



<http://www.GnoZtiK.com>

Le guide du référencement naturel

version 1.0 (1 juin 2009)

Ce document (01/06/09) est sous licence **Creative Commons**.

<http://creativecommons.org/licenses/by-nc-nd/2.0/fr/>



Paternité - Pas d'Utilisation Commerciale - Pas de Modification 2.0 France

Vous êtes libres :



de reproduire, distribuer et communiquer cette création au public

Selon les conditions suivantes :



Paternité. Vous devez citer le nom de l'auteur original <http://www.gnoztik.com> qui vous confère cette autorisation (mais pas d'une manière qui suggérerait qu'il vous soutienne ou approuve votre utilisation de l'oeuvre).



Pas d'Utilisation Commerciale. Vous n'avez pas le droit d'utiliser cette création à des fins commerciales.



Pas de Modification. Vous n'avez pas le droit de modifier, de transformer ou d'adapter cette création.

- A chaque réutilisation ou distribution de cette création, vous devez faire apparaître clairement au public les conditions contractuelles de sa mise à disposition. La meilleure manière de les indiquer est un lien vers cette page web (<http://creativecommons.org/licenses/by-nc-nd/2.0/fr/>).
- Chacune de ces conditions peut être levée si vous obtenez l'autorisation du titulaire des droits sur cette oeuvre.
- Rien dans ce contrat ne diminue ou ne restreint le droit moral de l'auteur ou des auteurs.

Sommaire

| | |
|--|----------|
| SOMMAIRE | 1 |
| AVANT-PROPOS..... | 6 |
| L'INDEXATION | 8 |
| L'INDEXATION, COMMENT ÇA MARCHE ? | 10 |
| Les crawlers | 10 |
| La soumission manuelle | 10 |
| Dois-je soumettre toutes les pages de mon site web ? | 10 |
| L'indexation naturelle | 10 |
| Comment me faire remarquer ? | 11 |
| L'indexation par le Google Toolbar | 11 |
| L'INDEXATION, OU VOTRE EXISTENCE SELON GOOGLE | 12 |
| Votre site web existe-t-il selon Google ? | 12 |
| Comment connaître le niveau d'indexation de mon site web ? | 12 |
| La commande « site: » de Google | 12 |
| La commande « site: » de Yahoo | 13 |
| VUE D'ENSEMBLE DES FACTEURS DECISIFS..... | 14 |
| Les facteurs bloquants l'indexation | 14 |
| Les facteurs favorisant l'indexation | 14 |
| LES FACTEURS BLOQUANTS | 15 |
| Les technologies à proscrire | 15 |
| Le format Flash | 15 |
| Le Javascript | 15 |
| Les Frames | 15 |
| Les facteurs bloquants au niveau du site..... | 16 |
| Le maillage..... | 16 |
| La page d'accueil | 16 |
| Le plan du site..... | 16 |
| Les facteurs bloquants au niveau des pages..... | 17 |
| Le nombre de liens | 17 |
| Le type des liens..... | 17 |
| Le Javascript | 17 |
| L'url rewriting..... | 17 |
| Le gabarit des pages..... | 17 |
| LES FACTEURS FAVORISANTS | 18 |
| Le sitemap | 18 |
| Le sitemap, Kezako ? | 18 |
| Petit rappel sur les crawlers..... | 18 |
| Le protocole sitemaps (avec un « s ») | 18 |
| Le fichier sitemap (sans « s ») | 18 |
| Les limites du sitemap | 19 |
| Comment le fichier sitemap aide mon site web ? | 19 |
| Est-ce qu'un fichier sitemap est indispensable à un bon référencement ? | 19 |
| Le fichier sitemap ne sert à rien, alors ? | 19 |
| Est-ce qu'un fichier sitemap améliore mon positionnement dans les moteurs de recherche ? | 19 |
| Le fichier sitemap en détail | 19 |
| Le sitemap, ça ressemble à quoi ? | 19 |
| Les différentes balises de votre fichier sitemap | 20 |
| Urlset | 20 |
| Url | 20 |
| Loc | 20 |
| Lastmod | 20 |
| Changefreq..... | 21 |
| Priority | 21 |
| Qu'est-ce que je fais de mon fichier sitemap ? | 21 |
| Le fichier robots.txt | 22 |
| Le fichier robots.txt, c'est quoi ? | 22 |
| « Robots Exclusion Protocol », kezako ? | 22 |

| | |
|---|-----------|
| Pourquoi s'embarrasser d'un fichier robots.txt ? | 22 |
| Mise en garde..... | 22 |
| Comment rédiger mon fichier robots.txt ? | 22 |
| La directive « User-agent » | 23 |
| La directive « Disallow » | 23 |
| Exemples..... | 23 |
| Les directives non standards..... | 24 |
| Qu'est ce que je fais de mon fichier robots.txt ? | 24 |
| L'EFFET SANDBOX..... | 25 |
| La SandBox, kezako ? | 25 |
| A quoi sert la SandBox ? | 25 |
| Mon site est-il dans la SandBox ? | 25 |
| Quel impact a la SandBox sur mon site web ? | 26 |
| Comment sortir de la SandBox ?..... | 26 |
| Votre site web est récent..... | 26 |
| Votre site web est ancien | 26 |
| Quelles mesures concrètes pour sortir de la SandBox ? | 26 |
| Mise en garde | 26 |
| LES MOTS-CLES | 27 |
| TYPOLOGIE DES MOTS-CLES | 29 |
| Un mot-clé concurrentiel, c'est quoi ? | 29 |
| Un mot-clé parasité, c'est quoi ?..... | 30 |
| Un mot-clé délaissé, c'est quoi ?..... | 31 |
| Un mot-clé de niche, c'est quoi ?..... | 32 |
| Comment faire concrètement pour déterminer le type d'un mot-clé ? | 33 |
| COMMENT CHOISIR DES MOTS-CLES EFFICACES ?..... | 35 |
| Ciblez des mots-clés liés à votre secteur d'activité | 35 |
| Choisissez précisément vos mots-clés | 35 |
| Préférez des mots-clés composés de plusieurs mots | 35 |
| Multipliez les bonnes combinaisons..... | 35 |
| Usez de synonymes..... | 36 |
| Optez pour des mots-clés à votre mesure | 36 |
| Visez les mots-clés les plus rentables..... | 36 |
| Soignez pragmatique | 36 |
| LA LONGUE TRAÎNE..... | 38 |
| La longue traîne, kezako ? | 38 |
| Longue traîne et longueur de mots-clés | 39 |
| La longue traîne, trafic et concurrence..... | 40 |
| Longue traîne et typologie des mots-clés | 40 |
| Comment se traduit la longue traîne pour mon site web ?..... | 41 |
| Comment tirer profit de la longue traîne ?..... | 41 |
| DES STATISTIQUES PRECISES POUR VOS MOTS-CLES | 42 |
| Suggestion de mots-clés à partir d'un ou plusieurs mots-clés | 42 |
| Suggestion de mots-clés à partir d'une page web | 43 |
| L'OPTIMISATION DU CONTENU | 45 |
| DEUX PRINCIPES CLES DE L'OPTIMISATION | 47 |
| Décompte | 47 |
| Densité..... | 47 |
| Exemple détaillé | 47 |
| Mise en garde | 47 |
| L'OPTIMISATION DES URL..... | 49 |
| Les urls..... | 49 |
| Anatomie d'une url..... | 49 |
| Décompte et densité dans une url | 49 |
| Décompte..... | 49 |
| Densité..... | 49 |
| Problème | 50 |

| | |
|--|----|
| Optimisation | 50 |
| Le nom de domaine..... | 50 |
| Hébergement gratuit ou payant ? | 50 |
| Choix du nom de domaine | 51 |
| Le TLD..... | 52 |
| Les sous-domaines..... | 52 |
| Les répertoires et les fichiers..... | 52 |
| Comment concrètement mettre en place cette organisation ? | 52 |
| Pourquoi m'embarrasser à organiser mes contenus ? | 53 |
| Les moteurs de recherche | 53 |
| Les internautes | 53 |
| Vous-même | 53 |
| L'url rewriting..... | 53 |
| Pourquoi bricoler mes urls ?..... | 53 |
| L'url rewriting, comment ça marche ? | 54 |
| Comment lire une URL à réécrire ? | 54 |
| Comment mettre en place la réécriture d'une URL ? | 55 |
| Activez d'abord la réécriture d'url sur votre site web | 55 |
| Identifiez les urls à réécrire | 55 |
| Comment définir mes règles de réécriture ? | 55 |
| Comment mettre en application mes règles de réécriture ? | 56 |
| Apache | 56 |
| Microsoft IIS | 57 |
| Vérifiez les liens hypertextes de votre site web | 57 |
| L'OPTIMISATION DU CONTENU HTML | 58 |
| Le HTML pour le référencement | 58 |
| HTML, le minimum..... | 58 |
| HTML, Kezako ? | 58 |
| Structure d'une page HTML | 58 |
| Aller plus loin..... | 59 |
| Le tag Title | 59 |
| Le tag Title, Kezako ? | 59 |
| Le tag Title, selon la norme HTML 4.01 | 60 |
| Le tag Title est-il vraiment important ? | 60 |
| Comment dois-je rédiger le tag TITLE de mes pages web ? | 61 |
| Les incontournables..... | 61 |
| Une balise TITLE différente pour chaque page | 61 |
| Un titre intelligible..... | 61 |
| Les mots-clés..... | 61 |
| La taille | 62 |
| Aller plus loin | 62 |
| En tête de fichier..... | 62 |
| La forme du titre..... | 62 |
| Mise en garde..... | 62 |
| Le tag Meta/Description..... | 63 |
| Le tag Meta/Description, Kezako ? | 63 |
| Le tag Meta/Description, selon la norme HTML 4.01..... | 63 |
| Le Meta/Description est-il vraiment important ? | 64 |
| Une balise Meta/Description différente pour chaque page web..... | 65 |
| Une description intelligible plutôt qu'un bourrage de mots-clés..... | 65 |
| Des mots-clés pas si importants | 66 |
| La taille..... | 66 |
| Les caractères accentués..... | 66 |
| Le tag Meta/Keywords | 67 |
| Le tag Meta/Keywords, Kezako ? | 67 |
| Le tag Meta/Keywords, selon la norme HTML 4.01..... | 67 |
| Le tag Meta/ Keywords est-il vraiment important ? | 67 |
| Les tags H1 à H6..... | 68 |
| Les tags H1 à H6, Kezako ? | 68 |
| Les tags H1 à H6 selon la norme HTML 4.01 | 68 |
| Quelle importance ont les balises H1 à H6 ? | 70 |
| Les balises H1 à H6 dans le corps d'une page web | 70 |
| Les balises H1 à H6 entre elles | 70 |

| | |
|---|-----------|
| Des balises stratégiques pour le référencement | 70 |
| Les tags H1 à H6 et les autres tags HTML..... | 70 |
| Les tags H1 à H6 entre eux | 70 |
| Comment bien rédiger mes balises H1 à H6 ? | 71 |
| SEO & internautes : le bon compromis ! | 71 |
| Quelle taille pour les balises H1 à H6 ? | 71 |
| Les images (IMG) | 71 |
| Les images selon la norme HTML 4.01 | 71 |
| Quelle importance ont les images ? | 72 |
| Comment dois-je construire les images de mes pages web ? | 72 |
| L'attribut ALT | 72 |
| Texte et Légendes environnantes..... | 72 |
| L'attribut SRC..... | 73 |
| Des images sans texte ! | 73 |
| Les liens (A)..... | 73 |
| Les liens, Kezako ?..... | 73 |
| Les liens selon la norme HTML 4.01..... | 73 |
| Quelle importance ont les liens ? | 74 |
| Comment dois-je construire les liens de mes pages web ? | 74 |
| Le texte | 74 |
| Un contenu ciblé..... | 74 |
| Densité et décompte avec modération | 74 |
| Une surface d'exposition démultipliée | 74 |
| Le texte du haut ! | 75 |
| Mise en forme mineure..... | 75 |
| LE DUPLICATE CONTENT | 76 |
| Pourquoi les moteurs de recherche luttent contre le duplicate content ? | 76 |
| Le duplicate content, c'est quoi ?..... | 76 |
| Comment les moteurs de recherche élisent la page web qui figurera dans les SERP ? | 76 |
| Comment éviter un duplicate content à mes pages web ? | 76 |
| Masquez vos contenus « en double » aux moteurs de recherche..... | 76 |
| Signalez à Google, Yahoo et Microsoft Bing vos urls canoniques ! | 77 |
| Exigez un backlink des sites web qui diffusent votre contenu | 77 |
| LE NETLINKING | 78 |
| LES BACKLINKS | 80 |
| Les backlinks, Kezako ?..... | 80 |
| Définition d'un lien hypertexte..... | 80 |
| Un backlink, c'est quoi ? | 80 |
| Comment connaître mes backlinks ? | 81 |
| La commande « link: » de Google | 81 |
| La commande « link: » de Yahoo | 81 |
| La commande « linkdomain: » de Yahoo..... | 82 |
| Les backlinks, à quoi ça sert ? | 83 |
| Plus de backlinks, c'est plus de visiteurs..... | 83 |
| Plus de backlinks, c'est être mieux placé dans les moteurs de recherche..... | 83 |
| A qui demander un backlink ? | 83 |
| Le backlink doit être indexé | 83 |
| Pas trop de liens sortants | 84 |
| Le thème doit être proche | 85 |
| Un PageRank proportionné | 85 |
| Quel lien pour un bon backlink ? | 85 |
| Un backlink « en dur » sinon rien !!!..... | 85 |
| Méfiez-vous des liens avec attribut « rel=nofollow » | 85 |
| Pas de backlink en bas de page..... | 86 |
| Variez les pages web ciblées par vos backlinks | 86 |
| Quel texte pour mes backlinks ? | 86 |
| Des mots-clés stratégiques | 86 |
| Des mots-clés variés | 86 |
| 2 à 3 mots au minimum | 87 |
| Attention aux fautes d'orthographe..... | 87 |
| Quelle stratégie adopter pour accroître mes backlinks ? | 87 |

| | |
|--|------------|
| Les annuaires | 87 |
| Le link ninja | 88 |
| Le link baiting | 88 |
| Link Ninja contre Link Baiting | 89 |
| PAGERANK & TRUSTRANK | 90 |
| Le PageRank | 90 |
| Le PageRank en clair | 90 |
| Le PageRank, c'est quoi ? | 90 |
| Le PageRank, ça sert à quoi ? | 90 |
| Comment est évalué le PageRank ? | 90 |
| Comment connaître mon PageRank ? | 91 |
| Avec la Google Toolbar | 91 |
| Par certains sites web | 91 |
| Comment augmenter mon PageRank ? | 92 |
| Logarithmique, Kesako ? | 92 |
| C'est quoi un bon PageRank ? | 92 |
| Le TrustRank | 92 |
| Le TrustRank en clair | 92 |
| Confusion autour du TrustRank | 92 |
| Le TrustRank, c'est quoi ? | 93 |
| Le TrustRank, ça sert à quoi ? | 93 |
| Comment est évalué le TrustRank ? | 93 |
| Comment connaître mon TrustRank ? | 93 |
| Comment augmenter mon TrustRank ? | 93 |
| LE LOGICIEL GRATUIT GNOZTIK | 94 |
| PRESENTATION | 95 |
| ANALYSE DETAILLEE | 95 |
| L'onglet Classement | 95 |
| Les données brutes | 95 |
| Les données résumées | 96 |
| Le contexte général | 97 |
| L'ambiance HTML | 98 |
| L'onglet Url | 98 |
| Les principaux indicateurs | 98 |
| Peut-on y croire ? | 99 |
| Conseils et pronostics | 100 |
| L'ambiance HTML | 100 |
| L'onglet HTML | 101 |
| Le différentiel HTML | 101 |
| Les mots-clés absents | 101 |
| Les mots-clés sous-représentés | 102 |
| La taille des mots des Tagclouds | 102 |
| Les effort et gain estimés | 102 |
| L'onglet Backlinks | 103 |
| Les effort et gain estimés | 103 |
| Les backlinks possibles | 103 |
| CONCLUSION | 105 |
| GLOSSAIRE | 106 |

Avant-propos

Le contenu de cet ebook est en grande partie issu du blog du logiciel **GRATUIT** de référencement GnoZtiK (<http://blog.gnoztik.com>). Le contenu a été revu, corrigé, mis à jour et étoffé de sorte à en faire un manuel concret d'apprentissage des techniques de référencement. Nous en sommes à la version 1.0. Des corrections, mises à jour, compléments et d'éventuels remaniements sont à prévoir dans l'avenir. Vos remarques, suggestions, corrections nous sont donc précieuses. N'hésitez pas à nous les communiquer à l'adresse <http://www.gnoztik.com/ebook/v1/feedback.php>.

Cette publication s'adresse à tous. Elle ne nécessite aucune connaissance au préalable en référencement. Elle se fixe deux objectifs : l'un pratique et l'autre théorique. L'objectif pratique consiste à vous **transmettre un savoir-faire opérationnel**. Au terme de cet ouvrage, vous devriez concrètement être en mesure de référencer votre site web dans les règles de l'art. Vous saurez le rendre accessible aux moteurs de recherche, cibler correctement vos mots-clés, optimiser vos pages web pour ces mots-clés et développer des réseaux de liens efficaces. Vous saurez ce qu'il ne faut pas faire, ce qu'il est bon de faire et dans quel ordre procéder. L'objectif théorique vise à vous donner un **socle solide et ample de connaissances en référencement**. Au terme de cette lecture, vous pourrez suivre et comprendre l'évolution des algorithmes des moteurs de recherche. Vous aurez acquis suffisamment pour mener à bien votre propre veille des techniques SEO. Les mots indexation, link ninja, duplicate content, longue traîne, optimisation html, backlinks et bien d'autres vous seront devenus familiers. Vous serez enfin autonome en matière de référencement.

Pas de pérennité hors du référencement dit « naturel ». Les techniques dites « black hat seo » sont vaines. Dans la durée, elles sont vouées à l'échec. Elles procurent parfois des avantages à court terme ... contre de sérieuses déconvenues dans la durée. Dès à présent, nous vous invitons à vous en écarter. Nous ne traiterons ici que de techniques reconnues saines, apportant des solutions pérennes à votre recherche de trafic qualifié.

