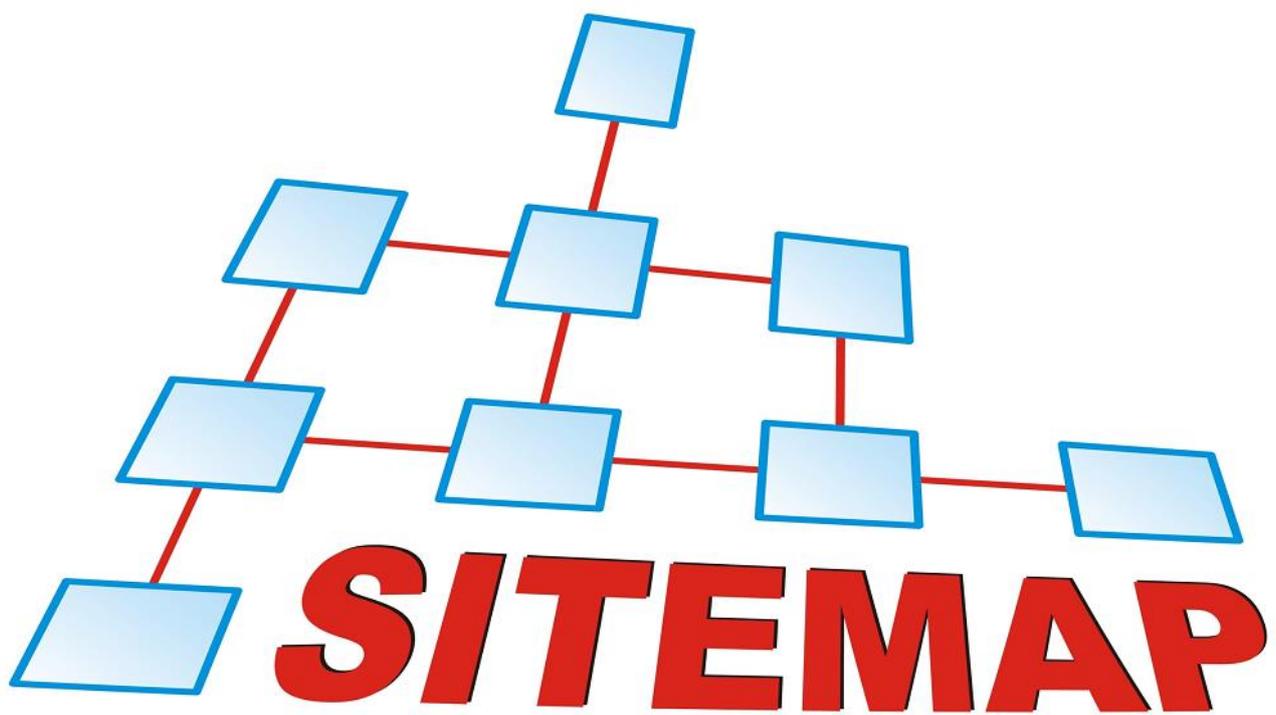


Sitemaps, qué son y cómo generarlos



Cualquier portal web, por pequeño que pueda ser, estará formado por varias páginas donde se muestran información sobre la empresa, noticias de interés o productos en el caso de tratarse de un **comercio online**. El problema que nos podemos encontrar es que hay ocasiones en que los buscadores no encuentran esas páginas que forman parte de nuestro sitio. En estos casos, el uso de los sitemaps se ha convertido en un recurso muy valioso para mejorar la indexación.

¿Qué es un sitemap?

Puede ser que muchos de vosotros hayáis escuchado hablar alguna vez del sitemap, pero para aquellos que no sepan de qué se trata, de forma sencilla se puede definir como un archivo escrito en lenguaje de marcas extensibles (eXtensible Markup Language o XML) que se encarga de proporcionar un listado de las páginas que forman parte del sitio, las cuales tienes que indicar tú.

Estos archivos sirven de guía a los **robots de los buscadores** para realizar un rastreo optimizado de la web. Un archivo sitemap.xml generado correctamente, junto con otras pautas, nos ayudará a conseguir que nuestro sitio web se indexe de una forma más adecuada.

