

Les techniques de référencement

Les techniques de référencement.....	1
Structure du site.....	3
Hébergeur	3
Nom De Domaine	3
Sous-domaines	3
Répertoires	3
Nom de page	3
Nombre de pages	4
Sitemap	4
Robots.txt	5
Structure des pages	7
Dans le HEADER.....	7
Doctype & Charset.....	7
Titre	7
Meta Description	8
Meta Keywords	9
Dans le BODY	9
Titres (Hx).....	9
Images	10
Liens Internes	10
Liens Externes	10
Contenu	11
Techniques "illégalles" (Black Hat)	12
Spamdexing :	13
Pages satellites	13
Texte caché (JS / CSS / Frames)	13
Cloaking.....	14
Link farms.....	14
Risques	15
Résumé	16

Structure du site

Hébergeur

Il est préférable que le serveur qui héberge le site soit localisé dans le même pays que les visiteurs ciblés dans les cas d'une extension non spécifique au pays : .com, .net, .biz etc. Il faut aussi éviter les serveurs lents ou qui tombent régulièrement en panne pour ne pas faire fuir les visiteurs.

Nom De Domaine

L'importance du mot-clef dans l'URL est de moins en moins décisive. Un terme mémotechnique ou un nom percutant est à privilégier à un nom de domaine à tiret du type www.ecole-ingenieur-paris.com. Le nom de domaine vous donnera un avantage certain sur un mot clef spécifique, mais la reconnaissance d'une marque est plus intéressante.

Cependant, la présence de mots-clés dans le nom de domaine est un critère pris en compte par les moteurs de recherches. Il faut dans ce cas choisir essayer de coller au plus près avec le thème du site.

Il ne faut pas que le nom soit trop long, ni trop compliqué à retenir. Il est préférable de ne pas dépasser les 3 mots dans le nom de domaine et si possible avoir une extension correspondant au pays ciblé (.fr pour la France), le .net, .org, .com, etc. sont aussi valables.

Sous-domaines

Il est d'usage de séparer les thèmes distincts en sous-domaines portant le mot-clé du thème traité.

Les sous-domaines sont presque considérés comme des sites indépendants les uns des autres.

Ex : <http://si.ece.fr/> - <http://tr.ece.fr/> - <http://se.ece.fr/>

Répertoires

On procédera à l'organisation d'un thème en répertoires portant le mot-clé du sous-thème traité

Ex : <http://www.ece.fr/formations/mastere-specialise/>

Nom de page

Tout comme le nom de domaine, le nom de pages est très important. Chaque nom de page doit évidemment être différent, mais doit surtout contenir des mots-clés (3 maximum) séparés par des tirets (-).

Ex : <http://www.ece.fr/formations/ecole-d-ingenieur/cursus-et-etudes/cycle-preparatoire-integre.php>

Il faut faire attention aux variables transmises dans l'URL des pages dynamiques, elles peuvent parfois gêner leur indexation.

C'est le cas des **variables de session** PHP qui empêchent l'indexation par de nombreux moteurs de recherche.

C'est le principe des sessions : un identifiant de session est attribué automatiquement à chaque visiteur qui arrive sur un site utilisant les sessions, ce qui peut par exemple servir à spécifier une langue différente selon le visiteur et garder cette langue dans tout le site.

Si cet identifiant de session est transmis dans l'URL (sous forme de variable `id` ou `sessionid`), alors à chaque fois qu'un robot d'indexation viendra visiter un site, toutes les pages auront une nouvelle URL. Plus précisément, une même page aura à chaque visite une nouvelle URL.

Le crawler pourra alors interpréter cela comme du *duplicate content*, c'est-à-dire un plagiat de pages, qui peut entraîner une baisse dans le classement des moteurs, voir même un *blacklistage* (suppression du site des index du moteur).

La solution est soit de se passer de sessions, soit de transmettre l'identifiant de session par cookies.

Nombre de pages

Le nombre de page est un facteur relativement important. Tout d'abord, plus le nombre de pages (distinctes) est grand, plus la quantité d'informations est importante et donc plus les possibilités de trouver le site sont grandes.

Un Wikipédia contenant des milliers de pages affichera bien plus de résultats sur de très nombreuses requêtes qu'une simple page perso ; ce qui ne veut pas dire que la page perso sortira forcément après dans l'ordre des résultats !

