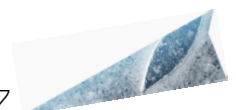


SITEMAP

CRISTINA GONZÁLEZ RECHE
17 de octubre de 2020

Diseño interface web Primer trimestre

PRÁCTICA 10:
Estandarización de
webs.
Sitemap



CONTENIDO:

- 1.- Introducción y recomendaciones W3C
- 2.- Qué es un sitemap
- 3.- Procedimiento para hacer un sitemap
- 4.- Resumen artículo sobre mapas de navegación (sitemap)
- 5.- Webgrafía

1.- Introducción y recomendaciones W3C

W3C son las siglas de Consorcio World Wide Web (W3C) es una comunidad internacional que trabaja para desarrollar estándares para páginas web.

Liderado por el inventor de la Web Tim Berners y el Director Ejecutivo (CEO) Jeffrey Jaffe , la misión del W3C es guiar la Web hacia su máximo potencial.

El trabajo de este equipo consiste en desarrollar protocolos y directrices que aseguren el crecimiento de las páginas WEB

Entre estas directivas encontramos:

- 📌 El uso de Web Semántica, refiriéndose a la visión del W3C sobre la web de datos enlazados (linked data), permitiendo la creación de almacenes de datos sobre la web, la construcción de vocabularios y regularización de reglas para manejar datos enlazados.
- 📌 La arquitectura web se centra en tecnologías que incluyen URLs y HTTP
- 📌 En cuanto a los estándares de construcción se incluye HTML5, CSS, SVG, Ajax y otras tecnologías para aplicaciones web, siendo otros estándares XML, XQuery, XML Schema, XSLT, etc.
- 📌 La web de servicios está basada en HTTP, XML, SOAP entre otros

Otro de los puntos prioritarios para este grupo de trabajo, es que se permita el acceso a las web's desde cualquier lugar, momento y dispositivo, incluyendo además de los teléfonos móviles, el uso de impresoras, televisión, automóviles, etc.

<https://html.spec.whatwg.org/multipage/sections.html#sections>



2.- Qué es un sitemap

Necesitamos un sitemap para que los motores de búsqueda de google no tengan que ir de hipervínculo en hipervínculo buscando por nuestro sitio web lo que alguien les ha preguntado.

Como su propio nombre indica es un mapa de un sitio web. Es un "documento" que contiene la información que indica el camino correcto que debe seguir un crawler, que no es otra cosa, que un robot que va rastreando internet, y va localizando páginas relevantes para cada búsqueda, que para el caso de Google es -> GoogleBot.

La función que hace un sitemap es facilitar el trabajo de este robot de Google, haciendo que los procesos de indexación sean más rápidos y eficientes, tanto en tiempo de rastreo como en carga de las páginas buscadas.

La información es almacenada en el servidor en el que está alojado el sitio web y contiene toda la estructura del sitio, es como un índice, de forma que al acceder a él se optimiza la eficiencia del buscador, para de esta manera obtener resultados más veloces y precisos.

En el contenido de este sitemap encontramos las URL principales del sitio web, que serán las que primero se van a leer, información también sobre jerarquía, enlaces y actualizaciones, por lo que es muy importante que contenga las fechas de las actualizaciones de nuestras páginas, así como la información multimedia de videos, audios, imágenes, etc.

En consecuencia con este sitemap conseguiremos que nuestro sitio web sea localizado por más gente y más rápido, haciendo de esta forma una labor de "marketing" de nuestro espacio en internet.

Una de las principales técnicas es el link building, que no es ni más ni menos que las referencias a otras páginas que redireccionan al usuario a tu sitio web. Bajo este concepto encontramos las siglas SEO (Searching Engine Optimization), que son una serie de técnicas que conducen a conquistar buenas posiciones en los resultados de mecanismos de búsqueda, sobre todo en Google.

3.- Procedimiento para hacer un sitemap

1.- Creación de un **sitio web** compuesto por varias páginas web enlazadas entre ellas. Este sitio web debe tener sus correspondientes archivos index.

