



BarcelonaTravelHacks.com

**Programmatically &
periodically Re-Build
sitemap.xml**

Problem



Maintaining a sitemap.xml file manually that contains 11,500 image URLs, 350 page URLs and about 100 pdf document URLs is too labour intensive and this that should be done automatically so that my website has a fresh updated sitemap.xml file once daily with the latest URLs and updated fields for each URL.

Google Research



Google Sitemap.xml specifications: <https://developers.google.com/search/docs/crawling-indexing/sitemaps/build-sitemap#xml>

Sitemap size limits: All formats limit a single sitemap to **50MB** (uncompressed) or **50,000 URLs**. If you have a larger file or more URLs, you must break your sitemap into multiple sitemaps.

Sitemap file encoding and location: The sitemap file must be **UTF-8 encoded**. You can host your sitemaps anywhere on your site, but unless you submit your sitemap through [Search Console](#), a sitemap affects only descendants of the parent directory. Therefore, **a sitemap posted at the site root** can affect all files on the site, which is where we recommend posting your sitemaps.

Referenced URLs' properties: Use fully-qualified, **absolute URLs** in your sitemaps. Google will attempt to crawl your URLs exactly as listed. For example, if your site is at `https://www.example.com/`, don't specify a URL such as `/mypage.html` (a relative URL), use the complete, absolute URL:
<https://www.example.com/mypage.html>.

Additional notes about XML sitemaps: As with all XML files, all tag values must be entity escaped. Google **ignores** `<priority>` and `<changefreq>` values.. Google **uses the** `<lastmod>` **value if it's consistently and verifiably (for example by comparing to the last modification of the page) accurate**

Insert the following line anywhere in your robots.txt file, specifying the path to your sitemap. We will find it the next time we crawl your robots.txt file: Sitemap: `https://example.com/my_sitemap.xml`

Sitemap format page URL

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
  xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>https://www.example.com/foo.html</loc>
    <lastmod>2022-06-04</lastmod>
  </url>
</urlset>
```

Sitemap format img URL



```
<?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>https://example.com/sample1.html</loc>
    <image:image>
      <image:loc>https://example.com/image.jpg</image:loc>
    </image:image>
    <image:image>
      <image:loc>https://example.com/photo.jpg</image:loc>
    </image:image>
  </url>
  <url>
    <loc>https://example.com/sample2.html</loc>
    <image:image>
      <image:loc>https://example.com/picture.jpg</image:loc>
    </image:image>
  </url>
</urlset>
```

<https://developers.google.com/search/docs/crawling-indexing/sitemaps/image-sitemaps>

Sitemap format pdf URL



```
<?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>https://www.example.si/good</loc>

    <image:image><image:loc>https://poster.domain.si/good.jpg</image:loc></image:image>
  </url>
  <url><loc>https://poster.example.si/good.PDF</loc></url>
</urlset>
```

<https://stackoverflow.com/questions/77625855/sitemap-xml-and-pdf>

Bing Webmaster Tools



I can submit a sitemap.xml to Bing Webmaster Tools using the same format as a Google sitemap:

<https://www.bing.com/webmasters/help/Sitemaps-3b5cf6ed>

Sitemap.xml Schema documentation

Sitemap.xml is a standard format that is documented here:

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

`<lastmod>2004-10-01T18:23:17+00:00</lastmod>`

Identifies the time that the corresponding Sitemap file was modified. It does not correspond to the time that any of the pages listed in that Sitemap were changed. The value for the lastmod tag should be in [W3C Datetime](#) format.

Desired Sitemap Layout



```
<urlset xmlns:xsi="http://www.w3.org/2001/XMLSchema-  
instance" xmlns:xhtml="http://www.w3.org/1999/xhtml" xmlns="http://www.sitemaps.org/schemas/sitemap/0.9" xmlns:image="https://www.google.com  
/schemas/sitemap-image/1.1" xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9  
http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd">
```

Header part

