

Saca el máximo partido de tu contenido online

*Guía para la edición de
contenidos en la web*

Google™

Índice

Introducción	2
Descripción breve de la búsqueda web	3
¿Qué novedades presenta la búsqueda web de Google?	4
¿Puede Google encontrar mi sitio?	5
¿Puede Google indexar mi sitio?	6
<i>Controlar lo que Google indexa</i>	7
<i>Robots.txt frente a metaetiquetas</i>	9
<i>Controlar el caché y los fragmentos de texto</i>	10
¿Incluye mi sitio contenido útil y único?	11
Aumentar la visibilidad: las mejores prácticas	12
Centro para webmasters	13
<i>Sitemaps</i>	14
Preguntas frecuentes	15
Glosario	19

Introducción

Si lo que buscas es visibilidad, Internet es tu sitio. Esto mismo te dirá cualquier anunciante que haya aumentado sus ventas mediante publicidad online, cualquier blogger cuya popularidad le haya llevado a publicar un libro, o cualquier periódico que haya ampliado su base de lectores internacionalmente gracias al aumento del tráfico online.

En Google nos preguntan muy a menudo cómo funciona la búsqueda web y cómo pueden los webmasters y editores de contenidos maximizar su visibilidad en Internet.

Hemos redactado este breve folleto para ayudarte a entender cómo el motor de búsquedas encuentra el contenido, y cómo puedes adaptar tu presencia en la Web para garantizar que se muestra aquello que deseas que se muestre, y que se mantiene oculto aquello que no deseas mostrar.

Este folleto está orientado tanto a webmasters de pequeños sitios web como a propietarios de grandes sitios, e incluye desde sugerencias y herramientas para webmasters hasta instrucciones paso a paso sobre preguntas frecuentes.

La propia visión de Google respecto a la búsqueda web y sus relaciones con los propietarios de sitios web se ha desarrollado de forma espectacular, del mismo modo que lo hizo Internet durante la última década. Hemos creado un número considerable de herramientas para facilitar que los webmasters maximicen la visibilidad de su contenido, así como también para controlar la indexación de sus páginas web. Pero siempre se puede mejorar. Esperamos que este folleto te anime a enviarnos comentarios y a opinar sobre lo que podemos hacer para convertir la web en un lugar mejor, tanto para usuarios en busca de información como para webmasters.

- El equipo de webmasters de Google

Descripción breve de la búsqueda en Internet:

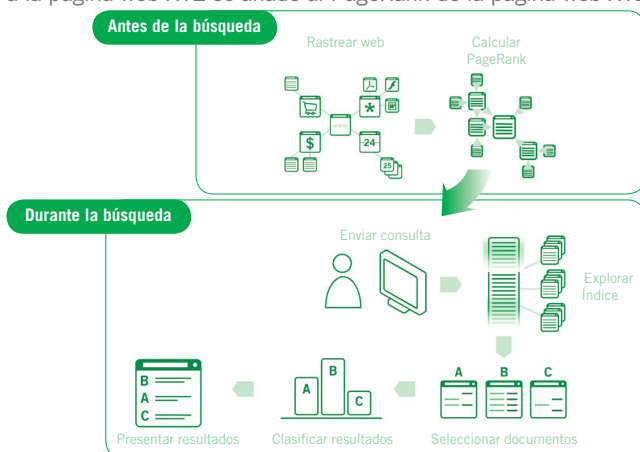
Funcionamiento

En términos muy sencillos, podría decirse que Internet es como un libro larguísimo, con un índice impresionante, que te dice exactamente en qué lugar se encuentra toda la información.

Google dispone de un conjunto de equipos informáticos, llamados “Googlebots”, que rastrean (buscan) continuamente miles de millones de páginas en la Web. Este proceso de rastreo es algorítmico: hay programas informáticos que determinan los sitios que deben rastrearse, la frecuencia de rastreo y la cantidad de páginas que recuperan de cada sitio web. Google no acepta pagos por rastrear un sitio más a menudo, por lo que nuestro servicio gratuito de búsqueda se mantiene así separado de nuestro servicio de publicidad, Google AdWords, que sí genera ingresos.

El proceso de rastreo que Google lleva a cabo empieza con una lista de diferentes URL. Cuando el Googlebot navega por los diferentes sitios web, detecta los vínculos de cada página y los añade a su lista de páginas pendientes de rastrear. El Googlebot hace una copia de cada una de las páginas que rastrea con el fin de compilar un índice masivo de todas las palabras que reconoce. Esta lista también indica dónde se localiza cada palabra dentro de cada página.

Cuando un usuario realiza una consulta, nuestras máquinas buscan en el índice las páginas que coinciden con la búsqueda y devuelven al usuario los resultados más relevantes. La relevancia se determina en función de unos 200 factores, uno de los cuales es el “PageRank” de cada página. El “PageRank” es la medida de la “importancia” de una página en función de los vínculos entrantes de otras páginas. Es decir, cada página web enlazada a la página web XYZ se añade al PageRank de la página web XYZ.



