

المرجع العربي لملفات خريطة الموقع

*www.syrdev.net*



Sitemap

إعداد المهندس: إسماعيل عنجريني

الإصدار 1.0

ismaeel.enjreny@gmail.com

## الفهرس

3.....	ماهي خرائط الموقع Sitemaps
3.....	صيغة ملف الـ XML لخريطة الموقع
4.....	مثال عن ملف XML لخريطة الموقع
4.....	تعريف واسمات ملف خريطة الموقع
6.....	المحارف الخاصة في ملف خريطة الموقع
7.....	عينة من ملف خريطة الموقع
8.....	استخدام ملفات فهرس خريطة الموقع لتجميع مجموعة من ملفات خريطة الموقع
9.....	عينة من ملف فهرس خريطة الموقع
10.....	تعريف واسمات فهرس خريطة الموقع
10.....	صيغ أخرى لخريطة الموقع
10.....	التغذية الراجعة Syndication feed
11.....	الملف النصي
12.....	مكان تواجد ملف خريطة الموقع
13.....	ملفات خرائط الموقع ضمن نطاقات مختلفة
14.....	التحقق من ملفات خريطة الموقع
15.....	توسعة الميفاق Sitemap
16.....	إرسال ملفات خريطة الموقع إلى محركات البحث
16.....	إرسال ملفات خريطة الموقع من خلال شاشات محركات البحث
16.....	تحديد ملفات خريطة الموقع من خلال الملف robots.txt
16.....	إرسال ملف خريطة الموقع من خلال طلب HTTP
17.....	استبعاد جزء من المحتوى
17.....	الأسئلة المتكررة

## ماهي خرائط الموقع Sitemaps

خرائط الموقع هي طريقة سهلة تسمح لمدراء مواقع الإنترنت بتعريف محركات البحث عن مواقعهم الجاهزة لجلب معلوماتها من قبل أدوات محركات البحث. وبأبسط أشكالها فإن خريطة الموقع هي عبارة عن ملف XML (اللغة القابلة للتوسع) يحتوي على قائمة بروابط الموقع إضافة إلى بيانات وصفية إضافية (مثل تاريخ آخر تعديل على الرابط, ومعدل تغير محتوى الرابط, وكم هو مهم, وصلة الربط مع روابط أخرى في الموقع), وتستخدم محركات البحث هذه المعلومات لجلب معلومات الموقع بطريقة أسهل وأدق.

عادة جالبات المعلومات Web Crawlers تكتشف الصفحات من الروابط الموجودة ضمن الموقع ومن مواقع أخرى. وخرائط الموقع تقدم هذه البيانات بحيث تسمح لجالبات المعلومات التي تدعم خرائط الموقع بالحصول على هذه الروابط من خلال خريطة الموقع مع معلومات إضافية متوفرة في البيانات الوصفية في نفس الملف. مع العلم أن استخدام ملف خريطة الموقع لا يضمن أن صفحات الموقع سيتم تضمينها في محركات البحث, لكنها تقدم تلميحات مفيدة لمحركات البحث للقيام بعمل أفضل لجلب معلومات موقعك وفهرسته.

الميثاق Sitemap 0.90 مقدم ضمن الترخيص " Attribution-ShareAlike Creative Commons License", ومعتمد بشكل واسع إضافة إلى الدعم المقدم من قبل شركات جوجل Google, ياهو! Yahoo! وشركة مايكروسوفت Microsoft.

## صيغة ملف الـ XML لخريطة الموقع

هذه الوثيقة تصف مخطط ملف الـ XML للميثاق Sitemap.

تتألف صيغة ميثاق Sitemap من واسمات XML (اللغة القابلة للتوسع). وكل قيم البيانات يجب أن تكون توافق قواعد الكتابة في ملفات XML, والملف نفسه (ملف خريطة الموقع) يجب أن يشفر بصيغة UTF-8.

ملف خريطة الموقع Sitemap يجب أن:

- يبدأ بالواسمة <urlset> وينتهي بالواسمة </urlset>.
- تحديد فضاء الأسماء (المواصفة القياسية للميثاق) ضمن الواسمة <urlset>.
- تضمين الواسمة <url> لكل عنوان على الإنترنت URL, كواسمة أب.
- تضمين الواسمة <loc> ضمن الواسمة <url>.

كل الواسمات الأخرى هي واسمات اختيارية أي يمكن بناء ملف خريطة الموقع بدونها, مع العلم أن دعم الواسمات الاختيارية الأخرى يتراوح بين محركات البحث المختلفة, لذلك يفضل مراجعة توثيق هذه المحركات للتأكد عن دعم هذه الواسمات. كما أن جميع العناوين

URL في نفس ملف خريطة الموقع يجب أن تكون من نطاق واحد فقط, مثال [www.example.com](http://www.example.com) أو [.stroe.example.com](http://stroe.example.com)

### مثال عن ملف XML لخريطة الموقع

المثال التالي هو عينة عن خريطة موقع تحتوي عنوان واحد فقط وتستخدم جميع الواسمات الاختيارية, والواسمات الاختيارية

محددة بخط مائل.

