

Référencement & Positionnement sur le WEB

Par : BOUBAKER Nobel El Houssine

Date : 24 & 25 Novembre 2008

Lieu : Hôtel Le PACHA

URL Rewriting : à quoi ça sert ?

L'URL Rewriting est une technique utilisée pour optimiser le référencement des sites dynamiques (utilisant des pages dynamiques). Les pages dynamiques sont caractérisées par des URL complexes, comportant en général un point d'interrogation, éventuellement le caractère & ainsi que des noms de variables et des valeurs.

Exemple : `article.php?id=12&page=2&rubrique=5`

Dans cet exemple, le fichier `article.php` est utilisé pour afficher un article dont le texte vient d'une base de données. C'est un fichier générique, qui peut afficher n'importe quel article, de n'importe quelle rubrique, page par page. Ici on cherche à afficher la page 2 de l'article numéro 12 qui fait partie de la rubrique 5.

URL Rewriting : à quoi ça sert ?

Le problème est que certains moteurs n'indexent pas les pages ayant des URL de ce type. Google par exemple n'indexe en général que les pages ayant au maximum 2 paramètres dans l'URL : il n'indexe donc pas une page comme celle de notre exemple.

Grâce à l'URL Rewriting, Google ainsi que n'importe quel robot va pouvoir indexer les pages dynamiques. L'intérêt est par exemple d'avoir plusieurs milliers de pages de votre site indexées dans Google (le catalogue de tous vos produits, votre forum, etc.).

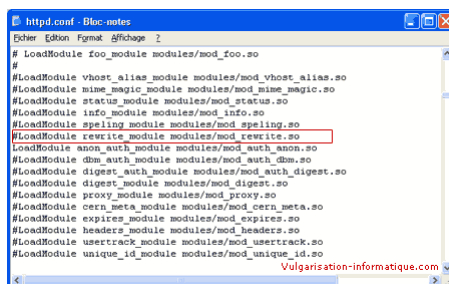
Le webmaster doit changer la façon dont il écrit les liens, selon des règles qu'il va se fixer lui-même. En reprenant l'exemple ci-dessus, on peut remarquer que les URL pour les pages d'articles ont toutes la même forme. On va donc changer nos scripts PHP et nos liens à l'intérieur du site selon le modèle suivant : `article_12_2_5.html`

URL Rewriting : comment ça marche ?

1. Vérifier que votre hébergeur permet l'utilisation de l'URL Rewriting. Si vous êtes chez un hébergeur gratuit qui ne le gère pas, c'est une très bonne raison pour franchir le cap et bénéficier de tous les avantages d'un hébergement professionnel (ça ne coûte pas grand chose...)
2. Identifier les pages dynamiques dont l'URL comporte des paramètres, et choisir un nouveau schéma d'URL "propre"
3. Ecrire les règles de réécriture dans le fichier `.htaccess`
4. Changer tous les liens vers chaque fichier dont l'URL a changé
5. Mettre à jour votre site et vérifier que tout fonctionne

Configurer apache pour la gestion de l'url rewriting

ouvrez le fichier de configuration. Il se trouve ici : /etc/apache/httpd.conf pour les utilisateurs de linux et dans le répertoire apacheconf pour les utilisateurs de Windows.



```
httpd.conf - Bloc-notes
[Edit] [Edit] [Format] [Page] ?
# LoadModule foo_module modules/mod_foo.so
#
#LoadModule vhost_alias_module modules/mod_vhost_alias.so
#LoadModule mime_magic_module modules/mod_mime_magic.so
#LoadModule status_module modules/mod_status.so
#LoadModule info_module modules/mod_info.so
#LoadModule spelling_module modules/mod_spelling.so
#LoadModule rewrite_module modules/mod_rewrite.so
LoadModule anon_auth_module modules/mod_auth_anon.so
#LoadModule dbm_auth_module modules/mod_auth_dbm.so
#LoadModule digest_auth_module modules/mod_auth_digest.so
#LoadModule digest_module modules/mod_digest.so
#LoadModule proxy_module modules/mod_proxy.so
#LoadModule cern_meta_module modules/mod_cern_meta.so
#LoadModule expires_module modules/mod_expires.so
#LoadModule headers_module modules/mod_headers.so
#LoadModule usertrack_module modules/mod_usertrack.so
#LoadModule unique_id_module modules/mod_unique_id.so
Vulgarisation-informatique.com
```