¿Qué novedades presenta la búsqueda web de Google?

Pese a que los principios básicos de la búsqueda web se han mantenido prácticamente constantes, Google trabaja a diario para mejorar sus resultados de búsqueda.

¿En qué ha cambiado la búsqueda web desde hace cinco años? Bien, sin ir más lejos, hoy es mucho más rápida.

Además, en comparación con los de hace cinco años, nuestros sistemas de rastreo e indexación son mucho más inteligentes. Por ejemplo, ahora se examinan páginas web continuamente y se programan las visitas a cada página de forma más inteligente con el fin de maximizar la actualización de los resultados. Con este enfoque más eficiente se tiene en cuenta el hecho de que, por ejemplo, el sitio web de un periódico online necesita un rastreo más frecuente mientras que un sitio estático que se actualiza una vez al mes no lo exige. De hecho, también estamos permitiendo que los webmasters controlen la frecuencia con la que rastreamos sus sitios, todo a través de nuestras herramientas para webmasters. En general, el resultado es un índice más actualizado y completo.

Aunque la búsqueda web actualmente es más rápida y eficiente que nunca, los factores clave que determinan la visibilidad de un sitio web en los resultados de búsqueda de Google han sido prioritarios desde que se estrenase nuestro motor de búsqueda.

[¿Puede Google encontrar mi sitio web?](#) (página 5)

[¿Puede Google indexar mi sitio web?](#) (página 6)

[¿Incluye mi sitio contenido útil y único?](#) (página 11)

¿Puede Google encontrar mi sitio web?

La inclusión en los resultados de búsqueda de Google es gratuita y fácil, dado que no es necesario enviar tu sitio web a Google. De hecho, la gran mayoría de los sitios incluidos en nuestros resultados no se han enviado manualmente, sino que el Googlebot los ha encontrado al rastrear la web, y su inclusión ha sido automática.

Pese a que Google rastrea miles de millones de páginas, es inevitable que algunos sitios se pasen por alto. Hay varias razones por las cuales el Googlebot puede no encontrar un sitio:

- El sitio no está correctamente enlazado mediante vínculos a otros sitios de la Web.
- El sitio se ha lanzado después del último rastreo realizado por Google.
- El sitio no estaba disponible o recibimos un mensaje de error cuando intentamos rastrearlo.

Mediante las herramientas para webmasters de Google (por ejemplo Sitemaps), puedes determinar si tu sitio ya está incluido en el índice de Google o si hemos recibido algún error al intentar rastrearlo (ver página 14). También puedes utilizar estas herramientas para añadir tu URL al índice de Google de forma manual, o bien para proporcionar un mapa de tu sitio web que nos dé información más detallada sobre el contenido de tu sitio. Esto nos ayudará a encontrar secciones y contenido nuevos en tu sitio web.

¿Puede Google indexar mi sitio web?

Es posible que, en alguna ocasión, los webmasters observen que sus sitios no aparecen en los resultados de búsqueda. El problema puede ser de “indexabilidad”, esto es, si el Googlebot puede o no realizar una copia de una página web para incluirla en nuestros resultados de búsqueda.

Estructura y contenido

Cuando una página no aparece incluida en los resultados de búsqueda, la razón suele estar ligada a la estructura y al contenido de la página. Por ejemplo, es posible que Google no pueda indexar una página que contenga un formulario para que lo rellenen los usuarios. Del mismo modo, tampoco los motores de búsqueda podrán indexar fácilmente las páginas que incluyan contenido dinámico, como por ejemplo, Flash, JavaScript, marcos o URL generadas dinámicamente. Si quieres saber si este es el problema de tu sitio, visualízalo con un navegador de texto como Lynx o con un navegador que tenga desactivadas las imágenes, JavaScript o Flash, y sabrás si todo el contenido es accesible.

Si tu sitio utiliza muchas imágenes, asegúrate de incluir una descripción del contenido importante de cada imagen en el texto. Esto no sólo permite que los motores de búsqueda indexen la imagen correctamente, sino que hace que la imagen sea accesible para usuarios con problemas de vista. También puedes utilizar el texto alternativo para la imagen, así como nombres de archivos descriptivos, como se muestra en el ejemplo siguiente, la imagen del logotipo de una empresa llamada “Pastelería Buñ”:

```

```

URL

Un obstáculo adicional podría constituirlo la propia URL. Si hay identificadores de sesión o varios parámetros en la URL, o bien si la URL se redirecciona varias veces, es posible que Google no pueda indexar la página.

Servidor y red

Ciertos problemas con el servidor y la red podrían impedirnos acceder a ciertas páginas de tu sitio. Mediante las herramientas disponibles en el Centro para webmasters de Google, podrás ver una lista de las páginas a las que el Googlebot no ha podido acceder. Para obtener más información acerca del Centro para webmasters, consulta la página 13.