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

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

  <url>

    <loc>http://www.example.com/</loc>

    <lastmod>2005-01-01</lastmod>

    <changefreq>monthly</changefreq>

    <priority>0.8</priority>

  </url>

</urlset>
```

### تعريف واسمات ملف خريطة الموقع

الخاصية		الوصف
<urset>	إجباري	تغلف ملف خريطة الموقع وتحدد مرجع المواصفات القياسية الحالية للميفاق.
<url>	إجباري	الواسمة الأب لكل واسمات العنوان الأخرى, وجميع الواسمات الأخرى هي أبناء لهذه الواسمة.
<loc>	إجباري	عنوان الصفحة على الإنترنت, والعنوان URL يجب أن يبدأ بميفاق (مثل http) وينتهي بالحرف /, إن كان خادم الويب يطلبها, وطول هذه القيمة لا يجب أن يتجاوز 2,048 حرف.
<lastmod>	اختياري	تاريخ آخر تعديل على الملف, والتاريخ يجب أن يكون بالصيغة <a href="http://www.w3.org/TR/2001/REC-W3CDTF-20010620/">W3C Datetime</a> . وتسمح هذه الصيغة بتجاهل الوقت عند عدم وجود ضرورة لذكره, عندها يمكن استخدام الصيغة -YYY.MM-DD.

<p>يجب الانتباه إلى أن هذه الواسمة لا علاقة لها بالترويسة -If (304) Modified-Since التي يمكن لمخدم الويب إعادتها, ويمكن لمحرك البحث استخدام هاتين المعلوماتين بشكل مختلف تماماً.</p>		
<p>تحدد معدل تغير معلومات الصفحة, حيث توفر هذه القيمة معلومات عامة لمحرك البحث وليس من الضرورة أن تحدد معدل جلب معلومات الصفحة من قبل محرك البحث, والقيم المقبولة هي:</p> <ul style="list-style-type: none"> <li>• always</li> <li>• hourly</li> <li>• daily</li> <li>• weekly</li> <li>• monthly</li> <li>• yearly</li> <li>• never</li> </ul> <p>القيمة "always" يجب أن تُستخدم مع الصفحات التي تتغير كل مرة تتم زيارتها, أما القيمة "never" يجب أن تُستخدم لتوصيف العناوين الموجودة في الأرشيف أي التي لا تتغير.</p> <p>مع العلم أن القيمة في هذه الخاصية يستخدمها محرك البحث على أنها تلميح للاستفادة منها وليست أمر يلتزم به محرك البحث, فحتى لو أن جالبات المعلومات أخذت بعين الاعتبار هذه القيم, فمن الممكن جلب معلومات الصفحات الموصفة بالقيمة "hourly" بمعدل أقل من ذلك, ويمكن أن تجلب معلومات الصفحات الموصفة بالقيمة "yearly" بمعدل أكبر من ذلك, وأيضاً من الممكن جلب معلومات الصفحات الموصفة بالقيمة "never" وذلك لفهرسة التغيرات الغير متوقعة على هذه الصفحات.</p>	<p>اختياري</p>	<p>&lt;changefreq&gt;</p>
<p>تحدد هذه القيمة أولوية العنوان بالنسبة للعناوين الأخرى في موقعك, والقيم المقبولة تكون ضمن المجال 1.0 - 0.0, ولا تؤثر هذه القيمة على كيفية مقارنة الصفحة بصفحات أخرى في مواقع أخرى, فقط تسمح هذه القيمة لمحرك البحث بمعرفة الصفحات التي تعتبرها أهم من الصفحات الأخرى, والقيمة الافتراضية هي 0.5.</p> <p>مع العلم أن هذه القيمة لا تؤثر على ترتيب الصفحة في صفحة</p>	<p>اختياري</p>	<p>&lt;priority&gt;</p>

نتائج البحث. ويمكن أن تستخدم محركات البحث هذه المعلومات عند الاختيار بين العناوين في نفس الموقع, لذلك يمكنك استخدام هذه المعلومات لزيادة أرحية الصفحة على الصفحات الأخرى في فهرس محرك البحث. أيضاً يجب الأخذ بعين الاعتبار بأن إعطاء قيمة عظمي لكامل عناوين الموقع لن يساعد في فهرسة الموقع بشكل أفضل, حيث أن الأولوية هي قيمة نسبية وتساعد على الاختيار بين عناوين الموقع ذاته.		
--	--	--

### المحارف الخاصة في ملف خريطة الموقع

تشفير ملف خريطة الموقع يجب أن يكون من نوع UTF-8 (بشكل عام يمكن تنفيذ ذلك عند حفظ الملف), وكما في جميع الملفات بصيغة XML فإن البيانات وحتى التي تحتوي عناوين صفحات على الإنترنت URLs يجب أن صيغ الهروب لبعض المحارف كما في الجدول التالي:

المحرف	رمز الهروب	
العطف Ampersand	&	&amp;
اقتباس مفرد Single Quote	'	&apos;
اقتباس مضاعف	“	&quot;
أكبر من	>	&gt;
أصغر من	<	&lt;

بالإضافة إلى أن جميع عناوين صفحات الإنترنت يجب أن تستخدم رموز الهروب لتمكين خادم الهروب من قراءة ما بشكل صحيح, وفي حال استخدام أداة توليد ملف خريطة الموقع فإن هذه العملية تتم بشكل تلقائي. ولمزيد من المعلومات الرجاء مراجعة [RFC-3986](#) من أجل عناوين الإنترنت URIs و [RFC-3987](#) من أجل المواصفات القياسية ل IRIs والمواصفات القياسية [XML](#).

المثال التالي يستخدم محارف ليست ضمن ASCII وهو المحرف (ü), إضافة إلى المحرف & الذي يحتاج إلى استخدام رمز الهروب &amp; بدلاً عنه:

<http://www.example.com/ümlat.php?q=name>

في المثال التالي نفس العنوان URL بتشفير ISO-8859-1 (حيث الاستضافة هي على خادم يعتمد هذا التشفير):