Configurer apache pour la gestion de l'url rewriting

Pour les utilisateurs de Windows repérez la ligne :
#LoadModule rewrite_module modules/mod_rewrite.so

pour les utilisateurs de linux (ou X est la version d'apache) :
#LoadModule rewrite_module /usr/lib/apache/X/mod_rewrite.so

Enlevez le signe # devant cette ligne

Repérez ensuite la ligne suivante :

#AddModule mod_rewrite.c si vous êtes sous Windows et enlevez le signe # placé devant

(on appelle cela **décommenter une ligne**)

URL rewriting

Les URL des deux pages de l'article 8125 seront donc sous la forme (si votre script se nomme article.php) :

`http://www.votresite.tld/article.php?numero=8125&page=1`
`http://www.votresite.tld/article.php?numero=8125&page=2`

Vous préféreriez, et cela se comprend, que vos visiteurs accèdent à cet article par :

`http://www.votresite.tld/article-8125-1.html`
`http://www.votresite.tld/article-8125-2.html`

Analysons point par point comment réécrire cette règle toujours simple :
Nous voyons dans ces URL qu'elles contiennent deux parties variables : le numéro d'article et le numéro de page, tout le reste étant fixe comme le nom du script et le nom des variables. la règle s'écrirait comme ceci :

```
RewriteEngine on
RewriteRule ^article-([0-9+)-([0-9+)]\.html$ article.php?numero=$1&page=$2 [L]
```

URL rewriting

`http://www.example.com/viewtopic.php?order=0&start=1256&message=50`

peut être transformée en :

`http://www.example.com/topic-1256-50.html`

On suppose ici que la valeur 'order' est à 0 par défaut.

Le .htaccess contiendra :

```
RewriteEngine on
RewriteRule ^topic-([0-9+)-([0-9+)]\.html$ viewtopic.php?order=0&start=$1&message=$2 [L]
```

Long Tail

L'effet long tail appliqué au référencement :

L'utilisation de nombreux mots clés moins importants peut avoir un impact parfois plus fort que le choix de quelques mots clés jugés primordiaux.

C'est l'effet long tail. (16/02/2007)

Pour sélectionner les bonnes expressions, il est nécessaire d'examiner les tendances à partir des statistiques issus du site afin de connaître les mots clés réellement tapés.

Impossible en effet d'inventer ou de deviner ces mots. Alors que les outils de suggestions de mots clés ne listent que les mots clés les plus tapés, seul le site peut éclairer sur ces expressions.

Reste ensuite à mettre en place le travail pour améliorer leur visibilité.

TrustRank : Autres critères utilisés par Google (suite)

Le Trust Rank (terme anglais signifiant confiance d'une page) ou TR désigne le degré de confiance que le moteur de recherche Google attribue, à une page web ou un site web pour le classement de ses sites, une note comprise entre 0 (qui équivaut à du spam) à 1 (page de confiance). La note est donc le TR. Ce nom a été déposé par la société Google le 16 mars 2005.

Le TrustRank remplace-t-il le pageRank ? Tout le monde n'est pas d'accord sur ce point mais une évolution du pagerank est clairement perceptible.

L'objectif du TrustRank est de lutter contre le web spamming.

Le terme est apparu lors de la publication d'un document écrit par Zoltán Gyöngyi, Hector Garcia-Molina (tous les 2 de Stanford University comme les 2 créateurs de Google) et Jan Pedersen de Yahoo! Inc. Le document s'appelle "Combating Web Spam with TrustRank" et a été publié le 1er mars 2004.

Spamdexing

Spamdexing est un néologisme anglais composé du substantif spam et du suffixe dexing pris sur le terme indexing signifiant référencement. En français, on traduit spamdexing par référencement abusif.

