

# DINT PRÁCTICA SITEMAPS

**Descripción breve**  
Investigar sobre sitemaps en la documentación dada y crear un sitemap “a mano” y autogenerar otro con PowerMapper. Enviar un sitemap a Google.

**NURIA BARIOS SANTOS**  
nuria.barsan@educa.jcyl.es  
18/10/2020

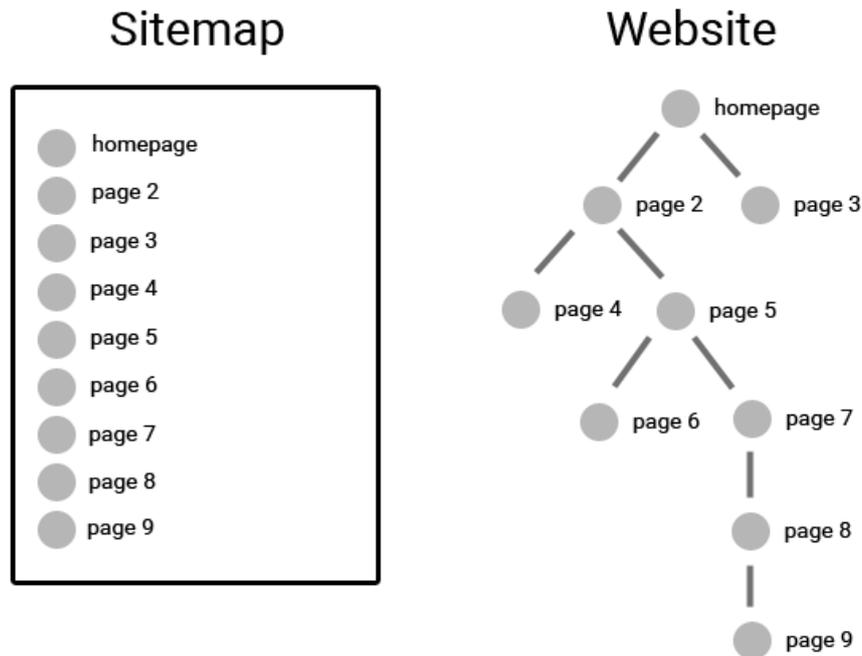


# Tabla de contenido

- ¿Qué son? ..... 2
- ¿Cómo crear un sitemap? ..... 2
  - Estructura del sitemap ..... 3
  - ¿Generadores sí o no? ..... 5
  - Validar sitemap ..... 6
  - Añadir el sitemap al sitio web ..... 7
  - Enviar sitemap a Google ..... 7
- Limitaciones de los sitemap ..... 9
- ¿Necesito un sitemap? ..... 9
- Referencias ..... 10
- Anexo – Herramientas de generación de sitemaps ..... 10
  - PowerMapper Desktop (trial) ..... 10
  - XML Sitemaps (Versión web) ..... 13

## ¿Qué son?

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, como Google, leen estos archivos para rastrear los sitios web de forma más eficaz.



**Figura 1.** Un motor de búsqueda encontrará las nueve páginas del sitemap de la imagen inmediatamente visitando un solo archivo. En cambio, en el sitio web tendría que recorrer 6 páginas enlazadas para llegar a la 9.

Los sitemaps proporcionan información sobre el contenido de las páginas, incluido el contenido multimedia, y su actualidad. De imágenes y vídeos pueden indicar, por ejemplo:

- **Imagen:** tema, tipo y licencia.
- **Vídeo:** duración, categoría y calificación por edad recomendada.

## ¿Cómo crear un sitemap?

Se pueden utilizar varios formatos para los sitemaps, entre ellos:

- XML: el más utilizado.
- RSS: se pueden usar las URL de los canales RSS como sitemap.
- Texto plano: si el sitemap incluye solo URL de páginas web, puede ser un archivo de texto simple que contenga una dirección URL por línea.

Además del formato a utilizar, hay que decidir si se usará una herramienta o se hará “manualmente”.

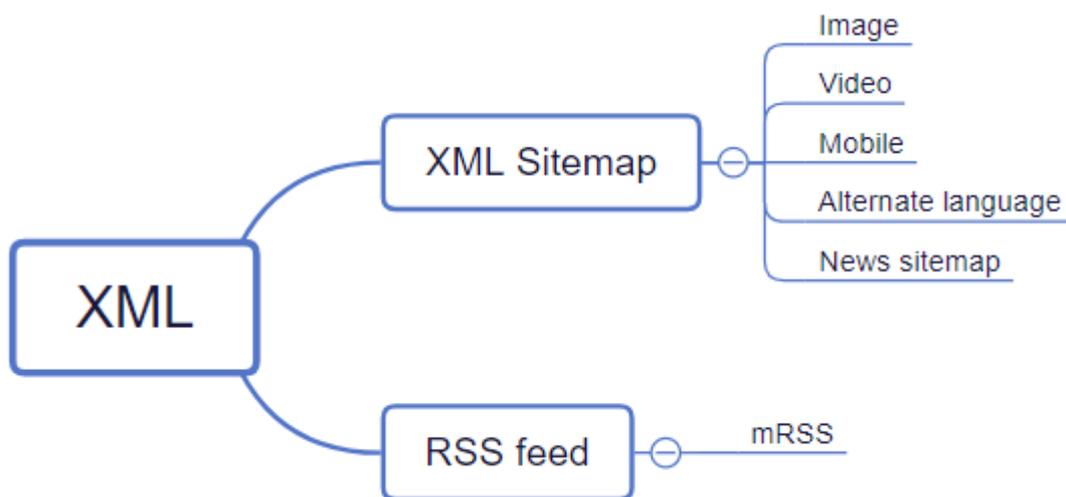
Por último, se debe decidir qué páginas incluir, habrá páginas que tengan varias versiones o que sean similares, en cuyo caso habría que elegir una como canónica— la más representativa de un conjunto de páginas duplicadas de un sitio web— o algunas que no sean significativas y se quiera excluir.