Sitemap

Le sitemap ou plan de site permet d'accéder rapidement au contenu d'un site web. Les moteurs de recherche s'en servent pour indexer l'ensemble des URL proposées dans cette ressource web. Auparavant Yahoo! avait son propre sitemap, nommé "urllist.txt", qui proposait l'ensemble des URL du site sous forme d'un listing (une URL par ligne).

Désormais, Google, Yahoo! et Live utilisent un fichier identique nommé "sitemap.xml" qui propose l'ensemble des URL du site formaté selon des règles XML.

Le fichier XML standard comporte outre l'URL:

Une périodicité pour scanner la page, la date de dernière mise à jour, l'importance relative de la page par rapport aux autres pages du site.

Exemple simple de fichier Sitemap :

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

```
<urlset xmlns="http://www.google.com/schemas/sitemap/0.84">
```

```
  <url>
```

```
    <loc>http://www.ece.fr/formations/groupe-ece/demande-de-documentation.php</loc>
```

```

</url>
<url>
    <loc>http://www.ece.fr/espace/candidat/rencontrez-nous.php</loc>
</url>
<url>
    <loc>http://www.ece.fr/formations/groupe-ece/demande-de-documentation.php</loc>
</url>
</urlset>

```

Il existe de nombreux script et outils qui permettent de générer automatiquement ce fichier Sitemap.

Il suffit ensuite d'envoyer ce fichier sur le site de Google :

<https://www.google.com/webmasters/sitemaps>.

Il est conseillé, lorsqu'une modification est faite dans votre SiteMap, de le soumettre de nouveau à Google via l'url suivante :

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

Il est aussi indispensable de créer un plan de son site manuellement afin de permettre aux moteurs de recenser tous les liens du site et aux visiteurs de se retrouver dans le site, en particulier si tous les liens ne sont pas accessible directement depuis la page d'accueil.

<http://www.ece.fr/formations/groupe-ece/plan-du-site.php>

Robots.txt

Le fichier Robots.txt est un standard reconnu par la plupart des moteurs de recherche qui permet de leur indiquer une liste de documents ou de répertoires à ne pas indexer.

Ce fichier peut être utilisé pour restreindre les points d'entrée vers le site depuis les pages de résultats des moteurs de recherche. En effet, dans certains cas, laisser indexables certaines pages internes d'un site ne sera pas judicieux pour des raisons de communication ou d'ergonomie de la navigation.

Il peut aussi permettre d'empêcher l'indexation de documents non-publics, mais laissés en accès libre. (par exemple, des fichiers logs).

Enfin, il est aussi utilisé pour empêcher l'indexation de pages dont le contenu ne présente pas d'intérêt du point de vue du référencement.

Le fichier Robots.txt est toujours placé à la racine du site. (Ex :

<http://www.webrankinfo.com/robots.txt>).

La structure d'un fichier robots.txt est la suivante :

un dièse (#) sert de ligne de commentaire

#

La requête s'applique à tous les moteurs. Si un robot est spécifié à la place de l'étoile, alors seul ce robot n'aura pas accès aux pages et répertoires du sites énumérés.

Une liste des robots est disponible sur ce site : <http://www.robotstxt.org/wc/active.html>

User-agent: *

Le sous-répertoire sousrep contenu dans le répertoire rep ne sera pas visité par le crawler

Disallow: /rep/sousrep/

Le fichier.html directement sous la racine ne sera pas visité par le crawler.

Disallow: /fichier.html

Le site complet ne sera pas visité par le crawler.

Disallow: /

Autorise un accès sans restriction au site.

Disallow:

Un site ne souhaitant cacher aucun fichier ni répertoire aux robots n'a pas besoin de fichier robots.txt.

Structure des pages

Dans le HEADER

Doctype & Charset

Dans les normes du HTML et du XHTML, un DOCTYPE informe le validateur de la version de (X)HTML que utilisée, et doit apparaître en première position dans chaque page web. Les DOCTYPEs sont des composants primordiaux pour les pages conformes aux normes, ils sont essentiels pour que les documents web s'affichent et fonctionnent correctement dans des navigateurs conformes aux normes.

Ex : `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">`

De même il est important de définir le charset (encodage) de la page (iso-8859-1, utf-8 etc.).