2.- Creación de un proyecto que aloje este sitio web, bien en una empresa de **hosting**, en un repositorio como el que vamos a usar nosotros para hacer el trabajo que es github. Una vez que tengamos nuestra estructura de directorios con nuestras páginas en un proyecto de github, por settings del proyecto obtendremos la url de enlace a nuestro sitio web



GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

✓ Your site is published at <https://cristinagreche.github.io/dint/>

Source

Your GitHub Pages site is currently being built from the `main` branch. [Learn more.](#)

Branch: main ▼

📁 / (root) ▼

Save

3.- Creación del sitemap, bien de forma manual o con alguna herramienta de ayuda de las múltiples que tenemos en internet obtendremos un documento XML de nuestro sitio web. Yo he utilizado **powermapper** (<https://www.powermapper.com>).

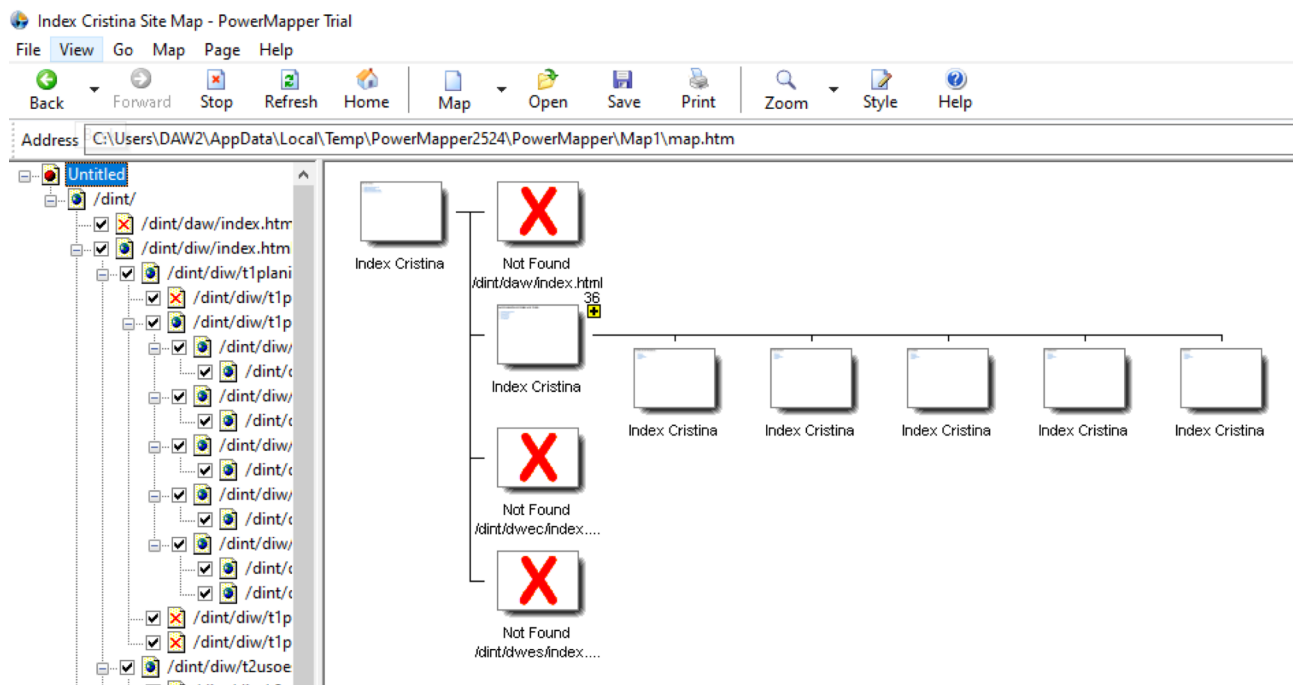
PowerMapper
Desktop

PowerMapper is a visual site mapping tool for Macintosh and Windows.

Windows and Mac

Que nos ofrece una prueba gratuita de 30 días o la obtención online del sitemap

Con esta herramienta además de obtener el XML del sitio web, tenemos otros tantos recursos como visualización de forma gráfica del contenido de nuestro espacio



4.- Subir nuestro documento sitemap en XML a github, y ya solo nos falta decir a google que tenemos todo preparado para que revise nuestro sitemap y nos meta en sus buscadores de forma eficiente.

