peligro de que el buscador Web nos penalice por intentos de “engañar” al buscador.

Estas penalizaciones suelen ser bastante severas un caso reciente es la penalización de Google a Google Japón por posicionamiento fraudulento. BMW fue también penalizado por crear páginas solo para enlazar su site (Doorway). En estas páginas utilizaba JavaScript invisible al crawler para hacer más pesada la página.

Entre las principales **penalizaciones** figuran:

Intentos de engaños al servidor Web

- Doorway y Cloaking (páginas especiales para el crawler con el fin de posicionarse en el buscador, pero distinta a la que se ofrece al usuario)
- Utilización de redirecciones (existen sin embargo excepciones a este punto como el redireccionamiento tipo 301).

Penalizaciones por Enlaces: Se trata de prácticas orientadas a aumentar fraudulentamente el número de enlaces entrantes

- Campañas para enlazar con un mismo texto una página (Bombing)
- Compra de enlaces (p.e. post pagados en blogs) o Intercambio de enlaces deshonestos (sobre todo con un mismo texto de enlace). Sobre todo del mismo segmento C de la IP o con páginas sin relación temática
- Granjas de enlaces o páginas con más enlaces de lo normal
- Enlaces a páginas penalizadas por haber incurrido en penalizaciones en el pasado. Existen listados de este tipo de páginas en cada segmento IP.

Contenido Invisible para el usuario pero no para el crawler

- Utilizar texto Oculto: de pequeño tamaño, texto mismo color que fondo o entre “<” o “>”
- Texto en etiquetas meta

Términos de consulta en el cuerpo del documento

- Repetir la misma palabra con mucha frecuencia (stuffing). La solución es evitar repetirla en la misma frase o línea, y repetirlas dentro de las reglas del lenguaje natural.

Plagio

- Contenido muy similar a otra página más antigua

Dado que este tipo de penalizaciones es variable es aconsejable consultar recursos como [Google Webmaster Guidelines](#). En caso de duda se puede consultar en help@google.com y en caso de confirmación de la penalización cumplimentar [Google re-inclusion request](#)

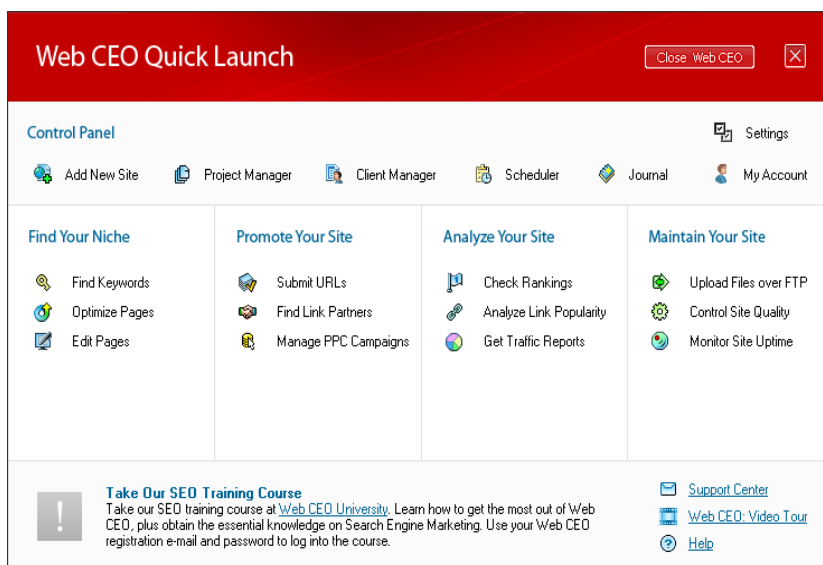
SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

4. Herramientas SEO

Hay que destacar la importancia que tienen las herramientas SEO (Search Engine Optimization). Es decir las herramientas que facilitan el estudio de los parámetros que afectan al posicionamiento de un Site. Estas herramientas nos sirven para:

- Estudiar las páginas que compiten con la mía (las que devuelve el buscador ante la misma consulta)
- Elegir palabras clave
- Analizar páginas propias
- Comportamiento del usuario
- Factores indirectos
- Análisis de Interacciones con Penalización

Existen aplicaciones que engloban prácticamente todos los aspectos que se verán a continuación. Ejemplos son **WebCEO**, **Internet Bussines Promoter**, **WebConfs**, **SEO Toolbox** o **SEOChat**.