Ce document se divise en quatre parties ordonnées dans le sens habituel d'une démarche référencement auxquelles s'ajoute une cinquième partie présentant le logiciel **GRATUIT** de référencement GnoZtiK (<http://www.gnoztik.com>) ainsi qu'un glossaire apportant des définitions aux jargons des référenceurs. La première partie aborde l'indexation, phase incontournable. L'indexation vise à faciliter le recensement de vos pages web par les moteurs de recherche. Si elles ne sont pas correctement recensées, n'escomptez pas obtenir du trafic en provenance des moteurs de recherche. Dans la deuxième partie, nous nous concentrons sur les mots-clés à cibler. Le choix des mots-clés à cibler est capital. Vous devez parler le langage de ceux que vous visez. Négligez cette étape majeure et vous vous condamnez à l'échec. Cette étape conditionne pour une large part la rédaction des contenus de votre site web. Dans la troisième partie, nous plongeons dans l'aspect le plus (mal ?) connu du référencement : l'optimisation des contenus. L'optimisation est une étape nécessaire mais elle ne constitue pas tout le référencement. Enfin dans la dernière partie, nous nous focalisons sur le NetLinking. Cette étape n'est pas toujours nécessaire. Sur certains marchés très ciblés, une optimisation efficace des contenus est parfois suffisante. Par contre, dès que la concurrence se fait sentir, le NetLinking devient une arme redoutable et la plupart du temps décisive.

GnoZtiK cherche une ou plusieurs personnes pour traduire cet ebook en anglais (éventuellement en d'autres langues aussi). Nous voulons, bien entendu, une traduction complète et de qualité. Une fois celle-ci réalisée, chaque contributeur aura droit à une mention, probablement en pied de page (à discuter), de son site web sur les pages qu'il aura

traduite. Si vous êtes intéressé, contactez nous via l'adresse email <http://www.gnoztik.com/ebook/v1/translation.php>.

Vous êtes nombreux à utiliser régulièrement le logiciel **GRATUIT** de référencement GnoZtiK (<http://www.gnoztik.com>). Depuis ses débuts, nous n'avons eu que de très bons retours. Viens maintenant cette publication **GRATUITE** que nous espérons de qualité et qui constitue un travail assez conséquent. Nous vous invitons donc à soutenir nos initiatives par un don Paypal (https://www.paypal.com/cgi-bin/webscr?cmd=s-xclick&hosted_button_id=5356201) du montant de votre choix. Par avance, merci.

Bonne lecture !

GnoZtiK Team

Le 1 juin 2009

L'Indexation

Pas de bon référencement sans bonne indexation !

L'indexation vise à faciliter le recensement de vos pages web par les moteurs de recherche. Si celles-ci ne sont pas recensées, n'escomptez pas être bien positionné sur vos mots-clés.

Son action est défensive car elle a pour principale tâche de lever les obstacles susceptibles de contrarier ce recensement. Elle s'articule autour de trois axes :

- multiplier les points d'entrée sur votre site web,
- faciliter la circulation entre les pages de votre site web (maillage interne),
- interdire l'accès à certaine partie de votre site web (partie privée, duplicate content, etc.).

Le premier chapitre de cette partie s'efforce d'expliquer comment se déroule l'indexation de vos pages web dans les moteurs de recherche. Une bonne connaissance de ces mécanismes permet de comprendre la logique des actions à entreprendre pour être mieux indexé.

Le deuxième chapitre s'attardera sur le moyen de déterminer où en est votre indexation dans les différents moteurs de recherche. Vous pourrez ainsi savoir si votre site web est correctement indexé et éventuellement décider d'entreprendre quelques actions correctrices, si ce n'est pas le cas.

Dans le troisième chapitre, nous vous donnerons une vue d'ensemble des facteurs décisifs à une bonne indexation. Nous serons alors au cœur du sujet. Nous y mentionnerons notamment les facteurs bloquants et les facteurs favorisants que nous développerons ensuite dans les deux chapitres suivants. Nous nous arrêterons notamment sur la technologie sitemap et les finesses du fichier robots.txt.

Enfin, dans le dernier chapitre, nous nous attarderons sur l'effet SandBox. Cet effet ne relève pas exactement de l'indexation mais les sites jeunes en sont souvent victimes à un moment où se pose le problème de l'indexation.

ATTENTION, ne négligez pas cette dimension du référencement.

L'indexation est **OBLIGATOIRE**.

L'indexation est nécessaire mais pas suffisante.

L'indexation, comment ça marche ?

Les crawlers

Un crawler est un robot (aussi appelé bot ou spider) qui parcourt le web et récupère automatiquement les pages web visitées. Les moteurs de recherche ont des crawlers qui se chargent du travail d'indexation des pages web.

*Le crawler de Google s'appelle **GoogleBot** alors que celui de Yahoo s'appelle **Yahoo Slurp**. Quant au crawler de Microsoft Bing (successeur de Microsoft Live Search), il s'appelle **MSNBot**.*

A chaque fois qu'un crawler visite une page web, il en fait l'analyse, en extrait les liens vers d'autres pages web et les indexe dans sa base de données.

Connaître leur fonctionnement, permet de lever les obstacles susceptibles de freiner l'indexation de votre site web. Le but ici est de leur faciliter la tâche.

La soumission manuelle

La manière la plus conventionnelle de signaler votre site web à un moteur de recherche est la soumission manuelle. Ci-dessous les adresses des formulaires de soumission pour :

| | |
|--------|---|
| Google | http://www.google.fr/addurl/ |
| Yahoo | http://fr.docs.yahoo.com/info/ajouter.html |
| Bing | http://www.bing.com/docs/submit.aspx |

Il n'y a plus qu'à remplir les formulaires correspondant en y indiquant l'url des pages de votre site web.

Dois-je soumettre toutes les pages de mon site web ?

Vous pouvez ne soumettre que votre page d'accueil (c'est ce que recommandent la plupart des moteurs de recherche sur leur formulaire de soumission). Dans ce cas, vous devrez compter sur le travail du moteur pour indexer le reste de vos pages web. En général, il ne tarde pas à la visiter, à en faire l'analyse, à en extraire les liens qui pointent vers l'intérieur (et également l'extérieur) de votre site web et à les indexer dans sa base. Si tout se passe bien, quelque temps plus tard, il fait de même avec les liens indexés à l'étape précédente, et ainsi de suite, jusqu'à crawler l'intégralité de votre site web. Avec un peu de patience, l'intégralité de votre site web sera ainsi crawlée.

Vous pouvez aussi soumettre toutes les pages de votre site web. Le travail est fastidieux mais vous aurez, en contrepartie, la certitude de leur indexation. Bon courage !

Faites comme bon vous semble !

En fait, ne faites ni l'un, ni l'autre !

L'indexation naturelle

Rien ne vaut une indexation naturelle. Nous verrons plus loin que ce n'est pas tout à fait vrai, qu'il est possible de parfaire l'indexation naturelle.

Le mieux reste d'être recommandé par un autre. Si un site web fait un lien vers le vôtre et que celui-ci est indexé par les moteurs de recherche, l'indexation de votre site web ne tardera pas à commencer. Il vous faut donc pour ce faire, vous faire remarquer par d'autres. C'est doublement efficace. D'une part, les liens signaleront votre site web aux moteurs de recherche de manière naturelle. D'autre part, ces mêmes liens vous octroieront un début de popularité. En effet, ces liens ne sont ni plus, ni moins que des backlinks.

Comment me faire remarquer ?

En publiant, par exemple, des articles sur d'autres sites populaires. En faisant des échanges de liens avec d'autres sites web thématiquement apparentés, en créant un buzz autour d'un événement, etc. Il existe de multiples méthodes qui ne sont pas vraiment l'objet de ce document. Ces méthodes relèvent souvent du web marketing.

L'indexation par le Google Toolbar

Google concentrant de nombreux intérêts, il est intéressant de noter le rôle particulier que peut avoir la Google Toolbar. Lorsque celle-ci affiche le PageRank d'une page, elle envoie à Google l'adresse de la page et la lui soumet indirectement par ce biais. En générant du trafic sur votre site web, vous augmenterez la possibilité d'avoir des visiteurs ayant la fameuse Toolbar. Ils soumettront alors, une à une, les pages de votre site à Google à leur insu. Ce n'est certes pas sa vocation première, néanmoins les spécialistes du référencement sont plutôt de cet avis.

L'indexation, ou votre existence selon Google

Votre site web existe-t-il selon Google ?

Inutile de faire de vains efforts pour vous référencer, si vous n'existez pas au sens des moteurs de recherche. Comme nous le disions en préambule à cette première partie, exister pour un moteur de recherche, ça veut dire être indexé dans sa base de données.

Comment connaître le niveau d'indexation de mon site web ?

La commande « site: » de Google

Pour connaître le nombre de pages indexées du site web <http://www.fnac.com> dans la base de données de Google, il suffit de taper dans Google **site: <http://www.fnac.com>**. ou **site: www.fnac.com**.

<http://www.google.fr/search?hl=fr&q=site%3Ahttp%3A%2F%2Fwww.fnac.com&btnG=Rechercher&meta=>



Apparemment, seules 11 200 pages sont indexées à la date du 30 avril 2009, ce qui est un peu surprenant pour un tel site web.

Le résultat diffère notablement en tapant **site:fnac.com**.

<http://www.google.fr/search?hl=fr&q=site%3Afnac.com&btnG=Rechercher&meta=lr%3D>



Plus de 2 000 000 de pages indexées à la date du 30 avril 2009. C'est pour le moins fameux !

Pourquoi cette différence ?

Remarquez en deuxième position une page issue d'un des sous-domaines. En fait, cette commande inclut l'ensemble des sous-domaines à la différence de la première qui ne se préoccupe que du seul sous-domaine **www**. Pour les sites web qui usent abondamment des sous-domaines, la différence peut donc se révéler très importante.

La commande « site: » de Yahoo

Yahoo propose la même commande que Google.

Pour connaître le nombre de pages indexées du site web <http://www.fnac.com> dans la base de données de Yahoo, il suffit de taper dans Yahoo **site:<http://www.fnac.com>**. ou **site:www.fnac.com**. Vous serez alors redirigé vers Yahoo Site Explorer.

<http://siteexplorer.search.yahoo.com/fr/siteexplorer/search?p=http%3A%2F%2Fwww.fnac.com&bwm=p&bwm=p&fr=yfp-t-501&fr2=seo-rd-se>

The screenshot shows the Yahoo Site Explorer interface for the URL <http://www.fnac.com>. The search bar contains the URL, and the 'Explorer l'URL' button is visible. On the left sidebar, there are links for 'Ajouter à Mes Sites', 'Préférences', 'Blog (US)', 'Module', 'API (site US)', and 'Commentaires (US)'. The main results area shows 'Résultats' with a lightbulb icon and a message: 'Astuce : Votre requête "site/domain:" a été redirigée vers Site Explorer. Vous pouvez continuer à utiliser le moteur de recherche.' Below this, it displays 'Pages (287 791)' and 'Liens entrants (145 959)'. The 'Montrer les pages de' section is set to 'Tous les sous-domaines'. The 'Détails des résultats' section shows three items: 1. Fnac : DVD, informatique, jeux vidéo, logiciel, CD, GPS, livres ... (text/html <http://www.fnac.com/> - 85k - en cache), 2. Livre, BD, manga et roman à découvrir et à acheter en ligne sur Fnac.com (text/html <http://www.fnac.com/livres.asp> - 84k - en cache), and 3. Musiques, CD, disques et nouveaux albums à découvrir et à acheter en ... (text/html <http://www.fnac.com/disques.asp> - 80k - en cache).

Presque 300 000 pages indexées à la date du 30 avril 2009.

Pour les mêmes raisons que Google, le résultat diffère aussi en tapant **site:[fnac.com](http://www.fnac.com)**.

<http://siteexplorer.search.yahoo.com/fr/siteexplorer/search?p=http%3A%2F%2Ffnac.com&bwm=p&bwm=p&fr=yfp-t-501&fr2=seo-rd-se>

The screenshot shows the Yahoo Site Explorer interface for the URL <http://fnac.com>. The search bar contains the URL, and the 'Explorer l'URL' button is visible. On the left sidebar, there are links for 'Ajouter à Mes Sites', 'Préférences', 'Blog (US)', 'Module', 'API (site US)', and 'Commentaires (US)'. The main results area shows 'Résultats' with a lightbulb icon and a message: 'Astuce : Votre requête "site/domain:" a été redirigée vers Site Explorer. Vous pouvez continuer à utiliser le moteur de recherche.' Below this, it displays 'Pages (5 101 336)' and 'Liens entrants (687)'. The 'Montrer les pages de' section is set to 'Tous les sous-domaines'. The 'Détails des résultats' section shows three items: 1. Fnac : DVD, informatique, jeux vidéo, logiciel, CD, GPS, livres ... (text/html <http://www.fnac.com/> - 85k - en cache), 2. Livre, BD, manga et roman à découvrir et à acheter en ligne sur Fnac.com (text/html <http://www.fnac.com/livres.asp> - 84k - en cache), and 3. Fnac.com - VOD - Vidéos à domicile, Téléchargement de vidéos (text/html <http://vod.fnac.com/> - 122k - en cache).

Plus de 5 000 000 pages indexées à la date du 30 avril 2009. Remarquez en troisième position une page issue d'un des sous-domaines.

Pouvez-vous en dire autant de votre site web ?

Vue d'ensemble des facteurs décisifs

Maintenant que vous savez où vous en êtes de votre indexation, nous pouvons aborder les techniques décisives à d'éventuelles actions correctrices.

Comme nous le disions précédemment, rien ne vaut une indexation naturelle. Deux types de facteurs sont à prendre en considération :

- les facteurs bloquants,
- et les facteurs favorisants.

Les facteurs bloquants l'indexation

Les facteurs bloquants l'indexation sont ceux qui concernent tout ce qui est susceptible de freiner l'indexation de votre site web. Autrement dit, il vous faut faciliter la tâche des moteurs de recherche, en évitant notamment certaines erreurs qui peuvent vous coûter très cher, au sens propre (si vous êtes e-commerçant) comme au sens figuré. Il est impératif de s'occuper d'eux en premier lieu.

On prendra notamment garde à éviter :

- de recourir à certaines technologies (Flash, Javascript),
- les erreurs à l'échelle du site (mauvais maillage),
- les erreurs au niveau des pages (liens sortants trop nombreux, poids des pages excessifs),
- les erreurs au niveau des urls (urls illisibles).

Nous détaillons chacun de ces points dans le chapitre suivant.

Les facteurs favorisants l'indexation

Les facteurs favorisants l'indexation sont ceux qui concernent tout ce qui est susceptible de faciliter l'indexation de votre site web. De manière générale, ils permettent une indexation personnalisée donc plus conforme aux besoins de votre site web. Ils sont toujours un atout supplémentaire mais ne sont pas incontournables pour la plupart des sites web. Ils répondent souvent à des problématiques particulières liées à l'emploi de technologies spécifiques ou encore aux sites web de plusieurs milliers de pages.

Nous nous arrêterons en particulier sur :

- le sitemap,
- le fichier robots.txt.

Nous abordons en détail ces techniques dans l'avant-dernier chapitre de cette première partie.

Les facteurs bloquants

Il existe trois catégories de facteurs bloquants :

- les technologies qui entravent l'indexation,
- les sites web mal organisés,
- les pages web mal calibrées.

Les technologies à proscrire

Le problème de ces technologies est qu'elles rendent difficile la lecture des urls qu'elles contiennent. Autrement dit, si une page de votre site web n'est accessible qu'à partir de ressources (urls) de ce type, il y a de fortes chances que cette page ne soit jamais indexée, donc jamais visitée par les moteurs de recherche.

Le format Flash

Le Flash est probablement la pire de toute, car illisible par la plupart des moteurs de recherche.

Le recours à Flash occasionne une double perte. D'une part, les liens ne seront pas suivis (car non recensés), donc les pages pointées par ces liens non indexées (à moins qu'elles ne soient pointées ailleurs via un lien « en dur »). Et d'autre part, la valorisation du texte du lien ne sera pas prise en compte ce qui représente une perte en matière d'optimisation de code HTML.

A cela s'ajoute la lourdeur de cette technologie. Nombreux sont les internautes qui quittent un site web avant d'avoir terminé le téléchargement de l'animation flash de la page d'accueil. N'oubliez pas que vous devez convaincre en moins de dix secondes.

Le Javascript

Le Javascript reste délicat pour la plupart des moteurs de recherche. En général, ils ne lisent pas son contenu. La prudence invite donc à s'en passer pour la mention des urls, car elles leurs seront illisibles, donc inaccessibles. Pour le reste, il n'y a pas de mal à y recourir.

A éviter, logiciel de référencement gratuit

Conseillé, logiciel de référencement gratuit

Comme pour le Flash, le recours au Javascript pour la mention des urls engendre une double perte. D'une part, les liens en Javascript ne seront pas suivis, donc les pages pointées par ces liens non indexées (à moins qu'elles ne soient pointées ailleurs via un lien « en dur »). Et d'autre part, la valorisation du texte du lien ne sera pas prise en compte ce qui représente une perte significative en matière d'optimisation de code HTML.

On évitera donc les obstacles que sont les roll-over, les pop-up, les menus dont les liens sont faits en Javascript, etc.

Les Frames

Les frames sont accessibles aux crawlers des moteurs de recherche. Cependant, chaque frame est perçue comme une page web différente. De ce fait, les crawlers indexeront autant de pages web qu'il y a de frames contenues dans vos pages web. Le contenu de vos pages

sera donc éclaté et les internautes risquent fort d'arriver sur des pages incohérentes parce qu'incomplètes.

L'usage de Frames est donc peu recommandé, néanmoins cette technologie étant obsolète, il nous semble vain de nous y attarder.

Les facteurs bloquants au niveau du site

Les facteurs bloquants l'indexation au niveau du site web sont ceux qui gênent la circulation des crawlers. Pour leur rendre la tâche facile, il faut faire en sorte que votre site web soit aisément parcourable en tout sens.