Protocolo de exclusión de robots

Algunas veces, el Protocolo de exclusión de robots, un estándar técnico que permite a los webmasters “ordenar” a los motores de búsqueda que no indexen el contenido del sitio, bloqueará las páginas. Si tu sitio web no aparece en los resultados de búsqueda de Google, deberás comprobar que ni el archivo robots.txt ni una metaetiqueta están bloqueando el acceso a nuestros rastreadores.

Controlar lo que Google indexa

Cada editor web tiene un objetivo diferente. Por ejemplo, algunos webmasters de periódicos desean ofrecer un acceso gratuito a sus artículos más recientes, mientras que el acceso a los archivos es un servicio de pago o especial. Unos ofrecen visibilidad plena en todas las propiedades de un motor de búsqueda, por ejemplo, Google para móviles o Búsqueda de imágenes de Google, mientras otros sólo desean aparecer en los resultados de búsqueda web.

Los motores de búsqueda deben respetar los deseos de los editores ya que, al fin y al cabo, son los propietarios de su contenido. Pero como no leemos el pensamiento, es muy importante que los webmasters nos digan cómo desean que indexemos su contenido. Pueden hacerlo mediante el Protocolo de exclusión de robots, una especificación técnica bien consolidada que ordena a los motores de búsqueda qué sitios o partes de un sitio web no deben examinar, y qué partes deben permanecer visibles en los resultados de búsqueda.

Robots.txt: Control de todo el sitio

La base del Protocolo de exclusión de robots es un archivo de texto llamado robots.txt, un estándar muy antiguo del sector. Con este archivo puede controlar el acceso en varios niveles, desde el sitio entero hasta directorios individuales, tipos concretos de páginas o incluso páginas individuales.

Hay ciertas páginas en mi sitio que no deseo que Google indexe. ¿Cómo puedo impedir que aparezcan en los resultados de búsqueda de Google?

En general, la mayoría de propietarios están interesados en que Googlebot acceda a su sitio para que los usuarios puedan llegar a sus páginas web mediante Google. Sin embargo, es posible que tengas páginas que no desees indexar: por ejemplo, registros internos o artículos de noticias cuyo acceso sea de pago.

Puedes excluir páginas del índice de Google creando un archivo robots.txt y colocándolo en el directorio raíz de tu servidor web. Este archivo robots.txt enumera las páginas que no desees que indexen los motores de búsqueda. Crear un archivo robots.txt es muy sencillo y proporciona a los editores un nivel de control sofisticado para gestionar el modo de acceso de los motores de búsqueda a sus sitios web.

Por ejemplo, si un webmaster desea impedir que los registros internos se indexen, el archivo robots.txt deberá incluir:

User-Agent: Googlebot La línea User-Agent especifica que la sección siguiente es un conjunto de instrucciones solo para Googlebot.

Disallow: /logs/ - La línea Disallow ordena a Googlebot que no acceda a los archivos del subdirectorio de registros de tu sitio

El propietario del sitio ha especificado que ninguna de las páginas del directorio de registros aparezca en los resultados de búsqueda de Google.

Todos los motores de búsqueda importantes leerán y obedecerán las instrucciones que proporciones en el archivo robots.txt y, si lo deseas, puedes especificar reglas diferentes para los distintos motores de búsqueda.

Metaetiquetas: control exhaustivo

Además del archivo robots.txt, que te permite especificar las instrucciones precisas para una gran cantidad de archivos de tu sitio web, puedes utilizar la metaetiqueta robots para controlar detalladamente las páginas individuales de tu sitio. Para hacerlo, solo debes añadir metaetiquetas concretas a una página HTML para controlar cómo se indexa ésta. La combinación de las metaetiquetas y el archivo robots.txt te dan la flexibilidad de expresar políticas de acceso complejas de una forma relativamente sencilla.

Tengo un artículo concreto en mi sitio al que sólo pueden acceder los usuarios registrados. ¿Cómo puedo hacer que no aparezca en los resultados de búsqueda de Google?

Para ello, simplemente debes añadir la metaetiqueta NOINDEX en la primera sección <head> del artículo. Sería algo así:

```
<html>
<head>
<meta name="googlebot" content="noindex">
[...]
```

Esto impedirá que Google indexe este archivo.

Sin embargo, vale la pena recordar que en algunos casos querrás que Google indexe este tipo de páginas, por ejemplo, una noticia de archivo que los usuarios puedan pagar para leerla online. Aunque este tipo de contenido especial no aparezca en los resultados de búsqueda de Google, habrá ciertos servicios de Google (como la Búsqueda de archivos de Google Noticias) que incluirán el artículo en sus índices, explicando claramente la información sobre el pago a los usuarios.