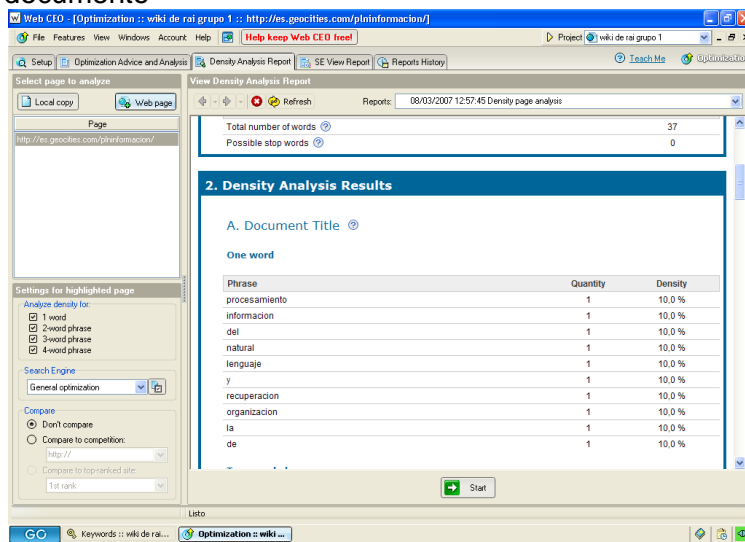
SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

a. Estudiar la competencia

Se entiende por competencia a las páginas que aparecen en un buscador ante determinada búsqueda (sobre todo las que aparecen por delante de la mía)

Realizan un análisis sobre frecuencia de la palabra clave en las páginas de la competencia

- **WebCeo**; Estudia la densidad de los términos en diferentes lugares del documento



- **Keyword Counter-Keyword Frequency Analyzer**
<http://www.keywordcount.com/>
- **Link Popularity Check**: Muestra bajo diferentes motores el número de enlaces entrantes a un conjunto de páginas Web competencia de la que está bajo estudio www.linkpopularitycheck.com

SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

URL	Total	AllTheWeb	AltaVista	Google/Tis...	MSN...	Yahoo
www.gioupm.com	238	105	122	11	0	0
www.ubvirtual.com/es/index.html	304	271	0	33	0	0
www.iede.edu/es/index.html	1.411	1.270	0	141	0	0
www.clase-ejecutiva.com	2.095	150	1.930	15	0	0
www.uniactiva.com	2.096	893	1.140	63	0	0
www.esic.es	3.984	992	2.910	82	0	0
exed.esade.edu	4.085	620	3.060	405	0	0
www.eoi.es	4.502	1.260	2.920	322	0	0
www.upcomillas.es	12.810	3.110	6.740	510	0	2.450
www.iese.edu/es/Home.asp	16.700	1.450	0	1.650	0	13.600
www.ie.edu	40.788	11.300	18.200	888	0	10.400
www.uoc.edu	110.220	20.200	75.300	2.920	0	11.800
www.uc3m.es	129.040	27.100	65.500	1.640	0	34.800

b. Elegir palabras clave

Se entiende por palabras clave los términos por lo que se debería poder encontrar mi página en determinado buscador. No todos los términos tienen la misma probabilidad de éxito, unos no se utilizan otros tienen mucha competencia, etc.

Las herramientas asociadas se basan normalmente en recursos online como:

- **Overture:** <http://inventory.overture.com/d/searchinventory/suggestion/>
- **WordTracker:** <http://www.wordtracker.com/>
- **GoogleSuggest:** herramienta de Google que indica el número de búsquedas que se ha realizado en un periodo de tiempo sobre un término. Lo óptimo es que pocas páginas para una consulta muy frecuente

SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN



Web [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

informa	
information	11,720,000,000 results
information technology	1,870,000,000 results
informatica	211,000,000 results
informatics	58,800,000 results
information systems	1,270,000,000 results
information commissioner	48,700,000 results
information clearing house	52,900,000 results
information security	1,220,000,000 results
informative speech topics	2,330,000 results
information system	2,760,000,000 results

Advanced Search
 Preferences
 Language Tools

sults. [Learn more](#)

©2007 Google

Google insights muestra una simbiosis entre Google Trends y Google Suggest.