Le maillage

Chaque page de votre site web est importante. Elles doivent toutes être facilement accessibles aux moteurs de recherche. Pour ce faire, il vous faut organiser le maillage de votre site. Faites donc en sorte que chacune de vos pages soit :

- à moins de trois clics de la page d'accueil,
- accessible à partir de plusieurs autres.

Ces deux mesures augmenteront ses chances d'indexation.

Travaillez la hiérarchie de votre site. N'hésitez pas à classer vos pages en catégories et sous catégories, vous arriverez ainsi à vos fins.

La page d'accueil

Votre page d'accueil est primordiale, mais ce n'est pas une raison suffisante pour négliger les autres pages. Vous pouvez considérer chaque page comme une porte d'entrée pour votre site, donc un point à partir duquel peut commencer l'indexation de votre site web.

Si on revient au point précédent, le maillage, on constate que la tâche se complique. En effet, si toute page peut-être virtuellement page d'accueil, il faut alors que toute autre page soit à moins de trois clics. Autrement dit, à partir de n'importe quelle page, on doit pouvoir atteindre n'importe quelle autre page en moins de trois clics.

Le cas des blogs est exemplaire. Ils réussissent tous cette tâche à merveille.

Le plan du site

Une solution pour prendre en considération les deux critères précédents, est de concevoir une page spécifique « plan du site ». Cette page référencera de manière cohérente, un peu comme une table des matières, chaque page de votre site web. Enfin, cette page sera référencée par chaque page de votre site web, dans le pied de page par exemple. Cette technique fait l'unanimité parmi les spécialistes du référencement.

ATTENTION, le lien vers votre plan du site doit être un lien en dur (facile à suivre par les crawlers des moteurs de recherche). Pareillement votre plan du site doit être constitué de liens en dur. Dans le cas contraire, tous vos efforts seront vains.

N'oubliez pas que cette technique est surtout à destination des moteurs de recherche pour faciliter l'indexation de votre site web. Néanmoins, elle est souvent très utile à l'internaute désireux de se repérer dans un site web. Il est donc préférable de la soigner.

Les facteurs bloquants au niveau des pages

Les facteurs bloquants l'indexation au niveau de la page web sont ceux qui amènent les crawlers à négliger certains liens. Pour leur rendre la tâche facile, il faut faire en sorte que vos pages web soient au gabarit attendu par les crawlers.

Le nombre de liens

Vos pages ne doivent pas avoir plus de cent liens, une vingtaine maximum est conseillée. Les robots estiment suspecte une page trop chargée en liens.

Le type des liens

Le Javascript

Comme dit précédemment, ces liens ne doivent pas être en Javascript, sous peine de ne pas être indexés par les moteurs de recherche. Ils doivent être en dur :

A éviter, logiciel de référencement gratuit

Conseillé, logiciel de référencement gratuit

L'url rewriting

Nettoyez l'url de chacune de vos pages. Les humains y gagneront aussi. Les robots préfèrent indexer une url qui ressemble à

http://www.monsite.com/baladeur-mp3-pas-cher.htm

plutôt qu'à

http://www.monsite.com/produit.php?id=123456§ion_id=123&subsection_id=456

Pour y arriver, on utilise une technique appelée url rewriting. Dans l'état actuel de vos connaissances, il est probable que vous ne sachiez pas en faire usage. Nous l'abordons donc dans la troisième partie de ce document.

Le gabarit des pages

Evitez les longues pages. Les robots préfèrent en général les pages plus concises. De cette manière, vous pourrez construire plus de pages à indexer, vous aurez donc d'autant plus de pages candidates à un bon positionnement dans les moteurs de recherche et ces pages du fait de leur contenu thématique plus homogène auront davantage de chances d'être bien positionnées dans les résultats des moteurs de recherche.

Les facteurs favorisants

Ces techniques ne sont pas incontournables. Elles répondent à des problèmes spécifiques. Néanmoins, elles ne peuvent en aucun cas vous nuire. Dans la mesure où elle demande assez peu de travail, il serait dommage de vous en priver alors qu'elles sont susceptibles d'apporter un gain notable en matière d'indexation. Nous nous concentrerons principalement ici sur deux techniques :

- le sitemap,
- et le fichier robots.txt.

Le sitemap

Le sitemap, Kezako ?

Le sitemap est une technique qui facilite le travail d'indexation des moteurs de recherche. Cette technique est approuvée officiellement par les moteurs de recherche majeurs : Google, Yahoo, Microsoft Bing et Ask.

Le fichier sitemap vous permet d'indiquer aux moteurs de recherche comment indexer votre site web. De cette manière, l'indexation se fera plus conformément à vos attentes.

Petit rappel sur les crawlers

Un crawler est un robot (aussi appelé bot ou spider) qui parcourt le web et récupère automatiquement les pages web visitées. Les moteurs de recherche ont des crawlers qui se chargent du travail d'indexation.

*Le crawler de Google s'appelle **GoogleBot** alors que celui de Yahoo s'appelle **Yahoo Slurp**.*

A chaque fois qu'un crawler visite une page web, il en fait l'analyse, en extrait les liens vers d'autres pages web et les indexe dans sa base de données. Quelques temps plus tard, il visite les pages nouvellement indexées et réitère l'opération.

C'est ce qu'on pourrait appeler l'indexation naturelle. Il est néanmoins possible de devancer ce processus, à l'aide du protocole sitemaps.

Le protocole sitemaps (avec un « s »)

En novembre 2006, Google, Yahoo et Microsoft se sont mis d'accord pour supporter le protocole sitemaps. Le protocole sitemaps est un standard qui peut vous aider à soumettre l'ensemble de vos pages web aux moteurs de recherche en une seule fois.

Ask a ensuite rejoint Google, Yahoo et Microsoft. Désormais les quatre principaux moteurs de recherche supportent le protocole sitemaps.

Le fichier sitemap (sans « s »)

Le protocole sitemaps s'implémente côté webmaster sous la forme d'un fichier sitemap. Les fichiers sitemap sont un moyen facile pour les webmasters d'informer les moteurs de recherche sur les pages de leurs sites web qui sont prêtes à être crawlées. C'est un format XML standard qui facilite la création d'un fichier sitemap compréhensible par tous les moteurs de recherche.

Les moteurs de recherche peuvent désormais détecter automatiquement votre fichier sitemap, nous verrons cela dans la suite de ce chapitre.

Les limites du sitemap

Comment le fichier sitemap aide mon site web ?

Habituellement, les crawlers du web découvrent les nouvelles pages web à partir de liens faits par d'autres sites web. Le fichier sitemap permet aux crawlers qui supportent ce protocole de collecter en une fois toutes ces informations ainsi que des meta-données associées.

Le fichier sitemap vous permet d'informer les moteurs de recherche à propos des pages de votre site web. Il est souvent plus facile pour les moteurs de recherche de trouver toutes vos pages web si vous utilisez un fichier sitemap.

Est-ce qu'un fichier sitemap est indispensable à un bon référencement ?

De manière générale un site web bien construit n'a pas besoin de fichier sitemap. Utiliser le protocole sitemap ne garantit pas que vos pages web seront intégrées aux index des moteurs de recherche, mais cela facilite grandement et améliore la qualité du travail des crawlers.

Le fichier sitemap ne sert à rien, alors ?

Non, en plus de favoriser une indexation optimale, il est souvent incontournable. Il répond à des problématiques particulières liées à l'emploi de technologies spécifiques (Flash, Javascript, etc.) ou encore aux sites web de plusieurs milliers de pages.

Est-ce qu'un fichier sitemap améliore mon positionnement dans les moteurs de recherche ?

Il faut bien prendre garde à ne pas confondre indexation et positionnement dans les résultats de recherche. Indexation veut dire présent dans la base de données des moteurs de recherche. Or, une page web ne peut en aucun cas sortir dans les résultats de recherche, si elle n'est pas indexée. L'indexation est donc une **condition nécessaire mais pas suffisante** à un bon positionnement.

Un sitemap aide les moteurs de recherche à trouver vos pages web mais ne vous aide pas à obtenir un meilleur positionnement dans les moteurs de recherche. Quand les moteurs de recherche indexent vos pages, ils doivent trouver des pages au contenu optimisé et de bons backlinks. Les moteurs de recherche vous positionneront bien uniquement si vous avez les deux. Ces éléments seront abordés dans les prochaines parties de ce document.

Le fichier sitemap en détail

Le sitemap, ça ressemble à quoi ?

Dans sa forme la plus simple, un fichier sitemap est un fichier XML qui liste les urls d'un site web avec en plus quelques meta-données renseignant chaque url :

- la date de sa dernière mise à jour,
- selon quelle fréquence change l'url,
- son importance relativement aux autres urls du site.

Ces informations aident les moteurs de recherche à crawler votre site web intelligemment.

Ce qui donne dans votre fichier XML sitemap.xml.

Code XML :

```
<?xml version='1.0' encoding='UTF-8' ?>
<urlset xmlns="http://www.google.com/schemas/sitemap/0.84">
  <url>
    <loc>http://www.monsite.com</loc>
    <lastmod>2009-06-03</lastmod>
    <changefreq>weekly</changefreq>
    <priority>1.0</priority>
  </url>
  <url>
    <loc>http://www.monsite.com/une-page.html</loc>
    <lastmod>2009-05-15</lastmod>
    <changefreq>monthly</changefreq>
    <priority>0.1</priority>
  </url>
  <url>
    <loc>http://www.monsite.com/autre-page.html</loc>
    <lastmod>2009-06-06</lastmod>
    <changefreq>daily</changefreq>
    <priority>0.1</priority>
  </url>
  ...
</urlset>
</xml>
```

Comme nous le verrons par la suite, la plupart des balises sont optionnelles.

Les différentes balises de votre fichier sitemap

Urlset

Le tag **urlset** précise la version du standard XML sitemap utilisée. La valeur renvoie à une url qui décrit le format utilisé dans le fichier courant. Dans la mesure où le standard peut évoluer, ce tag permet de faire cohabiter des anciennes versions du standard avec la dernière sans que cela ne pose problème. Il est donc inutile de mettre à jour votre fichier sitemap lorsque le standard évolue, à moins de vouloir profiter des nouvelles fonctionnalités que celui-ci propose.

Ce tag est obligatoire et il est unique.

Url

Le tag **url** regroupe l'ensemble des informations concernant une unique url ou page web. Il en faut donc un par url ou page web décrite dans le fichier sitemap (sitemap.xml).

Ce tag est obligatoire.

Loc

```
<loc>http://www.monsite.com</loc>
```

Le tag **loc** précise l'url courante, ici <http://www.monsite.com>.

Ce tag est, bien entendu, obligatoire.

Lastmod

```
<lastmod>2009-06-03</lastmod>
```

Le tag **lastmod** précise la date de dernière mise à jour de l'url courante.

Cette date doit nécessairement être au format AAAA-MM-JJ, ici 2009-06-03, soit le 3 juin 2009.

Ce tag est optionnel. Il est donc possible de s'en passer parfois ou même toujours.

Changefreq

```
<changefreq>weekly</changefreq>
```

Le tag **changefreq** précise avec quelle fréquence change l'url courante.

Il peut prendre les valeurs always, hourly, daily, weekly, monthly, yearly, never pour respectivement toujours (à chaque nouvel accès à l'url courante), chaque heure, chaque jour, chaque semaine, chaque mois, chaque année, ou jamais (pour les urls archivées), ici weekly, soit chaque semaine.

Ce tag est optionnel. Il est donc possible de s'en passer parfois ou même toujours.

Priority

```
<priority>1.0</priority>
```

Le tag **priority** précise l'importance de l'url courante **relativement aux autres urls du (le même) site web courant** (priority = 0.1, 1.0, etc.). Sa valeur varie de 0.0 (priorité faible) à 1.0 (priorité forte).

Ce tag est optionnel. Il est donc possible de s'en passer parfois ou même toujours. Si il est absent alors la valeur par défaut est 0.5. Dans le cas présent, la valeur 1.0 indique une forte priorité.

Qu'est-ce que je fais de mon fichier sitemap ?

Les quatre principaux moteurs de recherche Google, Yahoo, Microsoft Bing (successeur de Microsoft Live Search) et Ask vérifient régulièrement les fichiers robots.txt des sites web. L'ajout d'une ligne dans ce fichier suffit alors pour indexer votre fichier sitemap.

Ce qui donne dans votre fichier robots.txt.

Code robots.txt :

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

Attention, le fichier robots.txt doit se trouver à la racine de votre site web.

```
http://www.monsite.com/robots.txt
```

Vous pouvez néanmoins devancer leurs visites en soumettant directement vos fichiers sitemap à Google, Yahoo et Ask. Pour Google et Yahoo, il vous faudra ouvrir un compte. Alors que pour Ask, il suffit de « pinger » l'url qui suit avec l'adresse de votre sitemap en paramètre :

```
http://submissions.ask.com/ping?sitemap=http%3A//www.monsite.com/sitemap.xml
```

Le protocole sitemap peut vous aider à informer les moteurs de recherche à propos de vos pages web. Rappelez-vous, une indexation réussie est la première étape d'un référencement réussi. Vous trouverez davantage d'informations à propos du nouveau standard sitemap sur le site web officiel du sitemap (<http://www.sitemaps.org/fr/protocol.php>). Il peut également être intéressant de visiter la page sitemap de Google (<https://www.google.com/webmasters/tools/docs/fr/protocol.html>) et la page indexation de Ask (<http://sp.fr.ask.com/fr/docs/about/asksearch.shtml>).

Le fichier robots.txt

Le fichier robots.txt est une autre technologie qui permet de personnaliser l'indexation de votre site web par les crawlers des moteurs de recherche. Elle n'aura pas l'impact que peut avoir le fichier sitemap.xml, elle peut néanmoins vous rendre quelques petits services.

Le fichier robots.txt, c'est quoi ?

Le fichier robots.txt (avec un « s ») est le premier fichier que visitent les crawlers des moteurs de recherche lorsqu'ils commencent à indexer un site web. Dans ce fichier, sont consignés des directives écrites par le webmaster du site web à destination des robots et donc des crawlers des moteurs de recherche. La présence de ce fichier n'a rien d'obligatoire.

Le fichier robots.txt (ou fichier d'exclusion des robots) est le moyen côté webmaster d'implémenter le protocole d'exclusion des robots (Robots Exclusion Protocol).

« Robots Exclusion Protocol », kezako ?

Plus précisément, ce protocole permet aux webmasters de communiquer aux robots (et donc aussi aux crawlers des moteurs de recherche) :

- les parties (privées) de leurs sites web à ne pas indexer,
- la vitesse maximale d'indexation souhaitée (directive non standard),
- ainsi que l'emplacement de leur fichier sitemap.xml (directive non standard).

Ce protocole est issu d'un consensus, il n'a pas valeur d'obligation. Néanmoins, les crawlers des moteurs de recherche s'y astreignent.

Pourquoi s'embarrasser d'un fichier robots.txt ?

Le fichier robots.txt n'est pas une technologie incontournable. Elle peut cependant avoir son utilité. Si vous avez des milliers de pages qu'il n'est pas intéressant de voir indexées, le signaler aux robots via le fichier robots.txt vous permet d'économiser de précieuses ressources (bande passante, cpu, etc.). Les crawlers ne rapatrieront plus inutilement toutes ces pages web. Sachez que lorsque les robots se déchaînent, ils peuvent être gros consommateurs de ressources pour votre serveur web (hébergeur) et pénaliser lourdement dans leur navigation (ralentir) les internautes qui visitent votre site web au même moment.

Le fichier robots.txt n'est visité que par les robots. Si vous surveillez l'indexation de votre site web, c'est un moyen pour vous de savoir quels sont les robots qui vous rendent le plus souvent visite. La consultation de vos fichiers de logs (sur votre serveur web) vous fournira ces informations.

Mise en garde

Le protocole REP n'a pas valeur d'obligation pour les robots. Rendre inaccessible une partie de votre site web via le fichier robots.txt est donc une très mauvaise idée. Le fichier robots.txt est public, donc accessible à tout robot. N'oubliez pas qu'il en existe aussi de malveillants qui sont à la recherche de failles en tout genre. En signalant les parties privées de votre site web sans autre moyen que le fichier robots.txt pour empêcher d'y accéder, vous encouragez les actes de pirateries sur votre propre site web.

Comment rédiger mon fichier robots.txt ?

Le fichier robots.txt se compose d'un ou plusieurs blocs d'instructions. Une ou plusieurs lignes vides marquent la fin d'un bloc d'instructions. Chaque bloc contient des directives et

éventuellement des commentaires. Les commentaires commencent toujours par le caractère « # » (ils sont ignorés par les robots).

Le fichier robots.txt est construit à l'aide de deux directives standards : User-agent et Disallow. Chaque bloc débute par une ou plusieurs directives User-agent suivies par une ou plusieurs directives Disallow. Des commentaires peuvent être placés à n'importe quel endroit.

La directive « User-agent »

Chaque crawler est identifiable par son User Agent (identité qu'il s'est lui-même attribué). A chaque fois qu'il effectue une requête auprès d'un serveur web, il se présente avec son User Agent.

*Le crawler de Google a pour User Agent, **Googlebot**. Tandis que celui de Yahoo a pour User Agent **Yahoo Slurp**. Pareillement, le crawler de Microsoft Bing (successeur de Microsoft Live Search) a pour User Agent **MSNbot**.*

Cette directive est une **directive standard**, elle vous permet d'indiquer à quels crawlers s'adresse le bloc d'instructions qui la suit.

La directive « Disallow »

Cette directive est une **directive standard**, elle signale quels sont les répertoires ou fichiers qui ne doivent pas être visités.

Exemples

Le caractère « * » est un caractère spécial qui désigne n'importe quel User Agent.

Le bloc d'instructions suivant autorise tout robot à parcourir l'ensemble de votre site web. Ne pas avoir de fichier robots.txt produit le même effet.

Code robots.txt

```
User-agent: *  
Disallow:
```

Le bloc d'instructions suivant interdit à tout robot de parcourir la moindre partie de votre site web. N'espérez pas être bien positionné dans les SERP !