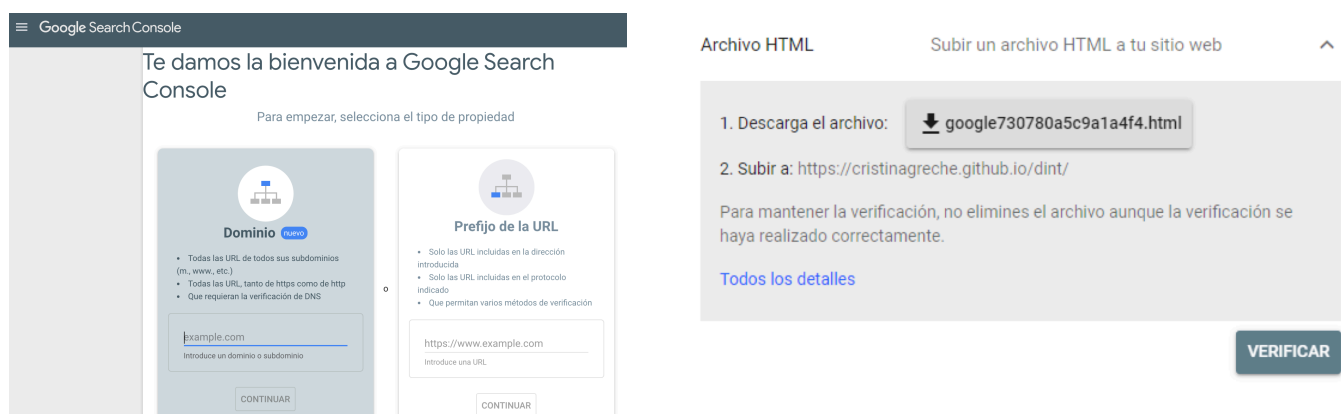
Primeras líneas de código de mi sitemap

```
79 líneas (79 sloc) | 3,19 KB
1  <? xml versión = " 1.0 " codificación = " UTF-8 " ?>
2  < urlset xsi : schemaLocation = " http://www.sitemaps.org/schemas/sitemap/0.9 http://www.sitemaps.org/schemas/sitemap/
3  < url >
4  < loc > https://cristinagreche.github.io/dint/ </ loc >
5  < lastmod > 2020-10-16T17: 44: 50 + 00: 00 </ lastmod >
6  </ url >
7  < url >
8  < loc > https://cristinagreche.github.io/dint/diw/index.html </ loc >
9  < lastmod > 2020-10-16T17: 44: 49 + 00: 00 </ lastmod >
10 </ url >
```

Tenemos que fijarnos que las URL que en él están no tengan referencia a nuestro C: del ordenador, si esto ocurriera, nunca vamos a obtener el resultado esperado de un sitemap, que es que cualquiera, desde cualquier SO pueda acceder a nuestro sitio web. Las URL que proporciona el sitemap deben de ser referencias, en nuestro caso, al repositorio donde hemos alojado nuestro proyecto, o a una empresa de hosting (000webhost, WordPress, etc)

5.- El último paso, como decía, es comunicar a Google que nuestro proyecto está preparado para que nos lo valide. Para ello, y gracias a una herramienta llamada Google Search Console, facilitando la url que nos proporciona github con CODE en prefijo de la URL, obtenemos un archivo html de google que nos tenemos que descargar y que también tenemos que incluir en nuestro proyecto de github.

Nos descargamos el enlace que nos da y tal y como nos indica lo subimos a github



Google Search Console

Te damos la bienvenida a Google Search Console

Para empezar, selecciona el tipo de propiedad

Dominio nuevo

- Todas las URL de todos sus subdominios (m, www, etc.)
- Todas las URL, tanto de https como de http
- Que requieran la verificación de DNS

Introduce un dominio o subdominio

Prefijo de la URL

- Solo las URL incluidas en la dirección introducida
- Solo las URL incluidas en el protocolo indicado
- Que permitan varios métodos de verificación

Introduce una URL

Archivo HTML

Subir un archivo HTML a tu sitio web

1. Descarga el archivo: [google730780a5c9a1a4f4.html](#)

2. Subir a: <https://cristinagreche.github.io/dint/>








Para mantener la verificación, no elimines el archivo aunque la verificación se haya realizado correctamente.

