

Bâtir

un référencement naturel de qualité

Un bon programmeur internet n'est plus seulement capable de concevoir un site de qualité, de produire du code irréprochable, et d'utiliser des outils de développement. Il lui est également demandé de fournir un site optimisé.

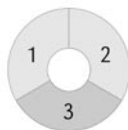
Cet article explique :

- Une méthodologie pour réaliser un site référencement naturel par les moteurs de recherche.

Ce qu'il faut savoir :

- Connaître le langage *xhtml*.
- Connaître le(s) langage(s) généré(s) par les outils de développement utilisés.

Niveau de difficulté



Un bon programmeur internet n'est plus seulement capable de concevoir un site de qualité, de produire du code irréprochable, et d'utiliser des outils de développement. Il lui est également demandé de fournir un site optimisé et exploitable par les moteurs de recherche, ainsi que de mettre en place des outils adaptés pour le suivi du référencement.

Pour ce faire, un grand nombre d'environnements de programmation internet sont maintenant à disposition. Très riche en fonctionnalités pour générer et gérer du contenu web, il est important de les paramétrer et de les enrichir avec les outils adéquats afin qu'ils soient en phase avec le référencement naturel d'un point de vue purement informatique. C'est-à-dire avec la manière dont le site va être indexé et présenté par les moteurs de recherche. Il est également important de prévoir l'exploitation des informations fournies en standard par le serveur http concernant le trafic sur le site. Et ceci afin de pouvoir analyser ces informations et réagir en conséquence.

Il peut donc être nécessaire de consacrer un peu de temps, lors de la conception du site d'intégrer des bases de programmation fluides pour un référencement naturel de qualité, ainsi que de vérifier le paramétrage de l'outil

et de mettre en oeuvre les plug-ins nécessaires afin d'obtenir :

- Une structure de site lisible,
- Un code de qualité,
- La mise en application de quelques règles spécifiques au référencement,
- Un suivi fidèle.

Une structure de site lisible

Bien souvent, pour démarrer un nouveau site, les trois matériaux de base sont :

- La charte graphique plus ou moins aboutie,
- Les fonctionnalités attendues sachant qu'au final, le site ne ressemblera pas toujours à ce qui était prévu initialement,
- Les outils de développement web.

En général, la première étape consiste à découper les maquettes graphiques en éléments de base. Et là, bien souvent, la première impulsion est de donner à ces éléments des noms purement fonctionnels.

Mais, bien qu'il soit plus aisé pour un programmeur de codifier ses éléments, notamment graphiques, il est utile de leur donner de vrais noms, traduisant la fonctionnalité ou l'objet, dans la langue du site. Le plus simple et le plus fonctionnel est d'utiliser des mots simples, séparés par un tiret et éviter les caractères accentués ou spéciaux.

En effet, les internautes et les moteurs associeront automatiquement et naturellement un poids plus important à, par exemple, un

bouton nommé *recherche.gif* que *but-rech.gif*. De même, pour un site en français, l'image d'entête est souvent nommée *header.gif* et non pas par exemple, *en-tete-site.gif*. Or, même une image peut être le résultat d'une requête. Et si, dans deux pages à contenu similaire de deux sites différents susceptibles de répondre à la même requête, la seule différence est une image nommée avec un mot composant la requête de l'internaute, cette différence sera un plus pour que cette page apparaisse avant l'autre dans les résultats de recherche.

Cette notion de *syntaxo-sémantico*, peut être déclinée par la suite à tous les niveaux de programmation, que ce soit dans le choix des noms de variables, de fonctions, de modules, de bibliothèques, de fichiers ou de pages. Si pour les premiers, il s'agit surtout de lisibilité, pour les noms de pages et des répertoires qui pourront former les urls pointant sur les pages web seront par contre triplement significatifs.

Les urls ainsi formées, guident non seulement la navigation. Mais elles sont également utilisées par les outils de suivi de référencement fournis par les moteurs ou par d'autres fournisseurs. En résumé, plus l'url est lisible, plus elle sera exploitable :

- Par le moteur pour présenter la page au sein des résultats car les moteurs accordent plus d'importance à une page dont le nom est en rapport avec le contenu,
- Par l'internaute pour le confirmer au premier coup d'œil dans le fait que cette page est celle qui correspond à sa demande car il sera plus enclin à cliquer sur une page dont le nom correspond à sa demande ou dont le nom signifie quelque chose dans sa propre langue,
- Par l'outil de suivi qui pourra intégrer cette page à ses résultats de suivi et dont ceux-ci seront analysés par la suite en vue d'améliorer le référencement.