Robots.txt frente a metaetiquetas

En general, el archivo robots.txt es un buen método para permitir un control del sitio, mientras que las metaetiquetas proporcionan un control más detallado sobre los archivos individuales. Las metaetiquetas son especialmente útiles si dispones de permisos para editar los archivos individuales, pero no para editar el sitio entero. También te permiten especificar políticas de control de acceso complejas página por página.

A veces, cualquiera de las dos herramientas puede solucionar el mismo problema:

¿Cómo puedo asegurarme de que se indexa el texto de una página, pero no las imágenes?

Una opción podría ser bloquear el acceso a las imágenes por la extensión de archivo en todo el sitio mediante robots.txt. En el archivo.txt, las líneas siguientes ordenan a Google que no indexe ninguno de los archivos con la extensión *.jpg o *.jpeg:

```
User-agent: Googlebot
Disallow: /*.jpg#
Disallow: /*.jpeg#
```

O bien, si tu Sistema de administración de contenido (CMS) almacena imágenes en un directorio diferente, puedes excluir el directorio completo. Si tus imágenes están en el directorio /imágenes, puedes excluir este directorio en todos los motores de búsqueda mediante las siguientes líneas:

```
User-agent: *
Disallow: /images/
```

Otra opción es añadir la etiqueta NOINDEX a cada archivo que incluya una imagen. Todas estas opciones impedirán que se indexen las imágenes. La única cuestión es qué grado de exclusión de imágenes deseas establecer.

Controlar la caché y los fragmentos de texto

Normalmente, los resultados de búsqueda muestran una página en caché y un fragmento de texto. Aquí, por ejemplo, puedes ver uno de los primeros resultados que encontramos cuando realizamos una búsqueda de “Pavo real”.

Pavo real

Pavo real Pavo cristatus. Fotos. bullet. Los pavos reales proceden de Asia. bullet. En el parque hay numerosos pavos reales que pasean por todo él y que se ...
centros3.pntic.mec.es/cp.clarin1/pavo%20real.htm - 4k - [En caché](#) - [Páginas similares](#)

Fragmento de texto: un extracto de texto de la página web.

Vínculo en caché: este vínculo dirige a los usuarios a una copia de la página almacenada en los servidores de Google.

¿Por qué un fragmento de texto? Es muy posible que los usuarios quieran visitar un sitio web si los resultados de la búsqueda muestran un fragmento de texto de un sitio. Esto se debe a que los fragmentos de texto facilitan que los usuarios vean la relevancia del resultado de su consulta. Si un usuario no puede hacer esta determinación rápidamente, normalmente pasará al siguiente resultado de búsqueda.

¿Por qué un vínculo en caché? El vínculo en caché es útil en ciertos casos: cuando un sitio no está disponible temporalmente; cuando sitios de noticias están sobrecargados por un acontecimiento importante, o cuando los sitios se borran accidentalmente. Otra ventaja es que la copia en caché de Google destaca las palabras que ha utilizado el usuario en su búsqueda, permitiendo una evaluación rápida de la relevancia de la página.

La mayoría de editores quieren que Google muestre tanto el fragmento de texto como el vínculo en caché. Sin embargo, en algunos casos el propietario de un sitio desea desactivar uno de ellos o ambos:

El contenido de mi periódico cambia varias veces al día. No parece que Googlebot esté indexando el contenido tan rápido como se actualiza, y el vínculo en caché apunta a una página que ya no está actualizada. ¿Cómo puedo impedir que Google cree un vínculo en caché?

El propietario del sitio de noticias puede impedir que el vínculo en caché aparezca en los resultados de la búsqueda añadiendo la etiqueta NOARCHIVE en su página de este modo:

```
<META NAME="GOOGLEBOT" CONTENT="NOARCHIVE">
```

Del mismo modo, puedes hacer que Google no muestre un fragmento de texto de una página mediante una etiqueta NOSNIPPET:

```
<META NAME="GOOGLEBOT" CONTENT="NOSNIPPET">
```

Nota: Añadiendo NOSNIPPET también puedes impedir que se muestre un vínculo en caché, de modo que si especificas NOSNIPPET automáticamente obtendrás también NOARCHIVE.

¿Incluye mi sitio contenido útil y único?

Una vez que el sitio se puede examinar e indexar, la pregunta final es saber si el contenido de las páginas web es único y útil.

Primero deberás observar el texto como un todo. ¿Son el título y los vínculos de texto descriptivos?; ¿Fluye el texto de forma natural, intuitiva y clara?

Cada página web debería estar centrada en un área o tema concreto, tal y como un capítulo de un libro está organizado. Las palabras clave y las frases emergen de forma natural en este tipo de texto, y es mucho más probable que los usuarios se queden en una página web que les proporciona contenido y vínculos relevantes.

Sin embargo, debes asegurarte de que las frases que escribas incluyan aquellas que los visitantes desean buscar. Por ejemplo, si tu sitio es para un club de simpatizantes de MG, asegúrate de que las palabras “MG” y “coches” aparezcan en el texto, en vez de sólo términos como “automóviles británicos”.