```
http://www.example.com/%FCmlat.php?q=name
```

في المثال التالي نفس العنوان URL بتشفير UTF-8 (حيث الاستضافة هي على خادم يعتمد هذا التشفير):

```
http://www.example.com/%C3%BCmlat.php?q=name
```

في المثال التالي نفس العنوان URL بتشفير UTF-8 مع استخدام رموز الهروب:

```
http://www.example.com/%C3%BCmlat.php& q=name
```

### عينة من ملف خريطة الموقع

المثال التالي للملف خريطة الموقع يحتوي على مجموعة من العناوين URLs وكل واحد منها يستخدم مجموعة من الواسمات

الاختيارية:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://www.example.com/</loc>
    <lastmod>2005-01-01</lastmod>
    <changefreq>monthly</changefreq>
    <priority>0.8</priority>
  </url>
  <url>
    <loc>http://www.example.com/catalog?item=12&desc=vacation_hawaii</loc>
    <changefreq>weekly</changefreq>
  </url>
  <url>
```

```
<loc>http://www.example.com/catalog?item=73&desc=vacation_new_zealand</loc>

<lastmod>2004-12-23</lastmod>

<changefreq>weekly</changefreq>

</url>

<url>

<loc>http://www.example.com/catalog?item=74&desc=vacation_newfoundland</loc>

<lastmod>2004-12-23T18:00:15+00:00</lastmod>

<priority>0.3</priority>

</url>

<url>

<loc>http://www.example.com/catalog?item=83&desc=vacation_usa</loc>

<lastmod>2004-11-23</lastmod>

</url>

</urlset>
```

### استخدام ملفات فهرس خريطة الموقع لتجميع مجموعة من ملفات خريطة الموقع

يمكن استخدام مجموعة من ملفات خريطة الموقع، ولكن كل ملف يجب ألا يتجاوز عدد العناوين التي يحتويها على 50,000 عنوان وحجمه لا يزيد عن 10 ميغا (10,485,760 بايت)، كما يمكن ضغط ملف خريطة الموقع باستخدام تقنية gzip لتقليل عرض الحزمة المطلوبة عبر الشبكة، ولكن عند فك ضغط الملف أيضاً يجب ألا يتجاوز حجمه 10 ميغا، وفي حال وجود أكثر من 50,000 عنوان يجب إضافة أكثر من ملف خريطة موقع.

عند استخدام أكثر من ملف خريطة موقع من الأفضل استخدام ملف فهرس خريطة الموقع، وطبعاً هذا الملف يجب ألا يحتوي على أكثر من 50,000 عنوان لملف خريطة موقع وحجمه يجب ألا يتجاوز 10 ميغا (10,485,760 بايت) كما يمكن استخدام

تقنية gzip لضغط الملف, ويمكن استخدام أكثر من ملف فهرس خريطة موقع, مع العلم أن صيغة الملف شبيهة بصيغة ملف خريطة الموقع.

ويجب أن تحقق صيغة ملف فهرس خريطة الموقع ما يلي:

- يبدأ بالواسمة <sitemapindex> وينتهي بالواسمة </sitemapindex>.
- استخدام الواسمة <sitemap> لكل ملف خريطة موقع.
- استخدام الواسمة <loc> كابين للواسمة <sitemap>.

كما أن الواسمة الاختيارية <lastmod> متوفرة ضمن صيغة ملف فهرس خريطة الموقع.

ملاحظة: ملف فهرس خريطة الموقع يمكن أن يحتوي فقط على ملفات خريطة موقع لنفس الموقع, فمثلاً الفهرس [http://www.yoursite.com/sitemap\\_index.xml](http://www.yoursite.com/sitemap_index.xml) يمكن أن يحتوي خريطة موقع للنطاق <http://www.yoursite.com> ولا يمكن أن يحتوي خريطة موقع للنطاق <http://www.example.com> أو <http://yoursite.yourhost.com>. ويجب تشفير الملف باستخدام UTF-8 كما هو ملف خريطة الموقع.