```
<url>  
  <loc>https://www.Barcelonatravelhacks.com</loc>  
  <lastmod>2025-05-06T22:37:06+02:00</lastmod>  
  <changefreq>monthly</changefreq>  
  <priority>1.0000</priority>  
</url>
```

Homepage

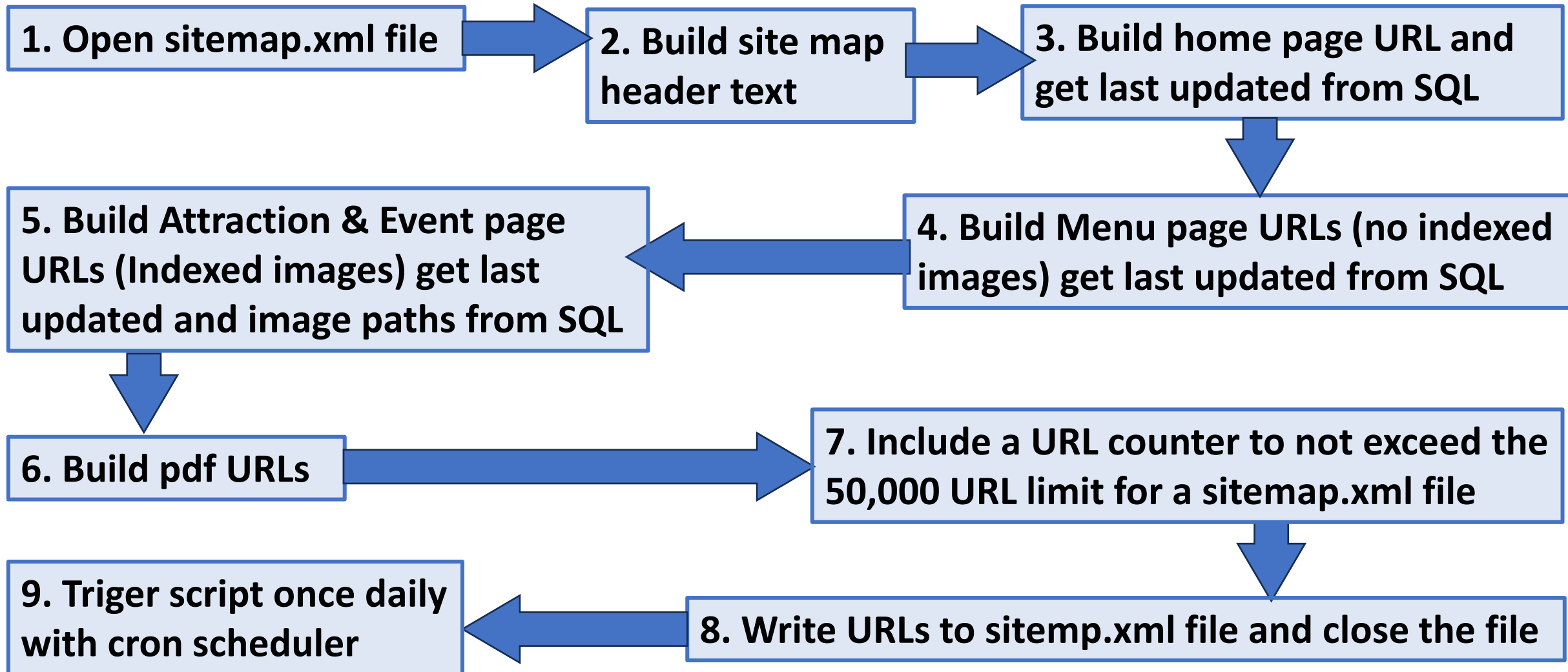
```
<url>  
  <loc>https://www.barcelonatravelhacks.com/barcelona/tibidabo-mountain</loc>  
  <lastmod>2024-06-28T00:00:00+02:00</lastmod>  
  <changefreq>monthly</changefreq>  
  <priority>0.8000</priority>  
  <image:image>  
    <image:loc>https://www.barcelonatravelhacks.com/assets/pub/webp-gal-imgs/a0001_ScnFull/a0001-07-la-sagrada-famila-construction-  
2011.webp</image:loc>  
    <image:caption>La Sagrada Famila Construction 2011</image:caption>  
    <image:geo_location>Barcelona, Catalunya, Spain</image:geo_location>  
    <image:title>La sagrada Familia</image:title>  
    <image:license>https://www.barcelonatravelhacks.com/info/i0002/terms-and-conditions-barcelonaTravelHacks</image:license>  
  </image:image>
```

Page URL with image URLs

```
<url>  
  <loc>https://www.barcelonatravelhacks.com/assets/pub/pdf-docs/t0004-001-train-map-sitges.pdf</loc>  
  <lastmod>2025-05-06T22:37:06+02:00</lastmod>  
  <changefreq>yearly</changefreq>  
  <priority>0.8000</priority>  
</url>  
</urlset>
```

PDF URL

High Level Script Design



The Sitemap.xml script



```
// 0.Database connection
try {
    $con = new PDO("mysql:dbname=database_name; host=localhost", "root", "");