Aumentar la visibilidad: las mejores prácticas

Los propietarios de sitios suelen preguntarnos cuáles son los mejores métodos para aumentar la visibilidad y la clasificación de los sitios en nuestros resultados de búsqueda. Nuestra respuesta es simple: “Piensa como un usuario, porque así es como nosotros intentamos pensar”.

Pero en la práctica, ¿en qué se traduce esto? Sobre todo, asegúrate de ofrecer la información que buscan los usuarios, ya que la relevancia es lo que dirigirá el tráfico a tu sitio y te ayudará a mantenerlo.

Muchos propietarios se obsesionan con la buena clasificación de sus páginas web, pero la clasificación está determinada por más de 200 criterios además del PageRank. Es mucho mejor concentrar tu tiempo en la calidad del contenido y su accesibilidad que intentar buscar formas de descubrir el algoritmo del motor de búsqueda. Si un sitio no cumple nuestras directrices de calidad, lo bloquearemos del índice.

Qué hacer:

1. Crear un contenido relevante y llamativo: los visitantes llegarán a tus páginas a través de varios vínculos, de modo que debes asegurarte de que atraerán su atención.
2. Implicar a los usuarios: ¿puedes añadir una sección de comentarios o blog en tu sitio web? Construir una comunidad ayuda a impulsar el uso habitual de tu sitio. Implicar a los visitantes te ayudará a aumentar la visibilidad y la fidelidad de los usuarios.
3. Controlar tu sitio: utiliza el Centro para webmasters (ver página 13) para ver las consultas que están dirigiendo a los visitantes a tu sitio o para hacer un seguimiento de los cambios de clasificación en los resultados de búsqueda en relación con las modificaciones grandes realizadas al sitio.
4. Buscar la calidad y vínculos entrantes.
5. Proporcionar vínculos de texto claros: coloca vínculos de texto adecuadamente en tu sitio y asegúrate de que incluyan términos que describen el tema.

Qué debes evitar:

1. No llenes la página de listas de palabras clave.
2. No intentes encubrir páginas escribiendo texto que puedan observar los motores de búsqueda pero no los usuarios.
3. No coloques páginas “sólo para rastreo” configurando páginas o vínculos cuyo único objetivo sea engañar a los motores de búsqueda.
4. No utilices imágenes para mostrar nombres, contenido o vínculos importantes, ya que los motores de búsqueda no pueden leer imágenes.
5. No crees varias copias de una página en distintos URL con la intención de despistar a los motores de búsqueda.

En caso de dudas, consulta nuestras directrices para webmasters, disponibles en: google.es/webmasters/guidelines.html

Centro para webmasters

Como empresa que desea proporcionar los resultados de búsqueda más relevantes y útiles en la Web, intentamos proporcionar una asistencia escalable y equitativa a todos los webmasters de todos los sitios web, independientemente de lo grandes que sean. Por ello, hemos creado el Centro para webmasters, que encontrarás en google.es/webmasters.

El Centro para webmasters es un recurso fantástico para todos los editores web. Responde extensamente a todas las preguntas sobre rastreo, indexación y clasificación, ofrece la posibilidad de compartir comentarios y problemas, e incluye herramientas de diagnóstico que ayudan a los webmasters a depurar los problemas potenciales del rastreo.

 **Centro para webmasters**

Le damos la bienvenida a este punto centralizado desde el que podrá obtener la información más exhaustiva en relación con los procesos de rastreo e indexación de sitios web realizados por Google. Aquí aprenderá el método para asegurarse de que el rastreo y la indexación de su sitio se realicen de la forma más sencilla. Dispondrá de herramientas que le facilitarán el diagnóstico de cualquier incidencia en el rastreo. Podrá valorar los datos estadísticos relativos al comportamiento de su sitio una vez incorporado al índice. Y también tendrá la posibilidad de indicarnos qué preferencias de rastreo e indexación estima más oportunas para su caso.

**Asistente informativo sobre el estado del sitio**
Descubra si su sitio forma parte ya del índice de Google.

**Blog de Google destinado a webmasters (inglés)**
Las últimas novedades y las noticias más recientes relacionadas con los sistemas de rastreo e indexación de sitios web utilizados por Google.

**Herramientas para webmasters (incluido Sitemaps)**
Estadísticas, diagnósticos y administración de los procesos de rastreo e indexación que realiza Google de su sitio web. Incluidas funciones como el envío de Sitemaps y la generación de informes.

**Foro de debate de Google para webmasters**
Entable conversación con otros webmasters y comparta sus ideas con nosotros.

**Envíe contenido a Google**
Descubra los métodos para enviar contenido a distintos servicios de Google, como Google Base y Google Libros.

**Centro de Asistencia para webmasters**
Consulte las respuestas a las preguntas más frecuentes formuladas acerca del rastreo, la indexación, la clasificación y otros muchos temas.

