



VISIBILIDAD WEB UNAM

SITEMAP XML



MAYO 2020

Sitemap XML

Un sitemap es una lista estructurada de todas las páginas contenidas en un sitio web, incluida la página principal, además de los vínculos hacia archivos o documentos importantes del sitio (PDFs, documentos, videos, etc). Existen dos versiones de sitemaps: la versión HTML y la XML.

¿Para qué sirve el sitemap?

La versión XML es utilizada para indicar a los motores de búsqueda como Google, un listado estructurado de todas las páginas contenidas en un sitio web, y así agilizar su proceso de indexación en los resultados de la búsqueda; también de esta manera nos aseguramos que se indexen todas las páginas que contenga el sitio web.

La versión HTML se utiliza para proporcionar al usuario un vistazo rápido y simple de toda la estructura de un sitio web con el objetivo de ubicar de manera sencilla alguna página o archivo.

¿Cómo se crea un Sitemap XML y dónde se ubica?

Un Sitemap XML se puede generar de manera manual o automática con herramientas de terceros. Se debe de ubicar en la raíz de nuestro directorio de archivos, es decir, en donde se encuentra nuestra página principal o index.html.

Si optamos por la forma manual, tendremos que crear un documento llamado sitemap con la extensión .xml mediante cualquier editor de texto y enlistar, una a una, la ubicación de todas las páginas de un sitio web.

A continuación podemos observar el ejemplo más básico de un Sitemap XML:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://www.unam.mx/directorio.html</loc>
    <lastmod>2019-06-04</lastmod>
  </url>
</urlset>
```

En la etiqueta <loc> que está dentro de la etiqueta <url> se escribe la dirección URL completa de la ubicación de una de las páginas que se requiere indexen los buscadores; mientras que en la etiqueta <lastmod> contendrá la fecha de la última modificación.

Se deberá seguir el protocolo de un sitemap estándar el cual se puede consultar en el siguiente vínculo, además de otras especificaciones opcionales:

<https://www.sitemaps.org/es/protocol.html>

Sin embargo existen herramientas de terceros que nos permiten generar de manera automática tanto el listado de todas las páginas (URLs) de un sitio web, así como el propio archivo XML.

Las siguientes herramientas permiten crear un Sitemap XML de manera automática en línea:

<https://www.xml-sitemaps.com/>

<http://www.web-site-map.com/index.php>

Una vez que se tenga listo el Sitemap XML y se haya ubicado en el directorio raíz de nuestro sitio web, los buscadores los rastrearán de manera automática; pero también podemos agilizar este proceso (en el caso de Google) y enviarlo mediante la herramienta [Google Search Console](#).

Buenas prácticas en el uso de un Sitemap XML

1. Un Sitemap XML siempre debe de estar ubicado en la raíz del servidor del sitio web.
2. No incluir identificadores de sesión de URLs en tu sitemap para evitar que esas URL se rastreen dos veces.
3. Si el sitemap es muy grande, se deberá dividir en pequeños sitemaps de máximo 50,000 URLs o 50 MB e incluir un [sitemapindex](#).
4. Incluir solo URLs canónicas.
5. Todos los sitemap XML deben estar codificados en UTF-8 y las URLs no deberán contener caracteres especiales como *, {}, ü, etc o en su caso utilizar caracteres de escape ASCII.
6. El orden de las URLs en un sitemap XML no tiene relevancia ya que Google u otros buscadores no las rastrean en el orden en que aparecen.

¿Por qué es importante el Sitemap XML para la visibilidad de nuestros contenidos?

1. Permite exponer el listado de todas las URLs del sitio web, permitiendo de esta forma incluir todas aquellas páginas que de otro modo no habrían sido encontradas por los motores de búsqueda.
2. Proporciona prioridad en los procesos de rastreo y ofrece la posibilidad de configurar el archivo para establecer qué páginas son las más importantes.
3. Podemos incluir etiquetas opcionales que pasarán datos relevantes a los motores de búsqueda, como la frecuencia de actualización o cuáles han sido las últimas modificaciones que se han hecho sobre la página.

Créditos

Coordinación del proyecto

Dra. Marcela Peñaloza Báez

Directora de Colaboración y Vinculación

Mtro. Juan Manuel Castillejos Reyes

Subdirector de Visibilidad Web

Lic. Irene G. Sánchez García

Enlace Institucional de Iniciativas de Visibilidad

Equipo de trabajo

L.C.G. Miguel Ángel Islas Flores

Autor del documento

Lic. Mario Alberto Arredondo Guzmán

MATIE. Juan Manuel Castillejos Reyes

Revisores del documento