Una vez creado, hay que publicar el sitemap, subiéndolo al sitio y, opcionalmente, a [Google](#) u otro motor de búsqueda. A parte de enviarlo a Google (mediante la herramienta Sitemaps de Search Console), otras formas para permitir a los buscadores recuperar el sitemap:

- Añadiendo la ruta del sitemap (<http://example.com/sitemap.xml>) como una línea en el archivo robots.txt.
- Solicitando a Google que rastree el sitemap mediante la función "ping". Enviando una solicitud HTTP GET como esta: <http://www.google.com/ping?sitemap=https://example.com/sitemap.xml>

⚠ Google **no comprueba los sitemaps siempre** que rastrea los sitios web; **solo lo hace la primera vez** que los detecta. Luego, solo los volverá a rastrear cuando se utilice la función "ping" para avisar a Google de que se han modificado. Se debe avisar a Google solo tras subir un sitemap nuevo o modificar uno existente; no enviar varias veces un sitemap no editado, ni utilizar la función "ping" en tal caso.

## Estructura del sitemap



**Figura 2.** Jerarquía de Sitemaps

### 1. Declaración

```

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

## 2. URL

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
```

```
<url>
  <loc>http://www.example.com/page1.html</loc>
</url>
```

```
</urlset>
```

Cada URL va entre etiquetas `<url>`. La única etiqueta indispensable es *location* (`<loc>`), gracias a la cual los buscadores sabrán todo lo importante acerca de las URL.

**⚠ IMPORTANTE:** las URL de las páginas deben estar escritas con su ruta completa, incluido el protocolo. Por ejemplo, <http://www.example.com/> es una URL válida para un sitemap, pero [www.example.com](http://www.example.com) no.

## 3. Fecha de la última modificación de las URL

Es muy recomendable incluir información sobre la fecha de la última modificación de las páginas. Se puede hacer añadiendo la etiqueta `<lastmod>` dentro de `<url>`:

```
<url>
  <loc>http://www.example.com/page1.html</loc>
  <lastmod>2017-10-20T17:30:00-02:00</lastmod>
</url>
```

Ejemplo básico completo:

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

```
<url>
  <loc>http://www.example.com/page1.html</loc>
  <lastmod>2017-10-20T17:30:00-02:00</lastmod>
</url>
```

```
</urlset>
```

## 4. Imágenes y otros elementos

Con lo anterior ya se tendría un sitemap básico, se puede extender y enriquecer añadiendo otros elementos, como imágenes o vídeos.

Por ejemplo, para añadir las imágenes habría que añadir el espacio de nombres para las imágenes:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
xmlns:image="http://www.google.com/schemas/sitemap-image/1.1">
  <url>
    <loc>http://www.example.com/page1.html</loc>
  </url>
</urlset>
```

Y las imágenes dentro de <url> mediante las etiquetas <image:image> e <image:loc>:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
xmlns:image="http://www.google.com/schemas/sitemap-image/1.1">
  <url>
    <loc>http://www.example.com/page1.html</loc>
    <image:image>
      <image:loc>http://example.com/photo.jpg</image:loc>
    </image:image>
    <image:image>
      <image:loc>http://example.com/image.jpg</image:loc>
    </image:image>
  </url>
</urlset>
```

## ¿Generadores sí o no?

Aunque los generadores de sitemap, como [XML-sitemaps.com](http://XML-sitemaps.com), son muy populares y fáciles de usar no son recomendables.

Para generar un sitemap con estas herramientas basta con introducir la URL del sitio web y el generador lo rastrea, captura las URL de las páginas y crea el sitemap. Con esto se crea una instantánea del sitio, es decir, un sitemap estático que no reflejará los cambios cuando se realicen.

### Problemas:

1. Los sitemaps generados quedan **desactualizados muy rápido**: si se actualiza una página o se añaden nuevas el sitemap no cambia.
2. Las **URL irrelevantes y/o duplicadas** se añaden. La mayoría de los generadores añaden todas las páginas del sitio al sitemap (o todas las que puedan recorrer según sus limitaciones). Se podrían eliminar las URL innecesarias, pero es fácil cometer errores.

- Los generadores **solo ven URL enlazadas**, mientras que al sitemap se pueden añadir también URL que no estén enlazadas, siendo uno de los intereses que puede tener, aunque no es necesario indexar todas las URL no enlazadas.

Salvo que el sitio sea nuevo y se necesite un indexado rápido es mucho mejor crear un sitemap dinámico, que se actualizará regularmente.

## Validar sitemap

Los sitemaps pueden ser muy complejos y es fácil cometer errores. Tras crear un sitemap es recomendable validarlo a través de un validador, como [XML Sitemap Validator](#), y, después, con la herramienta *Google Search Console*.

Para usar XML Sitemap Validator solo hay que añadir la URL del sitemap.

### XML Sitemap Validator

This tool will check if your xml sitemap is formatted correctly and optionally ping Google to inform them of your sitemap location.

Our xml sitemap validator will inform you straight away if there are any problems with your sitemaps allowing you to make changes prior to submitting it to Google.

Please note that we would recommend you inform Google of your sitemaps via the [Google Search Console](#).

Your Sitemap URL

Please enter the full http address for your sitemap, an example:

<http://www.domain.com/sitemap.xml>

<https://nuriadaw.github.io/dint/sitemap.xml>

VALIDATE SITEMAP

Figura 3. Validando sitemap

### XML Validation Results

✓ No issues detected	
Sitemap URL	<a href="https://nuriadaw.github.io/dint/sitemap.xml">https://nuriadaw.github.io/dint/sitemap.xml</a>
Character Set	UTF-8
Sitemap is valid	Yes
Warnings	-
Errors	-

Figura 4. Resultados de la validación

## Añadir el sitemap al sitio web

Es recomendable situar el sitemap en la carpeta raíz del servidor HTML. Siguiendo con el ejemplo: <http://example.com/sitemap.xml>.