Ex : `<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />`

Titre

La balise `<title>` est la balise la plus importante quant au référencement de la page. C'est en général le titre qu'affiche les moteurs de recherche lors de l'affichage des résultats.

`<title>Ecole d'Ingénieurs ECE Paris - Grandes Ecoles Post-Bac - Prépa Intégrée et Cycle Ingénieur</title>`



Web Images Groupes Actualités plus »

école d'ingénieur

Rechercher

Recherche avancée
Préférences

Rechercher dans : Web Pages francophones Pages : France

Web

Enseignement supérieur - liste des écoles d'ingénieurs habilitées

(1) Ecoles de spécialisation délivrant un diplôme d'ingénieur aux étudiants déjà titulaires d'un diplôme d'ingénieur. (2) Diplôme de spécialisation délivré ...

www.sup.adc.education.fr/supweb/etb/ei/ing.htm - 2k - [En cache](#) - [Pages similaires](#)

Ecole d'Ingénieurs ECE Paris - Grandes Ecoles Post-Bac - Prépa ...

Ecole d'Ingénieurs à Paris (ecole d'ingenieur), en informatique, électronique, systèmes embarqués, télécoms et réseaux, multimédia.

www.ece.fr/ - 11k - 26 nov 2006 - [En cache](#) - [Pages similaires](#)

Groupe ESIEE Paris - École Supérieure d'Ingénieurs - Accueil

"UN JEUNE, UN MÉTIER, UNE ÉCOLE" : Découvrez le métier d'ingénieur informatique - Ouverture d'une nouvelle filière Réseaux par la voie apprentissage à l' ...

www.esiee.fr/ - 9k - [En cache](#) - [Pages similaires](#)

Elle doit être constituée d'une ou deux phrases courtes contenant une dizaine de mots (environ 80 caractères) pour être optimale.

Il faut y mettre un maximum de mots-clés (en évitant les répétitions) en plaçant, si possible, les mots-clés les plus importants en début (le plus à gauche possible).

Il est donc préférable de mettre : `<title>Ecole d'Ingénieurs ECE Paris - ...</title>`

que : `<title>ECE Paris - Ecole d'Ingénieurs ...</title>` étant donné que le terme "Ecole d'Ingénieurs" est plus important que "ECE".

Cependant, nous ne devons pas oublier que le texte doit être accrocheur pour l'internaute ; si le titre est uniquement constitué de mots-clés, l'utilisateur sera plus tenté par un site situé en dessous dans les résultats avec une phrase claire et explicite.

Il est aussi important de mettre un titre différent sur chacune des pages du site afin d'élargir l'impact de cette optimisation.

Le titre de la page d'accueil doit être particulièrement bien étudié car c'est en général la page qui apparaît le plus souvent dans les résultats.

Meta Description

La balise `<meta name="Description" />` est aussi très important, bien que certains moteurs de recherche ne lui donne que peu de valeur.

Cette balise sert à donner une description de la page. Il est préférable d'y écrire une ou plusieurs phrases correctement construites qu'une série de mots-clés, tout en plaçant les mots-clés les plus importants le plus à gauche possible.

Ex : `<meta name="description" content="École d'Ingénieurs à Paris (ecole d'ingenieur), en informatique, électronique, systèmes embarqués, télécoms et réseaux, multimédia. Cycle préparatoire intégré; (prépa intégrée), et cycle ingénieur. L'ece est accessible post-bac sur concours, et à bac+1, bac+2, bac+3 et bac+4. Le group ECE propose également des mastères spécialisés." />`
C'est souvent la description affichée par les moteurs de recherche en dessous du titre sur les pages de résultat.

Il est conseillé de ne pas dépasser les 30 mots (200 caractères) et d'y placer un maximum de mots-clés tout en gardant des phrases syntaxiquement correctes, sur une seule ligne si possible.

Meta Keywords

La balise `<meta name="Keywords" />` est ignorée par de nombreux moteurs de recherche (Yahoo la prend encore en compte), mais ne doit pas non plus être négligée étant donné que certains moteurs l'utilisent encore.