Code robots.txt

```
User-agent: *  
Disallow: /
```

Le bloc d'instructions suivant interdit à tout robot l'accès aux répertoires /cgi-bin, /images, /tmp, /private (et à tout ce qu'ils contiennent).

Code robots.txt

```
User-agent: *  
Disallow: /cgi-bin/  
Disallow: /images/  
Disallow: /tmp/  
Disallow: /private/
```

Le nombre de blocs d'instructions n'est pas limité. Vous pouvez en utiliser autant que nécessaire.

Les deux blocs d'instructions suivants interdisent à tout robot, excepté le Googlebot, l'accès au répertoire /forum (et à tout ce qu'ils contient).

Code robots.txt

```
# interdiction à tous les robots d'indexer /forum
User-agent: *
Disallow: /forum

# excepté au googlebot
User-agent: googlebot
Disallow:
```

Les directives non standards

Il existe des directives non standards comme :

- **Crawl-delay** (fixe le délai minimum en secondes entre deux requêtes successives à un serveur web),
- **Allow** (accorde une permission à la différence de **Disallow**),
- ou **Sitemap** (indique l'emplacement du fichier sitemap.xml). Elles sont comprises par les principaux crawlers.

Le bloc d'instructions suivant :

- *fixe à 10 secondes le délai minimum entre deux requêtes successives au serveur web, pour tout crawler,*
- *autorise le googlebot à crawler l'intégralité de votre site web,*
- *interdit aux autres crawlers l'accès au répertoire /forum (et à tout ce qu'ils contient), excepté le fichier public.html.*

Code robots.txt

```
User-agent: *
Crawl-delay: 10

# restriction pour tous les robots
User-agent: *
Allow: /forum/public.html
Disallow: /forum

# excepté le googlebot
User-agent: googlebot
Disallow:
```

Diverses propositions vont dans le sens d'étoffer le standard avec de nouvelles directives.

Qu'est ce que je fais de mon fichier robots.txt ?

La présence du fichier robots.txt n'est pas obligatoire. Néanmoins si il existe, il doit nécessairement se trouver à la racine de votre site web pour pouvoir produire son effet. Les crawlers des moteurs de recherche n'iront pas le chercher ailleurs.

Par exemple, pour le site web du logiciel de référencement GRATUIT GnoZtiK, le fichier est accessible à l'url <http://www.gnoztik.com/robots.txt>.

L'effet SandBox

Dernier chapitre de cette première partie, l'effet SandBox ne relève pas exactement de l'indexation. Néanmoins, les sites jeunes le subissent souvent une fois la phase d'indexation achevée.

La réalité de l'effet SandBox ne fait pas l'unanimité chez les professionnelles. Nombreux parmi eux déclarent en avoir observé les effets, sans pour autant avoir été en mesure d'en cerner précisément les contours.

Plus qu'un mécanisme mis en place par Google, il serait du à un cumul de facteurs intervenants dans l'algorithme de classification des pages web. Dans la suite de ce chapitre, nous partirons du principe que la SandBox est une réalité voulue par Google.

La SandBox, kezaiko ?

L'effet SandBox semble affecter la plupart des nouveaux sites web. Ils apparaissent rapidement dans l'index de Google, mais sont souvent très mal classés dans les SERP, surtout (mais pas uniquement) sur les mots-clés compétitifs. L'effet SandBox peut aussi affecter certains sites anciens. Lorsqu'un site web ancien subit un fort déclassement dans les SERP, il est fort probable qu'il vient d'être placé en SandBox. Ces sites web, récents ou anciens, sont dits « SandBoxés » ou « dans la SandBox ». Cette situation est temporaire.

A quoi sert la SandBox ?

La SandBox inflige une période probatoire pour les nouveaux arrivants. Elle permet notamment d'éviter un bouleversement permanent des résultats de recherche de Google avec l'arrivée des nouveaux entrants très rapidement indexés.

La SandBox est également un filtre contre les tricheurs qui voudraient détourner l'algorithme de Google pour rapidement passer devant les autres. Elle vise donc à garantir une certaine stabilité des classements proposés par Google sur les différents mots-clés.

Attention, la SandBox n'est pas la blacklist. Un site web blacklisté n'apparaît plus dans l'index de Google (voir chapitre 2 de cette première partie).

Mon site est-il dans la SandBox ?

Tout dépend de l'ancienneté de votre site web.

Si votre site web est récent,

- indexé par Google,
- mal positionné dans les résultats de recherche Google,
- et qu'il est bien positionné sur d'autres moteurs de recherche,

alors il est fort probable qu'il se trouve dans la SandBox.

Si votre site web est ancien,

- indexé par Google,
- brutalement déclassé dans les résultats de recherche Google,
- et qu'il reste bien positionné sur d'autres moteurs de recherche,

alors il est possible qu'il vienne d'atterrir en SandBox.

Quel impact a la SandBox sur mon site web ?

Les effets de la SandBox durent entre quelques semaines et quelques mois selon les sites web. Durant cette période, Google ne vous amènera que peu de visites. Il faut alors vous armer de patience et savoir attendre votre heure. Dès que la pénalité SandBox sera levée, votre site web pourra être propulsé en haut des SERP, à condition qu'il soit jugé pertinent par Google.

Comment sortir de la SandBox ?

Votre site web est récent

La SandBox est souvent un passage obligé pour les nouveaux sites web. La plupart du temps, il n'y a rien à faire, si ce n'est patienter et mettre à profit cette période qui peut s'étendre sur plusieurs mois pour continuer votre travail de référencement (production de contenu original, optimisation html, collecte de backlinks).

Vous pouvez cependant vous concentrer sur des actions qui vous seront bénéfiques durablement et pourront peut-être réduire la durée de l'effet SandBox sur votre site web.

Votre site web est ancien

Recherchez les récentes modifications de votre site web qui ont pu contrarier Google. Toute manœuvre (ou cumul de manœuvres) artificielle visant un gain de positionnement peut être à l'origine de ce déclassement brutal. Dans ce cas, mieux vaut rebrousser chemin.

Vous pouvez également prendre des mesures positives. Réorientez vos objectifs vers des actions qui vous seront profitables dans la durée et pourront peut-être réduire le temps passé en SandBox par votre site web.

Quelles mesures concrètes pour sortir de la SandBox ?

Parmi les mesures que vous pouvez prendre, que votre site web soit récent ou ancien, les suivantes devraient vous être avantageuses (sans garantie) :

- évitez les inscriptions en masse dans des annuaires ou des sites d'échanges de liens (gonflement artificiel de la popularité - nombre de backlinks),
- obtenez des backlinks depuis des sites à forte notoriété (gagnez en crédibilité),
- obtenez des backlinks à sens unique de sites web de même thématique (backlinks naturels, autres que de simples échanges de liens).

Mise en garde

Prenez garde à ne pas tenter de forcer la sortie de la SandBox, vous risquez fort de voir votre situation empirer. Rappelez-vous que le comportement de votre site web (contenu html, réseau de backlinks, etc.) doit toujours apparaître normal aux moteurs de recherche. Si votre site web apparaît déviant, il risque fort d'être pénalisé par de sévères déclassements (SandBox), ou même une désindexation généralisée (blacklist) d'où il est souvent difficile de revenir.

Un comportement normal pour un site web, c'est, par exemple, ne pas avoir de contenu caché aux internautes, accumuler des backlinks progressivement (à un rythme crédible), avoir des backlinks à sens unique (autres que de simples échanges de liens), etc.

Soyez donc confiant et poursuivez la promotion de votre contenu par des voies saines.

Les mots-clés

Une fois les éventuels problèmes liés à l'indexation de votre site web levés, vous pouvez vous concentrer sur le choix des mots-clés à cibler.

Il est important de ne jamais perdre de vue que le référencement n'est pas une finalité en soi. Etre bien positionné sur des tas de mots-clés n'a aucun sens si vous n'atteignez pas vos objectifs. Mieux vaut être bien positionné sur peu de mots-clés, mais drainant un fort trafic qualifié, que sur une foule de mots-clés qui vous apporte des internautes qui quittent votre site web dès leur arrivée. N'oubliez pas, vous êtes à la recherche d'un trafic abondant (l'évaluation varie selon les secteurs) et qualifié (à la recherche de ce que vous offrez). Un bon référencement a pour but de favoriser la rencontre entre une offre (ce que vous proposez), pas nécessairement commerciale, et une demande (le motif de la visite de l'internaute). Le choix des mots-clés à cibler est donc capital. Vous vous devez donc de parler le langage de ceux que vous visez. Prenez garde à ne pas négliger cette étape majeure pour le référencement de votre site web.

Le premier chapitre de cette partie s'efforce de définir une typologie des mots-clés. A l'aide de cette typologie, vous pourrez identifier les mots-clés proposant un important retour sur investissement (temps et/ou argent). Les mots-clés nécessitant un effort trop élevé par rapport au trafic (qualité et/ou quantité) retourné seront alors écartés.

Dans le deuxième chapitre, nous présentons des conseils qu'il est important de mettre en œuvre pour optimiser votre référencement. Ces conseils permettent notamment de décupler l'efficacité de votre référencement. Nous proposons également une démarche pragmatique pour que votre site web gagne sûrement **mais progressivement** sa place sur la toile.

Dans le troisième chapitre, nous reviendrons sur le phénomène de la **longue traîne** mis en évidence récemment. Nous verrons notamment qu'elle synthétise beaucoup de notions et conseils que nous vous prodiguerons tout au long de cette deuxième partie.

Enfin dans le dernier chapitre, nous nous arrêtons sur un outil GRATUIT intéressant (pour ne pas dire incontournable) le générateur de mots-clés de Google. Cet outil dispense de précieuses informations, il permet notamment de suggérer des mots-clés en rapport avec vos ambitions, vos besoins, d'évaluer la concurrence et les volumes de recherche des internautes sur ces mots-clés.

ATTENTION, vous devez parler le langage de ceux à qui vous vous adressez.

Le choix des mots-clés est donc **CAPITAL**.

Comment pouvez-vous espérer toucher votre public cible si vous le négligez ?

Typologie des mots-clés

Tous les mots-clés n'ont pas la même valeur pour votre référencement. Trois critères majeurs sont à prendre en considération pour évaluer la pertinence de se positionner sur tel ou tel mot-clé :

- la quantité du trafic généré par le mot-clé,
- la qualité du trafic engendré par le mot-clé,
- et enfin la convoitise que le mot-clé suscite.

La quantité du trafic généré par un mot-clé correspond au nombre de fois qu'un internaute le saisit dans un moteur de recherche, tous moteurs confondus. Concrètement, cette quantité représente le maximum de visites que vous pouvez espérer recevoir en provenance des moteurs de recherche par l'optimisation de vos pages web pour ce mot-clé. C'est un maximum, et il faudra donc, en fait, partager (dans une large mesure) ce maximum possible avec les autres pages web bien positionnées dans les moteurs de recherche pour ce mot-clé.

La qualité du trafic engendré par un mot-clé représente l'adéquation entre l'offre (ce que vous avez à proposer – information, produit, etc.) et la demande (motif de la visite de l'internaute). Si un internaute ne cherche que de l'information sur un produit et que vous ne faites que de la vente, il n'y a pas adéquation. Autrement dit, ce trafic ne vous est pas utile puisque l'internaute en question n'a pas l'intention d'acheter (à moins que vous ne soyez particulièrement doué pour la vente). Ce paramètre est délicat à évaluer et relève dans une certaine mesure du webmarketing.

La convoitise qu'un mot-clé suscite, renvoie au nombre de webmasters (ou concurrents) qui font des efforts pour se positionner sur ce mot-clé. En effet, il n'y a en général que dix places sur la première page de résultats des moteurs de recherche. Plus il y aura de concurrents, plus la tâche sera ardue.

Sur la base de ce qui précède, il nous est possible de qualifier les mots-clés de :

- concurrentiels,
- parasités,
- délaissés,
- ou de niche.

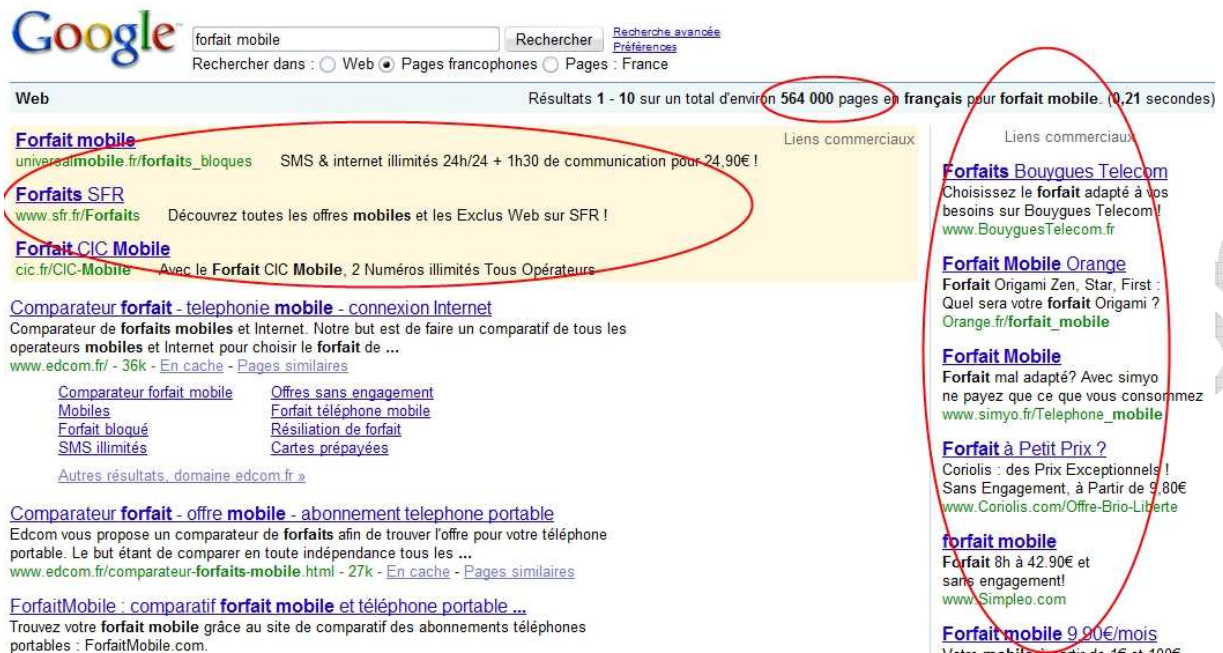
Cette classification permettra ensuite de vous concentrer sur le travail le plus efficace, en écartant certains mots-clés au profit d'autres plus rentables.

Un mot-clé concurrentiel, c'est quoi ?

Nombreux sont les webmasters qui font des efforts importants pour s'y positionner. De plus, ils ont, en général, des moyens importants (notoriété déjà acquise du site web, temps, argent). Vous risquez donc de rencontrer de grandes difficultés à vous placer parmi les meilleurs, car la lutte est âpre.

*Par exemple, le mot-clé **forfait mobile** est concurrentiel.*

http://www.google.fr/search?hl=fr&q=forfait+mobile&btnG=Rechercher&meta=lr%3Dlang_fr



A la date du 3 mai 2009, on constate que :

- Google retourne plus de 500 000 résultats (de nombreuses pages web en parle),
- la présence de nombreux liens sponsorisés (beaucoup d'annonceurs sont intéressés),
- les urls retournées sont courtes (surtout des pages d'accueil à l'exclusion du second lien).

En conséquence,

- il est difficile de s'y positionner,
- par contre le trafic en retour est de bonne qualité.

A moins d'avoir de sérieux atouts, il est préférable de passer votre chemin car vous risquez de dépenser beaucoup d'énergie pour un piètre retour sur investissement.

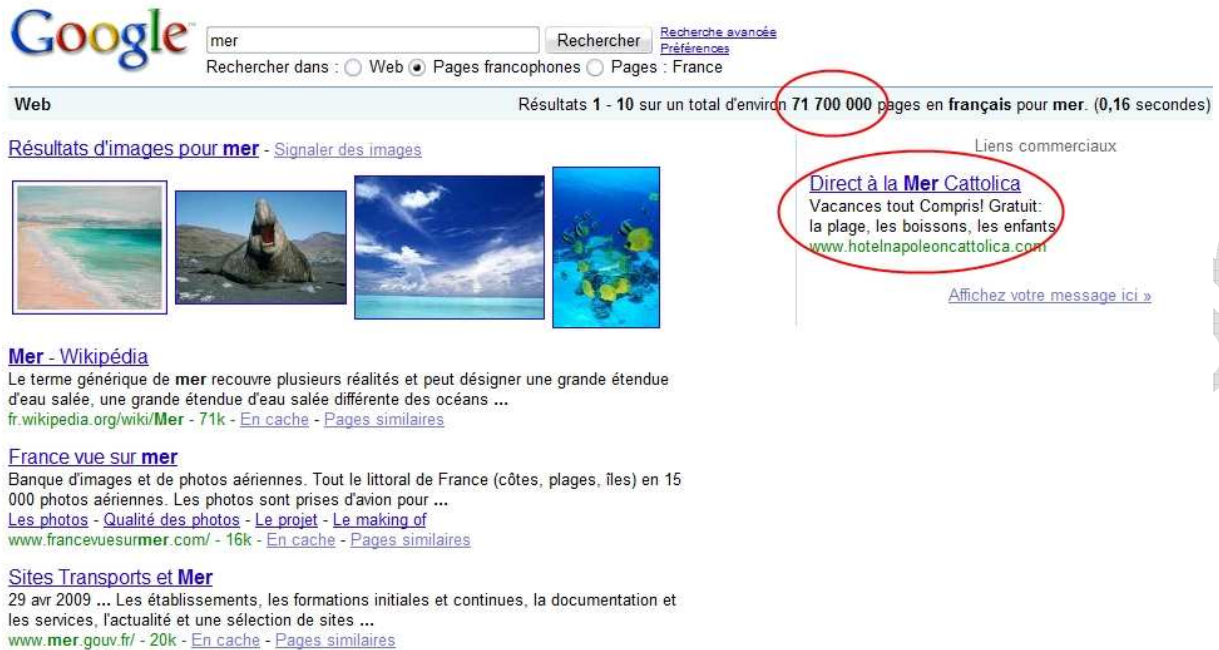
Un mot-clé parasité, c'est quoi ?

