

The background image shows a bright, modern office space with large windows. Several people are visible: one person is standing and pointing at a whiteboard, while others are seated at desks, some looking at laptops. The office has a clean, professional feel with wooden floors and modern furniture.

GRUPO
CLAVE

Curso “Creación y Diseño de Páginas Web”

Lugar, y fechas: CEFEM Ayuntamiento de Santander
Formador/a: Marián Toca

Un **sitemap** es un archivo **.xml** que se aloja en una URL de nuestro “site” y la visitan los **bots** de los distintos buscadores.

Los rastreadores web de los motores de búsqueda, por ejemplo, el **robot de Google**, leen este archivo para rastrear el sitio de forma más inteligente.

Nos proporcionan valiosos metadatos asociados a las páginas que enumeramos en él.



Los metadatos son información sobre una página web:

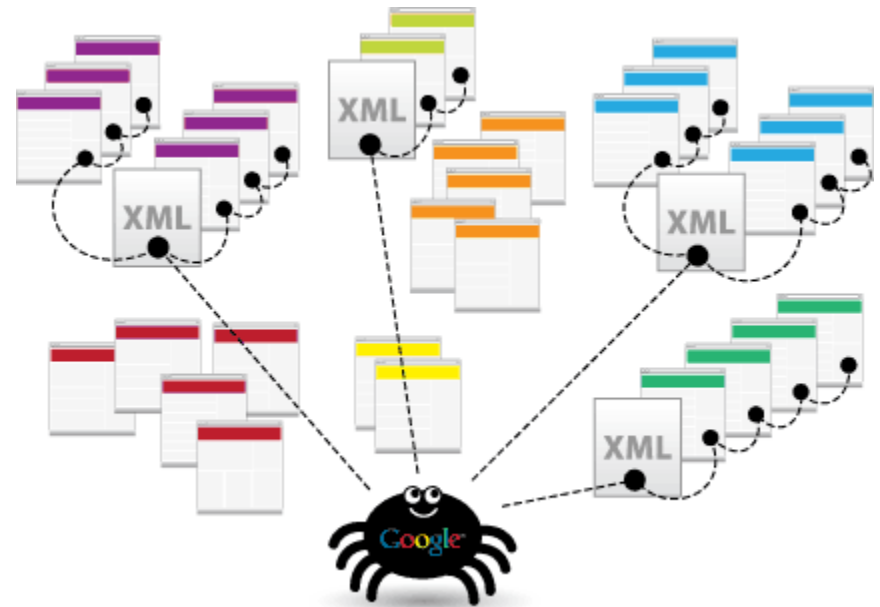
- Cuándo se ha actualizado por última vez
- Tipo de contenido incluido en las páginas

Vídeo

Imagen

¿Por qué es tan importante?

Google cuando entra en nuestro “site” para ver su contenido nuevo, lo único que hace, es **seguir enlaces** y, **cuanto más fácil le ofrezcamos el contenido nuevo, más va a aprovechar su tiempo en nuestro “site”**.



<http://www.virtualwebestudio.com/wp-content/uploads/2013/02/sitemap.png>

El camino que seguirá Google

1. Entra en nuestra página **robots.txt** para ver qué puede indexar y qué no.
(La última línea de nuestro robots.txt será un enlace al sitemap → el **bot de Google** seguirá)
2. Cuando el **bot de Google** entra en el **sitemap.xml**, seguirá la primera URL, que es el último cambio por ejemplo de nuestro blog.
3. Y después se va con la información, listo para indexarla.

GoogleBot

Robot de búsqueda usado por la empresa **Google** que colecciona documentos desde la web con el fin de construir una base de datos para el motor de búsqueda Google.

El camino que seguirá Google

Si no lo hace así, tendrá que empezar a seguir enlaces en nuestra home y las probabilidades de que se lleve contenido antiguo y deje contenido nuevo, son altas y habremos conseguido dos cosas:

Ralentizar la indexación del contenido nuevo de nuestro site.
Hacer perder el tiempo al bot de Google.

¿Necesito un sitemap?

Si las páginas del sitio web están bien enlazadas, nuestros rastreadores web pueden detectar la mayor parte del sitio.

Un **sitemap puede mejorar el rastreo del sitio**, sobre todo si este cumple uno de los siguientes criterios:

- Es muy grande.
- Tiene un gran archivo de páginas de contenido que están aisladas o no están bien enlazadas entre sí.
- Es nuevo y es destinatario de pocos enlaces externos.

Los **robots o arañas** son unos programas que utilizan los buscadores para rastrear la web.

