# Que es

Los sitemaps son archivos en los que se proporciona información sobre las páginas, los vídeos y otros archivos de un sitio web, así como sobre las relaciones que hay entre ellos. Los buscadores leen estos archivos para rastrear los sitios web de forma más eficaz. Los sitemaps informan de qué archivos de un sitio web son importantes según el webmaster y, además, incluyen datos importantes sobre ellos.

El buscador puede usar el sitemap para ver las paginas nuevas de una página web, o todas las páginas web a la vez, incluyendo imágenes, videos, etc.

Beneficios

* Indexación rápida. Los buscadores encontraran las paginas nuevas más rápidos, por lo tanto, el proceso de indexación y exhibición de la página en los resultados de búsqueda será más rápido como bueno.
* Mejor indexación de páginas profundas. Los buscadores pueden encontrar las páginas que no fueran descubiertas mientras rastrean la página web. Aunque no significa que necesariamente vayan a estar todas indexadas.
* Monitoreo de páginas indexadas. En combinación con la Consola de Búsquedas de Google (Google Search Console), es posible encontrar que URLs están cubiertas en el sitemap que indexa Google.

# ¿Cuándo necesitamos un sitemap?

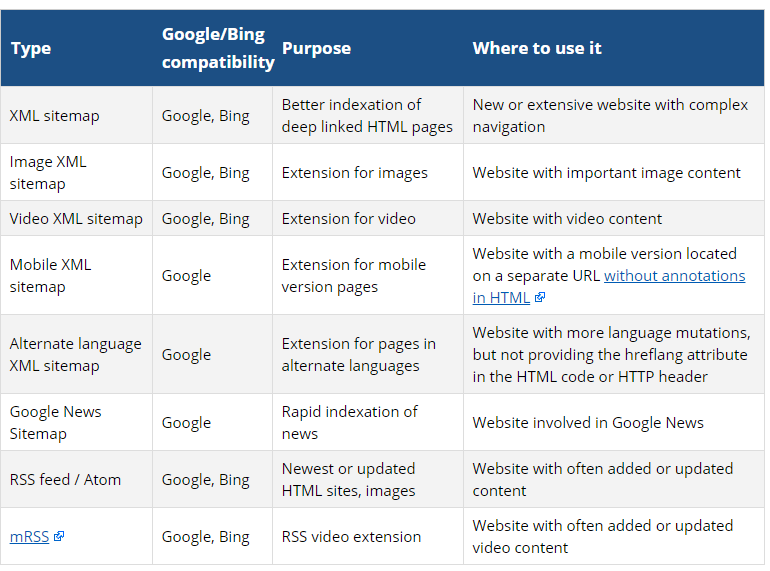
Tenemos una página web grande. Una página web con 100 paginas no es grande, incluso con 1000 paginas siguen siendo pequeña, por lo que un sitemap no es realmente necesario. En cambio, para páginas más grande, es necesario un sitemap.

El sitio web es nuevo y hay pocos enlaces externos que dirijan a él. El sitemap ayuda a los rastreadores web a buscar nuevas páginas web después de su lanzamiento y a indexar las nuevas páginas más rápido.

El sitio web tiene muchas páginas de contenido que están aisladas o no están bien enlazadas entre sí.

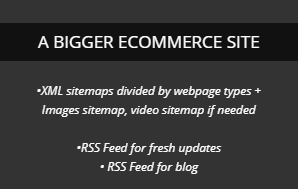
El sitio web incluye una gran cantidad de rich media, como video o imágenes, o aparece en Google Noticias.

# Tipos de Sitemap



# Tipos de sitemap según la página web

Comercio electrónico.



Blog



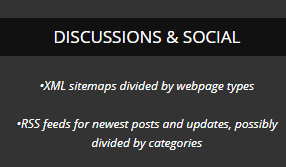
Noticias



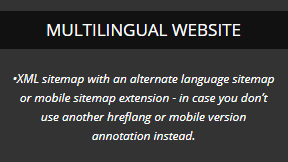
Directorio web.



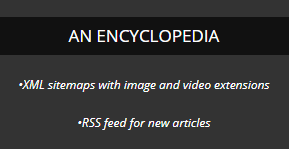
Sitios sociales y foros



Página web multilenguaje



Enciclopedia



# Generadores sitemap

Las herramientas de generadores sitemap como XML-sitemaps.com son muy populares. Simplemente metes la URL del sitio web, y después el generado empieza a rastrearlo y crear el sitemap.

Esto está bien en primera instancia, pero es desaconsejable porque a la hora de crear páginas nuevas, el sitemap no las va a reflejar, ya que solo nos da una instantánea del sitio web, por lo que estará desactualizado. También aparecerán URLs que en principio no son importantes, y otras, que puedan ser importante, pero no estén enlazadas no aparecerán.