Aquí tienes una muestra de lo que encontrarás y podrás hacer en el Centro para webmasters.

- Diagnosticar problemas potenciales para acceder a páginas y ofrecer soluciones.
- Solicitar la eliminación de páginas determinadas de nuestro índice.
- Garantizar que tu archivo robots.txt permite el acceso o bloqueo de las páginas que desees.
- Ver las estadísticas de la página y las consultas relacionadas con tu sitio web.
- Estadísticas de las consultas: determina cuáles de las consultas proporcionan más visitantes al sitio y qué temas podrías ampliar en el mismo para captar más tráfico.
- Análisis de la página: te permite ver tu página tal como lo hace Google. Visualiza las palabras más habituales de tu sitio, los vínculos entrantes y cómo describen otros tu sitio cuando enlazan a éste.
- Frecuencia de rastreo: Observa la frecuencia con que Googlebot rastrea tu sitio e indícanos si debería rastrearse con más o menos frecuencia.

Sitemaps

El Centro para webmasters también ofrece a los editores Sitemaps para búsqueda web, móvil y resultados de noticias.

Sitemaps es un protocolo compatible con otros motores de búsqueda para ayudar a los webmasters a que nos den más información acerca de sus páginas web. Complementa los mecanismos de rastreo web estándar existentes y los webmasters pueden utilizarlo para informarnos sobre las páginas de sus sitios, de manera que podamos mejorar tanto el rastreo como la visibilidad de sus páginas en los resultados de búsqueda de Google.

Además del protocolo Sitemaps para la búsqueda web, también ofrecemos Sitemaps para Google para móviles, que permite que los editores envíen las URL que muestran el contenido para dispositivos móviles en nuestro índice móvil.

Y para aquellos editores cuyo sitio de noticias esté incluido en Google Noticias, el Sitemaps de Noticias puede ayudarles a generar las estadísticas acerca de sus artículos, desde consultas hasta frecuencias de aparición. Junto con las herramientas de diagnóstico del Centro para webmasters, el Sitemaps de Noticias también puede proporcionar los informes sobre errores que te ayudarán a entender los problemas que pueda tener Google para rastrear o extraer los artículos de noticias de un sitio de edición de noticias. Además, el editor puede enviar un Sitemaps de Noticias que contenga las URL que desee que se consideren para incluirlas en Google Noticias. A diferencia del protocolo Sitemaps para la búsqueda web y Sitemaps para móviles, Sitemaps de Noticias actualmente sólo está disponible en inglés, aunque esperamos ampliarlo a otros idiomas muy pronto.

Preguntas frecuentes

¿Por qué no puedo recibir asistencia personal para mi sitio web?

Según cálculos aproximados, hay más de 100 millones de sitios en la Web. Todo los sitios son importantes para nosotros, ya que sin ellos, tanto los más pequeños como los más grandes, nuestro índice sería menos completo y, finalmente, menos útil para nuestros usuarios.

El Centro para webmasters es un recurso fantástico de asistencia para todo tipo de sitios web. Escribimos y respondemos a las preguntas de los editores para que todos puedan beneficiarse de la información proporcionada. En el Centro para webmasters también encontrará una comunidad de webmasters útil y amable con la que podrá compartir sugerencias y soluciones a los problemas.

¿Ejercen los anuncios que se muestran alguna influencia en las clasificaciones? ¿Son las listas de anuncios independientes completamente de los resultados de búsqueda?

Las clasificaciones sobre anuncios y búsqueda no están relacionadas de ninguna manera. De hecho, disponemos de equipos completamente separados que trabajan en cada uno de los sectores, de modo que no hay interferencia alguna entre ambos. Creemos que la objetividad de nuestros resultados de búsqueda es básica para poder proporcionar la mejor experiencia a los usuarios.

¿Qué debo hacer para añadir un sitio al índice de búsqueda de Google?

La inclusión en los resultados de búsqueda de Google es gratuita y sencilla, sin necesidad de enviar el sitio manualmente a Google. Google es un motor de búsqueda completamente automatizado; rastrea la Web regularmente y encuentra los sitios que debe añadir a nuestro índice. De hecho, la gran mayoría de sitios incluidos en nuestros resultados no se han enviado manualmente, sino que se han encontrado y añadido de forma automática cuando nuestros robots han rastreado la Web.

Además, las herramientas para webmasters de Google (en el Centro para webmasters) proporcionan una forma fácil de enviar un sitemap o sus URL al índice de Google y de obtener informes detallados acerca de la visibilidad de sus páginas en Google. Con las herramientas para webmasters de Google, los propietarios de un sitio pueden mantener informado a Google automáticamente acerca de las páginas actuales y de los cambios que hayan sufrido.

¿Cuánto tiempo tarda Google, aproximadamente, en encontrar un sitio web recién creado, y con qué frecuencia rastrea la Web en general?

No hay un tiempo definido para que Google encuentre un sitio nuevo.