### عينة من ملف فهرس خريطة الموقع

المثال التالي يحتوي على فهرس خريطة موقع يحتوي على ملفين لخريطة الموقع:

```
<?xml version="1.0" encoding="UTF-8"?>
<sitemapindex xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <sitemap>
    <loc>http://www.example.com/sitemap1.xml.gz</loc>
    <lastmod>2004-10-01T18:23:17+00:00</lastmod>
  </sitemap>
  <sitemap>
    <loc>http://www.example.com/sitemap2.xml.gz</loc>
    <lastmod>2005-01-01</lastmod>
  </sitemap>
</sitemapindex>
```

ملاحظة: عناوين ملفات خريطة الموقع مثل جميع العناوين في ملف XML يجب أن تستخدم رموز الهروب لبعض الحروف.

### تعريف واسمات فهرس خريطة الموقع

الوصف		الخاصية
تغلف جميع المعلومات حول جميع ملفات خرائط الموقع.	إجباري	<sitemapindex>
تغلف المعلومات حول ملف خريطة موقع واحد، وهو ضمن الواسمة <sitemapindex>.	إجباري	<sitemap>
تُعرف مكان ملف خريطة الموقع، ويمكن أن يكون ملف خريطة موقع Sitemap، أو ملف Atom، أو ملف RSS أو حتى ملف نصي.	إجباري	<loc>
تاريخ آخر تعديل على الملف، والتاريخ يجب أن يكون بالصيغة <a href="#">W3C Datetime</a> . وتسمح هذه الصيغة بتجاهل الوقت عند عدم وجود ضرورة لذكره، عندها يمكن استخدام الصيغة YYYY-MM-DD. وتوفير البصمة الزمنية timestamp، فأنت تسمح لمحرك البحث بجلب معلومات مجموعة جزئية من خرائط الموقع من الفهرس، أي أن جالب المعلومات يُعيد فقط خرائط الموقع التي طرأ عليها تغيير من تاريخ معين، وعملية جلب المعلومات المتدرجة تسمح بعملية اكتشاف سريعة للعناوين الجديدة ضمن المواقع الضخمة.	إختياري	<lastmod>

### صيغ أخرى لخريطة الموقع

إن الميفاق Sitemap يسمح لك بتقديم معلومات مفصلة عن صفحات موقعك إلى محركات البحث، ويُفضل استخدام هذا الميفاق لأنه يسمح بإضافة بيانات وصفية إضافية حول صفحات الموقع إضافة إلى عناوين صفحات الموقع، بالإضافة إلى الميفاق Sitemap ندعم تغذية RSS والملفات النصية txt، التي تقدم معلومات أقل عن العناوين.

### التغذية الراجعة Syndication feed

يمكن استخدام الميفاق (Real Simple Syndication) RSS الإصدار 2.0 أو Atom 0.3 or 1.0، وبشكل عام قد تستخدم هذه الطريقة عند توفر تغذية راجعة في موقعك، مع الانتباه إلى أن هذه الطريقة قد لا تسمح لمحركات البحث بمعرفة كامل عناوين الموقع، وذلك لأن هذه الطريقة تعرض العناوين التي أُضيفت مؤخراً، على الرغم من أن محركات البحث من الممكن أن تستفيد من هذه العناوين للبحث عن العناوين الأخرى في الموقع ضمن عملية جلب المعلومات العادية، كما يجب التأكد

من أن ملف التغذية الراجعة موجود في الجذر الرئيسي للمجلد الذي تريد محرك البحث فهرسته, وتعمل محركات البحث على استخراج المعلومات من التغذية الراجعة كما يلي:

- `<link>`: يُشير إلى عنوان الصفحة في موقعك.
- حقل تاريخ التعديل (الحقل `<pubDate>` للميفاق RSS والحقل `<modified>` في Atom): وتُشير متى تاريخ آخر تعديل على العنوان URL, مع العلم أن استخدام هذا الخيار هو اختياري.

### الملف النصي

- يمكن استخدام ملف نصي يحتوي على عنوان صفحة على الإنترنت في كل سطر, ويجب على الملف النصي اتباع ما يلي:
- يجب أن يحتوي الملف النصي على عنوان URL في كل سطر, ولا يمكن أن يحتوي العنوان على سطر جديد بشكل مضمن.
  - يجب تحديد عنوان الصفحة بشكل كامل بما فيها الميفاق http.
  - كل ملف نصي لا يمكن أن يحتوي على أكثر من 50,000 عنوان ويجب ألا يتجاوز حجم الملف 10 ميغا (10,485,760 بايت), وفي حال أن الموقع يحتوي على أكثر من 50,000 عنوان عندها يمكن توزيعهم على أكثر من ملف نصي وإضافته إلى محركات البحث كل على حدة.
  - يجب أن يستخدم الملف النصي التشفير UTF-8.
  - لا يجب أن يحتوي الملف النصي على معلومات أخرى غير عناوين صفحات الموقع.
  - لا يجب أن يحتوي الملف النصي على معلومات في الترويسة أو تذييل الملف.
  - يمكن استخدام تقنية gzip لضغط الملف بحيث تقلل عرض الحزمة المطلوب لقراءته.
  - يمكن استخدام أي اسم للملف النصي, ولمزيد من المعلومات الرجاء مراجعة [RFC-3986](#) من أجل عناوين الإنترنت URIs و [RFC-3987](#) من أجل المواصفات القياسية ل IRIs.