## Enviar sitemap a Google

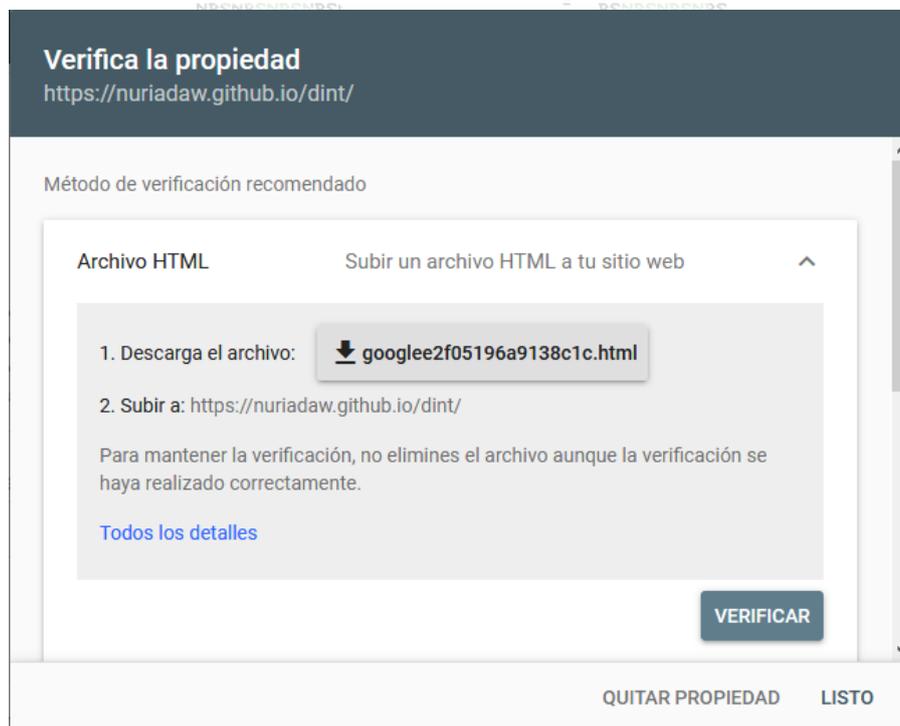
Vía [Google Search Console](#).

Lo primero que hay que hacer, una vez accedido con una cuenta de Google, es indicar a Google el dominio o la URL del sitio:



**Figura 5.** URL del sitio que buscará Google

Si es la primera vez Google requerirá la verificación de la pertenencia del sitio:



**Figura 6.** Verificación de la propiedad del sitio

## ✓ Propiedad verificada

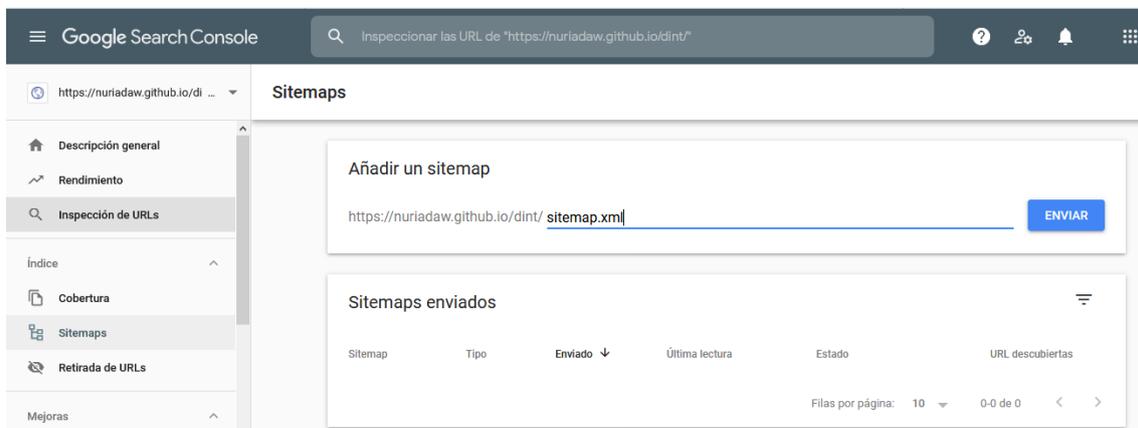
Método de verificación:  
Archivo HTML

Para mantener la verificación, no elimines el archivo. Para no perder la verificación, te recomendamos que añadas varios métodos de verificación en **Ajustes > Verificación de la propiedad**.

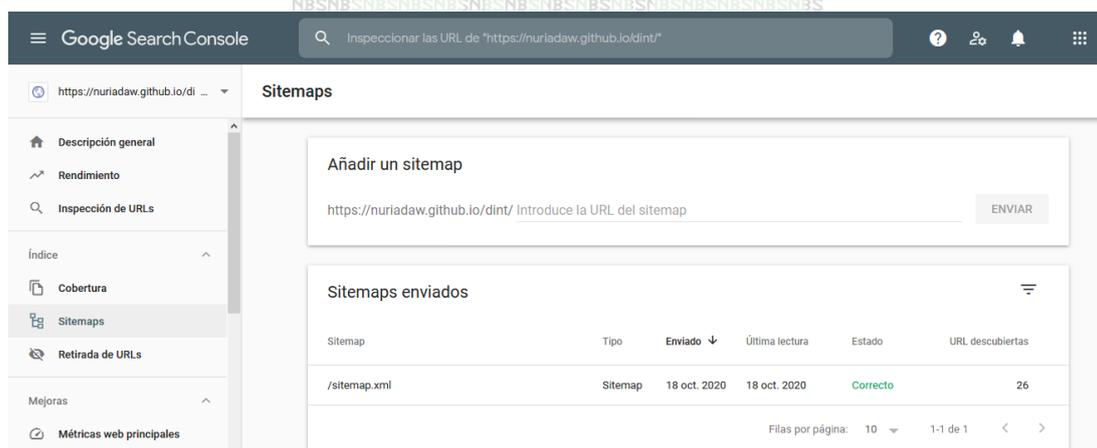
LISTO [IR A LA PROPIEDAD](#)