    $con->setAttribute(PDO::ATTR_ERRMODE, PDO::ERRMODE_WARNING);
} catch (PDOException $e) {
    echo "connection failed: " . $e->getMessage();
}

// 1. Open Sitemap.xml file
$filename = fopen("../../sitemap.xml", "w") or die("Unable to open file!");

// 2. Get last updated for homepage
$sql = "SELECT pageUpdated FROM allpages_metadata WHERE pageId = 'index'";
$stmt = $con->prepare($sql);
$stmt->execute();
$IndexPageArr = $stmt->fetchAll(PDO::FETCH_ASSOC);
$indexUpdated = $IndexPageArr['pageUpdated'];
$datetime = new DateTime($indexUpdated);
$indexLastMod = $datetime->format('Y-m-d\TH:i:sP');

// 3. initialise an image URL counter
$TotalImages = 0;

/*//////////////////////////////////// HOMEPAGE //////////////////////////////////////*/
```

The Sitemap.xml script



```
/*////////////////////////////////////// HOMEPAGE ////////////////////////////////////////*/

// 4. header
$txt = '<?xml version="1.0" encoding="UTF-8"?>
<urlset
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xhtml="http://www.w3.org/1999/xhtml"
xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd"
xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
xmlns:image="https://www.google.com/schemas/sitemap-image/1.1">

<!-- ***** -->
<!-- ***** HOMEPAGE ***** -->
<!-- ***** -->

<url>
<loc>https://www.Barcelonatravelhacks.com</loc>
<lastmod>' . $indexLastMod . '</lastmod>
<changefreq>monthly</changefreq>
<priority>1.0000</priority>
</url>
```

The Sitemap.xml script



```
<!-- ***** -->
<!-- ***** MENU_PAGES ***** -->
<!-- ***** -->

';

/*//////////////////////////////////// LOCATION, THEME, DISTANCE & BUDGET //////////////////////////////////*/

$sql = "SELECT
    uhn.urlparam0,
    uhn.urlparam1,
    uhn.urlparam2,
    md.pageUpdated
    FROM allpages_pageurl2 uhn
    JOIN allpages_metadata md ON md.pageId = uhn.pageId
    WHERE md.noIndexTag = 0
    AND md.pageType IN('v','t','b','w','u','x','y')";
$stmt = $con->prepare($sql);
$stmt->execute();
$menuPageArr = $stmt->fetchAll(PDO::FETCH_ASSOC);
$menuUrlCount = count($menuPageArr);
```

The Sitemap.xml script



```
foreach ($menuPageArr as $pageId) {
    $urlparam0 = $pageId['urlparam0'];
    $urlparam1 = $pageId['urlparam1'];
    $urlparam2 = $pageId['urlparam2'];
    $url = rtrim("https://www.barcelonatravelhacks.com/$urlparam0/$urlparam1/$urlparam2", "/");
    $updated = $pageId['pageUpdated'];
    $datetime = new DateTime($updated);
    $lastMod = $datetime->format('Y-m-d\TH:i:sP');

    $txt .= '<url>'
';
    $txt .= '    <loc>' . $url . '</loc>'
';
    $txt .= '    <lastmod>' . $lastMod . '</lastmod>'
';
    $txt .= '    <changefreq>monthly</changefreq>'
';
    $txt .= '    <priority>0.8000</priority>'
';
    $txt .= '</url>'

';
} // End of foreach loop
```

The Sitemap.xml script



```
/*////////////////////////////////////// INFORMATION ////////////////////////////////////////*/
```

```
$sql = "SELECT
    uhn.urlparam0,
    uhn.urlparam1,
    uhn.urlparam2,
    md.pageUpdated
FROM allpages_pageurl2 uhn
JOIN allpages_metadata md ON md.pageId = uhn.pageId
WHERE md.noIndexTag = 0
AND md.pageType = 'i'";

$stmt = $con->prepare($sql);
$stmt->execute();
$infoPageArr = $stmt->fetchAll(PDO::FETCH_ASSOC);
$infoUrlCount = count($infoPageArr);

foreach ($infoPageArr as $pageId) {
    $urlparam0 = $pageId['urlparam0'];
    $urlparam1 = $pageId['urlparam1'];
    $urlparam2 = $pageId['urlparam2'];
    $url = rtrim("https://www.barcelonatravelhacks.com/$urlparam0/$urlparam1/$urlparam2", "/");
    $updated = $pageId['pageUpdated'];
    $datetime = new DateTime($updated);
    $lastMod = $datetime->format('Y-m-d\TH:i:sP');
```

The Sitemap.xml script



