

# Les fichiers sitemap et leur utilité dans le référencement de votre site

Date de dernière mise à jour : 13/02/2014 à 14:28

Source : <http://www.vulgarisation-informatique.com/sitemap.php>.

Distribution interdite sans accord écrit d'Anthony ROSSETTO (<http://www.vulgarisation-informatique.com/contact.php>)

## Introduction

Après avoir proposé votre site aux nombreux moteurs de recherche de la toile, vous vous apercevrez peut-être que les pages situées les plus en profondeur dans l'architecture de celui-ci sont mal ou pas du tout référencées, avec une fréquence de mise à jour qui ne vous convient pas.

Les raisons à cela sont multiples et viennent principalement du manque de liens pointant sur les pages les plus en profondeur. Pour remédier à ça, il faut d'abord s'appliquer une règle simple et ne jamais utiliser plus de quatre niveaux hiérarchiques pour consulter une page.

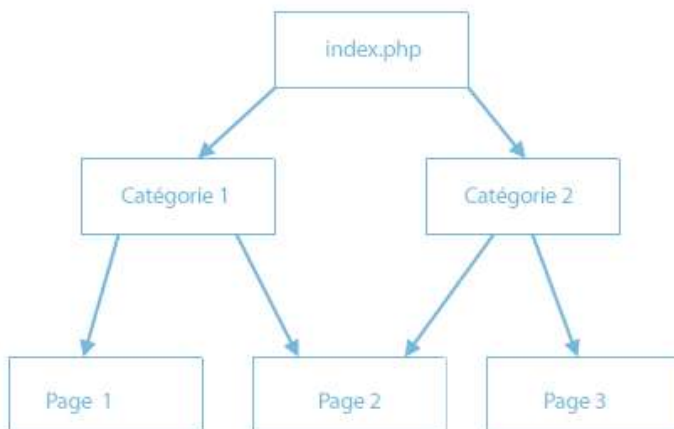
Tout internaute doit pouvoir accéder à toutes les pages de votre site en 4 clics maximum depuis la page d'accueil.

Pour guider les moteurs de recherche, il est recommandé de créer un **plan de site** (aussi appelé **sitemap**). Cela peut se faire de deux façons :

-En créant une page dédiée à cet effet (que vous nommerez par exemple **plan du site** ou encore **carte du site**). Cette solution est recommandée pour vos visiteurs (qui n'iront pas décortiquer le contenu d'un fichier sitemap).

-En créant un fichier contenant toutes les adresses des pages Web de votre site, ce fichier est appelé **sitemap**. Cette solution s'adresse aux robots des moteurs de recherche.

Vous pouvez combiner sans problèmes les deux solutions, c'est même recommandé.



## Arborescence basique d'un site Web

### Le fichier sitemap

Ce fichier est destiné uniquement aux moteurs de recherche qui parcourent votre site. Il recense toutes les URL (adresses) des pages Web présentes et évite ainsi aux bots (robots) de laisser de côté certaines pages. Sa construction obéit à un certain nombre de règles pré-établies. Google, Yahoo et Microsoft se sont mis d'accord pour utiliser un protocole commun que nous allons voir plus bas.

### Construction d'un fichier sitemap

Le sitemap est un fichier basé sur le **XML** qui est un langage à balises. Il s'agit d'un simple fichier texte, vous pouvez donc le créer dans le bloc-notes ou dans tout petit éditeur de texte qui se respecte (oubliez Word et compagnie !).

Seules les URL des pages que vous souhaitez référencer doivent figurer dans ce fichier. D'autres paramètres existent et sont facultatifs. Ils sont néanmoins très utiles car ils vont vous permettre de spécifier une fréquence de mise à jour des pages concernées ainsi qu'une date de dernière mise à jour. Cela aidera le bot (robot) du moteur concerné à scanner ou non vos pages et à privilégier celles qui ont besoin d'une mise à jour dans leur base de données.

Voici un exemple de sitemap basique :

```
<?
xml

version
=
"
1.0
"

encoding
=
"
ISO-8859-1
"

?>
urlset
xmlns
=
"
http
:
//
www.google.com
/schemas/sitemap/0.84
"
>
url
>
loc
>
http://www.vulgarisation-informatique.com/
loc
>

url
>
url
>
loc
>
http://www.vulgarisation-informatique.com/tutoriaux.php
loc
>

url
>

urlset
>
```

Ici nous avons un fichier sitemap composé de deux **URL**. Il est cependant très basique et quasiment inutile. Il vaut mieux le personnaliser comme

ceci :

<?

xml

version

=

"

1.0

"

encoding

=

"

ISO-8859-1

"

?>

urlset

xmlns

=

"

http

:

//

www.google.com

/schemas/sitemap/0.84

"

>

url

>

loc

>

http://www.vulgarisation-informatique.com

loc

>

lastmod

>

2014-02-12

lastmod

>

changefreq

>

daily

changefreq

>

priority

>

1.0

priority

>

url

>

```
url
>
loc
>
http://www.vulgarisation-informatique.com/tutoriels.php
loc
>
lastmod
>
2014-02-12
lastmod
>
changefreq
>
weekly
changefreq
>
priority
>
0.6
priority
>

url
>

urlset
>
```

On constate l'arrivée de nouvelles balises :

- : cette balise permet d'indiquer la date de dernière mise à jour de la page (format du [W3C](#)-). Vous pouvez ne pas mettre d'heure et laisser un format **AAAA-MM-JJ**-.

- : cette balise vous permet de spécifier une périodicité de mise à jour de la page. Si il s'agit d'une page destinée à changer très fréquemment, vous mettez **always**-, sinon **never**- pour une page statique qui restera telle quelle. Voici toutes les valeurs que vous pouvez spécifier :

- always
- hourly
- daily
- weekly
- monthly
- yearly
- never

- : Il s'agit ici d'un nombre compris entre 0.0 et 1.0 spécifiant la priorité de la page par rapport aux autres pages de votre fichier sitemap. Si vous ne spécifiez pas de priorité, elle sera égale par défaut à 0.5.

Envoyer votre sitemap aux moteurs de recherche

Votre sitemap est prêt ? bon, il ne reste plus qu'à le soumettre aux principaux moteurs de recherche (Google, Bing). Voici les principaux

emplacements de soumission :

-Google : [Google Webmasters Tools](#)

-Bing : [Bing Webmasters Tools](#)

Modifier le fichier robots.txt

Pour permettre une indexation à coup sûr de votre sitemap, spécifiez dans un fichier nommé **robots.txt** une ligne **sitemap: monsitemap.xml** (pour peu que votre fichier sitemap s'appelle **sitemap.xml** dans le cas contraire mettez le nom de votre fichier sitemap).

Source : <http://www.vulgarisation-informatique.com/sitemap.php>.

Distribution interdite sans accord écrit d'Anthony ROSSETTO (<http://www.vulgarisation-informatique.com/contact.php>)