Néanmoins, il peut être nécessaire d'utiliser des outils de développements qui génèrent des urls dynamiques. Dans ce cas, il faut de se pencher, dès la conception, sur la manière dont l'outil gère l'*url rewriting*, c'est à dire la technique permettant de transformer des urls dynamiques illisibles en urls lisibles. Malheureusement, tous les outils n'intègrent pas encore nativement la technologie de l'*url rewriting*. Afin de résoudre ce qui à l'usage pourrait se révéler un problème, il est important de bien étudier comment l'outil et le serveur vont intégrer cette *url rewriting*. Ceci permettra de prévoir les règles de génération des urls afin de les gérer nativement proprement.

A titre d'exemple, les dernières versions de Joomla! et de wordpress intègrent maintenant très naturellement l'*url-rewriting*. Dans ces deux cas, il suffit :

- d'adapter, si nécessaire, à la configuration du serveur http les fichiers *htaccess* fournis en exemple lors de l'installation du logiciel en *htaccess*,
- de les renommer en *.htaccess*,
- de les déposer à la racine du site,
- d'activer le module d'*url rewriting* au sein du logiciel (options *Global Configuration/SEO/Search Engine Friendly* et *Dynamic PageS Titles dans Joomla!*, option *Settings/Permalinks dans Wordpress*).

Ces deux outils illustrent bien la manière dont la plupart des outils autorisent maintenant l'*url-rewriting* : le logiciel génère les noms des pages en fonction :

- du titre de l'article dans le cas du système de gestion de contenu,
- du billet ou de la page, dans le cas du moteur de blogs.

Concernant les frameworks tels que Zend ou Symphony, l'*url rewriting* est également exploitable et les sites de documentations afférents à ces outils sont très bien documentés en fonction des versions des frameworks et des serveurs utilisés.

Un code de qualité

Outre que de s'appuyer sur une structure lisible, un autre point important est le respect de la syntaxe du langage de programmation utilisé et une utilisation appropriée de celui-ci.

Chaque langage s'appuie sur une grammaire qui décrit à la fois les règles d'écriture et les mots réservés du langage. C'est également sur cette grammaire que s'appuient les moteurs pour indexer les sites. Donc, plus le code produit respecte scrupuleusement la grammaire utilisée, plus le moteur lira aisément le site et donc plus il l'indexera proprement.

Certains mots réservés du langage de programmation traduisent la structure du site. Par exem-

ple, en `<x>html`, pour insister sur l'importance de certains mots ou concepts, les balises `h<x>` ou `` permettent, dans le premier cas, de hiérarchiser l'information, dans le second cas de mettre en valeur le texte mis en gras. Le contenu entre les balises `h1` correspondant au titre, peut être affiché par le moteur comme résultat. Le contenu des balises `h2` peut être utilisé comme description. Il est donc important de les utiliser avec soin. Concernant le graphisme, il est aisé de redéfinir ces balises, à l'aide de feuilles de styles css, sans alourdir le code, ni dénaturer la grammaire.

Concernant les langages de programmation, l'emploi des validateurs de code est très utile. La page principale des outils de validation de code du W3C donne accès à des outils qui produiront des informations très pertinentes sur la qualité du

code fourni. Le W3C fournit également des spécifications qui permettent de lever toute ambiguïté. Les moteurs de recherche, notamment Google-Yahoo-Msn, fournissent également des guides sur la manière dont leur moteur indexent les sites. Google et Yahoo fournissent aussi des guides de programmation par langages, et même par framework, autour desquels gravitent de nombreux blogs et sites spécialisés. Enfin, l'éditeur de code firefox (affichage/code source de la page), qui intègre le parser de validation du W3C, affiche les erreurs de code de la page en cours.

L'emploi des environnements de programmation tels que les éditeurs wysiwyg, les outils de gestion de contenu, les plateformes de blog, si ils se révèlent pratiques, peuvent être parfois générer un code très difficile à nettoyer, et qui

Déclaration de base du site

Bien évidemment, la déclaration auprès de Google est incontournable. Mais bien que Google est considéré comme le moteur de recherche de référence, l'utilité des autres outils de recherche est indéniable. En effet, la page d'accueil des navigateurs peut également être l'interface Yahoo ou Msn. Donc, l'internaute va parfois utiliser cette interface. Et si votre site est connu de ces deux moteurs, il pourra apparaître dans la page de résultats lorsqu'il correspond à la requête.