Certains annuaires Web proposent de récupérer le contenu de cette balise; ils se basent sur ce contenu pour fournir les résultats de leurs moteurs internes. Il peut donc être utile de l'utiliser sur la page d'accueil uniquement

Elle doit contenir une liste de mots-clés séparés par des espaces ou des virgules. Il faut penser à mettre tous le champ lexical du site, en ajoutant les mots au pluriel, les groupes de mots, voir même, les fautes d'orthographe.

Il est tout de même conseillé de ne pas dépasser les 1000 caractères.

Il est préférable que ces mots soient sur une seule ligne.

Ex : `<meta name="keywords" content="Ecole d'Ingénieurs, Ecole d'Ingenieur, prépa, prépa intégrée, prépa intégré, etc. />`

Dans le BODY

Le contenu de la page est l'aspect le plus important.

Il ne faut jamais oublier que le site est fait pour l'internaute, et non pas pour les moteurs de recherche.

C'est le principal défaut de nombreux webmasters.

Il faut trouver le juste milieu entre l'optimisation de la page et sa lisibilité.

Une page remplie de mots-clés dans de liens et remplies de titres de type `<h1>` `<h2>`, ... ne sera pas appréciée par un internaute et il ne risque pas de revenir dessus.

Titres (Hx)

Les balises de titres (`<h1>`, `<h2>`, ..., `<hn>`) ont récemment perdu de leur importance, mais restent tout de même utiles.

Il faut faire des titres en concordance avec la structure de la page (ne pas mettre trois balises `<h1>` à la suite), et ne pas oublier que ces balises sont faites pour afficher un titre à un paragraphe ou à une page.

L'occurrence des mots-clés dans le texte de ces balises reste un facteur non négligeable.

Il est possible de mettre une image dans une balise titre.

Ex : `<h1></h1>`

Dans ce cas, le moteur verra ce qui est affiché dans la balise ALT (cf. paragraphe suivant).

Images

Les images sont aussi lues par les moteurs grâce à la balise ALT qui affiche un texte alternatif.

Ce texte s'affichera à la place de l'image si elle n'est plus disponible. Il est lisible par les moteurs de recherche et permet aux aveugle d'en avoir une idée.

Il est préférable que la balise ALT de l'image contienne des mots-clés, mais elle doit tout de même décrire l'image.

Le nom de l'image est aussi lu par les moteurs, c'est pourquoi il est nécessaire de bien le choisir.

Il est préférable de nommer son image "logo-ecole-ingenieur.jpg" que "logo.jpg" par exemple.

Liens Internes

Les liens internes sont les liens pointant vers d'autres pages du même site.

Il faut essayer de respecter le format suivant :

```
<a href="http://www.ece.fr/" title="École d'Ingénieurs">L'ECE est l'école d'ingénieurs qui crée la différence</a>
```

Le texte affiché doit contenir des mots-clés et décrire la page vers laquelle il pointe. Le champ title est une option mais n'apporte pas de preuves officielles quant à son impact sur le référencement.

Il faut éviter au maximum les URL à rallonges avec des paramètres du type:

```
www.ece.fr/index.php?id=45&page=hdjqhdjkqdhqjkqs?lang=fr
```

Il est dans ce cas fortement conseillé d'utiliser l'*URL rewriting* afin d'obtenir des noms de page plus explicite du type : `www.ece.fr/ecole-centrale-electronique.php`.

Utiliser pour cela le tiret (-) et non pas l'underscore (_) comme séparateur de mot (cf. chapitre sur le nom des pages).

Il est très important de bien lier toutes les pages du site afin d'éviter les pages orphelines (aucun moyen d'y accéder par un lien). Il ne faut pas oublier que le PR est redistribué entre les pages via les liens. Il peut donc être très intéressant de lier une page "profonde" (3 ou 4ème niveau de la page d'accueil) directement sur la page d'accueil afin de redistribuer plus de PR aux pages plus profondes.

Les moteurs de recherche ne lisent pas (encore?) le Javascript, il ne faut donc pas utiliser ce langage pour faire des liens vers vos pages internes.

Liens Externes

Les liens externes (liens d'autres sites pointant vers le votre) apportent une augmentation du PageRank et donc un meilleur classement dans les résultats de Google.

Afin d'optimiser la qualité de ces liens et donc d'augmenter le PR, il faut que le site qui pointe vers le votre ait un PR élevé.