**Figura 7.** Google ha detectado su archivo de verificación

Una vez verificado se accede al panel de Google Search Console. Desde el apartado “Sitemaps” se puede enviar sitemaps a Google:



**Figura 8.** Enviando un sitemap a Google



**Figura 9.** Sitemap añadido a la lista de enviados

Se ha enviado el sitemap correctamente

Google lo procesará de forma periódica para saber si se ha modificado. Si detectamos algún problema, recibirás una notificación.

ENTENDIDO

Figura 10. Confirmación de envío del sitemap

## Limitaciones de los sitemaps

Los sitemaps no pueden exceder los 50MB ni contener más de 50,000 URLs. Si el sitio web los excediese habría que crear varios archivos sitemap y usar un index de sitemaps, que a su vez no podrá exceder los 50MB ni contener más de 50,000 URLs. Los sitemaps se pueden comprimir en gzip.

## ¿Necesito un sitemap?

Si las páginas de un sitio web están bien enlazadas los buscadores son capaces de detectarlas. Aun así, un sitemap puede mejorar el rastreo de los sitios web grandes o muy complejos, o de los archivos muy especializados.

Casos en los que sería recomendable tener un sitemap:

- Sitio **muy grande**.
- Sitio con **páginas aisladas** o **mal enlazadas**.
- Sitio **muy nuevo**, pocos enlaces externos dirigen hacia él.
- Sitio con **mucho contenido multimedia**, en los sitemaps se le puede incluir información adicional.
- El **contenido** del sitio **cambia con frecuencia**, un sitemap ayuda a acelerar la indexación de nuevas páginas una vez añadidas al sitemap.

Casos en que podría ser prescindible un sitemap:

- Sitio **muy pequeño** (500 o menos páginas).
- **Servicio de alojamiento** de sitios web **sencillo**, como Blogger o Wix (suelen crearlos de forma automática).
- Sitio **enlazado** totalmente **de forma interna**.
- No ver necesario que los buscadores de imágenes, vídeos o noticias recuperen el contenido multimedia.

## Referencias

*Ayuda de Search Console. Gestionar tu mapa del sitio.* (2020). Obtenido de Google:  
<https://support.google.com/webmasters/answer/156184?hl=es>

*Frequently asked questions* . (21 de noviembre de 2016 ). Obtenido de sitemaps.org:  
<https://www.sitemaps.org/faq.html>

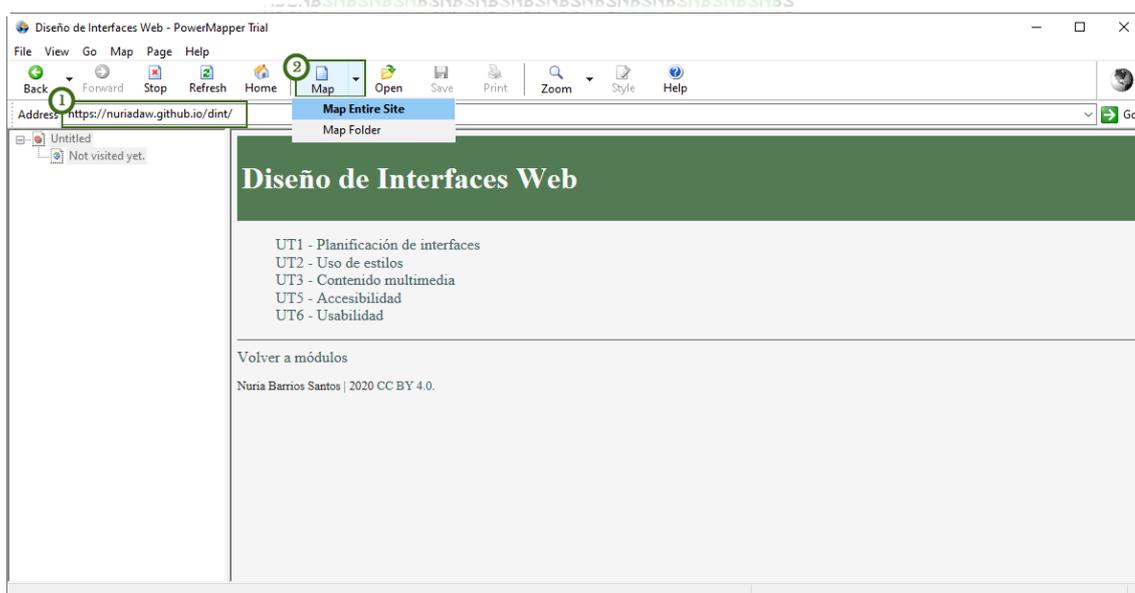
Januška, J. (12 de marzo de 2018). *SEO sitemap best practices in 2020 + cheat sheet.*  
 Obtenido de Spotibo: <https://blog.spotibo.com/sitemap-guide/>

## Anexo – Herramientas de generación de sitemaps

### PowerMapper Desktop (trial)

**Producto:** <https://www.powermapper.com>

- Disponible para Windows y Mac (prueba de 30 días)
- Mapea más de 10,000 página (versión de prueba limitada a 1,000)
- Opciones de rastreo personalizables
- No ofrece la posibilidad de incluir imágenes (al menos la versión de prueba)