Les annuaires et portails, lorsqu'ils sont de qualité, bénéficient également d'une bonne indexation par Google. Donc votre site peut être porté par ceux-ci si Google a suivi les liens menant à votre site.

Dès que la structure du site le permet, le minimum est de :

- déclarer le site auprès de *google-yahoo-msn*,
- déclarer le site auprès de moteurs en vérifiant la popularité de ces outils, en s'appuyant sur les listes de moteurs fournis par exemple par les méta-moteurs tels que *Copernic*,
- déclarer le site auprès de *dmoz*,
- déclarer le site auprès d'annuaires génériques, en s'appuyant par exemple sur la liste des annuaires génériques fournis par *Dmoz*,
- déclarer le site auprès d'annuaires spécialisés du domaine du site.

Listing 1. Jeu de méta-tags

```
<META name="title" content="titre de la page sur 5 mots">
<META name="description" content="description d'environ 2 lignes">
<META name="keywords" content="quelques mots clefs pour mémoire">
<META HTTP-EQUIV="CONTENT-LANGUAGE" CONTENT="French">
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=iso-8859-1">
<META name="Lang" content="FR">
<META name="Author" content="nom de l'auteur">
<META name="Owner" CONTENT="nom du site">
<META NAME="Indentifier-URL" CONTENT="http://url-du-site">
<META NAME="Category" CONTENT="catégorie la plus marquante du site à destination
des annuaires">
<META name="revisit-after" content="x days">
<META NAME="ROBOTS" CONTENT="ALL">
```