وفيما يلي عينة عن الملف النصي:

```
http://www.example.com/catalog?item=1
```

```
http://www.example.com/catalog?item=11
```

## مكان تواجد ملف خريطة الموقع

إن مكان وجود ملف خريطة الموقع يحدد العناوين التي يمكن تضمينها في ملف خريطة الموقع, فمثلاً ملف خريطة الموقع الموجود في المسار <http://www.example.com/catalog/sitemap.xml> يمكن أن يحتوي على أي عنوان يبدأ بـ <http://www.example.com/catalog> ولكن لا يمكن أن يحتوي على عناوين تبدأ بـ <http://example.com/images>. وإن كان لديك لتغيير محتويات المسار ملف خريطة الموقع في المسار <http://example.com/path/sitemap.xml> فمحرك البحث يفترض بأن لديك الصلاحيات المناسبة لتزويد معلومات حول العناوين التي تبدأ بـ <http://example.com/path>. وفيما يلي مثال عن العناوين التي تعتبر صحيحة في ملف خريطة الموقع ضمن المسار <http://example.com/catalog/sitemap.xml>:

```
http://example.com/catalog/show?item=23
http://example.com/catalog/show?item=233&user=3453
```

وكمثال عن العناوين التي لا تعتبر صحيحة:

```
http://example.com/image/show?item=23
http://example.com/image/show?item=233&user=3453
https://example.com/catalog/page1.php
```

هذا يعني أن جميع العناوين الموجودة في ملف خريطة الموقع يجب أن تستخدم نفس الميفاق (وهو <http> في مثالنا), والحفاظ على نفس اسم النطاق أيضاً, وعلى سبيل المثال فإن الملف الموجود ضمن <http://example.com/sitemap.xml> لا يمكن أن يحتوي على عناوين تبدأ بـ <http://subdomain.example.com>.

إن العناوين الغير صحيحة لأي سبب يتم إلغاؤها وعدم محاولة جلب معلوماتها أبداً, ويفضل وضع ملف خريطة الموقع في المجلد الرئيسي لمخدم الويب, فمثلاً إن كان خادم الويب هو [example.com](http://example.com) فإن ملف خريطة الموقع يجب أن يكون ضمن المسار <http://example.com/sitemap.xml>, وفي بعض الحالات قد يكون من المطلوب إنشاء ملفات خريطة الموقع مختلفة لمسارات مختلفة ضمن الموقع. وفي حال استخدام رقم المنفذ لإرسال مسار ملف خريطة الموقع عندها يجب تضمين رقم المنفذ في جميع العناوين الموجودة في ملف خريطة الموقع فمثلاً لو أن مسار خريطة الموقع هو

http://example.com:100/sitemap.xml فإن جميع العناوين يجب أن تبدأ بالمسار  
.http://example.com:100/

### ملفات خرائط الموقع ضمن نطاقات مختلفة

لتنفيذ عملية إرسال ملفات خريطة الموقع لنطاقات مختلفة, يجب أن تثبت أنك مالك هذه النطاقات, لكل العناوين التي تريد التعامل معها, وفيتم يلي مثال على ذلك:

```
www.host1.com with Sitemap file sitemap-host1.xml  
www.host2.com with Sitemap file sitemap-host2.xml  
www.host3.com with Sitemap file sitemap-host3.xml
```

وما تحتاجه هو وضع الملفات الثلاثة السابقة على نطاق واحد هو [www.sitemaphost.com](http://www.sitemaphost.com) كما يلي:

```
http://www.sitemaphost.com/sitemap-host1.xml  
http://www.sitemaphost.com/sitemap-host2.xml  
http://www.sitemaphost.com/sitemap-host3.xml
```

طبعاً هذا سيؤدي إلى خطأ حيث تحاول وضع خريطة الموقع [www.host1.com](http://www.host1.com) على الموقع [www.sitemaphost.com](http://www.sitemaphost.com) وهذا ينطبق على خريطتي الموقع الأخيرين, والحل الوحيد هو أن تثبت أنك مالك هذه النطاقات, أي تثبت أن لديك الصلاحيات المناسبة لتعديل هذه الملفات, ويمكن تنفيذ ذلك بتعديل الملف `robots.txt` الموجود على النطاق [www.host1.com](http://www.host1.com) بحيث يشير إلى ملف خريطة الموقع الموجود على [www.sitemaphost.com](http://www.sitemaphost.com). حيث سيحتوي الملف `http://www.host1.com/robots.txt` على السطر التالي:

Sitemap: <http://www.sitemaphost.com/sitemap-host1.xml>

بتعديل الملف `robots.txt` الموجود على [www.host1.com](http://www.host1.com) ليشير إلى ملف خريطة الموقع الموجود على [www.sitemaphost.com](http://www.sitemaphost.com) فأنت تُثبت بشكل ضمني أنه لديك الصلاحيات اللازمة على [www.host1.com](http://www.host1.com). وبكلمات أخرى أي شخص يستطيع التحكم بالملف `robots.txt` الموجود على [www.host1.com](http://www.host1.com) فهو يستطيع أن يثق

بخرطة الموقع في <http://www.sitemaphost.com/sitemap-host1.xml> لتحتوي على عناوين الموقع [www.host1.com](http://www.host1.com)

### التحقق من ملفات خريطة الموقع

يمكن استخدام المخططات التالية التي تعرف الخصائص والعناصر التي يجب أن يحتوي عليها ملف خريطة الموقع وملف فهرس خريطة الموقع, للتحقق من صحة الملفات التي تم توليدها لعناوين موقعك.

Sitemap: <http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd>

Sitemap index: <http://www.sitemaps.org/schemas/sitemap/0.9/siteindex.xsd>

لكن للتحقق من صحة ملفات خريطة الموقع وملفات فهرس خريطة الموقع, يجب إضافة فضاءات الأسماء التالية إلى ملف خريطة الموقع:

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

<urlset xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

      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">

  <url>

    ...

  </url>

</urlset>
```

وإلى ملف فهرس خريطة الموقع:

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

<sitemapindex xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

      xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
http://www.sitemaps.org/schemas/sitemap/0.9/siteindex.xsd"
```

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

<sitemap>

    ...

</sitemap>

</sitemapindex>
```

## توسعة الميفاق Sitemap

يمكن توسعة الميفاق Sitemap وذلك بإضافة فضاءات الأسماء الخاصة بك إلى العنصر الجذر `urlset` كما يلي:

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

<urlset xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

        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:example="http://www.example.com/schemas/example_schema"> <!--
namespace extension -->

    <url>

        <example:example_tag>

            ...

        </example:example_tag>

        ...

    </url>

</urlset>
```

## إرسال ملفات خريطة الموقع إلى محركات البحث

بعد إنشاء ملفات خريطة الموقع وملفات فهرس خريطة الموقع وإضافتها إلى خادم الموقع, يجب الآن إخبار محركات البحث التي تدعم هذا الميثاق عن هذه الملفات, ويمكن تنفيذ ذلك بعدة طرق:

### إرسال ملفات خريطة الموقع من خلال شاشات محركات البحث

توفر محركات البحث أدوات مختلفة لإرسال هذه الملفات مباشرة إلى محرك البحث, ومن خلال هذه الشاشات يمكن الحصول على معلومات إحصائية ومعلومات عن الأخطاء وغيرها, ويمكن مراجعة وثائق محرك البحث المستهدف, وبشكل عام يمكن البحث عن الأدوات WebMasterTools في محرك البحث.

### تحديد ملفات خريطة الموقع من خلال الملف robots.txt

يمكن تحديد مكان ملفات خريطة الموقع من خلال الملف robots.txt, ولتنفيذ ذلك أضف السطر التالي:

```
Sitemap: http://www.example.com/sitemap.xml
```

ولا يهم في أي سطر تم وضع هذه العبارة, وفي حال وجود ملف فهرس خريطة الموقع يمكن تضمينه ضمن الملف robots.txt وعندها لا داعي لذكر المسار الكامل لكل ملف خريطة موقع. كما يمكن ذكر مسار أكثر من ملف خريطة موقع:

```
Sitemap: http://www.example.com/sitemap-host1.xml
```

```
Sitemap: http://www.example.com/sitemap-host2.xml
```

### إرسال ملف خريطة الموقع من خلال طلب HTTP

لإرسال ملف خريطة الموقع من خلال طلب HTTP استبدل <searchengine\_URL> بالعنوان الذي يوفره محرك البحث:

```
<searchengine_URL>/ping?sitemap=sitemap_url
```

Google :http://www.google.com/webmasters/sitemaps/ping?sitemap=

Bing: http://www.bing.com/webmaster/ping.aspx?siteMap=

Yahoo:

http://search.yahooapis.com/SiteExplorerService/V1/updateNotification?appid=Yah  
ooDemo&url=

فمثلاً لو اردت إرسال ملف gz موجود ضمن المسار http://www.example.com/sitemap.gz:

```
<searchengine_URL>/ping?sitemap=http://www.example.com/sitemap.gz
```

عند إرسال ملف خريطة موقع باستخدام طلب HTTP بشكل ناجح عندها سيعيد محرك البحث الشيفرة 200 للدلالة على نجاح العملية, وهذا يعني نجاح استقبال الملف ولا يعني أن ملف خريطة الموقع صحيح أو غير صحيح. ويمكن جدولة عمل لتنفيذ هذه المهمة حيث يعمل على توليد ملف خريطة موقع ومن ثم إرساله إلى محرك بحث واحد أو أكثر.