De plus il est très apprécié que ce même site soit dans la même thématique (même sujet) que votre site ; Google associe des mots-clés à chaque site indexé, s'il trouve des termes communs entre les sites pointant vers le votre et le votre, ce lien aura plus de poids quant à votre classement.

Enfin, le nombre de liens présents sur le site "partenaire" joue un rôle important. Plus le site

comporte de liens, plus la diffusion de son PR sur ces liens sera diluée.

En conclusion, il est préférable d'obtenir un lien à partir d'un site traitant d'école d'ingénieurs ayant un PR4 et 5 liens externes, qu'un site à PR7 traitant de cuisine albanaise comportant 30 liens externes.

Plus le nombre de liens pointant vers un site est élevé, plus son PR est élevé, et donc mieux il sera classé.

Il faut aussi vérifier que le webmasters du site partenaire n'ait pas rajouté l'attribut `rel="nofollow"` dans le lien, car dans ce cas, les moteurs de recherche ne le prendront pas en compte, de même avec un lien en Javascript (non lu par les moteurs).

Il est conseillé d'éviter les fermes de liens (link farming), c'est-à-dire les pages remplies de 50 liens à la suite. Ces pages sont très "mal vues" par les moteurs de recherche et n'apporte que peu (ou pas) de PR, et de visiteurs.

Les mots clés présent sur le lien jouent un rôle primordial quant au référencement. C'est pourquoi il est important d'y placer des mots-clés et de les faire varier sur les sites partenaires afin d'étendre les possibilités de résultats.

Utiliser ce format de lien pour les partenaires : `L'ECE est l'école d'ingénieurs qui crée la différence`

ou dans le cas d'une bannière : ``

Les échanges de liens entre sites sont facilement détectables par les moteurs. Si le site A pointe vers les site B C D E F que ces sites ont tous un lien vers le site A, il y a de forte chance que certains moteurs se "rendent compte" du partenariat et ne feront donc pas bénéficier le site A d'un meilleur classement.

Il est donc préférable d'utiliser des techniques de partenariats a plusieurs.

Par exemple le site A fait un lien vers le site B ; le site B fera un lien vers un troisième site C qui lui fera un lien vers le site A. Cette méthode est pour l'instant difficilement détectable et permet de garder la valeur des liens.

Il ne faut cependant pas oublier que la principale cible n'est pas le moteur de recherche, mais l'internaute. Un lien situé au milieu d'une page, dans un contexte approprié sera beaucoup plus visité qu'un lien en pied de page ou sur un page partenaire.

Contenu

Le point le plus important à retenir, c'est qu'un site mis à jour régulièrement attirera plus souvent les visiteurs, ainsi que les moteurs.

Un site sans contenu n'a que peu d'intérêt, que ce soit pour un moteur de recherche ou pour un internaute.

A l'inverse, une page trop chargée fait diminué le poids des mots-clés. De plus elle fera fuir les visiteurs souvent très pressés.

En général il ne faut pas concevoir de page trop longue : il est conseillé de ne pas dépasser les 50 Ko sur la pages d'accueil et 80 Ko pour les autres pages.

Il existe quelques balises (X)HTML qui permettent de mettre en évidence du texte.

La balise `` sert à mettre légèrement en valeur un texte. Les navigateurs l'affichent en

italic.

La balise sert à mettre en avant un texte important. Les navigateurs l'affichent en **gras**.

Ces balises donnent un peu plus d'important aux mots contenus dedans aux yeux des moteurs de recherche qui savent les interpréter.

Cependant, il est (pour l'instant) inutile de mettre du texte en avant à l'aide des feuilles de style css (ex : texte en gras rouge et avec police de 26px) car les moteurs ne lisent pas les feuilles CSS.

Le nombre de mots-clés dans une page joue un rôle important. Il vaut mieux que la densité (le rapport entre le nombre de mots dans la page et le nombre de mots-clés) ne doit pas dépasser les 10% (certains disent 20%), mais ce qui compte le plus c'est que le texte reste cohérent pour les visiteurs. Il ne faut pas perdre de vue que le site est fait pour être lu par des internautes et non pas par un moteur.

Cependant, il ne faut pas se soucier de cette densité, elle reste un critère parmi d'autres. A condition de maîtriser les multiples autres critères, une densité de quelques % suffit amplement pour être éligible en mot-clé par les moteurs.