Listing 2. Structure du fichier sitemap

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.google.com/schemas/sitemap/0.84">
  <url>
    <loc>http://www.url-de-la-page</loc>
    <lastmod>YYYY-MM-DD</lastmod>
    <changefreq>hourly/daily/weekly/monthly/yearly</changefreq>
    <priority>0.5<x<1.0</priority>
  </url>
</urlset>
```

parfois, après nettoyage, peut se révéler impropre à une migration de l'outil. Il est donc important, de bien étudier celui-ci, de se tenir informé des nouveautés en matière de nouvelles fonctionnalités et de plug-ins, en créant par exemple des alertes sur le sujet visé.

Un gros travail *au fil de l'eau* d'amélioration des pages et des modèles est un atout non négligeable en matière de référencement naturel. En effet, les moteurs reviennent plus souvent sur des pages dont le contenu est modifié et amélioré que sur des pages statiques.

La mise en application de quelques règles spécifiques au référencement

Quelques règles simples servent également à amorcer le référencement naturel et concernent la gestion des méta-tags, du fichier sitemap et la déclaration du site au sein des moteurs de recherche, des annuaires et d'autres sites.

Les méta-tags renseigneront les moteurs sur le contenu global du site, ainsi que sur la page, étant entendu qu'ils doivent être en adéquation. Il faut donc que les personnes en charge d'alimenter le site et/ou de le maintenir puissent les gérer facilement et intelligemment.

Le fichier sitemap aide au pilotage de l'indexation des pages par les moteurs reconnaissant ce format tels que Google et Yahoo. Plus il sera en adéquation avec la vie du site, plus le moteur utilisera son temps à indexer les pages et à revenir en fonction des indications données. Là aussi, il faut donc qu'il puisse être mis à jour en même temps que le site.

La déclaration du site auprès des moteurs les informe de l'existence du site et leur permet donc de le prendre en compte. La déclaration auprès d'annuaires ou de sites, par exemple portails dans un domaine particulier, augmente la notoriété du site, sous entendu que l'annuaire ou le portail utilisé soit également de bonne qualité.

Outre la qualité du site, la mise en application de ces trois principes contribuera plus que fortement à la construction d'un référencement naturel de qualité.

Les méta-tags

Les méta-tags, intégrés dans l'entête des pages, sont utilisés comme moyen de dialogue entre le site, les moteurs et l'internaute. Le minimum vital est le couple *titre-description*. Ces deux informations peuvent être utilisées par le moteur pour présenter le site au sein de ses pages de ré-

sultat, par les autres sites pour présenter cette page, ou par les flux d'actualités. Lorsque le moteur n'utilise pas les informations fournies, c'est qu'il peut y avoir plusieurs problèmes :

- le moteur n'arrive pas à accéder à la page (problème de navigation ou d'*url-rewriting*),
- le moteur n'arrive pas à lire correctement la page (problème de lisibilité de code),
- le moteur juge que ces informations ne correspondent pas à la réalité, c'est à dire, au contenu du site et de la page.

Dans ce cas précis, il est donc nécessaire de faire une adéquation, par exemple entre le titre de la page et le méta *title*, ainsi qu'entre le premier paragraphe de la page et le méta *description*. Lorsqu'un site va être trop volumineux pour particulariser chaque page, il peut exister un paramétrage dans l'outil ou un plug-in permettant de générer automatiquement des méta à partir du contenu de la page. Par contre, le fait de générer le même titre et la même description pour l'ensemble du site n'est pas du tout optimal pour les moteurs et même pénalisant. Si l'outil utilisé n'a pas cette fonctionnalité de gestion des méta-tags en standard, ou n'a pas de plug-in pour le faire, il est toujours judicieux de programmer un petit add-on pour remplir cette tâche.

En outre, l'utilisation d'un jeu de méta génériques, c'est à dire réutilisables pour l'ensemble du site est intéressant. Par exemple, le jeu de méta suivant, pour un site français peut être utilisé après avoir été complété (Listing 1).

Les deux dernières méta, peuvent être particularisées en fonction des pages. Par exemple, une page actualités peut être revisitée tous les jours. Ou si une page ne doit pas être indexée pour des raisons de maintenance, le remplacement de `ALL` dans la méta Robots par `noindex` permet de ne pas indexer la page en question.

Concernant le couple titre-description, la plupart des outils permettent de les gérer très aisément lors de l'écriture, sous entendu évidemment que le logiciel ait été paramétré pour cela.

Par exemple, pour Joomla!, l'option *Global Configuration/MetaData* est à utiliser et permet en sus, de gérer la description.

Par contre, pour Wordpress, il est nécessaire de télécharger et d'installer en sus un plug-in, tel que all-in-seo-pack.

Sitemap

Ce fichier xml est utilisé pour indiquer au trio google-yahoo-msn quelles sont les pages à indexer. Après l'avoir généré, il est à déposer à la racine du site, puis à être déclaré auprès de Google. Ce fichier doit être la représentation exacte du site car, si des urls indiquées ne sont pas accessibles aux moteurs lors de l'indexation, le moteur peut ne pas revenir de suite. En conséquence, il faut donc le mettre à jour à chaque fois que le site est modifié.

Sur Internet

- <http://www.w3.org/QA/Tools/> – Page principal du W3C des validateurs de code,
- <http://www.w3.org/> – Page principal du W3C,
- <http://http://www.google.fr/webmasters/> – Page principale Google d'aide aux programmeurs,
- <http://http://www.google.com/webmasters/> – Page principale US Google d'aide aux programmeurs,
- http://code.google.com/apis/gadgets/docs/dev_guide.html – Guide de Google pour développer des widgets ou des api web,
- <http://developer.yahoo.com/> – page principale US Yahoo des développeurs,
- <http://xml-sitemaps.com/> – Générateur gratuit de sitemap,
- <http://www.google.fr/intl/fr/webmasters/> – Interface du gestion de sitemap de Google.

Suivi de site

Tout le trafic du site est noté dans les fichiers logs du serveur. Pour chaque connexion au site, les informations fournies concernent la date et l'heure, la provenance de l'url et l'url complète de requête. Ce fichier peut être exploité de trois manières parallèles :