Beaucoup de webmasters s'y trouvent positionnés « par hasard » (effet de bord). Ils n'ont pour la plupart jamais fait l'effort pour se positionner sur ce mot-clé, mais leur site ayant d'une façon ou d'une autre un rapport avec le sujet, ils s'y trouvent.

Pareillement, il y a beaucoup de motifs pour lesquels un internaute peut faire cette recherche dans un moteur de recherche. Le trafic qui en résulte est donc peu qualifié.

Par exemple, le mot-clé **mer** est parasité.

http://www.google.fr/search?hl=fr&q=mer&btnG=Rechercher&meta=lr%3Dlang_fr



Google mer Rechercher Recherche avancée Préférences

Rechercher dans : ☐ Web ☒ Pages francophones ☐ Pages : France

Web Résultats 1 - 10 sur un total d'environ 71 700 000 pages en français pour mer. (0,16 secondes)

Résultats d'images pour mer - Signaler des images

Liens commerciaux

Direct à la Mer Cattolica
Vacances tout Compris! Gratuit:
la plage, les boissons, les enfants
www.hotelnapoleoncattolica.com

Affichez votre message ici »

Mer - Wikipédia
Le terme générique de mer recouvre plusieurs réalités et peut désigner une grande étendue d'eau salée, une grande étendue d'eau salée différente des océans ...
fr.wikipedia.org/wiki/Mer - 71k - [En cache](#) - [Pages similaires](#)

France vue sur mer
Banque d'images et de photos aériennes. Tout le littoral de France (côtes, plages, îles) en 15 000 photos aériennes. Les photos sont prises d'avion pour ...
[Les photos](#) - [Qualité des photos](#) - [Le projet](#) - [Le making of](#)
www.francevuesurmer.com/ - 16k - [En cache](#) - [Pages similaires](#)

Sites Transports et Mer
29 avr 2009 ... Les établissements, les formations initiales et continues, la documentation et les services, l'actualité et une sélection de sites ...
www.mer.gouv.fr/ - 20k - [En cache](#) - [Pages similaires](#)

A la date du 3 mai 2009, on constate que :

- Google retourne plus de 71 000 000 résultats (de nombreuses pages web en parle),
- l'absence de liens sponsorisés (les annonceurs ne sont pas intéressés),
- les urls retournées sont courtes (surtout des pages d'accueil, peu de spécialisation),
- les contenus retournés sont peu homogènes (gouvernement, tourisme, actualité, sport, culture, etc.).

En conséquence,

- il n'est pas simple de s'y positionner,
- et le trafic en retour est de mauvaise qualité, car les motifs des internautes, qui tapent ce mot-clé, sont très variés.

Passez votre chemin, il n'y a que des inconvénients : positionnement difficile et trafic mal ciblé.

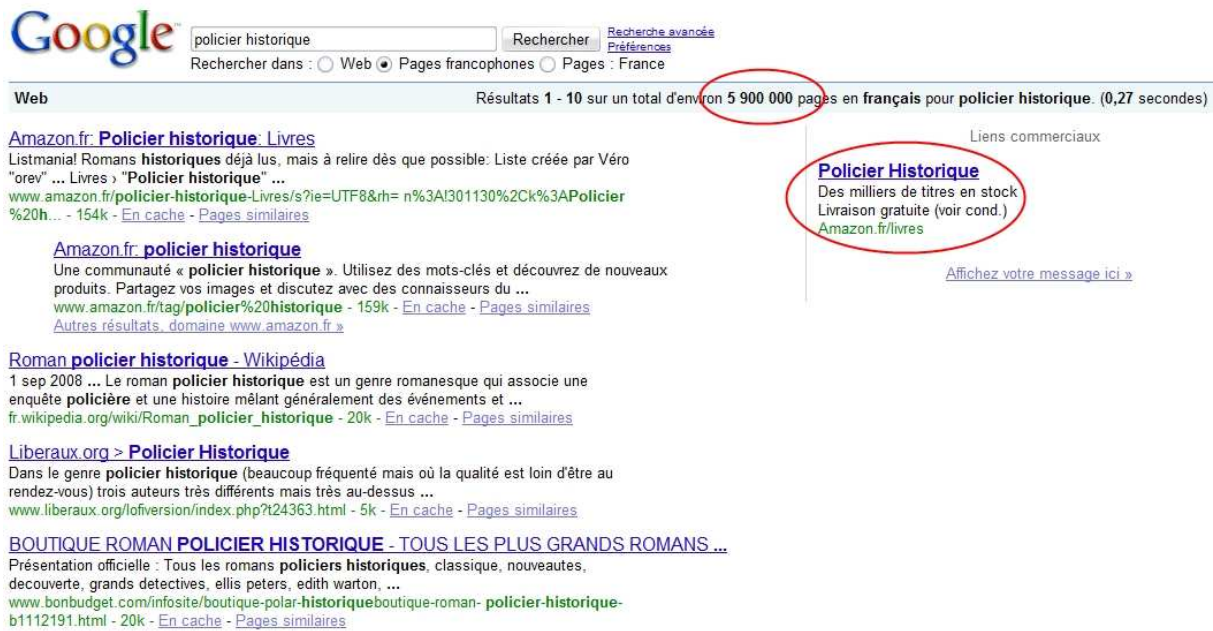
Un mot-clé délaissé, c'est quoi ?

A la différence du mot-clé parasité, ce mot-clé renvoie souvent à une thématique plus ciblée. Les sites qui s'y trouvent positionnés bénéficient, en général, de moins de notoriété que dans le cas du mot-clé parasité. La plupart du temps, beaucoup de webmasters s'y trouvent positionnés par effet de bord. Ils n'ont pour la plupart jamais fait de réels efforts sur ce mot-clé en particulier. Autrement dit, les pages qui s'y trouvent sont souvent peu optimisées sur un thème assez ciblé. Il n'est donc pas trop difficile de faire parti des meilleurs moyennant un effort raisonnable.

Par exemple, le mot-clé **policier historique** est délaissé.

http://www.google.fr/search?hl=fr&q=policier+historique&btnG=Rechercher&meta=lr%3Dlang_fr

Les mots-clés



A la date du 3 mai 2009, on constate que :

- *Google retourne plus de 5 000 000 résultats (de nombreuses pages web en parle),*
- *l'absence de liens sponsorisés (les annonceurs ne sont pas intéressés),*
- *les urls retournées sont longues mais ne contiennent que rarement la séquence policier historique (spécialisées uniquement en apparence),*
- *les contenus retournés manquent d'homogénéité.*

En conséquence, si votre site web colle à cette thématique, c'est un mot-clé intéressant,

- *car un effort abordable vous permettra de vous hisser parmi les meilleurs,*
- *avec en retour un trafic variablement qualifié. Très qualifié si les résultats retournés par Google sont homogènes en thème.*

Faites l'effort, les avantages compensent les inconvénients : positionnement accessible donc effort raisonnable (pas de vrais concurrents) et trafic variablement qualifié.

Un mot-clé de niche, c'est quoi ?

Peu nombreux sont les webmasters qui font des efforts importants pour s'y positionner. Par contre ceux qui y sont, n'y sont pas par hasard, ils ont fait des efforts conséquents. Le trafic qui en ressort est hautement qualifié. La lutte est sérieuse, mais le nombre de prétendants est limité.

*Par exemple, le mot-clé **gite ribeauville**, est un mot-clé de niche.*

http://www.google.fr/search?hl=fr&q=gite+ribeauville&btnG=Rechercher&meta=lr%3Dlang_fr

The screenshot shows a Google search for "gite ribeauville". The search bar contains the text "gite ribeauville" and the "Rechercher" button is visible. Below the search bar, it says "Rechercher dans : Web Pages francophones Pages : France". The results show "Résultats 1 - 10 sur un total d'environ 16 000 pages en français pour gite ribeauville. (0,15 secondes)". A map on the left shows the location of Ribeauvillé in Alsace. The search results list various gites with their websites and phone numbers. A red circle highlights the text "16 000 pages en français" in the search results. Another red circle highlights a link titled "Gites ruraux en Alsace" with the description "Loc entre particuliers de 2-6 pers Vallée Kaysersberg-Vigne & Montagne lapepinere.fr".

A la date du 3 mai 2009, on constate que :

- Google retourne moins de 20 000 résultats (peu de pages web en parle),
- l'absence de liens sponsorisés (les annonceurs ne sont pas intéressés),
- les urls retournées sont longues (spécialisation),
- les urls retournées contiennent presque toutes la séquence gite ribeauville (très spécialisées),
- les contenus retournés sont très homogènes (location tourisme).

En conséquence,

- peu de webmasters tentent de s'y positionner,
- mais ceux qui tentent, font des efforts importants,
- et le trafic en retour est hautement qualifié.

Si votre site colle à cette thématique, c'est un mot-clé idéal. Car un effort abordable (nombre limité de concurrents) vous permettra de vous hisser parmi les meilleurs avec en retour un trafic hautement qualifié.

Faites l'effort, les avantages sont plus nombreux que les inconvénients : positionnement abordable donc effort raisonnable et trafic très, très qualifié.

Comment faire concrètement pour déterminer le type d'un mot-clé ?

Les éléments à observer pour pouvoir trancher sont multiples. Des exemples qui précèdent, vous noterez que les facteurs suivants, lorsqu'ils sont recoupés, sont souvent des indices probants :

- la longueur du mot-clé, en général, plus il est long plus il est spécialisé,
- le nombre de pages retournées par Google,
- la présence ou l'absence de liens sponsorisés,
- la forme des urls, généralement courtes ou longues, retournées par Google,
- la présence du mot-clé dans les urls retournées par Google,

Les mots-clés

- la présence de sites institutionnels,
- l'homogénéité des thèmes abordés par les urls retournées par Google,
- ainsi que d'autres paramètres plus difficiles à évaluer « à la main ».

Le **logiciel GRATUIT GnoZtiK** (<http://www.gnoztik.com>) recoupe toutes ces informations avec d'autres paramètres plus difficiles à évaluer « à la main ». Pour davantage d'informations, consultez la page http://www.gnoztik.com/seo-faq.htm#chart_typology.

Comment choisir des mots-clés efficaces ?

Pour être visible sur Internet, il convient d'apparaître parmi les premiers résultats des moteurs de recherche. Mais attention, tous les mots-clés n'ont pas la même valeur. Cette valeur n'est pas absolue, elle est relative à vos ambitions et à vos moyens.

Il est donc impératif de faire des choix proportionnés à votre capacité actuelle. Le choix des mots-clés à cibler est capital, car ce choix influencera le contenu de vos pages web, de vos backlinks, les éventuels achats de liens sponsorisés, etc.

Dans ce chapitre, nous vous présentons des conseils pragmatiques qu'il est facile et important de mettre en œuvre pour optimiser votre référencement. Ces conseils permettent notamment de multiplier les mots-clés efficaces visés et, par conséquent, de décupler l'efficacité de votre référencement.

Ciblez des mots-clés liés à votre secteur d'activité

Dans un premier temps, il vous faut dresser une liste de mots-clés reliés à votre activité. Le dictionnaire des synonymes peut être très utile dans cette tâche. Concentrez-vous sur ce que proposent les pages de votre site web. C'est du bon sens, mais cet écueil revient souvent.

*Par exemple, un site qui parle de voyage au Maroc pourra choisir les mots-clés **voyage, Maroc, tourisme**, et les combiner entre eux, **voyage Maroc, tourisme Maroc**, etc.*

Choisissez précisément vos mots-clés

Plus un mot-clé est ciblé, plus vous aurez de chances de bien vous y positionner. La concurrence sera moindre et le trafic qui en résultera, sera du trafic hautement qualifié, autrement dit, des internautes qui cherchent exactement ce que vous proposez. Des mots-clés trop vagues diluent votre site web dans une multitude de sites au contenu très divers.

*Par exemple, le mot-clé **table** est trop vague. L'internaute obtiendra aussi bien des sites sur la vente de mobilier que des recettes de cuisine ou encore sur le design à travers les siècles. Autrement dit, l'internaute qui arrive sur votre site par le biais de ce mot-clé ne vient pas nécessairement pour acheter. En ciblant un mot-clé tel que **table d'extérieur en tek** vous obtiendrez de meilleurs résultats.*

Préférez des mots-clés composés de plusieurs mots

La plupart des recherches des internautes comptent deux, trois, voire quatre mots. Mieux vaut donc en tenir compte dans le choix de vos mots-clés.

*Reprenons l'exemple du site de voyage au Maroc, les mots-clés à plusieurs mots donneront : **tourisme Maroc, location vacances Marrakech, treck dans atlas marocain**, etc.*

Multipliez les bonnes combinaisons

En utilisant des mots-clés à plusieurs mots correctement « assemblés », vous pourrez vous positionner sur davantage de mots-clés. Vous augmenterez par là même le trafic potentiel de votre site web.

*Par exemple, à partir mot-clé **location vacances Marrakech**, vous ciblez les mots-clés **location vacances Marrakech** mais aussi **location vacances** et **vacances Marrakech**.*

Usez de synonymes

En usant de synonymes, à contenu de volume identique, vous cibleriez davantage de mots-clés. Sans pénaliser votre référencement, votre visibilité sera plus grande.

Optez pour des mots-clés à votre mesure

Tous les mots-clés ne sont pas à votre portée. Cette portée est relative et varie en fonction de vos moyens. Ces moyens sont bien entendu les efforts que vous êtes prêt à consentir (temps, argent) pour le résultat visé mais également le potentiel du site web que vous cherchez à positionner. Un site web récent a moins de poids qu'un site web ancien. De même, un site à forte notoriété a plus de poids qu'un site web classique. Il est donc vain de vous acharner sur certains mots-clés. Vous risquez fort de perdre du temps, de l'argent et de n'en tirer que découragement.

Dans la durée, vous pourrez réviser à la hausse vos objectifs. Le référencement est incrémental, les marches se montent les unes après les autres. Quand votre site web aura gagné en poids, vous pourrez viser des objectifs (mots-clés) plus ambitieux.

Visez les mots-clés les plus rentables

Les mots-clés les plus rentables (parmi ceux qui sont à votre mesure) sont ceux qui vous apporteront le plus de visites de qualité pour un moindre effort (ou un moindre coût), ceux pour lesquels le retour sur investissement (ROI) est optimal. Idéalement, les mots-clés doivent être faiblement convoités et apporter beaucoup de trafic qualifié (de bonne qualité, c'est-à-dire à fort taux de transformation). En pratique, c'est chose rare. Il vous faut donc trouver le bon compromis entre la convoitise que suscite le mot-clé chez les concurrents (importance de l'effort à fournir) et ce que recherchent les internautes (qualité du trafic).

*Le mot-clé **baladeur mp3** est certes très recherché... mais il vous faudra d'importants moyens pour espérer vous y placer. A l'inverse, il est plus aisé de bien se positionner sur le mot-clé **joint d'étanchéité pour climatiseur** mais est-il recherché par les internautes ?*

Soignez pragmatique

La notion de rentabilité n'est pas la même pour un site web naissant avec quelques dizaines de pages et un site web âgé de quelques années et pourvu de quelques milliers de pages. Il est impératif de toujours contextualiser les notions que nous abordons dans ce document.

Pour mener à bien votre tâche, il vous faudra avancer étape par étape. Chaque étape validée devra,

- marquer une avancée mesurable en matière de trafic transformant,
- accroître la notoriété de votre site web dans son secteur d'activité,
- améliorer votre savoir-faire en référencement,
- et préparer l'étape suivante dans laquelle les objectifs pourront alors être révisés à la hausse.

A court terme, les mots-clés les plus accessibles sont les mots-clés délaissés. La concurrence est presque inexistante. En conséquence, ils nécessitent un faible effort pour un trafic souvent appréciable. De plus, les résultats sont observables à court terme, ce qui est encourageant.

A moyen terme, si vous validez l'étape précédente, vous pourrez aborder les mots-clés de niche. Il y existe une vraie concurrence, mais celle-ci est limitée à des sites de faible à

moyenne envergure en nombre restreint. L'effort sera immanquablement plus conséquent que pour les mots-clés délaissés mais votre site aura déjà les bases suffisantes (acquises à l'étape précédente) pour que votre effort ne soit pas vain. En contrepartie, ils vous apporteront un trafic très ciblé à fort pouvoir transformant.

Lorsque vous aurez validé avec brio les deux étapes précédentes, vous pourrez alors prétendre à davantage. Votre site web drainera du trafic transformant et aura acquis en notoriété. Fort de cette expérience acquise, vous pourrez alors partir à la conquête des mots-clés concurrentiels. L'effort sera encore plus important qu'aux étapes précédentes et les résultats mettront plus de temps à se faire sentir, mais vous aurez acquis l'expérience suffisante pour mener à bien cette tâche.

La longue traîne

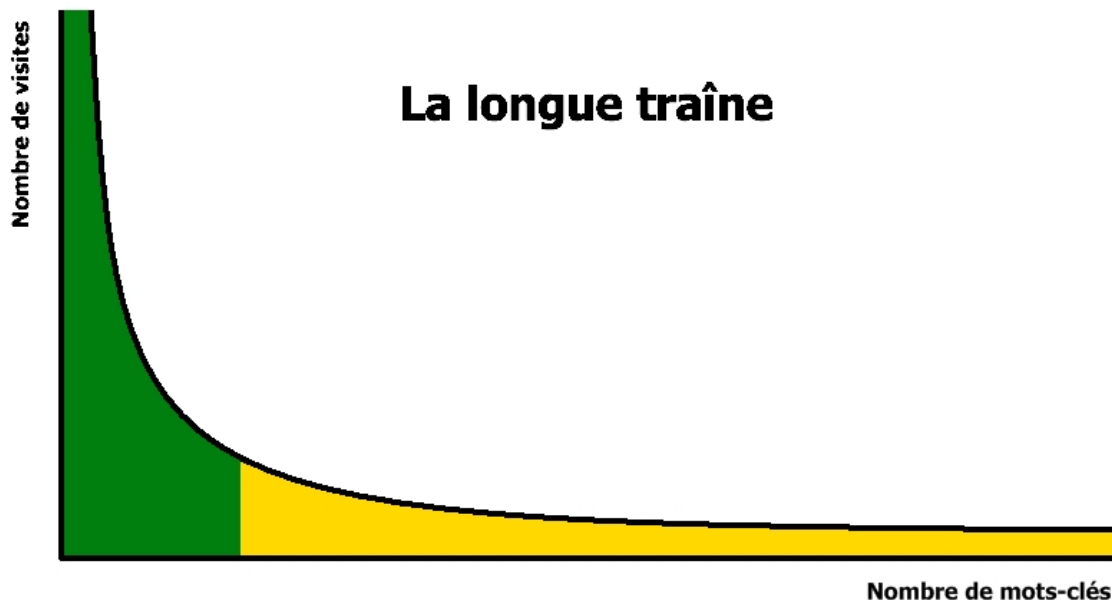
L'expression longue traîne (long tail en anglais) n'est pas spécifique au référencement.