N'oublions pas que les moteurs de recherche ne lisent ni le Flash ni Javascript, il est donc préférable d'éviter de mettre du texte, des liens ou des redirections à l'aide de Javascript ou de faire un site 100% en Flash. Il vaut mieux utiliser Flash pour quelques animations isolées, mais surtout pas pour le menu principal, sinon les moteurs ne parcourront jamais la totalité du site. De même il est préférable d'éviter les frames, elles ne sont pas lues par les moteurs.

En ce qui concerne l'AJAX, il est à utiliser avec parcimonie également. En effet, le contenu de la page étant modifié après son chargement, il n'y a aucune garantie que les moteurs seront en mesure de s'adapter. Pour l'instant, ils n'interprètent pas du tout le langage Javascript, ce qui leur occulte la totalité du contenu affiché au moyen d'AJAX.

Il est donc judicieux d'appliquer à l'utilisation d'AJAX des règles similaires à celles adoptées pour Flash : ne l'utiliser que pour ce qui est le plus utile, pour améliorer l'expérience utilisateur. Il faut donc éviter de charger du contenu au moyen d'AJAX car les bots seront incapables de le référencer.

Il faut de même éviter les pages vides ou en construction : si le moteur indexe une page en construction (ou vide) , il ne risque pas de repasser de si tôt, or une indexation lente empêche la prise en compte des mises à jour du site par les moteurs.

Au sujet de la structuration il est préférable d'écrire d'entrée le sujet principal d'une phrase ou d'un paragraphe (on parle de "front-loading"). Cette recommandation est avant tout ergonomique, mais elle influence le référencement, en forçant à rédiger un contenu dense en mots-clés. Si on l'applique à une page, cette recommandation est d'autant plus intéressante, puisque elle force à placer les mots-clés en tout début du texte, ce qui est bénéfique au référencement, les moteurs favorisant le contenu placé haut dans le code. Il est donc important d'introduire le contenu, que ce soit au niveau de la page entière ou de ses sous parties.

Techniques "illégalles" (Black Hat)

Le terme illégal est un abus de langage étant donné qu'aucune loi n'interdit ces pratiques. Cependant les moteurs de recherche peuvent interpréter ces techniques comme "non conforme" à leurs principes et en cas de détection de telles méthodes mettre le site sur black list, c'est-à-dire retirer le site de la table d'index, et donc ne plus apparaître sur les résultats de recherche.

Voici les principales techniques utilisées par certains webmasters. Elles ne seront pas détaillées étant donné le risque qu'elles peuvent entraîner.

Spamdexing :

Le spamdexing consiste à tromper l'algorithme du moteur de recherche pour que le site reçoive davantage d'audience qu'il le mérite.

C'est en fait du référencement abusif destiné aux robots d'indexation de moteur de recherche qui consiste à modifier des pages Web en utilisant des mots-clés d'une façon abusive pour améliorer le classement dans les moteurs de recherche.

Ex : La mise en place de mots-clés invisibles.

L'ajout de mots-clés dans les méta tags sans aucun rapport avec la page.

La répétition de mots-clés (appelé aussi *bourrage de mots-clés*, en anglais *keywords stuffing*).

Le détournement de pages web (**pagejacking**).

L'utilisation abusive d'alias de domaine ou de sous domaines. La soumission de contenus sous différents noms de domaine n'est pas tolérée par la plupart des moteurs de recherche et des annuaires.

L'indexation intempestive de pages auprès des moteurs de recherche.

Certaines de ces techniques sont décrites ci-dessous.

Pages satellites

Les pages "satellites" contiennent un texte entièrement optimisé pour le référencement et n'ayant que peu de rapport avec celui de la page d'origine. Ce texte consiste le plus souvent en une répétition mécanique des termes et des synonymes des termes sur lesquels on souhaite obtenir un bon classement. Les pages satellites ne figurent pas forcément sur le site qu'elles sont censées promouvoir. Elles contiennent une redirection automatique qui renvoie le visiteur vers la page "visible" sans même qu'il s'en aperçoive. Le résultat obtenu est très proche de celui du cloaking. Il y a quelques années, une simple balise meta "refresh" permettait de gérer efficacement cette redirection. Aujourd'hui, cette technique est inefficace. Elle est généralement remplacée par une redirection en Javascript. Plusieurs sites utilisant des pages satellites redirigées en javascript ont toutefois été récemment pénalisés par Google, tandis que d'autres sont toujours en tête des classements.