**Figura 11.** Pantalla principal de PowerMapper

Por defecto incluye todas las páginas del sitio y los pdf. Se pueden establecer filtros desde el menú “View/Options”:

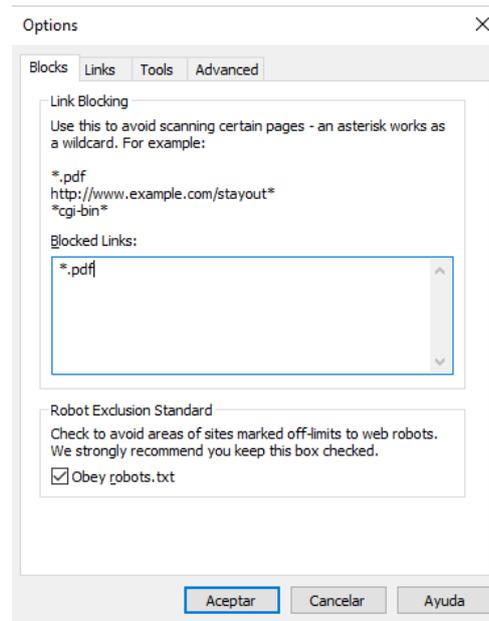


Figura 12. Opciones de PowerMapper

Con esta configuración genera un sitemap sin los pdf.

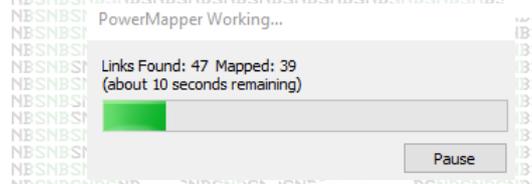


Figura 13. PowerMapper en acción

Al terminar lanza un aviso con los archivos que no se han añadido por estar bloqueados:

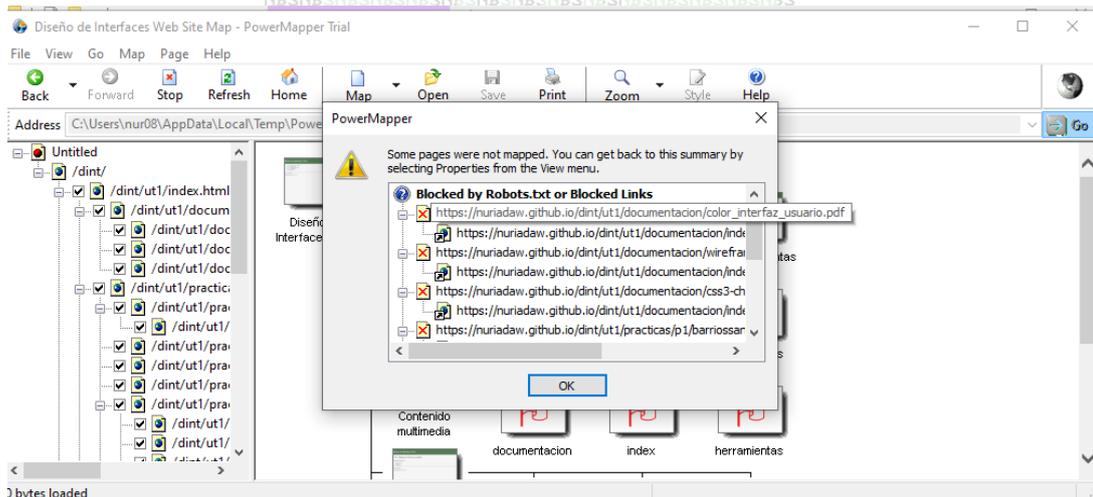


Figura 14. Aviso de PowerMapper

**Resultado:**

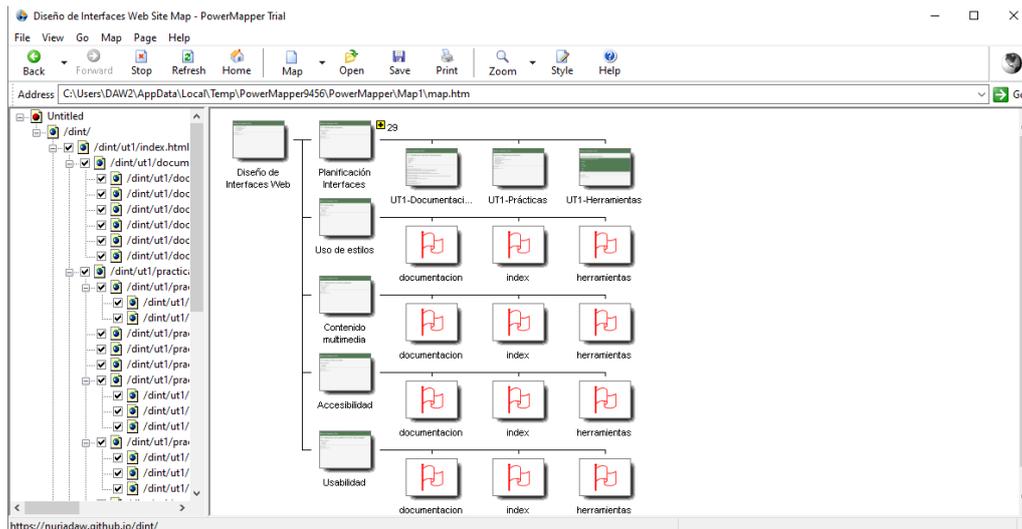


Figura 15. Mapa del sitio

### Exportar a XML:

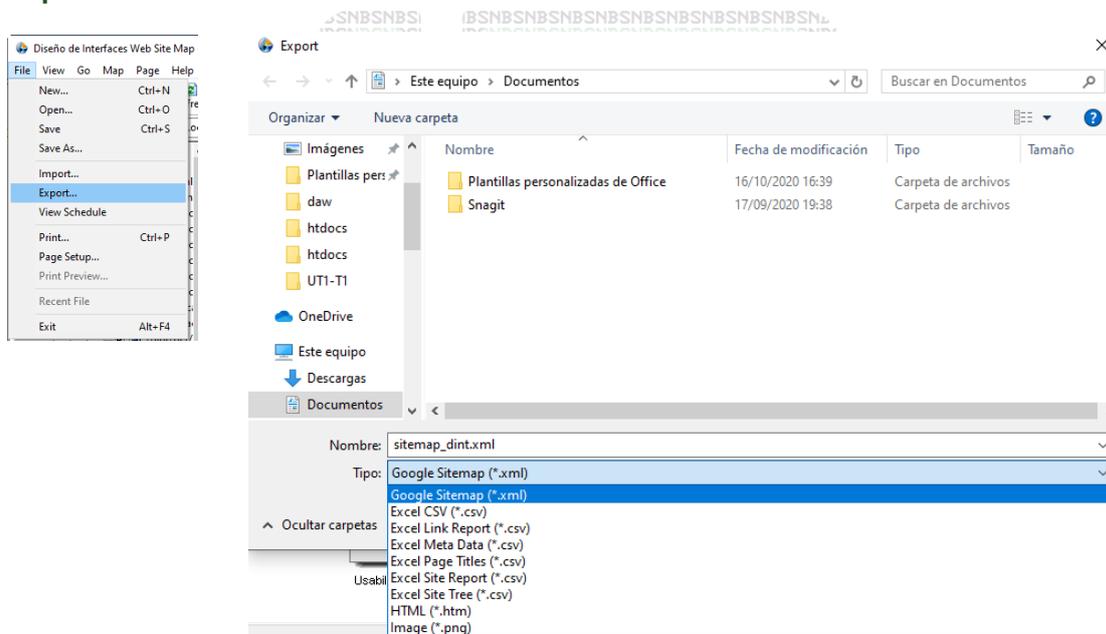
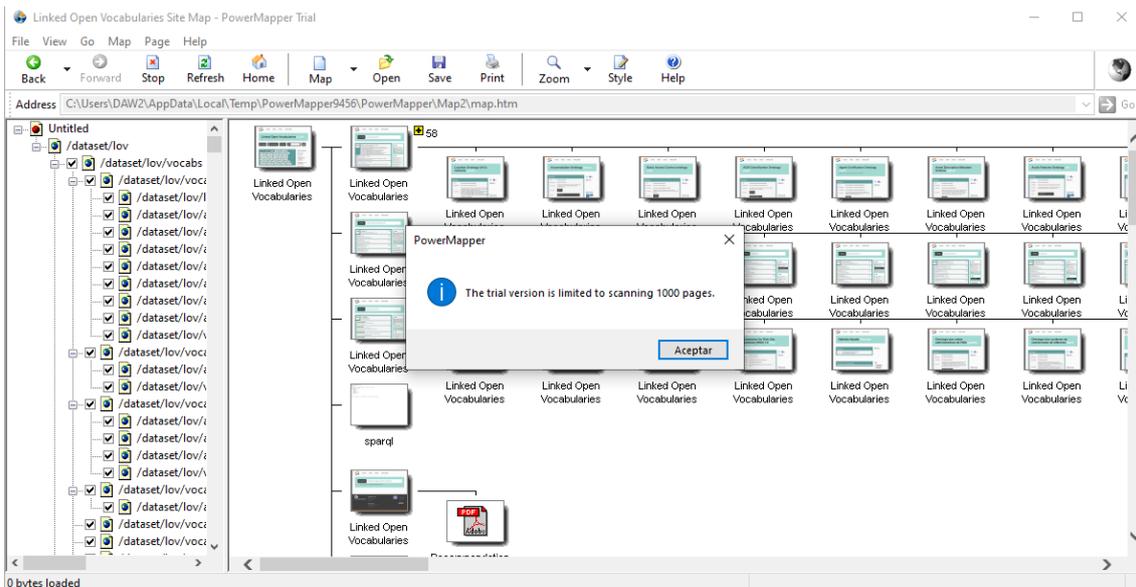


Figura 16. Exportación a XML

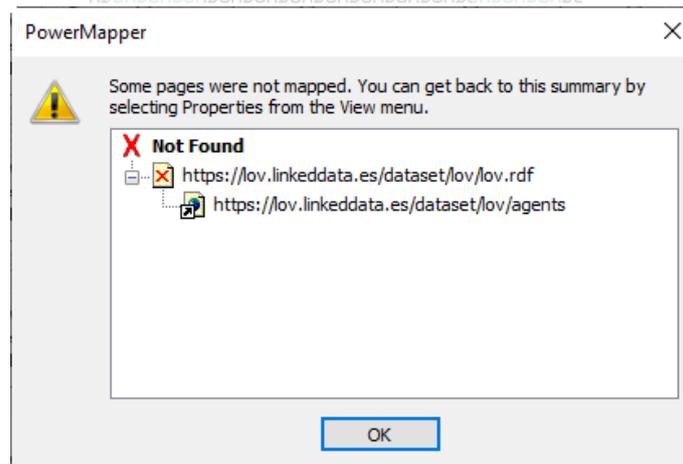
Además del sitio web del curso, se ha realizado otra prueba con sitio con mayor profundidad y altamente hiperenlazado, [Linked Open Vocabularies](#).