En 2004, une analyse des parts de marché de grands sites e-commerce (Amazon, etc.) met en évidence un phénomène jusque là ignoré : la longue traîne. Les produits les moins demandés (très nombreux) engendrent un chiffre d'affaires plus élevé que les produits les plus populaires (peu nombreux). La diversité de l'offre vient compenser le peu de ventes qu'occasionne chaque produit.

Ce phénomène vient contrecarrer la célèbre loi de Pareto (loi des 80/20) qui dit qu'en vente, 20% des produits engendrent 80% du chiffre d'affaires.

La longue traîne, kezako ?

En collectant le nombre de visites qu'apporte chaque mot-clé à un site web, il est possible de tracer le graphique suivant.



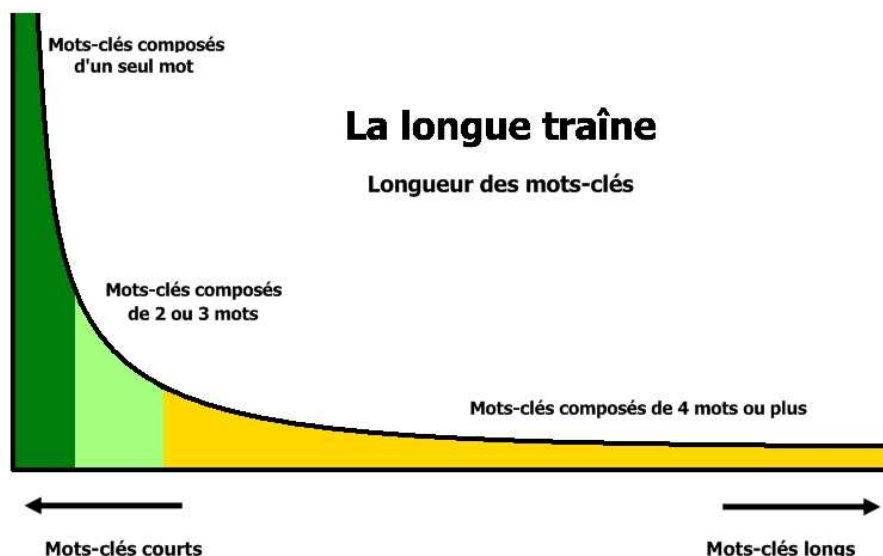
Ce graphique nous dit que :

- les mots-clés à fort trafic sont peu nombreux,
- les mots-clés à trafic faible sont très nombreux.

La partie en jaune sur le graphique correspond à ce qu'on appelle la longue traîne. Ici elle est coupée, mais en réalité, elle s'étire très loin vers la droite.

Ce graphique nous dit donc que le trafic total engendré par les mots-clés à trafic faible est supérieur au trafic total engendré par les mots-clés à fort trafic. En pratique, ce trafic issu des mots-clés à trafic faible peut constituer plus de 80% du trafic total.

Longue traîne et longueur de mots-clés

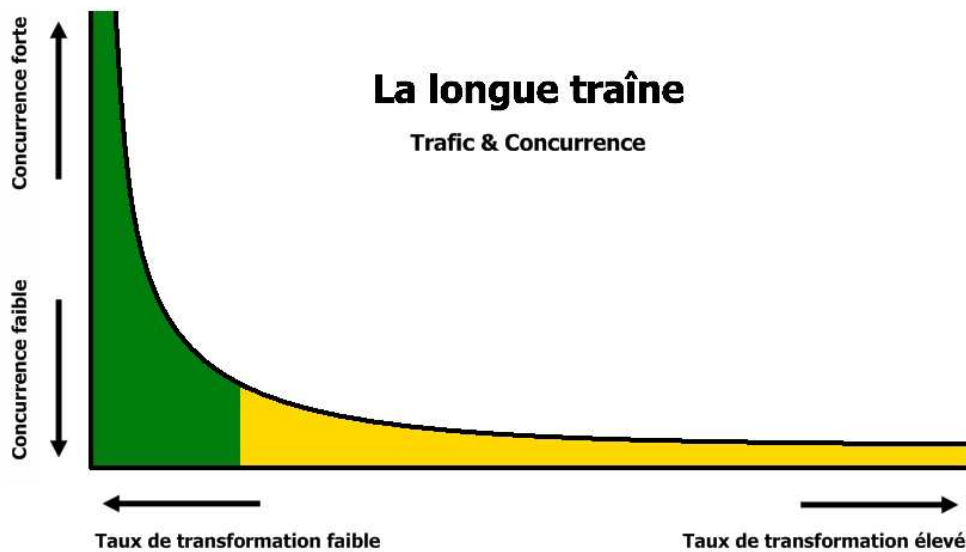


En regardant la longue traîne d'un peu plus près, nous constatons ce qui suit.

De manière générale, les mots-clés appartenants à la longue traîne se composent de plusieurs mots, quatre et souvent davantage. De la même manière, les mots-clés à fort trafic sont habituellement composés de peu de mots, un et rarement plus.

Attention, il est important de pondérer ce constat, qui repose sur une moyenne. Dans un marché très concurrentiel, les mots-clés concurrentiels pourront parfois aller jusqu'à trois mots. Alors que dans un marché peu disputé les mots-clés de longue traîne pourront quelquefois commencer à trois mots.

La longue traîne, trafic et concurrence



En poursuivant notre analyse de la longue traîne, des tendances se dégagent en matière de trafic et de concurrence.

Les mots-clés qui génèrent un important trafic sont aussi souvent les plus concurrentiels autrement dit, ceux pour lesquels l'effort nécessaire pour bien se positionner dans les SERP sera le plus grand. Les mots-clés qui génèrent un faible trafic (ceux de la longue traîne) sont sans concurrence. Ils sont également ceux qui présentent le taux de transformation le plus élevé.

Attention, lorsqu'un mot-clé amène trois visites mensuelles et qu'une de ces visites est transformée (achat, inscription newsletter, etc.), le taux de transformation est d' $\frac{1}{3} > 33\%$, ce qui est certes très élevé, mais qui demande à être relativisé vu le nombre total de visites (ici, trois). Un mot-clé qui vous apporte 300 visites mensuelles avec un taux de transformation de 5% vous permet de transformer à 15 reprises, donc 15 fois plus que dans le premier cas.

Autre mise en garde, la longue traîne n'est pas l'eldorado : tout ce qui s'y trouve n'est pas merveilleux. Elle se compose aussi de beaucoup de mots-clés déchets (fautes d'orthographe, expressions dépourvues de sens, etc.) qui n'amènent qu'un trafic sans intérêts. Des observations qui précèdent, il n'est donc pas possible de tirer de règle systématique.

Longue traîne et typologie des mots-clés

La longue traîne est simple en apparence. Mais en matière de référencement, elle n'est pas aussi triviale qu'elle le semble au premier abord. Dans le premier chapitre de cette deuxième partie, nous avons défini une typologie des mots-clés : concurrentiels, parasités, de niche ou délaissés. Voyons maintenant comment cette typologie se répartit sur notre graphique de longue traîne.



Attention, pour bien interpréter ce graphique, il est nécessaire d'avoir en mémoire les mises en garde que nous avons faites dans les deux sections précédentes. Le graphique de longue traîne ne dissocie pas les mots-clés concurrentiels des mots-clés parasités ce qui rend délicat l'interprétation de la partie grisée. Ces mots-clés apportent un trafic important, mais seule une partie d'entre eux possède un taux de transformation intéressant : les mots-clés concurrentiels. Alors que les mots-clés parasités drainent un fort trafic mais peu qualifié.

Rappelez-vous, nous vous invitons à la fin du chapitre précédent à commencer votre travail de référencement par les mots-clés délaissés, puis de niche et enfin les mots-clés concurrentiels. Ce n'est ni plus, ni moins que le parcours du graphique de longue traîne en sens inverse.

Comment se traduit la longue traîne pour mon site web ?

Tout ce qui précède est valable pour tous les sites web, donc également pour le vôtre. En consultant les statistiques de votre site web et les mots-clés qui vous amènent des visiteurs, vous devez pouvoir observer une foule de mots-clés, en général assez long (quatre mots et plus), qui chacun vous amène chaque mois quelques visites. Vous pouvez même contrôler mois après mois si votre longue traîne s'étend (quantité de mots-clés souvent longs apportant peu de visites). Si c'est le cas, cela indique que votre site web colonise progressivement les SERP, c'est bon signe.

Comment tirer profit de la longue traîne ?

Ne faites pas de fixation sur vos mots-clés favoris, la part de trafic qu'ils vous apportent n'est pas aussi importante que souvent vous le pensez. Il vous faut immanquablement travailler sur un ensemble de mots-clés représentatifs de votre offre : c'est votre base de travail. Mais il vous faut songer à décliner cet ensemble de mots-clés au maximum.

Comment ? En veillant à varier les formulations, les synonymes, etc. De cette manière, votre référencement ne perdra pas en qualité puisque votre contenu restera ciblé et homogène. Bien au contraire, avec un contenu de même volume, simplement plus varié votre surface d'exposition ne sera que plus grande.

Des statistiques précises pour vos mots-clés

Nous terminons notre partie sur les mots-clés par la présentation d'un outil incontournable pour mener à bien votre tâche.

Google propose pour la gestion des campagnes Adwords un outil de suggestion de mots-clés. Ce générateur de mots-clés est gratuit et peut être utilisé en dehors d'une campagne Adwords pour votre référencement.

Il est accessible à l'adresse suivante : <https://adwords.google.fr/select/KeywordToolExternal>.



Générateur de mots clés

Utilisez le Générateur de mots clés pour obtenir de nouvelles suggestions de mots clés. Sélectionnez une des deux options proposées ci-dessous selon que vous souhaitez entrer des mots clés ou bien l'adresse URL de votre site. [Conseils d'utilisation du Générateur de mots clés](#)

Important : Le Générateur de mots clés recense de façon automatique une liste de requêtes courantes effectuées sur le moteur de recherche de Google, en rapport avec le mot clé que vous avez saisi. Les termes énumérés ne vous sont pas conseillés ni recommandés et Google ne peut garantir que ces mots clés amélioreront les performances de votre campagne. Nous nous réservons également le droit de désapprouver tout nouveau mot clé sélectionné. Vous êtes seul responsable des mots clés que vous sélectionnez et devez vous assurer que leur utilisation n'enfreint aucune législation ni ne porte atteinte aux droits de tiers, notamment au regard du droit des marques et de la concurrence déloyale. Avant de sélectionner un mot clé, vous devriez vérifier qu'il ne s'agit pas d'un terme protégé (marque, nom commercial, dénomination sociale) en consultant un registre des marques (ex: www.icimarkes.com) et des sociétés (ex: www.euridile.com).

Les résultats sont adaptés au paramètre suivant : Français, France [Modifier](#)

L'intérêt de cet outil est qu'il vous propose des données chiffrées précises et précieuses pour déterminer les mots-clés sur lesquels vous positionner. Vous y trouverez notamment :

- des listes de mots-clés suggérés proches des vôtres,
- des listes de mots-clés suggérés à partir de vos pages web,
- des chiffres sur la concurrence,
- des chiffres sur les volumes de recherche des internautes.

Suggestion de mots-clés à partir d'un ou plusieurs mots-clés

Pour vous en servir,

- renseignez un ou plusieurs (un par ligne) mots-clés,
- élargissez éventuellement la suggestion aux synonymes en cochant la case **Utiliser des synonymes**,
- renseignez le captcha,
- puis cliquez sur Trouvez des idées de mots-clés.

Les mots-clés

Comment souhaitez-vous générer des idées de mots clés ?

☒ Expressions ou termes descriptifs
(exemple : thé vert)

☐ Contenu de site Web
(exemple : www.exemple.fr/produit?id=74893)

Entrez un mot clé ou une expression par ligne :

baladeur mp3
lecteur mp3 pas cher

☒ Utiliser des synonymes

[Filtrer mes résultats](#)

[Trouver des idées de mots clés](#)

Sélectionnez les colonnes à afficher : [Afficher/masquer les colonnes](#)

Type de ciblage : [Large](#)

| Mots clés | Concurrence entre annonceurs | Volume de recherche locale : avril | Volume de recherche mensuel global | |
|---|------------------------------|------------------------------------|------------------------------------|-------------------------|
| Mots clés en rapport avec le(s) terme(s) entré(s) - trié par pertinence | | | | |
| baladeur mp3 | <div></div> | 165 000 | 165 000 | Ajouter |
| baladeurs mp3 | <div></div> | 8 100 | 12 100 | Ajouter |
| lecteur mp3 pas cher | <div></div> | 9 900 | 5 400 | Ajouter |
| baladeur mp3 usb | <div></div> | 2 400 | 1 600 | Ajouter |
| lecteur baladeur mp3 | <div></div> | 40 500 | 1 900 | Ajouter |
| baladeur mp3 fm | <div></div> | 18 100 | 12 100 | Ajouter |
| achat baladeur mp3 | <div></div> | 210 | 390 | Ajouter |
| prix baladeur mp3 | <div></div> | 390 | 390 | Ajouter |
| acheter baladeur mp3 | <div></div> | 210 | 260 | Ajouter |
| comparatif baladeur mp3 | <div></div> | 590 | 880 | Ajouter |
| baladeur mp3 1go | <div></div> | 2 400 | 1 900 | Ajouter |
| baladeur mp3 creative | <div></div> | 4 400 | 1 900 | Ajouter |
| baladeur mp3 tuner | <div></div> | 1 600 | 590 | Ajouter |
| meilleur baladeur mp3 | <div></div> | 480 | 590 | Ajouter |

L'outil vous affichera alors un tableau présentant les mots-clés susceptibles de vous intéresser. Vous trouverez notamment comme information pour chaque mot-clé :

- la concurrence entre annonceurs,
- le volume de recherche locale et saisonnière,
- et le volume de recherche mensuel global.

Il est possible de trier ces informations par colonne et de les exporter vers une feuille Excel.

Suggestion de mots-clés à partir d'une page web

Pour vous en servir,

- renseignez une page de votre site web,
- élargissez éventuellement la suggestion aux pages proches en cochant la case **Inclure d'autres pages de mon site auxquelles cette URL permet d'accéder**,
- renseignez le captcha,
- puis cliquez sur Trouvez des idées de mots-clés.

Les mots-clés

Comment souhaitez-vous générer des idées de mots clés ?

☐ Expressions ou termes descriptifs
(exemple : thé vert)

☒ Contenu de site Web
(exemple : www.exemple.fr/produit?id=74893)

Entrez une URL de page Web pour rechercher des mots clés en rapport avec le contenu de la page. [?](#)
<http://www.fnac.com>
Inclure d'autres pages de mon site auxquelles cette URL permet d'accéder

► Ou bien saisissez votre propre texte dans la zone de texte ci-dessous (facultatif).

► [Filtrer mes résultats](#)

Sélectionnez les colonnes à afficher : [?](#)
Afficher/masquer les colonnes
☒ Regrouper les mots clés par termes communs

Affichage des mots clés regroupés en fonction des termes suivants :
[appareil photo numérique](#) (5), [jeux gratuits](#) (7), [xbox 360](#) (5), [jeux pc](#) (5), [jeux pour](#) (5), [jeu](#) (70), [jeu](#) (39), [appareil](#) (10), [wii](#) (5), [dvd](#) (5), [Mots clés divers](#) (44)

| Mots clés | Concurrence entre annonceurs ? | Volume de recherche locale - avril ? | Volume de recherche mensuel global ? | Type de ciblage : ? Large |
|--|--|--|--|--|
| Mots clés concernant appareil photo numérique - trié par pertinence ? | | | | |
| appareil photo numérique pas cher | <div><div></div></div> | 4 400 | 1 600 | Ajouter v |
| vente appareil photo numérique | <div><div></div></div> | 590 | 880 | Ajouter v |
| appareil photo numérique | <div><div></div></div> | 246 000 | 201 000 | Ajouter v |
| achat appareil photo numérique | <div><div></div></div> | 1 600 | 1 900 | Ajouter v |
| appareille photo numérique | <div><div></div></div> | 2 400 | 1 900 | Ajouter v |
| Tout ajouter - 5 » | | | | |
| Mots clés concernant jeux gratuits - trié par pertinence ? | | | | |
| jeux gratuits pour mobile | <div><div></div></div> | 8 100 | 8 100 | Ajouter v |
| jeux gratuits | <div><div></div></div> | 5 000 000 | 4 000 000 | Ajouter v |

Vous y trouverez alors les mêmes informations et possibilités que pour la suggestion à partir des mots-clés. Les mots-clés suggérés seront regroupés autour des expressions les plus fréquentes de votre page web et les plus précises : expressions composées de trois mots, puis de deux mots et enfin d'un seul mot. Ceci confirme la nécessité de cibler précisément vos mots-clés (expressions longues plutôt que courtes) en rapport avec le contenu de vos pages web.

L'optimisation du contenu

Dans cette troisième partie, nous abordons le versant optimisation de contenu sous tous les angles.

Vous avez levé tous les obstacles liés à l'indexation de votre site web. Vous avez défini votre marché, ou plutôt vous l'avez formulé avec le langage du public que vous visez et qui est à votre portée. Il est temps maintenant de passer au travail d'optimisation de votre contenu.