Cette technique sera évidemment inopérante pour les utilisateurs qui ont désactivé l'exécution du javascript dans leur navigateur.

Texte caché (JS / CSS / Frames)

Le contenu étant un facteur important dans le référencement, il peut être utile d'en afficher en quantité sur le site. Il est possible de faire cela sans que celui-ci nuise à la lisibilité du site en le cachant aux visiteurs, mais non aux moteurs.

A l'aide de Javascript :

Le texte entré entre les balises `<noscript>` et `</noscript>` est caché grâce à Javascript. Il suffit de désactiver Javascript pour voir apparaître ce texte caché. Les moteurs n'utilisant pas ce langage voient donc le texte caché.

A l'aide des CSS :

Il est possible de mettre son texte dans une balise invisible à l'écran en cachant le contenu de cette balise via les CSS.

Le paramètre `display:none` par exemple n'affiche pas la balise du tout (pour un internaute, pas pour un moteur), de même on peut cacher le texte en le plaçant "hors de l'écran" avec `position:absolute` puis `left:-9999px` pour l'envoyer tout à gauche par exemple.

On pourrait aussi réduire la police du texte ou utiliser `overflow:hidden`, etc.

La technique d'afficher du texte de la même couleur que le fond de la page pour le masquer est facilement détectable par les moteurs de recherche.

Les moteurs ne lisent pas (encore ?) les feuilles de styles CSS, il est donc préférable pour ces méthodes "illégales" de mettre le code CSS dans une feuille de style à part et non dans le code HTML.

A l'aide de frames

Une autre technique consiste à utiliser les "frames" (cadres). Dans une page de ce type, il est en effet possible de définir une frame aux dimensions microscopiques ou située "hors écran" qui sera invisible par les internautes.

Cloaking

Le cloaking consiste à fournir au moteur de recherche une version différente du site par rapport à ce que voient les visiteurs.

Un site web qui utilise un langage côté serveur (PHP, ASP, CGI, SSI, ...) peut très facilement adapter (cloaker) la page qu'il renvoie en fonction des informations transmises dans la requête du Header HTTP.

C'est sur les informations `USER_AGENT` (navigateur) et `REMOTE_ADDR` (adresse IP) que repose le cloaking à destination des moteurs de recherche.

Si l'adresse IP permet en effet de déterminer dans quel pays se trouve le visiteur, elle permet aussi de reconnaître un robot d'indexation (Certains sites tiennent à jour la liste des IP des principaux robots d'indexation pour faciliter leur détection par les cloakers). Il ne reste plus qu'à générer des pages remplies de mots clés pour les présenter aux robots et leur faire croire que le contenu du site visité est d'une richesse extraordinaire.

Il faut néanmoins se méfier de ne pas faire de cloaking sans s'en rendre compte.

Par exemple, un site voulant s'afficher dans une langue différente selon l'origine des utilisateurs via leur adresse IP s'affichera différemment si l'IP source vient d'Espagne ou des États-Unis.

Étant donné que les bots de Google sont presque tous aux EU, il n'indexera jamais la version française du site. De plus il pourra interpréter cela comme du cloaking, car sur la même URL il risque de voir 2 versions différentes du site. Attention aux Blacklistage !

Link farms

Les fermes de liens ont pour but de construire un réseau de sites qui s'échangent des liens, de manière à augmenter leur quantité de backlinks.

Risques

Google a mis en place un service de dénonciation destiné aux webmasters qui s'estiment lésés par les pratiques de certains de leurs concurrents ou aux utilisateurs qui s'estiment trompés par le contenu de certains sites : <http://www.google.com/intl/fr/webmasters/spamreport.html>

Résumé

Voici un petit résumé énonçant les 10 points les plus importants quant au référencement d'un site :

- 1) Contenu
- 2) Nombreux liens externes
- 3) Liens externes de qualité
- 4) Bons titres
- 5) S'inscrire dans les grands annuaires
- 6) Pas de frames, et une bonne adresse de site
- 7) Un site toujours disponible
- 8) Interconnexion de vos pages
- 9) Pas de site sans texte
- 10) Mises à jour régulières du site