Tras recorrer 1,000 páginas la aplicación lanza un mensaje indicando que la versión de prueba está limitada a 1,000 páginas:



**Figura 17.** Límite de 1,000 páginas

Si hubiese algún otro problema lanza un aviso. En este caso no ha podido encontrar una de las páginas (devuelve un error 404):



**Figura 18.** Aviso de PowerMapper

Manual de usuario: <https://www.powermapper.com/download/mapper/powermapper-user-manual.pdf>

## XML Sitemaps (Versión web)

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

- La versión web es adecuada para crear rápidamente un sitemap XML de un sitio web pequeño (hasta 500 páginas)
- No requiere registro y es muy rápido
- Se puede descargar el sitemap o recibirlos vía email
- No ofrece opciones de rastreo personalizables

- No ofrece la posibilidad de incluir imágenes (al menos la versión web)

Para utilizarlo basta introducir la URL del sitio web para generar el sitemap:

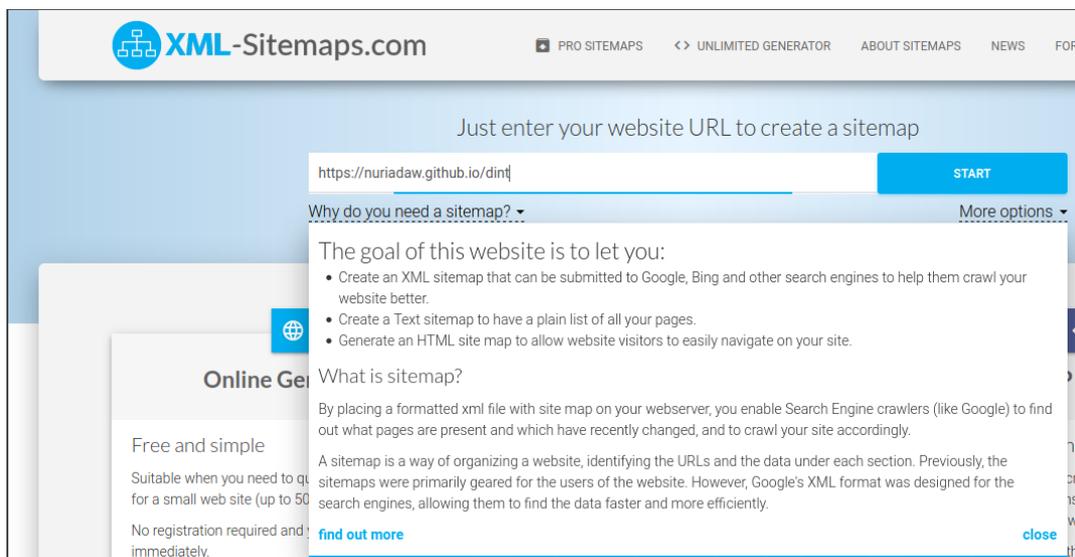


Figura 19. Generación del sitemap

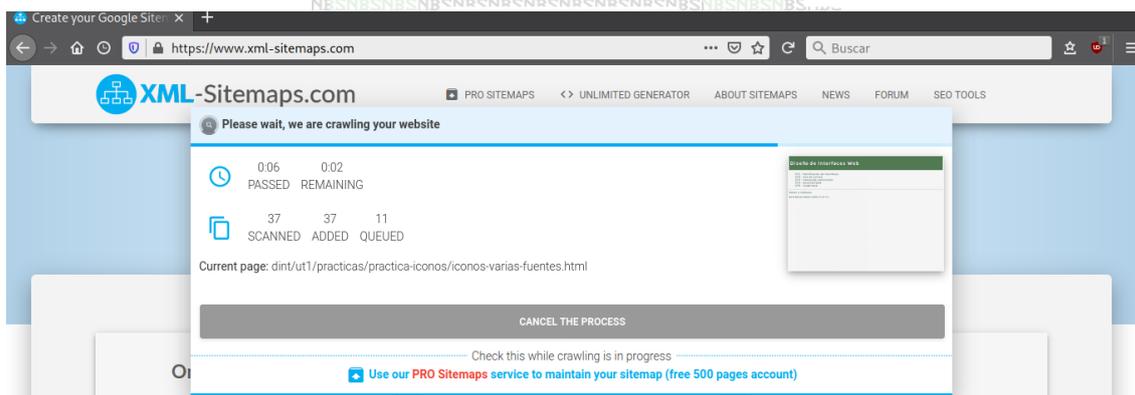


Figura 20. XML Sitemaps en acción

**Resultado:**

The screenshot shows the XML-Sitemaps.com interface for a specific sitemap. At the top, the navigation bar includes 'XML-Sitemaps.com', 'PRO SITEMAPS', 'UNLIMITED GENERATOR', 'ABOUT SITEMAPS', 'NEWS', 'FORUM', and 'SEO TOOLS'. The main heading is 'Sitemap Details' for the URL 'https://nuriadaw.github.io/dint'. Below this, there are several sections:

- What's next?:** A section with a blue button labeled 'DESCARGA' (Download) and the text 'Download your XML Sitemap file'. Below it, instructions state: 'Upload it into the domain root folder of your website. Then open Google Search console and add your sitemap URL.'
- Sitemap Details:** A box containing metadata: 'Created on: 16 October 2020, 20:08', 'Processing time: 0:02:00s', 'Pages indexed: 44 (crawled: 48)', 'Pages size: 0.10Mb', and 'Broken links: None'.
- Sitemap Preview:** A table with columns 'URL', 'Last Mod', and 'Priority'. It lists several URLs from 'https://nuriadaw.github.io/dint/' to 'https://nuriadaw.github.io/dint/ut1/herramientas.html' with their respective last modification dates and priorities.
- Looking for an easy way to keep your sitemap up-to-date?:** A section with a red button 'CREATE FREE PRO-SITEMAPS ACCOUNT' and text encouraging users to create a free account for 'nuriadaw.github.io'.
- Other Downloads:** A section with a button 'DOWNLOAD ALL SITEMAPS IN A ZIP FILE' and an email subscription form with the text 'OR RECEIVE ALL GENERATED SITEMAP FILES VIA EMAIL' and 'Your email address'.
- View Sitemap:** A section with buttons for 'VIEW XML SITEMAP' and 'VIEW HTML SITEMAP'.
- Feedback:** A section with a 'SEND IT VIA EMAIL' button and text asking for testimonials and feedback.