L'optimisation du contenu HTML est la face la plus connue du référencement. Il est vrai qu'aux premiers temps du web, il fallait peu de choses pour être bien positionné. Un simple travail au niveau du contenu HTML suffisait quasiment une bonne fois pour toute.

Les choses ont bien changé ! Le référencement est devenu un travail de forçat qui demande aujourd'hui à être actualisé régulièrement.

Dans un premier chapitre nous revenons sur les principes clés de l'optimisation de contenu : le décompte et la densité des mots-clés. Impossible de faire l'impasse, ils interviennent à tous les niveaux de l'optimisation de contenu.

Le deuxième chapitre sera consacré à l'optimisation des urls de vos pages web, dimension souvent négligée. Vous y trouverez des conseils pour le choix de votre nom de domaine, de votre hébergement, sur le bon format des urls (celui qui plaît aux moteurs de recherche). Une sous-partie sera également consacrée à l'url-rewriting, technologie souvent incontournable en matière de référencement.

Dans le troisième chapitre, nous nous intéresserons au code HTML de vos pages web. Nous passerons en revue la plupart des balises HTML importantes pour le référencement. Vous saurez ce que disent les standards, en quoi ces balises infléchissent votre référencement, comment les calibrer, etc.

Enfin dans le dernier chapitre, nous nous attarderons sur un problème de plus en plus fréquent sur Internet : le duplicate content. Les sites web qui reprennent tout ou partie de votre contenu peuvent pénaliser votre référencement. Vous saurez comment vous en prémunir.

ATTENTION, en référencement l'optimisation n'est pas tout.

L'optimisation est **NECESSAIRE**.

L'optimisation n'est pas **SUFFISANTE**.

Deux principes clés de l'optimisation

Le décompte et la densité font partie des paramètres utilisés par les moteurs de recherche pour juger de la pertinence d'une page web par rapport à un mot-clé. La densité ou IDM (Indice de Densité d'un Mot-clé) semble plus importante que le décompte. Il ne faut néanmoins pas négliger le décompte.

Décompte

Le décompte renvoie simplement au nombre d'occurrence d'un mot-clé dans un texte.

Densité

La densité est un peu plus subtile, mais reste accessible. C'est le rapport entre le décompte et le nombre total de mots. Elle mesure l'importance de la présence d'un mot-clé par rapport à l'ensemble d'un texte. Si un mot-clé apparaît 2 fois sur un total de 400 mots ($2/400 = 0,5\%$), il est moins présent que s'il apparaît 1 fois pour un total de 30 mots ($1/30 = 3,3\%$).

Exemple détaillé

Dans le texte suivant,

« Le football, appelé soccer aux États-Unis et au Canada par opposition au football américain et au football canadien, est un sport collectif qui oppose deux équipes de onze joueurs, dont le but est de mettre un ballon sphérique dans le but adverse, sans utiliser les bras. Le football est le sport le plus populaire dans le monde. »

Le mot-clé football apparaît à 4 reprises et il y a au total 57 mots. Nous avons donc un décompte de 4 et une densité de $4/57 = 7\%$, ce qui est très élevé.

Pour améliorer la densité d'un mot-clé, on peut soit augmenter le décompte, soit réduire le nombre total de mots.

Transformons le texte précédent comme suit :

« Le football, appelé soccer aux États-Unis et au Canada par opposition au football américain et au football canadien, est un sport opposant deux équipes de onze joueurs, ayant pour but de mettre un ballon dans le but adverse, sans utiliser les bras. Le football est un sport très populaire. »

Le mot-clé football apparaît toujours à 4 reprises et il y a au total 49 mots. Nous avons donc un décompte de 4 et une densité de $4/49 = 8,2\%$, soit un gain de $1,2\%$, ce qui est encore plus élevé. La signification est très proche mais la densité est meilleure, ce qui favorise davantage un bon référencement.

Mise en garde

Le référencement n'est pas une science exacte. Les moteurs de recherche se gardent bien de livrer le secret de leurs algorithmes. Nous nous trouvons face à des boîtes noires et seule l'expérimentation fournit des pistes sérieuses. La plupart des conseils relèvent donc de l'expérience, d'observations et de tests.

Ces deux critères de décompte et densité sont valables pour toutes les balises HTML : title, a/href, img/alt, h1, strong, etc. Ils le sont également pour les urls. De manière générale pour un bon référencement, il faut un décompte et une densité supérieurs à la moyenne (des

meilleurs, soit le top 10) pour les mots-clés visés, sans toutefois tomber dans l'excès (spamdexing, parfois pénalisé en terme de positionnement).

<http://www.gnoztik.com>

L'optimisation des urls

Les urls

L'optimisation des urls est un critère majeur. Le choix de bonnes urls est capital pour un bon référencement.

Anatomie d'une url

Une url se compose de plusieurs éléments. Nous ne verrons ici qu'une présentation simplifiée. Pour notre propos, nous nous intéresserons seulement aux éléments qui nous importent :

- le sous-domaine (facultatif),
- le nom de domaine (obligatoire),
- le tld (Top Level Domain) (obligatoire),
- le nom du répertoire (facultatif),
- et le nom du fichier (facultatif).

Dans l'url http://sport.ados.fr/news/guy-roux-de-retour-sur-les-terrains_article5263.html on a,

- le sous-domaine **sport**,
- le nom de domaine **ados.fr**,
- le tld (Top Level Domain) **fr**,
- le nom du répertoire **news**,
- et le nom du fichier **guy-roux-de-retour-sur-les-terrains_article5263.html**.

Décompte et densité dans une url

Comment fonctionnent décompte et densité avec les urls ?

Décompte

Le mot-clé **sport** apparaît

- 0 fois dans l'url suivante : <http://www.lequipe.fr/>,
- 1 fois dans les urls suivantes : <http://www.sports.fr/>,
<http://fr.wikipedia.org/wiki/Sport>, <http://sport.aol.fr/>.
- 2 fois dans les urls suivantes : <http://www.sports.fr/sport.html>,
<http://sport.aol.fr/sport/index.html>.

Densité

Le mot-clé **sport** apparaît

- 0 fois dans l'url suivante : <http://www.lequipe.fr/> et l'url contient 3 mots : **www**, **lequipe**, **fr**, la densité est donc de $0/3 = 0\%$.
- 1 fois dans l'url suivante : <http://fr.wikipedia.org/wiki/Sport> et l'url contient 5 mots : **fr**, **wikipedia**, **org**, **wiki**, **Sport**, la densité est donc de $1/5 = 20\%$.

- 2 fois dans l'url suivante : <http://sport.aol.fr/sport/index.html> et l'url contient 6 mots : **sport**, **aol**, **fr**, **sport**, **index**, **html**, la densité est donc de $2/6 = 33\%$.

Pour améliorer la densité d'un mot-clé, on peut augmenter le décompte ou réduire le nombre total de mots. Rappelez-vous, de bonnes valeurs en décompte et densité sont nécessaires pour optimiser vos urls.

La densité passe à $3/6 = 50\%$ lorsqu'on transforme <http://sport.aol.fr/sport/index.html> en <http://sport.aol.fr/sport/sport.html>.

En transformant <http://sport.aol.fr/sport/index.html> en <http://sport.aol.fr/sport.html> la densité passe à $2/5 = 40\%$.

Problème

Qu'en est-il de l'url <http://www.sportweek.fr> pour le mot-clé sport ?

Le décompte est-il à 0 ou 1 ? Et la densité est-elle à 0/3 ou 1/3 ?

Apparemment Google détecte le mot sport (en gras) dans l'url.

[Sportweek : Foot, Rugby, Tennis, Formule 1 - lives, résultats et ...](#)
Sportweek : le site de tous les passionnés de **sport**. Retrouvez toute l'actualité sportive en direct du foot, rugby, tennis, formule 1, les interviews, ...
www.sportweek.fr/ 238k - [En cache](#) - [Pages similaires](#)

Pour pouvoir définitivement trancher, il faudrait faire des tests complémentaires. Comme nous le disions dans la partie sur le décompte et la densité, le référencement n'est pas une science exacte.

Optimisation

Optimiser une url sur un mot-clé, c'est améliorer les décompte et densité de ce mot-clé dans l'url. Modifier les urls n'est pas toujours chose facile. Aussi, il est impératif de faire les bons choix dès le départ.

Pour faire cette optimisation, trois points importants sont à prendre en considération :

- le nom de domaine,
- la structure des répertoires et fichiers (arborescence) de votre site web,
- la réécriture d'urls (url-rewriting).

Nous nous proposons d'aborder ces points en détail dans la suite de ce document.

Le nom de domaine

Pour aller droit au but, nous dirons qu'un nom de domaine à soi est chose quasi-incontournable pour un référencement digne de ce nom.

Hébergement gratuit ou payant ?

Il existe foule d'hébergeurs gratuits avec des niveaux de qualité de service variables. Il est possible d'y recourir au début d'un projet « pour voir ». Néanmoins votre crédibilité est sérieusement entachée. En effet, un hébergement professionnel est un investissement peu coûteux (moins de cinquante euros à l'année). Si vous n'êtes pas prêt à faire cet effort, n'attendez pas de vos visiteurs qu'ils en fassent davantage et achètent vos produits. Vous montrez par là que vous ne croyez pas à ce que vous faites.

Enfin, votre référencement s'en trouvera nettement amélioré. Vous aurez un nom de domaine qui colle à votre activité, vous pourrez créer vos propres sous-domaines, organiser vos fichiers et répertoires à votre guise, mettre en place l'url-rewriting. En bref, vous pourrez faire tout ce qui profite à un bon référencement.

*Pour vous en convaincre à la date du 8 mai 2009, Google renvoyait pour le mot-clé **sport** en **Pages francophones** les résultats suivants :*

- | | | |
|----|---|-----------------------|
| 1 | http://www.sport.fr/ | domaine |
| 2 | http://www.sport.fr/Football/ | domaine |
| 3 | http://www.lequipe.fr/ | ABSENT |
| 4 | http://www.sports.fr/ | domaine |
| 5 | http://fr.sports.yahoo.com/ | sous-domaine |
| 6 | http://www.sport24.com/ | domaine |
| 7 | http://sportplus.canalplus.fr/ | sous-domaine |
| 8 | http://www.sportweek.fr/ | domaine |
| 9 | http://www.wat.tv/guide/sport | répertoire ou fichier |
| 10 | http://www.eurosport.fr/ | domaine |

9 sur 10 contiennent le mot sport dans l'url, 6 sur 10 l'ont dans leur nom de domaine, 2 sur 10 dans un sous-domaine, 1 sur 10 en répertoire ou en fichier. Eloquent, non ?

Apparemment, le nom de domaine à plus de poids que les sous-domaines et les répertoires ou fichiers. Ce n'est pas surprenant, ils sont plus difficiles à manipuler. Le seul site web qui ne comporte pas sport dans son url est un site à très forte notoriété en matière de sport, ce qui compense largement.

Attention, avoir le mot-clé dans l'url ne suffit pas à être bien positionné. Il est nécessaire d'avoir un bon contenu, de bons backlinks, etc. Mais à optimisation égale, l'efficacité est radicale.

Choix du nom de domaine

Il faut lever un maximum d'obstacles dans la promotion de vos produits. Le nom de domaine se doit donc d'être simple et facile à retenir. Dans la mesure où vous disposez de peu de caractères (maximum 20 et c'est déjà beaucoup), vous ne pourrez pas y placer beaucoup de mots-clés, il faut donc bien les choisir. Il est impératif que votre nom de domaine soit centré sur votre activité.

Les sites web sur le sport du classement précédent sont instructifs <http://www.sport.fr/>, <http://www.sport24.com/>, <http://www.sportweek.fr/>, <http://www.eurosport.fr/>, etc.

Nous avons vu avec <http://www.sportweek.fr> que Google détecte les mots-clés « collés » dans les urls. Cependant, la prudence reste de rigueur car il n'en est pas forcément de même pour tous les moteurs de recherche.

Enfin, le séparateur le plus approprié est le tiret « - », le séparateur underscore « _ » est à proscrire.

Par exemple, <http://www.sport-plus.fr/> est préférable à http://www.sport_plus.fr/.

Le TLD

Le TLD (Top Level Domain) sont les extensions du type .com, .org, .net, .fr, etc. Ces extensions ne sont pas sans valeurs. De manière générale, il est préférable de choisir une extension adaptée à vos besoins. Si vous vous adressez à un public français, le .fr est adapté et sera probablement un facteur favorisant votre référencement sur les versions françaises des moteurs de recherche. Le .com est très intéressant pour les sites multilingues ou pour la simple raison que l'internaute a tendance à rajouter spontanément un .com.

On notera, parmi les sites web sur le sport du classement précédent, le manque de cohérence de <http://www.sportweek.fr/> qui associe l'anglais à un tld en .fr.

Les sous-domaines

Il est peu probable que vous trouviez un hébergement gratuit vous offrant la possibilité de créer des sous-domaines à votre guise. Prenez néanmoins garde à l'hébergement payant que vous choisirez, car tous ne le permettent pas.

Les sous-domaines sont intéressants à plus d'un titre. Ils sont une occasion supplémentaire de placer des mots clés dans vos urls. Apparemment, les mots-clés en sous-domaine sont plus efficaces que ceux dans les répertoires et les fichiers. Certains moteurs de recherche considèrent que chaque sous-domaine est un site différent. De ce fait, les liens entre ces sous-domaines sont considérés comme des backlinks externes (venant d'un autre site web, ce qui est meilleur qu'un backlink interne, venant du même site web).

Pour un site web sur le sport, on pourra définir un sous-domaine spécifique par sport : football.sport-plus.fr, natation.sport-plus.fr, golf.sport-plus.fr, etc.

Les répertoires et les fichiers

Aussi bien pour les moteurs de recherche que pour les internautes, une organisation cohérente et hiérarchique est préférable.

Prenons l'exemple de la page d'un catalogue en ligne vendant des chaussures de sport homme pour le basket. Laquelle des deux urls qui suivent vous semble la plus appropriée ?

<http://www.monsite.com/catalogue/produit.php?rubrique=17&reference=247>

ou

<http://www.monsite.com/chaussures-de-sport/homme/basket/>

Comment concrètement mettre en place cette organisation ?

Il est en premier lieu nécessaire de classer vos contenus, puis de définir une arborescence logique. Cette arborescence doit aller dans le sens de la spécialisation des contenus.

Une organisation hiérarchique cohérente répond aux exigences suivantes :

- chaque niveau (répertoire) est l'équivalent d'une catégorie pour ses sous-niveaux (répertoires ou fichiers),
- chaque niveau (répertoire ou fichier) doit être nommé à l'aide de mots-clés représentatifs des contenus qu'il dispense,
- chaque sous-niveau (répertoire ou fichier) a une désignation plus précise (spécialisation) que le niveau (répertoire) dont il est issu,

- le nombre de niveaux est limité (trois semble une bonne limite).

Pourquoi m'embarrasser à organiser mes contenus ?

Les moteurs de recherche, les internautes et même vous y trouveraient un avantage.

Les moteurs de recherche

Côté référencement, vos urls

- sont optimisées au mieux par la présence des mots-clés,
- décrivent exactement ce qu'elles contiennent ce qui renforce l'homogénéité de leur contenu,
- sont contextualisées à l'extrême puisque chaque sous-niveau précise le niveau dont il est issu.

Les internautes

Les sites web sont souvent des dédales dans lesquels il est difficile de se repérer.

La présence des catégories et sous-catégories dans l'url fournit une ébauche de plan aux internautes. Ils sauront apprécier. Cette boussole leur facilitera le parcours de votre site web et sera pour eux une incitation à cheminer plus profondément au sein des différentes catégories de votre site web qui désormais se laissent deviner.

Enfin, la simple lecture de l'url les renseignera sur le contenu que la page web correspondante dispense. Vous leur ferez gagner du temps. N'oubliez pas que vous devez rapidement transformer votre trafic entrant sous peine de ne rien transformer du tout. Faire perdre du temps aux internautes parce que la navigation que vous leur proposez est chaotique peut, par agacement, les pousser hors de votre site web. Ne les faites pas fuir alors que vous avez tant peiné pour les attirer chez vous !

Vous-même

Vous y verrez incontestablement plus clair !

L'url rewriting

Manipuler les urls n'est pas chose facile. De manière générale, mieux vaut y songer avant de mettre en place un nouveau site web. Car après, le travail se complique sérieusement. Il n'y a néanmoins pas de fatalité. Si vos urls ne sont pas très optimisées coté référencement, il vous est possible de recourir à une technique appelée url-rewriting.

Pourquoi bricoler mes urls ?

En naviguant sur le web, vous avez probablement déjà remarqué que certaines pages ont une url assez compliquée.

Par exemple,

<http://www.monsite.com/catalogue/produit.php?rubrique=32&reference=2647>

De telles urls

- ne facilitent pas le travail d'indexation des moteurs de recherche (bien qu'ils s'en sortent quand même dans la plupart des cas),
- n'aident pas les internautes à la mémorisation,

- et surtout négligent un aspect souvent décisif en matière de référencement, à savoir l'insertion de mots-clés dans les urls.

Elles sont donc une entrave de taille. La réécriture d'url est un procédé qui apporte une solution à ce problème. Elle organise la mise en place d'urls « propres » pour accéder à vos pages web, ce qui répond aux trois objections soulevées plus haut.

Prenons l'exemple de la page d'un catalogue en ligne vendant un ballon de football dont l'url serait <http://www.monsite.com/catalogue/produit.php?rubrique=32&reference=2647>

La réécriture d'url permet d'accéder à la même page via une URL telle que, <http://www.monsite.com/catalogue/ballon-football-32-2647.html>

Cette url est plus lisible (pour les moteurs de recherche comme pour les internautes) et elle est optimisée pour un référencement sur les mots-clés ballon et football.

L'url rewriting, comment ça marche ?

Désormais lorsque cela sera nécessaire, les pages de votre site web auront deux urls :

- une interne (« sale » au sens du référencement),
- et une externe (« propre » au sens du référencement).

Les moteurs de recherche et les internautes ne connaîtront plus dorénavant que les urls externes.