[Todos los detalles](#)

VERIFICAR

Una vez que tengamos también en nuestro repositorio el archivo que nos ha facilitado la herramienta **Google Search Console**, por la que avisamos de que tenemos nuestro sitemap preparado, ya tenemos todo listo para conseguir nuestros fines



 .vscode	commit2 insistiendo
 diw	commit2 insistiendo
 ._index.html	commit2 insistiendo
 README.md	Compromiso inicial
 google730780a5c9a1a4f4.html	commit4 aviso a google
 index.html	commit2 insistiendo
 sitemap.xml	commit3 subo mapa del sitio

4.- Resumen artículo sobre mapas de navegación (sitemap)

Un sitemap ayuda a realizar una indexación de un sitio web para facilitar a los motores de búsqueda su trabajo.

Un mapa del sitio es una lista de páginas web creada para rastreadores web de forma que puedan encontrar su contenido web de forma rápida y sencilla

Hay que tener en cuenta que un mapa del sitio no es un factor de clasificación, no vale con enumerar URL's

1.- Existen varios tipos de mapas de sitio

En XML: básico, de imagen, de video, móvil, en idioma alternativo

Otros: de noticias, de fuente RSS

2.- Cómo hacer el mapa del sitio

Existen dos formatos básicos: XML y RSS. Vamos a centrarnos en un sistema en XML

2.1. Declaración, forma que tienen los rastreadores de reconocer un sitemap

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
</urlset>
```

2.2. Agregar URL

La etiqueta <urlset> contiene además de la <url>, que irán colocadas una seguida de otra, la etiqueta <loc> para añadir las URL las URL importantes



2.3. Agregar lastmod a la URL

Con la etiqueta <lastmod> dentro de <url> podemos acelerar el rastreo con contenido de últimas actualizaciones.

3.- Directrices del mapa del sitio

Un sitemap debe cumplir unas reglas, como usar codificación UTF-8, usar URL absolutas, no haciendo referencia a directorios que cuelgan de nuestro disco C:, existen caracteres como < > ' " Y que necesitan escapar para no confundirse con los caracteres propios del lenguaje utilizado, escapando con < > ' " y &erio, así como también los diacríticos y caracteres no ASCII requieren de escape del carácter

Los sistemas tienen también unas limitaciones, no pudiendo contener más de 50.000 URL, o no pudiendo superar el tamaño del archivo los 50 Mb sin comprimir.

No se deben incluir todas las URL del sitio, sino sólo las más importantes

Si se incluyen imágenes, se pueden agregar hasta 1000 imágenes por cada página web, no teniendo que estar necesariamente en el mismo dominio que el sitio web. Y se incluyen con la etiqueta <image: image>, teniendo como atributos: loc, titile, license, geo_location, caption, etc.

En cuanto a los videos, éstos deben tener un tamaño mínimo de 160 x 90 px y máximo de 1920 x 1080 px, la descripción del video debe coincidir con la descripción que se muestra en la página web, agregándose con la etiqueta <video: video>, teniendo como atributos: thumbnail_loc, titile, descripción, content_loc, etc.

Las fuentes RSS son archivos que contienen las últimas diez actualizaciones. Esta información se pasa en el sitemap con la etiqueta <channel> y cada URL en la etiqueta <item>. Si el contenido es de video en vez de RSS utilizaremos mRSS

Finalmente se debe validar el mapa del sitio generado con Yandex Sitemap Validator y luego con Google Search Console

A partir de aquí los motores de búsqueda encuentran y utilizan el sitemap con varios métodos como robots.txt, registro como editor para utilizar Google News Sitemap, etc.

5.- Webgrafía

<https://rockcontent.com/es/blog/sitemap/> (info sobre sitemap)

<https://www.powermapper.com> (herramienta que hace sitemap en XML de forma automática)

<https://search.google.com/search-console/about?hl=es> (google search console)

<https://blog.spotibo.com/sitemap-guide/> (artículo a comentar)