- **Google Adwords:** se trata del sistema por el que Google obtiene sus ingresos. Para animar a los usuarios a patrocinar sus páginas para determinada consulta propone un conjunto de palabras con posibilidades de competir con éxito.
- **Search-Based Keyword Tool**

Google Search-based Keyword Tool

Keywords: My draft keywords (0)

Website: Word or phrase: More filters Or browse all keywords

Categories for this search: All categories

Brands: All keywords Unrelated keywords (0) systems (3) oracle (1)

Help: Learning the basics, Filtering and saving keywords, Incorporating new Address, Troubleshooting

Keyword ideas

This website is not associated with your AdWords account. You're seeing a limited set of keyword ideas, which may include those already in your account and may or may not be currently active for this website.

Keyword	Monthly searches	Competition	Sugg. bid	All Search share	Extracted from webpage
keywords related to uc3m.es (85)					
masters programs	6,200	Low	USD7.45	—	UC3M
reduced registration fees	20	Low	USD7.33	—	CCMP V
online conference registration form	25	Low	USD6.44	—	DC-2005, Registration form
phd programs in economics	52	Low	USD6.16	—	PhD in Economics
dc hotel accommodation	12	Low	USD3.87	—	DC-2005, Accommodation form
law registration form	12	Low	USD3.39	—	DC-2005, Registration form
business consultants	12	Low	USD3.32	—	Linux Consultants HOWTO, USA - Virginia
3 year phd	125	Low	USD3.24	—	PhD in Economics
meeting registration form	82	Low	USD3.20	—	DC-2005, Registration form
hotels in madrid city centre	540	Low	USD3.17	—	DC-2005, Accommodation form
hotels in madrid city	230	Low	USD2.76	—	DC-2005, Accommodation form

- Wordtracker Labs
- Yahoo!
- Technorati
- iBoogie: clusters para identificar competencia de palabras clave
- Competencia de palabras clave: Analiza los 10 primeros resultados en Google

SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

- **Delorie** <http://www.delorie.com/web/ses.cgi>
- **Search Engine Spider Simulator** <http://www.webconfs.com/search-engine-spider-simulator.php>
http://www.searchenginewold.com/cgi-bin/sim_spider.cgi

d. Comportamiento del usuario

Es recomendable estudiar a los usuarios que acceden al sitio para optimizar las páginas según sus características. Ejemplos de estas características pueden ser el navegador que utilizan, la zona del mundo desde donde acceden, el tipo de conexión, etc.

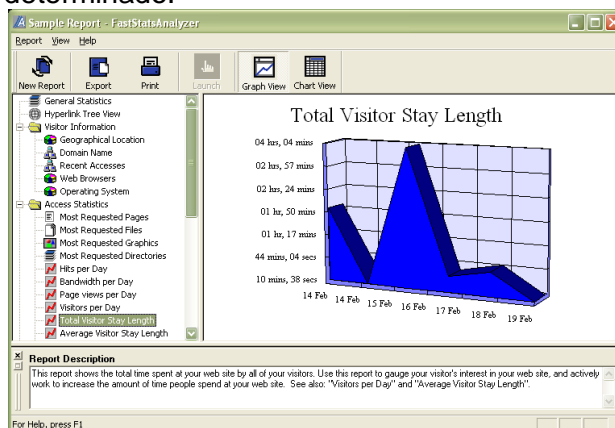
Análisis del Site

Una vez creado el site es aconsejable monitorizar el comportamiento de los usuarios. Normalmente esta monitorización se hace mediante los logs que se encuentran en el servidor. Existe información especialmente sensible en este apartado, por ejemplo:

- Por qué palabras me buscan los usuarios
- Qué tipo de hardware y de conexión tienen
- Cuáles son las páginas en las que se detienen, cuánto tiempo están, ...

FastStatsAnalyzer

Aplicación para analizar logs. Aunque en un principio pueda parecer parecido a Analytics, el conjunto de datos a analizar es diferente. Analytics analiza datos de acceso desde cualquier servidor a una página concreta. Los logs reflejan los accesos a las páginas alojadas en un servidor determinado.



SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

Google Analytics

Programa de Google para monitorizar una página Web. Requiere poner un código JavaScript en la página, por lo que no permite monitorizar páginas sobre las que no tengamos control. Permite saber datos sobre: cómo acceden a nuestra página, desde qué buscadores, desde qué páginas, configuración de los ordenadores que acceden, qué sistema operativo, qué tipo de conexión, el tipo de monitor, etc

**Calcular tiempo de descarga de un WEB**

El tiempo de descarga de un Web influye en los grupos de usuarios con redes de conexión lenta. Hay que tener en cuenta el usuario objetivo. Un usuario del tercer mundo necesitará probablemente una página más ligera para adecuarse a redes de baja velocidad.