Consideraciones antes de generar el sitemap

Antes de meternos de lleno con la explicación de estas etiquetas, es necesario tener presente una serie de consideraciones:

1. Las URLs que forman parte del archivo sitemap.xml deben pertenecer al mismo **dominio**. No se aceptan subdominios que cuelguen del dominio principal.
2. Hay que utilizar siempre el mismo protocolo. Es decir, si la web utiliza el protocolo HTTP, no se pueden indicar las URLs haciendo uso del protocolo HTTPS.
3. El tamaño del archivo no puede superar los 10 MB.
4. El número máximo de URLs que pueden aparecer en el fichero, no pueden superar el valor de 50.000.
5. Debe estar ubicado en el directorio raíz de nuestro sitio.

Posteriormente, explicaremos cómo hacer en el caso de superar el tamaño de 10 MB y las 50.000 URLs. En esos casos, se puede utilizar más de un sitemap.

¿Cómo se generan los archivos sitemap.xml?

Como cualquier archivo de formato XML, hay que hacer uso de una serie de etiquetas y de las que algunas son obligatorias de utilizar para generar este tipo de archivos. Al tratarse de un archivo XML, lo podemos generar con un simple editor de texto, aunque recomendamos utilizar algún editor tipo **Notepad++** que resalta las etiquetas utilizadas.

Como hemos comentado con anterioridad, hay una serie de etiquetas obligatorias que debemos usar para la correcta generación de nuestro archivo. Estas etiquetas son:

- **<urlset>**: Se trata de la etiqueta utilizada para la apertura del archivo. A su finalización, hay que utilizar la etiqueta **</urlset>**.

- **<url>**: Indica que el bloque que contendrá esta etiqueta es una url, tendremos una por cada una de las urls que vayamos a especificar en el sitemap. Para indicar el final del bloque, se utiliza la etiqueta **</url>**.
- **<loc>**: El contenido de esta etiqueta define la ruta de la url en la que se debe indicar el protocolo utilizado: HTTP o HTTPS. Su etiqueta de cierre es **</loc>**.

Un ejemplo de sitemap creado a partir de las anteriores etiquetas podría ser el siguiente:

```
1 <urlset>
2 <url>
3   <loc>http://www.misitio.com/</loc>
4 </url>
5
6 <url>
7   <loc>http://www.misitio.com/quienes-somos</loc>
8 </url>
9
10 <url>
11   <loc>http://www.misitio.com/contacto</loc>
12 </url>
13 </urlset>
```

Pero además de las etiquetas obligatorias, también nos podemos encontrar otras opcionales que aportan más información a los buscadores. Veamos cuáles son estas etiquetas.

- **<lastmod>**: Sirve para indicar la última modificación del archivo. Además de la fecha, también se puede indicar la hora. El formato de la fecha debe ser: AAAA-MM-DD. Su etiqueta de cierre es **</lastmod>**.
- **<changefreq>**: Especifica la frecuencia con la que la página se modifica. Su etiqueta de cierre es **</changefreq>**. Este campo puede contener los siguientes valores.
 - always: Cada vez que se accede al documento cambia.
 - hourly: Cada hora.
 - daily: Cada día.
 - weekly: Cada semana.
 - monthly: Cada mes.
 - yearly: Cada año.
 - never: Nunca se modifica.
- **<priority>**: Se utiliza esta etiqueta para definir la prioridad de una URL respecto a las demás direcciones que forman parte de la web. Su valor puede ir desde 0.0 hasta 1.0. El valor por defecto es de 0.5. Su etiqueta de cierre es **</priority>**.

Una vez que hemos visto estas etiquetas opciones, podríamos hacer nuestro sitemap de una forma más completa.

```

▼<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
    http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd">
  ▼<url>
    <loc>https://www.hostalia.com/</loc>
    <changefreq>weekly</changefreq>
    <priority>1.00</priority>
  </url>
  ▼<url>
    <loc>https://www.hostalia.com/acceso-clientes.html</loc>
    <changefreq>weekly</changefreq>
    <priority>0.85</priority>
  </url>
  ▼<url>
    <loc>https://www.hostalia.com/actualidad/</loc>
    <changefreq>weekly</changefreq>
    <priority>0.56</priority>
  </url>
  ▼<url>
    ▼<loc>
      https://www.hostalia.com/actualidad/boletin-mensual.html
    </loc>
    <changefreq>weekly</changefreq>
    <priority>0.46</priority>
  </url>
  ▼<url>
    ▼<loc>