**ملاحظة:** عند إرسال ملف فهرس خريطة موقع لا داعي لإعادة إرسال ملفات خريطة الموقع بشكل منفرد كل على حدة.

### استبعاد جزء من المحتوى

تُستخدم ملفات خريطة الموقع لإخبار محركات البحث عن العناوين التي ترغب في فهرستها, ولتخبر محرك البحث عن العناوين التي تريد استبعادها من عملية الفهرسة يمكنك استخدام الملف robots.txt أو الواسمات المرتبطة به. ويمكن مراجعة الموقع [www.robotstxt.org](http://www.robotstxt.org) لمزيد من المعلومات حول ذلك.

### الأسئلة المتكررة

• كيف أمثل العناوين في ملف خريطة الموقع؟

○ كما في جميع ملفات ال XML (اللغة القابلة للتوسع) فإن أي قيمة بما فيها عناوين الصفحات على شبكة الإنترنت يجب أن تستخدم رموز الهروب, للمحارف التالية: العطف &, الاقتباس الفردي ' , الاقتباس المضاعف " , أصغر من <, أكبر من >. ويجب التأكد بأن جميع العناوين تتبع المواصفة القياسية [RFC-3986](http://www.rfc-editor.org/rfc/rfc3986) من أجل العناوين URIs, والمواصفة القياسية [RFC-3987](http://www.rfc-editor.org/rfc/rfc3987) للعناوين IRIs, والمواصفات القياسية لملف [XML](http://www.w3.org/TR/2006/REC-xml-20060816/). إن كنت تستخدم شيفرة لتوليد العناوين, يمكنك استخدام رموز الهروب من خلال الشيفرة, وكمثال فإن شيفرة بايثون Python التالية تطبق رموز الهروب على الرابط <http://www.example.com/view?widget=3&count>2>

```
$ python
```

```
Python 2.2.2 (#1, Feb 24 2003, 19:13:11)
>>> import xml.sax.saxutils
>>> xml.sax.saxutils.escape("http://www.example.com/view?widget=3&count>2")
```

والنتيجة هي:

```
http://www.example.com/view?widget=3&amp;count&gt;2
```

- هل يهم أي تشفير أطبقه على ملفات خريطة الموقع؟
  - نعم, يجب استخدام التشفير UTF-8.
- كيف أحدد الوقت؟
  - استخدم الترميز [W3C Datetime](#) للبصمة الزمنية في الخاصية lastmod وجميع الأوقات والتواريخ في هذا الميثاق. على سبيل المثال 2004-09-22T14:12:14+00:00. ويسمح الترميز بإمكانية تجاهل الوقت من خلال الصيغة ISO8601, على سبيل المثال 2004-09-22 أيضاً صحيح. وفي حال أن معلومات موقعك تتغير باستمرار فيفضل استخدام جزء الوقت عندها يصبح لدى محرك البحث معلومات كافية عن موقعك.
- كيف أحسب تاريخ الخاصية lastmod!
  - من أجل الملفات الثابتة, فهي تاريخ تعديل الملف, ويمكن استخدام الأمر date في UNIX للحصول على هذا التاريخ:

```
$ date --iso-8601=seconds -u -r /home/foo/www/bar.php
>> 2004-10-26T08:56:39+00:00
```

ومن أجل معظم العناوين الديناميكية فيمكن حساب قيمة الخاصية lastmod اعتماداً على تاريخ تعديل البيانات المتغيرة في الصفحة أو حساب تقديري اعتماداً على التعديلات الدورية (إن كان ممكناً). واستخدام تاريخ تقريبي أو بصمة زمنية تساعد محركات البحث على تجنب إعادة قراءة معلومات العناوين التي لم تتغير, وهذا يقلل استهلاك عرض الحزمة الخاصة بموقعك وإمكانات المعالج المطلوبة لخادمت الويب الخاصة بالموقع.

- أين أضع ملفات خريطة الموقع؟

○ من المفضل أن تضع ملفات خريطة الموقع على الجذر الرئيسي لخدم الويب <http://example.com/sitemap.xml>, وفي بعض الحالات قد تحتاج إنشاء ملفات خريطة موقع مختلفة لمسارات مختلفة على موقعك, مثلاً تم تجزئة صلاحية الكتابة على مجلدات مختلفة. ونفترض أنه إن كان لديك صلاحية لرفع ملف إلى المسار <http://example.com/path/sitemap.xml> فإنه لديك صلاحيات أيضاً لإرسال بيانات وصفية عن المسار <http://example.com/path>. ويجب أن تنتمي كل العناوين في نفي ملف خريطة الموقع إلى نفس اسم النطاق, مثلاً إن كانت خريطة الموقع موجودة تحت المسار <http://www.example.com/sitemap.xml> فإن هذا الملف لا يمكن أن يحتوي على عناوين من المسار <http://subdomain.exempl.com>. وإن كان تحت المسار <http://www.example.com/myfolder/sitemap.xml> عندها لا يمكن أن يحتوي على عناوين من <http://www.example.com>.