Podemos utilizar el archivo robots.txt para:

- Evitar que ciertas páginas y directorios de nuestro sitio sean accesibles a los buscadores.
- Bloquear el acceso a archivos de código o utilidades.
- Impedir la indexación de contenido duplicado en nuestro sitio, como copias de prueba o versiones para imprimir.
- Indicar la localización de los mapas del sitio en XML.

Cosas que debemos tener en cuenta sobre robots

- El archivo es público, lo que significa que cualquier puede verlo con solo teclear:

<https://www.uneatlantico.es/robots.txt>

```
#
# robots.txt
#
# This file is to prevent the crawling and indexing of certain parts
# of your site by web crawlers and spiders run by sites like Yahoo!
# and Google. By telling these "robots" where not to go on your site,
# you save bandwidth and server resources.
#
# This file will be ignored unless it is at the root of your host:
# Used: http://example.com/robots.txt
# Ignored: http://example.com/site/robots.txt
#
# For more information about the robots.txt standard, see:
# http://www.robotstxt.org/robotstxt.html

User-agent: *
# CSS, JS, Images
Allow: /core/*.css$
Allow: /core/*.css?
Allow: /core/*.js$
Allow: /core/*.js?
Allow: /core/*.gif
Allow: /core/*.jpg
Allow: /core/*.jpeg
Allow: /core/*.png
Allow: /core/*.svg
Allow: /profiles/*.css$
Allow: /profiles/*.css?
Allow: /profiles/*.js$
Allow: /profiles/*.js?
Allow: /profiles/*.gif
Allow: /profiles/*.jpg
Allow: /profiles/*.jpeg
Allow: /profiles/*.png
Allow: /profiles/*.svg
# Directories
Disallow: /core/
Disallow: /profiles/
# Files
Disallow: /README.txt
Disallow: /web.config
# Paths (clean URLs)
Disallow: /admin/
Disallow: /comment/reply/
Disallow: /filter/tips/
Disallow: /node/add/
Disallow: /search/
Disallow: /user/register/
Disallow: /user/password/
Disallow: /user/login/
Disallow: /user/logout/
# Paths (no clean URLs)
Disallow: /index.php/admin/
Disallow: /index.php/comment/reply/
Disallow: /index.php/filter/tips/
Disallow: /index.php/node/add/
Disallow: /index.php/search/
Disallow: /index.php/user/password/
Disallow: /index.php/user/register/
Disallow: /index.php/user/login/
Disallow: /index.php/user/logout/
```


Como generar el archivo robots.txt

El archivo robots.txt no es obligatorio.
Solo necesitamos crearlo cuando queramos impedir que ciertas páginas o directorios de nuestro sitio aparezcan en los resultados de búsqueda.

Para generar el archivo, basta con crear un documento de texto con el nombre de **robots.txt** y **subirlo a la raíz de nuestro dominio** que es el lugar donde los buscadores esperan encontrarlo.

- **Bloc de notas** o Editor de código como **Notepad++**.
- Para subir el archivo usaremos un **cliente FTP** como **FileZilla** o **Cyberduck**, ...

Comandos más importantes:

- **User-agent:** indica que tipo de robot debe cumplir con las directivas que se indiquen a continuación.
- **Disallow:** Deniega el acceso a un directorio o página concreta.
- **Allow:** Funciona al contrario que la directiva Disallow, permitiendo el acceso a directorios o páginas.
- **Sitemap:** Indicar la ruta donde se encuentra un mapa del sitio en XML.
- **Crawl-delay:** Indica al robot el número de segundos que debe esperar entre cada página.

Restricciones más comunes

Comodines:

(*) secuencia cualquiera de caracteres

(\$) indica el final de una URL.

Ex: cualquier archivo que acabe con la extensión .php

/.php\$*

- **Incluir todos los robots:** *User-agent: **
- **Especificar el robot de Google:** *User-agent: Googlebot*
- **Especificar el robot de Bing:** *User-agent: Bingbot*
- **Denegar todo el sitio:** *Disallow: /*
- **Denegar un directorio:** *Disallow: /directorio/*
- **Denegar una página:** *Disallow: /pagina-web.htm*
- **Denegar la extensión .gif:** *Disallow: /*.gif\$*
- **Permitir un subdirectorio:** *Allow: /directorio/subdirectorio*
- **Señalar el mapa del sitio:** *Sitemap: http://www.example.com/sitemap.xml*

Modificar el Robots.txt en WP

Para modificar el archivo Robots.txt en WordPress: (**plugins WP Robots txt**)

1. Dentro del escritorio de WP, pulsaremos en Ajustes / Lecturas
2. Modificaremos las líneas de código por el Robots.txt que queramos y guardaremos los cambios.