```

Índice de sitemaps

En un punto anterior hemos visto cómo estos archivos no podían superar los 10 MB de peso o las 50.000 URLs. Pero esto no significa que **páginas web** grandes no puedan hacer uso de los ficheros sitemaps. En el caso de sobrepasar estos límites, lo que se puede hacer es dividirlos en archivos más pequeños y subirlos por separado. Podemos darle el nombre que queramos a cada uno, pero debemos utilizar un archivo índice para indicar a los buscadores cada uno de los sitemaps que tendrán que rastrear.

En el caso del archivo índice, debemos utilizar las siguientes etiquetas a la hora de generarlo.

- **<sitemapindex>**: Se trata de la etiqueta de apertura. El fichero se finaliza con la etiqueta **</sitemapindex>** de cierre.
- **<sitemap>**: Es la etiqueta que tendremos que utilizar para indicar cada uno de los sitemaps que deberá de analizar los robots de los buscadores. Su etiqueta de cierre es **</sitemap>**.
- **<loc>**: Aquí indicaremos la dirección del sitemap. Se termina con **</loc>**.
- **<lastmod>**: Etiqueta opcional para indicar la última fecha en la que se modificó ese sitemap. Se cierra con la etiqueta **</lastmod>**.

Un ejemplo de archivo índice lo podéis ver en la siguiente imagen.

```

1 <sitemapindex>
2   <sitemap>
3     <loc>http://www.misitio.com/sitemap1.xml</loc>
4     <lastmod>2015-12-01</lastmod>
5   </sitemap>
6   <sitemap>
7     <loc>http://www.misitio.com/sitemap2.xml</loc>
8     <lastmod>2016-12-01</lastmod>
9   </sitemap>
10 </sitemapindex>

```

Cómo enviar tu Sitemap al buscador

Una vez que hemos creado nuestro sitemap, el siguiente paso será avisar a los buscadores de su existencia para que se pasen a examinarlo. Esta tarea la podemos hacer de tres formas.

1) Mediante el archivo robots.txt

Consiste en incluir una línea al final del archivo robots.txt de nuestro sitio, haciendo referencia a nuestro sitemap o en el caso de tener varios, hacia el archivo índice. La línea a incluir sería parecida a la siguiente:

```
Sitemap: http://www.misitio.com/sitemap.xml
```

2) Método Ping

Hacer ping es como decirle al buscador la ubicación de nuestro archivo sitemap. Es tan sencillo como cargar una dirección web en nuestro navegador. Veamos cómo se haría en Google y en Bing:

- **Google:** <http://www.google.com/webmasters/tools/ping?sitemap=URL>.
- **Bing:** <http://www.bing.com/webmaster/ping.aspx?sitemap=URL>

Donde URL debería ser cambiado por la dirección hacia nuestro archivo sitemap.

3) Utilizando WebMaster Tools de Google

Google ofrece una herramienta para los administradores con información valiosa para la gestión del sitio. Entre las opciones está la de enviar nuestro sitemap. Para ello, lo primero que deberemos hacer será entrar en [WebMaster Tools](#).

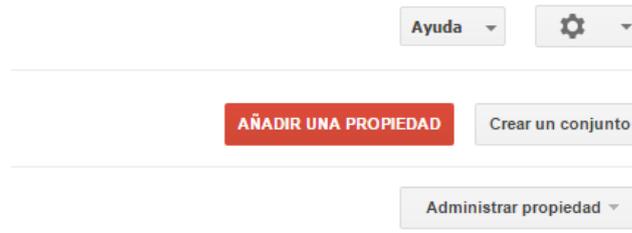
Una cuenta. Todo Google.

Accede a la cuenta para continuar a Google Search Console



The image shows a login interface for Google. At the top, there is a large, light gray circle representing a user profile picture. Below this is a text input field with the placeholder text "Ingresa tu dirección de correo electr". Underneath the input field is a blue button with the text "Siguiente". At the bottom right of the form area, there is a link that says "Buscar mi cuenta".