At the bottom of the page, there are links for 'PRIVACY POLICY', 'CONTACT', and 'PRO SITEMAPS', along with a footer: 'Our sitemap XML, HTML, Images, Video, News © 2005-2020 XML-Sitemaps'.

Figura 21. Detalles y descarga

Si hay algún error lo muestra también:

**XML-Sitemaps.com** PRO SITEMAPS > UNLIMITED GENERATOR ABOUT SITEMAPS NEWS FORUM SEO TOOLS

### Sitemap Details

<https://lov.linkeddata.es/dataset/lov>

**What's next?**

↓

DOWNLOAD YOUR XML SITEMAP FILE

↓

Upload it into the domain root folder of your website.

↓

Then open [Google Search console](#) and add your sitemap URL.

**Sitemap Details**

Created on: 16 October 2020, 20:16  
 Processing EL: 0:04:21s  
 Pages indexed: 500 (crawled: 502)  
 Pages size: 11.89Mb  
 Broken links: 1

**Maximum 500 pages Limit Exceeded**

Don't let your website get left behind - improve your site's visibility in Google and other search engines by creating and submitting **full XML Sitemaps**.

Try our **Standalone XML Sitemap Generator** or create a **Pro Sitemaps account** for your website (it includes also Images, Video, News sitemap and many other features!)

**Sitemap Preview**

URL	Last Mod	Priority
<a href="https://lov.linkeddata.es/dataset/lov">https://lov.linkeddata.es/dataset/lov</a>	2020-10-16T23:16:12+03:00	1.00
<a href="https://lov.linkeddata.es/dataset/lov/vocabs">https://lov.linkeddata.es/dataset/lov/vocabs</a>	2020-10-16T23:16:12+03:00	0.80
<a href="https://lov.linkeddata.es/dataset/lov/terms">https://lov.linkeddata.es/dataset/lov/terms</a>	2020-10-16T23:16:12+03:00	0.80
<a href="https://lov.linkeddata.es/dataset/lov/agents">https://lov.linkeddata.es/dataset/lov/agents</a>	2020-10-16T23:16:12+03:00	0.80
<a href="https://lov.linkeddata.es/dataset/lov/suggest">https://lov.linkeddata.es/dataset/lov/suggest</a>	2020-10-16T23:16:12+03:00	0.80
<a href="https://lov.linkeddata.es/dataset/lov/about">https://lov.linkeddata.es/dataset/lov/about</a>	2020-10-16T23:16:12+03:00	0.80
<a href="https://lov.linkeddata.es/dataset/lov/qg">https://lov.linkeddata.es/dataset/lov/qg</a>	2020-10-16T23:16:12+03:00	0.80
<a href="https://lov.linkeddata.es/dataset/lov/vocabs?tag=Metadata">https://lov.linkeddata.es/dataset/lov/vocabs?tag=Metadata</a>	2020-10-16T23:16:12+03:00	0.80
<a href="https://lov.linkeddata.es/dataset/lov/vocabs?tag=Methods">https://lov.linkeddata.es/dataset/lov/vocabs?tag=Methods</a>	2020-10-16T23:16:12+03:00	0.80
<a href="https://lov.linkeddata.es/dataset/lov/vocabs?tag=Geography">https://lov.linkeddata.es/dataset/lov/vocabs?tag=Geography</a>	2020-10-16T23:16:12+03:00	0.80
<a href="https://lov.linkeddata.es/dataset/lov/vocabs?tag=Thema-Concepts">https://lov.linkeddata.es/dataset/lov/vocabs?tag=Thema-Concepts</a>	2020-10-16T23:16:12+03:00	0.80

**Broken Links Detected**

While creating your sitemap, we've found **1 broken link(s)** at your site. Please read our article that [explains the broken links problem](#) in details.

You can find out **exact URLs of broken links** using our **PRO Sitemaps service**. Note, that you will not only see the dead links, but also the pages that refer to them (i.e., contain the bad links). You can easily locate the problem and fix it using this feature.

That is one of many features available for a small \$4.99/mo fee. You will get automatically updated sitemap of up to 1,000 pages, including images, videos, news sitemaps, broken links report and many other features!

[CREATE FREE PRO-SITEMAPS ACCOUNT](#)

**Other Downloads**

↓

DOWNLOAD ALL SITEMAPS IN A ZIP FILE

OR RECEIVE ALL GENERATED SITEMAP FILES VIA EMAIL

Your email address:

[SEND SITEMAPS](#)

**View Sitemap**

[VIEW XML SITEMAP](#) [VIEW HTML SITEMAP](#)

**Feedback**

We'll be happy to receive your **testimonial** for the sitemap service we provide.

[SEND IT VIA EMAIL](#)

If you found our service helpful please consider **recommending** our site to your friends or place a link on your site to [XML-Sitemaps.com](#).

**Your feedback is very important** for us to build the best service for creating of XML sitemaps. If you are not satisfied with generation results and/or have comments, do not hesitate to [contact us](#).

Figura 22. Detalles y descarga

Linked Open Vocabularies - Mozilla Firefox

Linked Open Vocabularies: x

https://lov.linkeddata.es/dataset/lov/lov.rdf

VOCABS TERMS AGENTS SPARQL/DUMP

# 404 - This page does not exist - 404

Not found

Linked Open Vocabularies

DOCUMENTATION  
 About  
 API documentation  
 Source code  
 Contact

PUBLICATION  
 Semantic Web Journal '16  
 ERCIM News '14  
 Library Hi Tech '13

Hosted by the Ontology Engineering Group - UPM

Figura 23. Página no encontrada por los rastreadores