```
$txt .= '<url>
';
$txt .= '    <loc>' . $url . '</loc>
';
$txt .= '    <lastmod>' . $lastMod . '</lastmod>
';
$txt .= '    <changefreq>monthly</changefreq>
';
$txt .= '    <priority>0.8000</priority>
';
$txt .= '</url>

';
} // End of foreach loop

/*////////////////////////////////////// ATTRACTION //////////////////////////////////////*/

$txt .= '<!-- ***** -->
<!-- ***** ATTRACTION & EVENT PAGES ***** -->
<!-- ***** -->

';
```

The Sitemap.xml script



```
$totalPageImages = 0;

$sql = "SELECT
    uhn.pageId,
    uhn.urlparam0,
    uhn.urlparam1,
    uhn.urlparam2,
    md.pageUpdated
FROM allpages_pageurl2 uhn
JOIN allpages_metadata md ON md.pageId = uhn.pageId
WHERE md.noIndexTag = 0
AND md.pageType = 'a'
OR md.pageType = 'e'";

$stmt = $con->prepare($sql);
$stmt->execute();
$attEvtPageArr = $stmt->fetchAll(PDO::FETCH_ASSOC);
$attEvtUrlCount = count($attEvtPageArr);
```


The Sitemap.xml script



```
foreach ($attEvtPageArr as $pageArr) {
    $pageId = $pageArr['pageId'];
    $urlparam0 = $pageArr['urlparam0'];
    $urlparam1 = $pageArr['urlparam1'];
    $urlparam2 = $pageArr['urlparam2'];
    $url = rtrim("https://www.barcelonatravelhacks.com/$urlparam0/$urlparam1/$urlparam2", "/");
    $updated = $pageArr['pageUpdated'];
    $datetime = new DateTime($updated);
    $lastMod = $datetime->format('Y-m-d\TH:i:sP');

    $txt .= '<url>
';
    $txt .= '    <loc>' . $url . '</loc>
';
    $txt .= '    <lastmod>' . $lastMod . '</lastmod>
';
    $txt .= '    <changefreq>monthly</changefreq>
';
    $txt .= '    <priority>0.8000</priority>
';
}
```

The Sitemap.xml script



```
$sql = "SELECT
    ai.imgWebpScnFull,
    ai.imgCity,
    ai.imgState,
    ai.imgCountry,
    ai.iptcTitle,
    ai.iptcDescription
FROM asset_images2 ai
JOIN pagecontent_gallery pcg ON pcg.imgId = ai.imgId
WHERE pcg.pageId = :pageId";
$stmt = $con->prepare($sql);
$stmt->execute(array(':pageId' => $pageId));
$pageImgArr = $stmt->fetchAll(PDO::FETCH_ASSOC);
$totalImages = count($pageImgArr);
$totalPageImages = $totalPageImages + $totalImages;
```

The Sitemap.xml script



```
foreach ($PageImgArr as $img) {
    $imgLoc = $img["imgWebpScnFull"];
    $imgCity = $img["imgCity"];
    $imgState = $img["imgState"];
    $imgCountry = $img["imgCountry"];
    $imgTitle = $img["iptcTitle"];
    $imgCaption = $img["iptcDescription"];
    $imgGeoLoc = $imgCity . ", " . $imgState . ", " . $imgCountry;
    $imgLicense = "https://www.barcelonatravelhacks.com/info/i0002/terms-and-conditions-barcelonaTravelHacks";

    $txt .= '        <image:image>
';
    $txt .= '            <image:loc>https://www.barcelonatravelhacks.com/' . $imgLoc . '</image:loc>
';
    $txt .= '            <image:caption>' . $imgCaption . '</image:caption>
';
    $txt .= '            <image:geo_location>' . $imgGeoLoc . '</image:geo_location>
';
    $txt .= '            <image:title>' . $imgTitle . '</image:title>
';
    $txt .= '            <image:license>' . $imgLicense . '</image:license>
';
    $txt .= '        </image:image>

';
}
```

The Sitemap.xml script



```
}
$txt .= '</url>

';
}

/*//////////////////////////////////// PDF FILES //////////////////////////////////////*/

$txt .= '<!-- ***** -->
<!-- ***** PDF_DOCUMENTS ***** -->
<!-- ***** -->

';

$sql = "SELECT asset_pdf.pdfPath FROM asset_pdf WHERE noIndexTag = 0";
$stmt = $con->prepare($sql);
$stmt->execute();
$totalPdfs = $stmt->rowCount();
```

The Sitemap.xml script



```
while ($row = $stmt->fetch(PDO::FETCH_ASSOC)) {

    $PdfDocumentPath = $row["pdfPath"];