Por estas razones, solamente se debería de usar un generador sitemap si necesitamos indexarlo rápidamente.

Principalmente hay dos formatos (XML y RSS).

# Que URLs deberían estar en un sitemap

Incluir: páginas que queramos que aparezcan en el ranking de Google y queramos enseñar a otros usuarios (las importantes y de calidad).

Excluir: páginas de utilidad (páginas que son útiles para los usuarios como de cuentas, listas de deseos, etc.), redirecciones, de errores, URLs de otros subdominios, duplicadas, etc.

# Validadores sitemap

Una vez hemos creado o generado nuestro sitemap, es una buena opción validarlo con Yandex Sitemap Validator y después con Google Search Console.

Yandex es una buena herramienta para testear sitemaps inicialmente o para revisar sitemaps que no nos pertenezcan.

Google Search Console es la mejor solución para testar el sitemap. A la consola de búsqueda le lleva algún tiempo para revisar (desde unas horas hasta días, según lo grande que sea el sitemap).

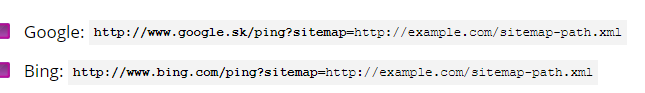
# Robots.txt

Un método que podemos utilizar para indexar los sitemap, es el de robots.txt. Si definimos más de un sitemap, podemos meter sus URLs absolutas en un fichero robots.txt. Esto ayuda al motor de búsqueda para que encuentre todos los sitemap de una forma más rápida, pero hay que tener en cuenta una cosa, que son públicos.

Tanto los sitemap en un fichero robots.txt como el código fuente son públicos, esto implica que, de una forma sencilla, los competidores pueden coger todos esos datos (lista de categorías, tags o productos) y abusar de ellos.

# Hacer ping a los motores de búsqueda después de cada actualización.

Cuando añadimos un nuevo sitio al sitemap, podemos alertar al motor de búsqueda a través de “pinging”. De esta forma, van a conocer los cambios inmediatamente y después pueden indexar el contenido nuevo.



Para cualquier otro motro de búsqueda se puede usar:

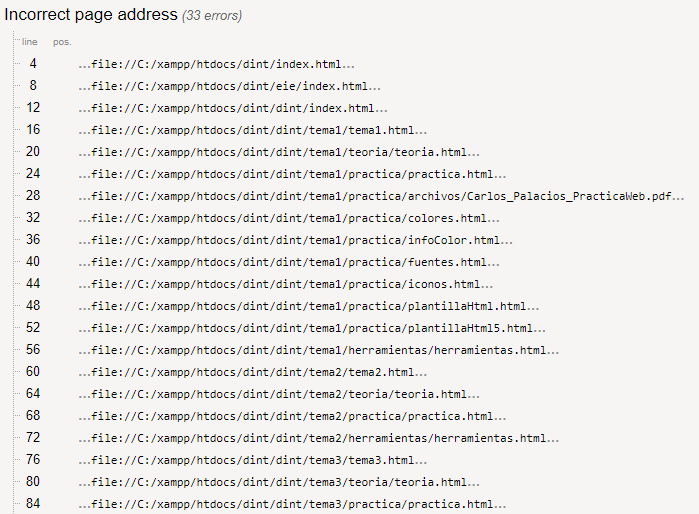


# Herramientas

PowerMapper. Con esta herramienta podemos generar el código XML de un sitemap de nuestro sitio web. En mi caso, lo he usado para el sitio web en local y el que está en github (ambos iguales).

Después, al pasarlo por el validador nos muestra los siguientes errores:

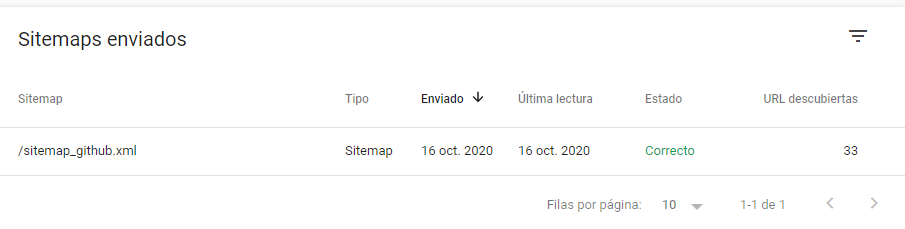
Fichero



URL. Por lo que podemos ver, el validador Yandex, al validar el fichero del sitio web no nos da ningún error.



Ahora lo valido con Google Search Console. Al validarlo puede tardar un poco más, aunque si es un sitio web pequeño, como el mio, tarda apenas unos segundos.



# Ejemplos sitemaps

PowerMapper

XML-sitemaps.com

Screaming Frog

Manual

Fuentes

<https://blog.spotibo.com/sitemap-guide/>

https://support.google.com/webmasters/answer/156184?hl=es