● ماهو الحجم الأعظمي لملف خريطة الموقع؟

○ كل ملف يجب ألا يتجاوز عدد العناوين التي يحتويها على 50,000 عنوان وحجمه لا يزيد عن 10 ميغا (10,485,760 بايت), كما يمكن ضغط ملف خريطة الموقع باستخدام تقنية gzip لتقليل عرض الحزمة المطلوبة عبر الشبكة, ولكن عند فك ضغط الملف أيضاً يجب ألا يتجاوز حجمه 10 ميغا, وفي حال وجود أكثر من 50,000 عنوان يجب إضافة أكثر من ملف خريطة موقع. وعند استخدام أكثر من ملف خريطة موقع من الأفضل استخدام ملف فهرس خريطة الموقع, وطبعاً هذا الملف يجب ألا يحتوي على أكثر من 50,000 عنوان لملف خريطة موقع وحجمه يجب ألا يتجاوز 10 ميغا (10,485,760 بايت) كما يمكن استخدام تقنية gzip لضغط الملف, ويمكن استخدام أكثر من ملف فهرس خريطة موقع, مع العلم أن صيغة الملف شبيهة بصيغة ملف خريطة الموقع.

● موقعي يحتوي على عشرات الملايين من العناوين, هل أستطيع إرسال العناوين التي تم تغييرها مؤخراً؟

○ يمكن سرد العناوين التي تتغير بشكل مستمر ضمن ملفات خريطة موقع صغيرة الحجم, ومن ثم استخدام الواسمة lastmod في ملف فهرس خريطة الموقع لتحديد ملفات خريطة الموقع التي تغيرت, عندها محركات البحث ستعيد فهرسة ملفات خريطة الموقع التي تغيرت فقط.

● ماذا أفعل بعد إنشاء ملف خريطة الموقع؟

○ بعد إنشاء ملف خريطة الموقع, يجب إرسالها إلى محركات البحث وذلك من خلال الأدوات المباشرة أو من خلال تقنية ping أو من خلال إضافتها إلى الملف robots.txt

● هل يجب تحديد كامل مسار العنوان في ملف خريطة الموقع؟

- نعم يجب. يجب تضمين الميثاق (مثل http) ضمن معلومات العنوان, كما يجب إنهاء العنوان بالمحرف / إن كان خادم الويب يحتاج هذا المحرف, ومثال فإن العنوان http://www.example.com/ هو عنوان صحيح في ملف خريطة الموقع, بينما www.example.com ليس صحيحاً.
- يحتوي موقعي على إصدار http و إصدار https من عناوين الصفحات, هل يجب سرد الإثنين في خريطة الموقع؟
  - لا يجب ذلك. يجب ذكر إصدار واحد فقط من العنوان, لأن ذكر عدة إصدارات من نفس العنوان قد تؤدي إلى عدم اكتمال عملية الفهرسة بشكل صحيح.
- عناوين الصفحات URLs في موقعي تحتوي على رقم الجلسة Session ID, فهل يجب إزالته قبل إضافته إلى خريطة الموقع؟
  - نعم يجب إزالته. إضافة رقم الجلسة إلى عنوان الصفحة URL قد يؤدي إلى عدم اكتمال عملية الفهرسة بشكل صحيح.
- هل يؤثر مكان العنوان URL في خريطة الموقع على استخدامه؟
  - لا يؤثر. لا يؤثر مكان العنوان URL في ملف خريطة الموقع على الطريقة التي يتعامل فيها محرك البحث مع الصفحة.
- بعض الصفحات في موقعي تستخدم الإطارات, فهل يجب تضمين عناوين URLs الإطارات في خريطة الموقع؟
  - نعم يجب إدراج جميع العناوين URLs.
- هل أستطيع استخدام تقنية zip لضغط ملف خريطة الموقع أم يجب استخدام تقنية gzip؟
  - يجب استخدام تقنية gzip لضغط ملف خريطة الموقع, مع الانتباه إلى أن حجم الملف لا يجب أن يتعدى 10 ميغا (10,485,760 بايت) سواء كان الملف مضغوطاً أم لا.
- هل تؤثر قيمة الواسمة priority على تقييم الصفحة في صفحة نتائج البحث؟
  - تحدد قيمة الواسمة priority أهمية الصفحة بالنسبة لك مقارنة مع الصفحات الأخرى في موقعك, ولا تؤثر أبداً على تقييم الصفحة في نتائج البحث.
- هل يوجد مخطط XML للتأكد من صحة ملف خريطة الموقع؟
  - نعم يوجد, حيث يوجد ملف مخطط XML للتحقق من صحة ملف خريطة الموقع وهو <http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd> وملف مخطط XML للتحقق من صحة ملف فهرس خريطة الموقع وهو <http://www.sitemaps.org/schemas/sitemap/0.9/siteindex.xsd>.
- ماذا لو كان لدي سؤال آخر حول استخدام ملف خريطة الموقع وإرساله إلى محركات البحث؟

○ الرجاء قراءة الوثائق المتوفرة لدى كل محرك بحث حول استخدام وإرسال ملفات خريطة الموقع.

نهاية الوثيقة.