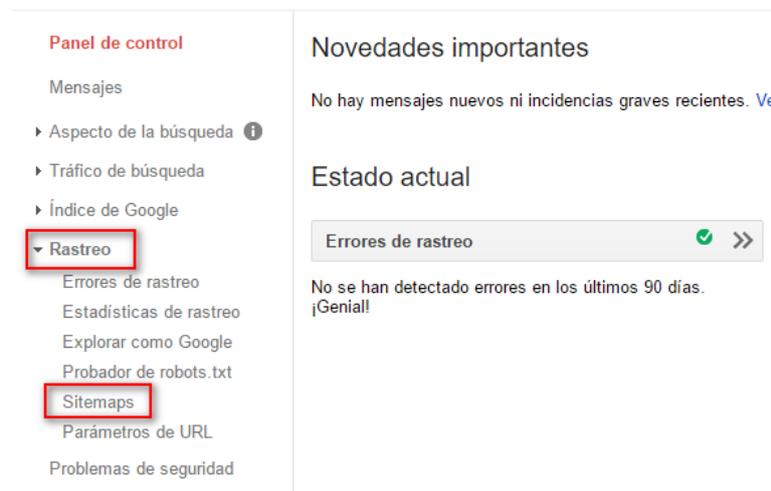
Una vez dentro, deberemos añadir nuestra página web y validarla. Para ello, debemos pulsar en la opción que pone "**Añadir una propiedad**".



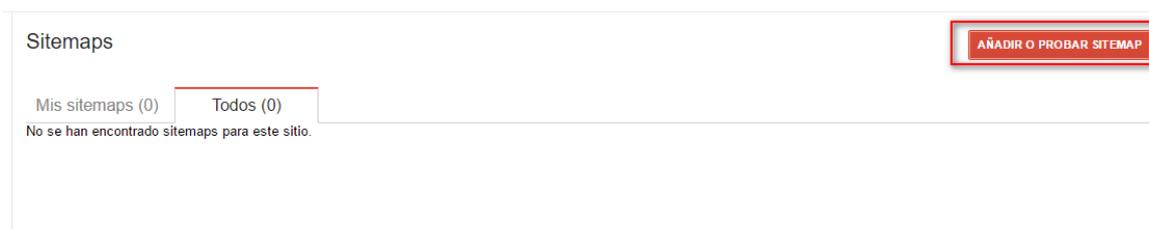
Tras haberla validado, entraremos en sus opciones pulsando sobre su nombre.



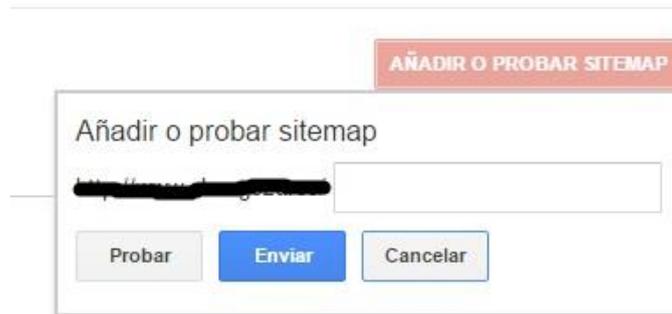
En el menú que nos aparece en la parte izquierda, pulsamos en la opción **Rastreo -> Sitemaps**.



Por último, en la esquina superior derecha, pulsaremos sobre el botón "**Añadir o probar sitemap**" para enviar la ubicación de nuestro archivo.



En la ventana que se nos abrirá, será donde habrá que indicar la ubicación. Pulsaremos en el botón enviar y solo nos quedará esperar a que los buscadores empiecen a rastrear nuestro sitemap.



The image shows a modal dialog box titled "Añadir o probar sitemap" (Add or test sitemap). The dialog has a red header bar with the text "AÑADIR O PROBAR SITEMAP". Below the title, there is a text input field containing a URL that has been redacted with black bars. At the bottom of the dialog, there are three buttons: "Probar" (Test), "Enviar" (Send), and "Cancelar" (Cancel). The "Enviar" button is highlighted in blue.

A lo largo de este libro blanco hemos visto cómo generar el sitemap para nuestro portal web e informar a los buscadores de su existencia.