Googlebot rastrea de forma regular la Web para volver a crear el índice. Mediante el Centro para webmasters, estos podrán ver la frecuencia con la que Googlebot rastrea su sitio y comentarnos si desean que la frecuencia sea mayor o menor.

¿Qué pasa si quiero que mi sitio web aparezca en los resultados de búsqueda web, pero no en otros servicios como Google Noticias o Búsqueda de imágenes de Google?

Google siempre permite que los editores web rechacen sus servicios, de modo que pueden ponerse en contacto con el equipo de asistencia del producto correspondiente para hacerlo.

Como hemos mencionado anteriormente, el Protocolo de exclusión de robots puede utilizarse para bloquear la indexación tanto de imágenes como de páginas web. Para ello, también se puede utilizar la función “eliminar URL” del Centro para webmasters, ya que incluye la búsqueda web y búsqueda de imágenes.

Además, dado que Googlebot cuenta con varios robots diferentes, usted puede orientar lo que desee bloquear.

- Googlebot: rastrea páginas para nuestro índice web y de noticias.
- Googlebot-Mobile: rastrea páginas para nuestro índice móvil.
- Googlebot-Image: rastrea páginas para nuestro índice de imágenes.
- Mediapartners-Google: rastrea páginas para determinar el contenido de AdSense. Sólo utilizamos este robot para rastrear su sitio si publica anuncios de AdSense en él.
- Adsbot-Google: rastrea páginas para medir la calidad de la página de destino de AdWords. Sólo utilizamos este robot si usted utiliza AdWords de Google para anunciar su sitio.

Por ejemplo, para bloquear Googlebot completamente, puede utilizar los siguientes comandos:

User-agent: Googlebot

Disallow:

¿Puedo seleccionar el texto que deseo que se vea en el fragmento de texto?

No. No es buena idea, ni desde la perspectiva del usuario ni desde la del creador de contenido. Nosotros seleccionamos un fragmento de texto del sitio que muestra la consulta realizada en contexto, que a su vez demuestra la relevancia del resultado.

Los estudios demuestran que es más posible que los usuarios visiten un sitio web si los resultados de búsqueda muestran el fragmento de texto. Esto se debe a que los fragmentos de texto hacen que el usuario visualice mejor por qué el resultado es relevante para su búsqueda. Si un usuario no puede determinar la relevancia de forma rápida, se irá hacia el siguiente resultado de búsqueda.

Los editores web pueden incluir una metaetiqueta en sus páginas para proporcionar información adicional a Google en los casos en los que no podamos generar un fragmento de texto útil del contenido de la página de forma algorítmica. Para ello, simplemente debes añadir lo siguiente a la sección <head> de la página:

```
<meta name="description" content="¿Por qué a Ana le gustan los gatitos? Estamos intentando averiguarlo.">
```

Cualquier webmaster que no desee que sus páginas generen un fragmento de texto puede utilizar una etiqueta NOSNIPPET tal como aparece a continuación:

```
<meta name="robots" content="nosnippet">
```

Finalmente, a veces utilizamos la descripción de un sitio para sacar el fragmento de texto de los resultados de búsqueda. Si no deseas que utilicemos esta descripción, sólo debes añadir la siguiente metaetiqueta:

```
<meta name="robots" content="noodp">
```

Las noticias más recientes de mi sitio sólo aparecen durante unas cuantas horas antes de actualizarse y colocarse en la sección de artículos estándar. Quiero que aparezca el artículo entero en el índice de Google, pero no este tipo de artículos recientes.

Una opción es colocar todos los artículos de noticias recientes en un directorio y utilizar robots.txt para impedir que Googlebot acceda a este directorio.

Otra opción es añadir la etiqueta NOFOLLOW en la sección <HEAD> del html de tus artículos de noticias recientes. Esto hace que Googlebot no siga ningún vínculo que encuentre en esa página. Ten presente que, pese a ello, esta etiqueta sólo hace que Googlebot no siga los vínculos de una página a otra. Pero si otra página web enlaza con ese artículo, Google podrá encontrarlo cuando lo indexe.

Si dispongo de varios dominios y publico el mismo contenido a través de cada uno de ellos, ¿pueden expulsarme de los resultados de búsqueda?

Pese a que algunos webmasters pueden intentar engañar a los motores de búsqueda duplicando el contenido y mostrando sitios espejo, también hay contenido que puede duplicarse por un buen motivo. No deseamos penalizar esos sitios. Por ejemplo, no tratamos el mismo contenido publicado en varios idiomas (supongamos que un sitio está en inglés y otro en francés) como contenido duplicado.

El mismo contenido en sitios web diferentes, por ejemplo, la sindicación de artículos, no dará pie a que uno o algunos de esos sitios se eliminen completamente de los resultados de búsqueda. Sin embargo, recuerda que es probable que cada aparición del artículo se incluya en una clasificación inferior, ya que sólo tiene una fracción de los vínculos entrantes que podría tener una sola copia. En general, una única copia de un artículo se clasificará en mejor posición y, por lo tanto, la verán más usuarios que las copias múltiples del mismo contenido.