Robots.txt Content

Robots.txt Content

```
User-agent: *  
Disallow: /wp-admin/  
Allow: /wp-admin/admin-ajax.php  
  
Sitemap: https://pruebasdeformacionmariantoca.000webhostapp.com/wp-sitemap.xml
```

The content of your [robots.txt](#) file. Delete the above and save to restore the default.

Guardar cambios

Google Search Console

Google Search Console es un servicio gratuito para webmasters de Google que permite a los creadores de páginas web comprobar el estado de la indexación de sus sitios en internet por el buscador, y optimizar su visibilidad.

Google Search Console

Inspeccionar las URL de "https://pruebasformacionmariantoca.000webhostapp.com/"

https://pruebasform... ▼

Sitemaps

Añadir un sitemap

https://pruebasformacionmariantoca.000webhostapp.com/ Introduce la URL del sitemap

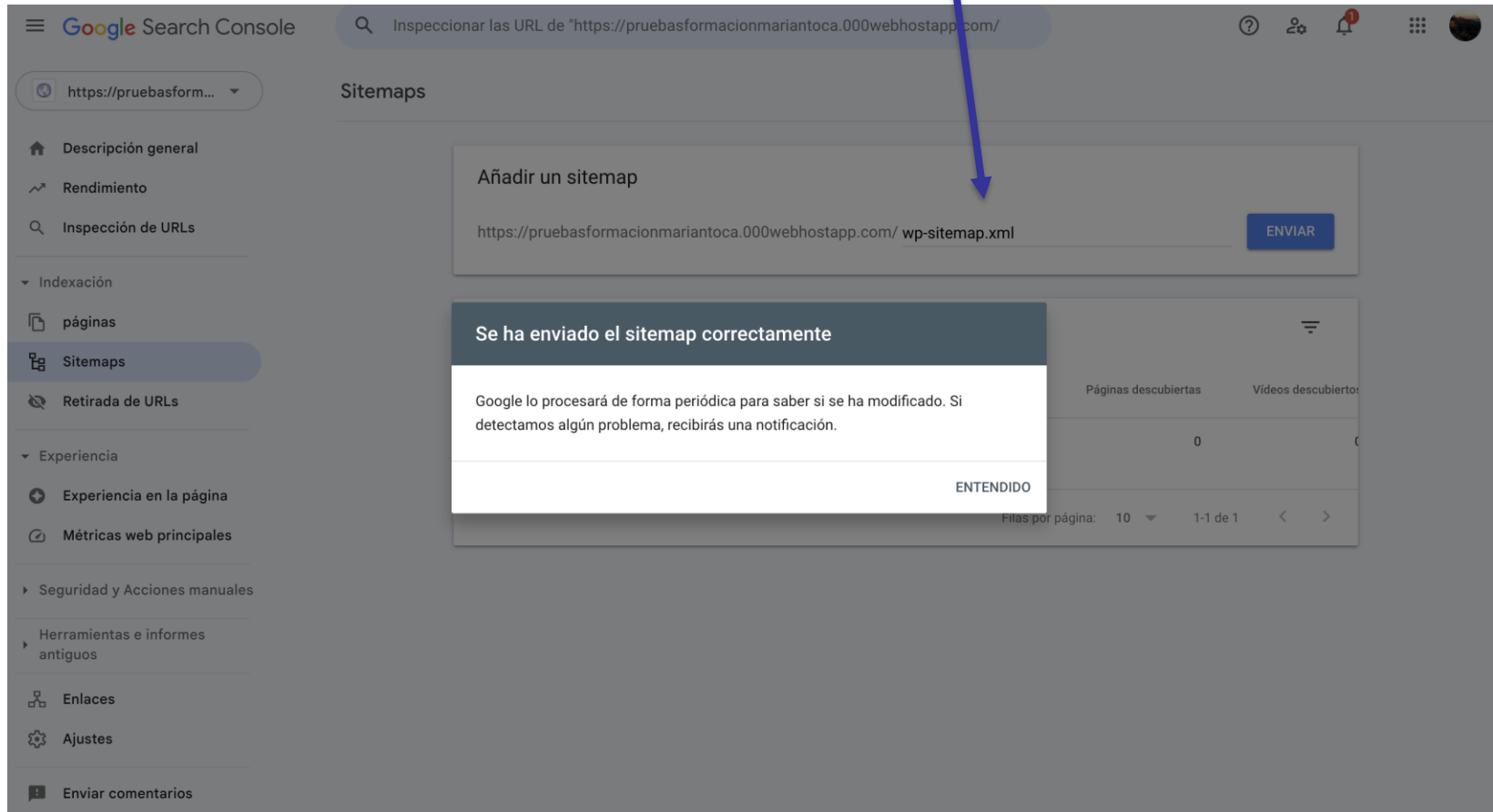
Sitemaps enviados

Sitemap	Tipo	Enviado ↓	Última lectura	Estado	Páginas descubiertas	Videos descubiertos
/wp-sitema p.xml	Desconocido	30 mar 2023		No se ha podido obtener	0	0

Filas por página: 10 ▼ 1-1 de 1 < >

Google Search Console

Escribiremos el nombre del archivo del sitemap de WP y pulsaremos en el botón “Enviar” para que lo reciba Google.



Pequeños archivos de texto almacenados automáticamente por un sitio web durante nuestra visita

Originalmente, la cookie se desarrolló como un servicio para ofrecer memoria a los sitios web, **mejorando así la experiencia del usuario** y logrando una **interacción más intuitiva** y sin incidencias entre el sitio web y el usuario.



<https://en.wikipedia.org/wiki/Cookie#/media/File:2ChocolateChipCookies.jpg>

Cookies. ¿Qué saben de nosotros?

Esta página web usa cookies

Las cookies de este sitio web se usan para personalizar el contenido y los anuncios, ofrecer funciones de redes sociales y analizar el tráfico. Además, compartimos información sobre el uso que haga del sitio web con nuestros partners de redes sociales, publicidad y análisis web, quienes pueden combinarla con otra información que les haya proporcionado o que hayan recopilado a partir del uso que haya hecho de sus servicios. Usted acepta nuestras cookies si continúa usando esta página web.