- par la mise en place de Google Analytics qui donnera des informations très précises sur la provenance des requêtes et les pages vues, par périodes,
- par l'hébergeur lui-même qui fournit parfois une interface d'exploitation de ce fichier en présentant des statistiques périodiques sur la provenance géographique, par moteurs, par domaines...,
- par l'emploi d'outils annexes tels que Phpmysite, outil gratuit d'analyse d'audience qui, correctement paramétré, permet un suivi assez fin de ce trafic.

Concernant ces trois types d'outils, les informations qui sont les plus pertinentes à exploiter, par périodes (par exemple mensuelles) sont :

- le trafic global,
- la décomposition du trafic par domaine,
- la décomposition du trafic en provenance des moteurs par moteur,
- le nombre de pages vues globalement,
- le nombre de requêtes et les requêtes ayant permis d'aboutir sur votre site.

pour chaque page vue, le nombre de requête et les requêtes ayant permis d'aboutir.

L'analyse de ces requêtes permettra d'affiner par exemple, l'adéquation du langage utilisé, la pertinence des pages...

Structure du sitemap

Pour beaucoup d'outils de développement web ou de gestion de contenu, il existe des add-ons sous forme de plug-in pour le générer. En outre, un grand nombre de sites proposent des outils pour le générer gratuitement, tel par exemple xml-sitemap qui, à partir d'une url, génère ce fichier. Mais la génération est globale. Alors qu'il peut être intéressant de paramétrer chaque page, ou des ensembles de pages d'une manière spécifique.

Par exemple, sachant que pour chaque page à indexer, il est possible de définir :

- la date de la dernière mise à jour,
- la fréquence de demande d'indexation en fonction de la fréquence de mise à jour de la page,
- l'importance de la page au sein du site par rapport au site lui-même (0.5 est la moins importante, 1.0 est la plus importante) suivant la structure suivante (Listing 2).

On peut alors jouer sur les paramètres de fréquences d'indexation en fonction des mises à jour, changer les dates de modification lorsque c'est nécessaire et prioriser différemment les pages.

Déclaration du sitemap

Pour déclarer le sitemap via l'interface de google (<http://www.google.fr/intl/fr/webmasters/>) dédié à cet effet, il suffit de :

- déposer le fichier sitemap.xml à la racine du site,
- créer un compte en utilisant une adresse email valide et en créant un mot de passe,
- se positionner sur le tableau de bord,
- ajouter l'url en se positionnant sur *cliquer ici pour ajouter un site*,
- valider le site avec soit la méthode *d'ajout d'une balise méta* qui devra donc être intégrée dans l'entête des méta-tags de la page principale du site, soit la méthode *transfert d'un fichier* dont le nom vous est fourni par Google et qui doit être déposé à la racine du site après création
- Après avoir intégré la balise ou déposé le fichier, cliquer sur *vérifier*,
- se repositionner ensuite sur le tableau de bord,
- cliquer sur *ajouter* en face du nom du site,
- donner le nom du fichier et sélectionner par exemple *ajouter un plan sitemap web générique*
- cliquer sur *ajouter un plan sitemap web générique*,
- se repositionner ensuite sur le tableau de bord,
- cliquer sur *vérifier* en face du nom du site.

Il est important ensuite de suivre régulièrement l'évolution de l'indexation du site à l'aide de ce tableau de bord car il donne de précieuses informations sur la manière dont Google intègre votre site. Toujours en se connectant à l'interface de ges-

tion des fichiers de sitemap et en se positionnant sur le site, le tableau de bord vous donne, entre autres, les informations suivantes :

- *exploration/exploration du web* : sont recensées toutes les urls qui n'ont pas pu être explorées. Elles sont classées par catégorie (inaccessibles, inexistantes, restreintes par le fichier robots.txt...) ce qui vous oriente pour corriger les liens,
- *exploration/analyse du contenu* : ce sont tous les problèmes rencontrés par Google, par exemple en terme d'interprétation de code,
- *statistiques* : concerne par exemple le nombre de requêtes et les requêtes ayant permis d'accéder à votre site, la date à laquelle googlebot est venu sur votre site la dernière fois, le nombre de liens pointant vers vos pages et quelles sont ces pages,
- *liens* : ce sont les liens externes, internes ou générés par Google pointant vers votre site.

Ce premier niveau d'information vous permet ainsi de suivre périodiquement l'évolution de l'indexation et les accès à votre site.

Conclusion

Bien entendu, ces indications sont purement théoriques et doivent être adaptées en fonction du contexte, du langage et des outils qui n'offrent pas toutes les mêmes possibilités et la même souplesse. Ces règles sont simplement un cadre indicatif à suivre. Et si certaines d'entre elles ne peuvent pas être suivies à cause de l'outil employé, il est toujours possible de retomber sur ses pieds en mettant l'accent sur un autre point ou en utilisant un palliatif. Chacun de ces points est documenté génériquement par des sites ou blogs spécialisés dans l'information relative au référencement tels que les excellents sites *Abondance* ou *Search Engine Land*. Les sites éditeurs ou leader d'un logiciel opensource fournissent également toutes les informations nécessaires et font souvent l'objet de forums et de blogs dédiés.

En un mot, plus la structure d'un site et de ses éléments seront lisibles sémantiquement et naturellement, et plus le site aura de la valeur. Le site doit donc pouvoir se parcourir comme une sorte de livre. Ainsi, plus la structure du site ainsi que les éléments d'une page seront en rapport avec le contenu de la page, plus cette page aura de chance de répondre aux interrogations d'un internaute via les pages de recherches des moteurs. Et plus l'internaute se retrouvera dans la structure du site, plus il aura envie d'y rester.

ISABELLE LUPI

Spécialisée dans les domaines de la gestion documentaire, du traitement du langage naturel et du référencement. Elle travaille depuis 10 ans dans ce dernier domaine et a créé, début 2007, sa propre société, B-Find, spécialisée dans le domaine du référencement.