Además, para garantizar la calidad de la búsqueda, Google no incluye varias copias de una página en los resultados de búsqueda. Todo lo contrario, normalmente escogemos sólo una versión de la página para mostrarla. Sin embargo, los webmasters pueden indicarnos la versión que prefieren mediante robots.txt o una metaetiqueta para bloquear las copias que no deseen mostrar en nuestros resultados de búsqueda.

¿Por qué está bloqueado mi sitio en el índice de Google?

En primer lugar, tu sitio puede no estar bloqueado. Hay muchas razones por las que un sitio puede no aparecer en nuestros resultados de búsqueda (ver páginas 5-11).

Si tu sitio no presenta ningún obstáculo para la detección o indexación, entonces sí es posible que esté bloqueado. Los sitios pueden estar bloqueados en nuestro índice porque no cumplen los estándares de calidad expresados en la Guía para webmasters, disponible en el Centro para webmasters. Esto suele suceder cuando un sitio web utiliza métodos desleales para intentar aparecer en posiciones más elevadas en las clasificaciones de búsqueda. Entre las infracciones habituales de las políticas se incluye el encubrimiento (escribir texto de forma que lo lean los motores de búsqueda, pero no los usuarios) o la configuración de páginas o vínculos para engañar a los motores de búsqueda y manipular los resultados de éstos.

Cuando los webmasters sospechen que su sitio infringe nuestras políticas de calidad, pueden modificarlo para cumplir estas políticas y, a continuación, hacer clic en el vínculo “solicitar reinclusión” en la interfaz de las herramientas para webmasters para solicitar que volvamos a evaluar el sitio.

Glosario

Contenido dinámico

Imágenes, animaciones o vídeos que dependen de Flash, JavaScript, marcos, o URL generadas dinámicamente.

Directorio raíz

Directorio principal del sistema de archivos de un equipo.

Encubrimiento (cloaking)

Mostrar a los motores de búsqueda un contenido diferente al que se muestra a los usuarios.

Extensión de archivo

Parte del nombre de un archivo (.doc, .txt, .pdf, etc.) que normalmente se utiliza para indicar el tipo de datos almacenados en el archivo.

HTML (Hypertext Markup Language)

Lenguaje con códigos que se utiliza en la Web para estructurar el texto.

Indexar

Proceso por el que el contenido de un sitio se añade a un motor de búsqueda.

Metaetiquetas

Etiqueta HTML que describe el contenido de una página web. Se pueden utilizar para controlar la indexación de páginas individuales en un sitio web.

Page Rank

Función de Google que ayuda a determinar la clasificación de un sitio en nuestros resultados de búsqueda. PageRank cuenta con la naturaleza democrática exclusiva de la Web al utilizar su enorme estructura de vínculos como indicador del valor de una página individual. Los sitios importantes y de calidad reciben un PageRank superior, que Google recuerda cada vez que realiza una búsqueda. Google combina PageRank con técnicas de concordancia de texto sofisticadas para buscar páginas que sean tanto importantes como relevantes para las búsquedas.

Palabra clave

Palabra que se introduce en el cuadro de búsqueda de un motor de búsqueda. De este modo, el motor de búsqueda busca las páginas que incluyen esta palabra o frase.

Protocolo de exclusión de robots

Especificación técnica que ordena a los motores de búsqueda qué sitio o partes del sitio no deben buscar, y qué partes deben permanecer visibles en los resultados de búsqueda.

Robots.txt

Archivo de texto que permite a los editores web controlar el acceso a sus sitios a diferentes niveles, desde el sitio entero a directorios individuales, tipos de páginas concretas o incluso páginas individuales. Este archivo indica a los rastreadores los directorios que pueden o no rastrear.

Sistema de administración de contenido (CMS)

Sistema de software que se utiliza para gestionar el contenido de los archivos, imágenes y audio en la web.

Sitio espejo

Página web duplicada; a veces se utiliza para engañar a los motores de búsqueda e intentar optimizar la indexación y clasificación de un sitio web.

URL (Uniform Resource Locator)

Dirección de un sitio web en Internet que consta del protocolo de acceso (http), el nombre del dominio (*www.google.es*) y, en algunos casos, la ubicación de otro archivo (*www.google.es/webmaster*).

Vínculo en caché

El vínculo en caché instantánea de cómo apareció una página la última vez que Google la visitó. Una copia en caché permite a los usuarios visualizar una página aunque su versión actual no esté disponible. No obstante, el contenido puede variar ligeramente. Para visualizar una copia en caché, haz clic en el vínculo “en caché” que aparece bajo los resultados de búsqueda.

Para obtener más información acerca del Centro para webmasters, visita la página

google.es/webmasters/