C'est un ensemble de techniques consistant à tromper les moteurs de recherche sur la qualité d'une page ou d'un site afin d'obtenir, pour un mot-clef donné, un bon classement dans les résultats des moteurs.

Elle est parfois sévèrement punie par les moteurs. Les techniques habituelles de référencement abusif consistent par exemple à truffer une page satellite de listes de mots-clefs (pour attirer les utilisateurs de moteurs qui font une recherche sur ces mots), ou à créer des dizaines de sites qui pointent les uns vers les autres (link farms ou pépinières de liens) pour améliorer leur classement dans les moteurs qui jugent la qualité d'une page en fonction du nombre de liens pointant vers elle.

Dissimulation du spamdexing

Pour ne pas donner des soupçons à l'utilisateur qui verrait sur son écran une longue liste de mots, les nombreux termes placés dans une page pour « piéger » les moteurs sont souvent camouflés par différents procédés:

- Relégation de ces listes de mots en bas de page
- Ecriture en caractères minuscules
- Mots placés dans une section " noframes ", " noscript " ou " display:none " (généralement non affichée par le navigateur, mais lue par les robots des moteurs)
- Caractères de même couleur que le fond de la page (ce qui rend le texte invisible)

Dissimulation du spamdexing (suite)

- Une page satellite (doorway), truffée de mots-clefs, est lue par les robots des moteurs de recherche; mais quand un humain la consulte, il est redirigé vers une autre page (et donc il ne voit pas la page bidon).
- Le cloaking (l'occultage) consiste à présenter des résultats différents selon le logiciel utilisé pour afficher la page : une page anodine pour un navigateur web, une page optimisée, remplie de mots-clefs, réservée aux robots des moteurs

Sandbox

Saviez-vous que Google gérait un bac à sable ? C'est en tout cas le terme utilisé par les anglo-saxons (sandbox) pour désigner une nouvelle forme de filtrage ou de pénalité qui touche des sites sur Google.

Le sandbox est une forme de pénalité qui semble toucher les nouveaux sites sur Google. Alors qu'ils n'ont pas de problème particulier à être indexés, il est difficile de bien les positionner pour des expressions qui ne semblent pas hors de portée.

Il semblerait que Google pénalise ces nouveaux sites pendant plusieurs semaines ou mois, le temps sans doute qu'ils fassent leur preuve, puis quand la quarantaine est terminée les sites gagnent soudainement beaucoup de places dans Google alors qu'aucune optimisation n'a été effectuée.

Sandbox (suite)

Sur quels critères Google se base-t-il pour infliger cette pénalité ? Cela semble toucher tout nouveau site. L'ancienneté d'un site devient donc un critère qui compte...

Nombreux sont ceux qui pensent que la pénalité ne s'applique pas au site directement, mais aux backlinks de ce site qui ne seraient pris en compte que très tardivement par Google.

Sandbox et Trustrank :

1. La confiance se construit (sur la durée)
2. Elle se perd facilement
3. Elle se reconstruit difficilement (beaucoup de temps).

Référencement : news et updates des derniers outils

Les annonces faites par Google via Mat Cutts - le chef technique du référencement de Google - pour parler un peu de référencement, de ses news et des nouveaux outils mis à disposition des webmasters.

Tout d'abord Google met à disposition des webmasters un outil indispensable pour valider et vérifier si le site est bien référencé: c'est le webmaster tool.

En revanche, ce dernier vient d'être amélioré et permet désormais, à partir de la zone "Diagnostics" et "Content analysis" de l'interface, d'obtenir des indications diverses sur votre site web:

- Titres manquants / Titres dupliqués
- Titres trop longs / Titres trop courts
- Titres pas assez descriptifs
- Balises meta description dupliquées / Balises meta description trop longues
- Balises meta description trop courtes
- Contenu non indexable (Flash, etc.)

Centre Google pour les webmasters

Google Centre pour les webmasters

Ce centre regroupe toutes les informations sur la façon dont Google explore et indexe les sites Web. Apprenez à créer un site que Google est en mesure d'explorer et d'indexer facilement. Vous trouverez également des outils conçus pour identifier les problèmes liés à l'exploration et pour consulter les statistiques relatives à l'indexation de votre site. Enfin, vous pourrez nous faire part de vos souhaits en matière d'exploration et d'indexation des sites.