Gracias a la cookie, el sitio web de mejor manera, ofreciéndonos **sugerencias de acuerdo con nuestros intereses y necesidades.**

El sitio web es capaz de rastrear nuestras acciones exactas mientras navegamos por la red.

¿Sobre qué artículos se ha desplazado nuestro cursos?

¿Ante cuáles nos hemos detenido y nos hemos puesto a leer?

¿Qué es aquello que nos lleva a hacer clic o a compartir?

Clases de Cookies

Cookies de sesión: son temporales y vencen una vez se abandona el sitio.
(Tiendas virtuales para la cesta de la compra).

Cookies permanentes: pueden permanecer incluso terminada la sesión. Almacenan información como **datos de acceso, de contacto y números de cuentas** para que no las tengas que escribir de nuevo cada vez que hagas uso del sitio.

Cookies de primeros: sirven para facilitar memoria al sitio web en torno a nuestros datos y preferencias.

Cookies de terceros: configuradas por otro sitio web distinto al sitio en el que nos encontramos. La finalidad de estas es la de recopilar cierta información sobre nuestro **comportamiento, demografía y publicidad...**

☐ **CookieYes | Consentimiento de cookies y RGPD**

[Ajustes](#) | [Soporte](#) | [Desactivar](#)

Un modo sencillo de mostrar que tu sitio cumple con la ley europea de cookies / RGPD.

Versión 3.0.8 | [Por CookieYes](#) | [Ver detalles](#)

[Activar las actualizaciones automáticas](#)

Pruebas de Formación

Pruebas de Formación

Pruebas de formación para las clases

We value your privacy

We use cookies to enhance your browsing experience, serve personalized ads or content, and analyze our traffic. By clicking "Accept All", you consent to our use of cookies.

[Customize](#)

[Reject All](#)

[Accept All](#)

Marián Toca
-Formación Área Digital-



GRUPO CLAVE

A CORUÑA

BILBAO

LISBOA

MADRID

SANTANDER

SANTIAGO DE COMPOSTELA

C/ Manuel Guzmán, 1

C/ María Díaz de Haro ,36 1º

Rua José Estevão, 17 - 1º

C/ de Génova, 11 1ºD

C/ Joaquín Costa, 13

C/ Restollal, 37, Bajo.

15008 A Coruña

48010 Bilbao

1169 - 300 Lisboa

28004 Madrid

39005 Santander

15702 Santiago

Telf: 981 14 51 59

Telf. 944 79 09 95

Telf. 351 213304538

Telf. 915 912 950

Telf. 942 36 14 04

Telf: 981 59 49 70

Muchas gracias por su atención