Ejemplo: WebWait

A screenshot of the WebWait website. The page has a pink 'webwait' logo at the top. Below it, there's a search bar where 'www.google.es/search?hl=es&btnq=...' is entered. A large box displays the result '0.89s'. To the right, there's a preview of a Google search results page for 'La Web', with a red box highlighting the '0.89s' loading time. The website also includes a 'Multiple Call Options' section with input fields for '# Calls' and 'Call Interval', and a 'What's WebWait?' section.

SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

e. Análisis Factores Indirectos

La accesibilidad incide en el número máximo de usuarios potenciales que tengo, una mala accesibilidad puede disminuir mis visitas y por tanto recibir menos enlaces

Ejemplos:

TAW

Aplicación online para evaluar la accesibilidad. Prestar más atención a resolver los errores detectados automáticamente, pues suelen ser más graves, algunos de los manuales son subjetivos

- Con test con varios niveles: A, AA, AAA: www.w3c.org/TR/WCAG10/full-checklist.html
- Con Software: Bobby, TAW www.tawdis.net, Cynthia www.cynthiasays.com

Validadores de código

Un código no estándar puede hacer empeorar tanto la visualización, como la accesibilidad y la interoperabilidad

Algunos ejemplos de validadores del código de la página son:

- Para XHTML: <http://validator.w3c.org> Valid XHTML 1.0
- Para CSS: <http://jigsaw.w3c.org/css-validator> Valid CSS



SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

f. Analisis de interacciones con penalización

Existen herramientas que muestran si nuestra página enlaza a páginas penalizadas por los buscadores. Por ejemplo, [Bad Neighborhood detection tool](#)

Glosario

- **BLOOP (BackLinkOver Optimisation Penalty):** Penalizaciones por malas prácticas en enlaces.
- **Bombing:** campaña para que un gran número de páginas Web enlacen determinada página Web con un mismo texto, con la intención de que el buscador interprete que esa página trata del concepto que representa ese texto. Puede ser despectivo, pero debe ser masivo para que funcione.
- **Cloaking:** Páginas falsas con la finalidad de engañar al motor de búsqueda o al usuario. Por ejemplo hacer una versión de la página especial para posicionarse en determinado motor, u ocultando parte de la página al usuario.
- **Doorway:** páginas que tienen la única finalidad de tener un enlace que nos lleve a la página que tiene el contenido. En principio no pertenecen al site y no son accesibles por la navegación normal dentro del site. Perjudica el posicionamiento.
- **Efecto de la Reina Roja:** es el efecto que se observa en posicionamiento por el cual un cambio en el conocimiento de los algoritmos de posicionamiento implica un cambio automático en el diseño de dichos algoritmos y en su sistema de penalizaciones.
- **Geotargeting:** mostrar diferente idioma o información local según la ubicación de quien hace la consulta. No está penalizado.
- **Granja de enlaces (Linkfarm):** Creación, normalmente de forma automático, de páginas cuya finalidad es contener enlaces que apunten a páginas Web. Estas páginas suelen tener más de un centenar de enlaces.
- **Herramientas SEO:** es el software que sirve para analizar, estudiar y optimizar la posición de una página ante determinada consulta.
- **Optimización Web:** Aplicación de técnicas de posicionamiento a una página Web con el objetivo de mejorar su posicionamiento.
- **PageRank:** Es el cálculo que hace Google a priori sobre la relevancia de una página Web.
- **Penalización:** es aquella práctica que el motor de recuperación castiga debido a su previsible intención de manipular su política de posicionamiento. La penalización puede ser una peor puntuación en la creación del ranking de posicionamiento o incluso la expulsión del índice.
- **Posicionamiento:** Orden en el que se devuelven los resultados.
- **SEO (Search Engine Optimization):** Disciplina que estudia los algoritmos por los que se puede optimizar una página Web para determinada consulta. Normalmente está unida al concepto de SEM (Search Engine Management) que es la gestión periódica de la página Web para optimizar el posicionamiento.
- **SERP (Search Engine Result Page);** Página de resultados que devuelve el buscador.