Assistant d'état du site

Déterminez si Google procède actuellement à l'indexation de votre site.



Blog Google pour les webmasters (anglais)

Les toutes dernières informations sur les techniques d'exploration et d'indexation des sites Web utilisées par Google.



Outils pour les webmasters (notamment Google Sitemaps)

Statistiques, diagnostics et gestion de l'exploration et de l'indexation par Google de votre site Web (avec accès aux fonctionnalités d'envoi de plans Sitemap et de création de rapports Sitemap).



Groupe de discussion Google pour les webmasters

Discutez avec vos confrères et communiquez-nous vos commentaires.



Envoi de votre contenu à Google

En savoir plus sur la façon de transférer du contenu vers des sites Google tels que Google Base et Google Recherche de Livres.



Centre d'aide des webmasters

Consultez les réponses aux questions les plus fréquentes sur l'exploration, l'indexation, le classement et autres tâches de l'administrateur Web.

Centre Google pour les webmasters

Google Outils pour les webmasters

[Aide](#)

[État du site](#) » Étape 1 sur 2

Cette procédure détaillée vous permet d'en savoir plus sur l'état de votre site dans l'index Google. Pour plus d'informations sur la manière dont Google explore et indexe votre site, connectez-vous à votre compte [Outils Google pour les webmasters](#). Ce service est gratuit.

À propos de quel site souhaitez-vous obtenir des informations complémentaires ?

Entrez l'URL de votre site ci-dessous :

Exemple : <http://www.google.com>

Centre Google pour les webmasters

Google Outils pour les webmasters

[État du site](#) > Étape 2 sur 2

<http://www.immobilierejabbes.com/>

- ✓ Des pages de votre site figurent dans l'index Google. ⓘ
- ✓ Le robot Googlebot est parvenu à accéder à votre page d'accueil.

Problèmes d'indexation éventuels :

Nous n'avons pas indexé toutes les pages de votre site. Vous pouvez nous envoyer un sitemap pour nous donner davantage d'informations concernant votre site.

Possibilité de découvrir de nouvelles informations relatives à votre site

En utilisant les outils Google pour les webmasters, vous pouvez obtenir des informations réservées aux propriétaires de sites. Par exemple :

- erreurs que Googlebot a rencontrées lors de l'exploration de votre site
- requêtes les plus fréquentes qui renvoient votre site

[Découvrir les outils Google pour les webmasters](#)

[Non merci](#)

Conclusion

- Le Référencement touche les moteurs et les annuaires
- La soumission se fait de façon manuelle pour les moteurs les plus connus et par logiciel pour le reste des moteurs pour augmenter la popularité du site et améliorer les résultats dans les Meta Moteurs
- Il faut penser à exploiter les moteurs thématiques
- Le contenu doit être optimisé : texte, images, vidéos, url, nom de fichier, balises, nom de domaine
- Bien choisir les mots clés : selon le trafic, l'utilisation et le contenu : long tail
- Vérifier l'existence des fichiers : robots, Allway, sitemap, .htaccess
- Vérifier de façon continue la présence de l'intégralité du site au niveau des différents serveurs des outils de recherche
- Assurer une suivie permanente des statistiques de visite et du trafic

Conclusion (suite)

- Communiquer de façon continue sur le site : on line et off line
- Augmenter la popularité du site et des différents pages
- Veuillez à réaliser de façon continue des analyses concurrentiels du trafic
- Veuillez à réaliser de façon continue des analyses de la visibilité du site
- Suivi du positionnement et du ranking
- Eviter le spamdexing
- Tenir compte des spécificités des moteurs, de leurs historiques et de leurs complémentarités et divergences : googledance, sandbox, load balancing
- Utiliser les différents outils mis par les moteurs pour faciliter l'indexation
- Suivre de près les nouveautés et les news des différents moteurs et annuaires

Référencement & Positionnement sur le WEB

Merci pour votre attention