    $url = "https://www.barcelonatravelhacks.com/" . $PdfDocumentPath;

    $tz_object = new DateTimeZone('Europe/Madrid');
    $datetime = new DateTime();
    $datetime->setTimezone($tz_object);
    $lastMod = $datetime->format('Y-m-d\TH:i:sP');

    $txt .= '<url>'
';
    $txt .= '    <loc>' . $url . '</loc>'
';
    $txt .= '    <lastmod>' . $lastMod . '</lastmod>'
';
    $txt .= '    <changefreq>yearly</changefreq>'
';
    $txt .= '    <priority>0.8000</priority>'
';
    $txt .= '</url>'
';
}
```

The Sitemap.xml script



```
} // End of while loop

$txt .= '</urlset>

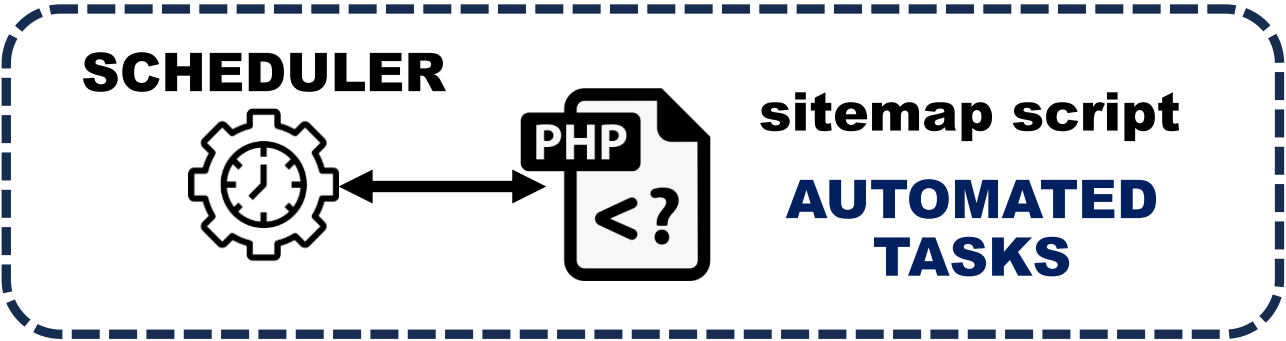
';
$menuAndHomepageUrlCount = $menuUrlCount + $infoUrlCount + 1;
$pagesTotal = $menuAndHomepageUrlCount + $attEvtUrlCount;
$assetsTotal = $totalPageImages + $totalPdfs;
$urlTotal = $pagesTotal + $assetsTotal;

$txt .= '<!-- ***** -->
<!-- *** TOTAL_HOME_&_MENU_URLs:' . $menuAndHomepageUrlCount . ' *** TOTAL_ATTRACTION_&_EVENT_URLs:' .
$attEvtUrlCount . ' *** TOTAL_PAGES:' . $pagesTotal . ' **** -->
<!-- ***** -->
<!-- ***** TOTAL_IMAGES:' . $totalPageImages . ' ***** TOTAL_PDFs:' . $totalPdfs . ' *****
TOTAL_ASSETS:' . $assetsTotal . ' ***** -->
<!-- ***** -->
<!-- ***** URL_TOTAL:' . $urlTotal . ' ***** -->
<!-- ***** -->
';

fwrite($filename, $txt);
fclose($filename);

echo "BTH site map updated: " . $datetime->format('D d-m-yy H:i');
```

Configure the Cron scheduler



The scheduler will automatically trigger the sitemap generator script once daily at midnight, server time (CET)

Current Cron Jobs

Minute	Hour	Day	Month	Weekday	Command	Actions
0	0	*	*	*	<code>/usr/local/bin/php [REDACTED]/public_html/[REDACTED] [REDACTED].php >/dev/null 2>&1</code>	Edit Delete
Sitemap.xml Scheduler						
0	20	*	*	*	<code>/usr/local/bin/php [REDACTED]public_html/[REDACTED] [REDACTED].php >/dev/null 2>&1</code>	Edit Delete
IndexNow API scheduler						

Verification of sitemap.xml



<https://www.barcelonatravelhacks.com/sitemap.xml>

Is a public document on my website.

Robots.txt updated to point to the sitemap.xml

<https://www.barcelonatravelhacks.com/robots.txt>

Future developments:

The site map has a URL counter in it and if the sitemap ever passes 50,000 URLs, then I will need to break down my site map into multiple sitemap files and add a sitemap index file to the robots.txt.