Votre serveur web ne comprend véritablement que les urls internes. Vous devez donc lui fournir des règles de traduction des urls externes vers les urls internes associées. Il sera alors en mesure d'honorer les requêtes externes faites par les crawlers des moteurs de recherche et les internautes. Lorsqu'ils demanderont à votre serveur web la page renvoyant à l'url externe, il opérera une traduction en url interne et retournera la page correspondante.

Bien entendu, à aucun moment, les moteurs de recherche et les internautes n'en auront connaissance.

Comment lire une URL à réécrire ?

Une url à réécrire se présente souvent sous la forme suivante :

<http://www.monsite.com/mapage.php?variable1=valeur1&variable2=valeur2>

Elle se compose comme suit :

- la partie située avant le point d'interrogation (?) est l'url de base,
- la partie située après le point d'interrogation (?) est une liste de couples (variable = valeur) séparés par un & et transmises à la page web comme paramètres en vue d'un traitement spécifique.

L'url <http://www.monsite.com/catalogue/produit.php?rubrique=32&reference=2647> se découpe donc en deux parties :

- l'url de base <http://www.monsite.com/catalogue/produit.php>,
- et deux variables *rubrique=32* et *reference=2647*.

Comment mettre en place la réécriture d'une URL ?

Activez d'abord la réécriture d'url sur votre site web

La mise en place de l'url-rewriting dépend du serveur qui héberge votre site web (Apache, Microsoft IIS) et de votre hébergeur. La plupart proposent cette fonctionnalité. Néanmoins, avant de faire de vains efforts, contactez le vôtre pour savoir si c'est le cas pour votre hébergement.

Identifiez les urls à réécrire

Il n'est pas nécessaire de réécrire l'ensemble des urls de vos pages web. Les urls doivent remplir les deux conditions suivantes :

- être complexes au sens de ce qui précède (toutes ne le sont pas forcément),
- avoir un intérêt pour votre référencement.

Une boutique en ligne se doit d'avoir des urls optimisées pour les pages qui présentent ses offres promotionnelles, ses fiches produits, etc. Par contre, l'optimisation des conditions générales de vente ne présente pas d'intérêt particulier pour le référencement. L'url

<http://www.monsite.com/maillot-equipe-france-76-42-63543.html>

est préférable à l'url

<http://www.monsite.com/produit.php?rubrique=76&sousrubrique=42&reference=63543>

Comment définir mes règles de réécriture ?

Une règle de réécriture permet d'établir la correspondance entre deux urls. Elle peut être générique, lorsqu'elle s'applique à plusieurs urls de votre site web, ou particulière lorsqu'elle ne concerne qu'une seule url. Cette technique repose sur l'utilisation des expressions régulières (http://fr.wikipedia.org/wiki/Expression_rationnelle). Une seule règle suffit rarement à régler tous les cas d'un site web. Pour votre besoin, vous pouvez donc en écrire autant que nécessaire.

Ce document n'a pas pour but de présenter un cours sur les expressions régulières mais d'illustrer la méthode au moyen d'un exemple. Le langage des expressions régulières est très sophistiqué. Néanmoins, son usage est la plupart du temps élémentaire dans le cadre de l'url-rewriting. Vous trouverez ci-dessous les motifs les plus fréquents utilisés pour l'élaboration de ces expressions régulières.

| Motif | Désignation |
|----------|---|
| [0-9] | un caractère numérique |
| [a-z] | un caractère minuscule (sans accent) |
| [A-Z] | un caractère majuscule (sans accent) |
| [a-zA-Z] | un caractère minuscule ou majuscule (sans accent) |
| [0-9]+ | une suite d'au moins un caractère numérique |
| ^ | signale le début de l'url réécrite |
| \$ | signale la fin de l'url réécrite |

- () les parenthèses capturent la séquence de l'url qu'elles « attrapent »
- \$N dans l'url à réécrire, indique la valeur capturée par la N^{ème} paire de parenthèses selon l'ordre de lecture

Par exemple, décryptons la règle suivante,

`^([0-9a-zA-Z-]+)-([0-9]+)-([0-9]+)\.html$` /produit.php?rubrique=\$2&reference=\$3

Avec cette règle, l'url externe (appelée par les moteurs de recherche et les internautes)

`http://www.monsite.com/maillot-equipe-france-42-63543.html`

devient l'url interne (traitée par le serveur web),

`http://www.monsite.com/produit.php?rubrique=42&reference=63543`

Motif **Effet**

`([0-9a-zA-Z-]+)` capture la séquence **maillot-equipe-France** et la stocke dans **\$1**
\$1 n'est pas utilisé ensuite

`([0-9]+)` capture la séquence **42** et la stocke dans **\$2**

`([0-9]+)` capture la séquence **63543** et la stocke dans **\$3**

`\.` interprète le point littéralement
(sinon, il représente un caractère quelconque)

Attention, le choix des caractères séparateurs de mots est très important. Les séparateurs conseillés sont : le tiret « - », la virgule « , », le slash « / » ou le point « . ». Les séparateurs déconseillés sont : l'underscore « _ », le dièse « # » et l'espace.

Comment mettre en application mes règles de réécriture ?

La mise en application des règles de réécriture varie selon le type de serveur web qui héberge votre site web. Nous ne traitons ici que le cas des serveurs Apache et Microsoft IIS (les plus répandus sur le web).

Apache

Vous devez consigner vos règles de réécriture dans un fichier

- nommé .htaccess (avec un point en premier caractère),
- et situé à la racine de votre site web.

Ce fichier permet d'infléchir la configuration du serveur Apache.

Ci-dessous un exemple :

Code Apache :

```
RewriteEngine on
RewriteRule ^([0-9a-zA-Z-]+)-([0-9]+)-([0-9]+)\.html$
    /produit.php?rubrique=$2&reference=$3
RewriteRule . . .
RewriteRule . . .
```

| Instruction | Signification |
|------------------|--|
| RewriteEngine on | ne doit apparaître qu'une seule fois doit être placée avant toutes les règles de réécriture signale au serveur Apache l'existence de règles de réécriture |
| RewriteRule | signale une règle de réécriture il y en a autant que nécessaire (une par ligne) au format : url à réécrire [espace] url réécrite à chaque url, ne s'applique que la première règle qui peut s'appliquer |

Microsoft IIS

Si vous utilisez une version d'ASP sans la plate-forme .NET, il vous faudra avoir recours à des composants payants. Leur mode de configuration varie. La définition de vos règles de réécriture reste valable. En revanche, leur mise en application est spécifique au composant que vous utiliserez.

Si vous utilisez .NET, il existe plusieurs solutions pour mettre en application vos règles de réécriture :

- <http://www.isapirewrite.com>,
- <http://urlrewriter.net>,
- <http://www.codeproject.com/aspnet/URLRewriter.asp>,
- etc.

Une fois encore, la définition de vos règles de réécriture reste valable mais la mise en application est spécifique à la technologie utilisée.

Vérifiez les liens hypertextes de votre site web

Une fois les règles de réécriture mises en place, mettez à jour dans vos pages web tous les liens hypertextes concernés. Désormais dans chacune de vos pages web, vous ne désignerez plus vos autres pages web qu'avec des urls externes. Ce travail est capital, sous peine de ruiner tous vos efforts.

En effet, si vos pages web sont accessibles à partir de deux urls, les moteurs de recherche considéreront qu'il existe deux pages web différentes au contenu identique. Il y aura alors cas flagrant de duplicate content (voir plus loin). Or le duplicate content peut être pénalisant pour votre référencement.

Il faut savoir que les pages web dont l'url est réécrite restent accessibles à partir de leur url interne (à condition que cette url soit mentionnée quelque part sur une de vos pages web). D'où la nécessité de supprimer toute référence aux urls internes.

Par exemple, la page web dont l'url à réécrire est <http://www.monsite.com/produit.php?rubrique=42&reference=63543> et l'url réécrite <http://www.monsite.com/maillot-equipe-france-42-63543.html> est accessible via ces deux urls. Ne présentez que l'url <http://www.monsite.com/maillot-equipe-france-42-63543.html> dans vos liens hypertextes pour éviter le piège du duplicate content. Votre référencement y gagnera sans aucun doute.

L'optimisation du contenu HTML

Le HTML pour le référencement

Nous abordons maintenant l'optimisation du code HTML proprement dite. Avant d'entrer dans le vif du sujet, nous devons faire quelques rappels à propos du langage HTML.

HTML, le minimum

Ces rappels représentent le minimum de connaissances nécessaires en HTML sans lesquelles il est vain d'espérer pouvoir optimiser vos pages web. Ceux qui sont familiers du HTML n'y apprendront rien de nouveau et y trouveront même quelques approximations à portée purement pédagogique.

HTML, Kezako ?

A chaque page web, que vous visitez sur Internet, est associé un fichier au format HTML que votre navigateur (Internet Explorer, Opera, Firefox ou autre) reçoit et a la charge d'afficher. Ce fichier au format texte (donc humainement lisible) décrit l'apparence que doit avoir la page web correspondante (texte, images, mise en forme, etc.).

Par exemple, en vous rendant sur la page d'accueil du journal du Net <http://www.journaldunet.com>, vous pouvez afficher le fichier qu'a reçu votre navigateur (clic droit + Afficher la source dans Internet Explorer). Dans un éditeur de texte s'affiche alors le texte suivant (code HTML) de la page d'accueil du journal du Net.

```
<head>
<title>Le Journal du Net</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
<meta name="Description" content="Analyses, tendances, interviews : tous les jours, le Journal du Net vous j
<meta name="Keywords" content="Internet, news, business, t&eacute;acute;l&eacute;coms, publicit&eacute;e, commerce
<meta name="robots" content="all,noodp">
<meta name="Revisit-after" content="2 days">
<link rel="stylesheet" href="/html_externe3/style/journaldunet/magazine/home.css">
<link type="text/css" rel="stylesheet" href="/maa/html_include/style/microapplicationajax/global.css">
<link rel="alternate" type="application/rss+xml" title="Le Journal du Net : A la Une" HREF="http://www.joi
```

Cette description est faite en HTML. Le HTML (Hyper Text Markup Language) est un langage standard qui repose sur des balises (de mise en forme, en général). Il existe des balises pour afficher du texte, pour le mettre en gras, pour insérer des images, etc. Les balises (tags en anglais) sont toujours présentées entre chevrons <nom de la balise>. Le texte mis en forme est la plupart du temps encadré par une balise ouvrante et une balise fermante.

Par exemple,

pour afficher en gras le texte **baladeur mp3**, on utilise la balise **b** (pour bold, gras en anglais) comme suit,

```
<b>baladeur mp3</b>
```

Autre exemple, pour afficher en **titre de premier niveau** le texte **voiture de sport**, on utilise la balise **h1** comme suit,

```
<h1>voiture de sport</h1>
```

Structure d'une page HTML

Une page HTML est divisée en deux sections : un entête (HEAD) et un corps (BODY). On trouve dans l'entête des balises spécifiques d'entête qui fournissent des informations d'ordre

général sur le document à afficher. Et on trouve dans le corps le document à afficher et ses mises en forme par des balises spécifiques de corps.

Voici comment s'organise une page HTML.

Code HTML :

```
<HTML>
<HEAD>
... Balises d'entête -> informations d'ordre général ...
</HEAD>
<BODY>
... Balises de corps -> page web et mise en forme ...
</BODY>
</HTML>
```

Aller plus loin

Les connaissances qui précèdent sont succinctes et sont le minimum nécessaire à la compréhension de l'optimisation du code HTML de vos pages web. Davantage de connaissances en HTML est néanmoins préférable.

Vous trouverez sur le site du zéro <http://www.siteduzero.com/tuto-3-6-0-apprenez-a-creer-votre-site-web.html> un excellent tutorial sur le XHTML, un « HTML amélioré » (compris par les moteurs de recherche). Vous devriez en peu de temps faire d'immenses progrès.

Pour aller encore plus loin, il peut être intéressant de consulter les standards

HTML 4.01 <http://www.la-grange.net/w3c/html4.01/cover.html>

XHTML 1.1 <http://www.la-grange.net/w3c/xhtml11>

Le tag Title

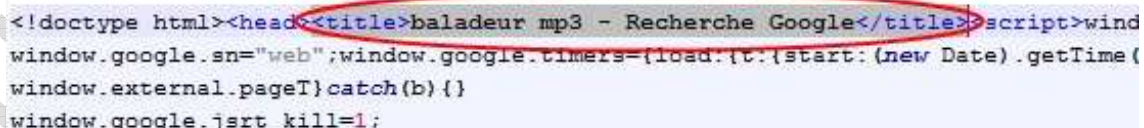
Nous abordons maintenant l'optimisation du code HTML proprement dite. Nous commencerons par la balise Title.

Le tag Title, Kezako ?

Lorsqu'on examine le code html de la page web

http://www.google.fr/search?hl=fr&q=baladeur+mp3&btnG=Rechercher&meta=lr%3Dlang_fr

(clic droit + Afficher la source dans Internet Explorer), on peut y voir en début de fichier le petit bout de code suivant.

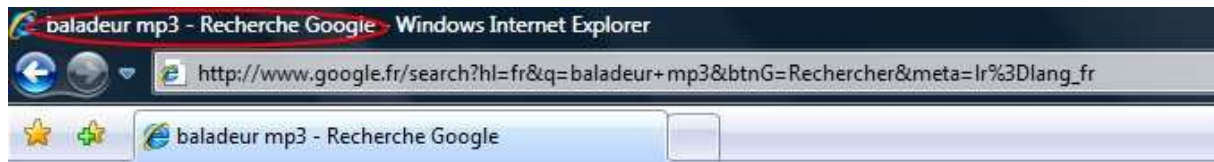


```
<!doctype html><head><title>baladeur mp3 - Recherche Google</title><script>wind
window.google.sn="web";window.google.timers={load:{t:{start:(new Date).getTime(
window.external.pageT} catch(b){}
window.google.jsrt_kill=1;
```

Ce bout de code

```
<title>baladeur mp3 - Recherche Google</title>
```

permet d'afficher dans la barre des titres (d'où le nom de balise Title) de la page de résultats de Google la mention **baladeur mp3 - Recherche Google**.



Le tag Title, selon la norme HTML 4.01

La norme HTML 4.01 dit que :

- chaque document HTML doit avoir un et un seul élément TITLE,
- cet élément TITLE doit se trouver dans la section HEAD (l'entête) du document,
- l'élément TITLE ne fait pas partie du flux de texte (il devrait être affiché comme titre de la page ou de la fenêtre),
- l'élément TITLE peut contenir des caractères accentués ou spéciaux,
- enfin, l'élément TITLE nécessite obligatoirement une balise ouvrante et une balise fermante.

Ci-dessous un exemple d'élément TITLE correctement formé.

Code HTML :

```
<HEAD>
<TITLE>Catalogue de baladeurs mp3</TITLE>
... autres éléments d'en-tête ...
</HEAD>
<BODY>
... corps du document ...
</BODY>
</HTML>
```

Il est vivement conseillé de se conformer à cette norme car les moteurs de recherche apprécient les documents correctement formés, c'est-à-dire formés selon la norme HTML.

Par abus de langage on désigne souvent l'élément TITLE par balise TITLE ou tag TITLE. Pour être clair, l'élément TITLE se compose d'une balise ouvrante (<TITLE>), d'une balise fermante (</TITLE>) et du texte associé, ici **Catalogue de baladeurs mp3**.

Le tag Title est-il vraiment important ?

Cette balise HTML est unique. Il n'y en a qu'un seul exemplaire par page HTML, en général en début de fichier. Dans la mesure où celle-ci ne fait pas réellement corps avec la page web, elle ne se voit quasiment pas. Il est donc légitime de se demander si elle est vraiment importante pour le référencement et donc si il est vraiment nécessaire d'y travailler.

Cette balise HTML est probablement la plus importante de toutes. Les moteurs de recherche y accordent une grande importance. Si il ne fallait en optimiser qu'une seule, ce serait celle-là. Il est donc impératif pour améliorer le référencement de vos pages web de parfaire leur balise TITLE. Car une balise TITLE correctement calibrée, c'est déjà une grande partie du travail de thématisation et de positionnement réalisé.

Comment dois-je rédiger le tag TITLE de mes pages web ?

Les incontournables

Une balise TITLE différente pour chaque page

A chaque page, son titre. Ce titre doit absolument refléter le contenu réel de la page. Il est donc impératif de ne pas répéter le même titre sur toutes les pages de votre site web. Car vous perdriez une occasion de varier vos mots-clés. Dans l'idéal, vous ne devriez pas avoir deux pages de votre site web avec la même balise TITLE.

Un titre intelligible

Il est capital de ne pas faire de vos balises TITLE des successions inintelligibles de mots-clés. Vous devez donner envie à l'internaute de cliquer. Lorsqu'on examine les résultats Google http://www.google.fr/search?hl=fr&q=baladeur+mp3&btnG=Rechercher&meta=lr%3Dlang_fr pour le mot-clé **Baladeur mp3**, on constate que c'est le TITLE des pages (cerclé de rouge dans la capture ci-dessous) qui est rapporté par Google dans sa liste de résultats.



Vous devez donc vous attacher à faire des balises TITLE en forme de slogans attractifs, sans pour autant être racoleurs.

Par exemple,

A conseiller, <TITLE>Lecteur MP3 - Baladeur MP3 - Achat au meilleur prix !</TITLE>.

A éviter, <TITLE>Baladeur Baladeurs, petit prix, achat, vente, comparatif, meilleur prix, baladeur mp3, baladeurs mp3</TITLE>.

Les mots-clés

La balise TITLE de vos pages web doit absolument contenir vos mots-clés les plus importants. De manière générale, ce sont les mots-clés situés le plus à gauche qui auront l'impact le plus fort sur le référencement de vos pages web.

*Par exemple, pour le mot-clé **Baladeur mp3**, le titre*

<TITLE>Baladeur mp3 - Achat baladeurs et lecteurs mp3</TITLE>

est préférable au titre,

`<TITLE>Achat baladeurs et lecteurs mp3 - Baladeur mp3</TITLE>.`

La taille

La taille optimale de la balise TITLE varie d'un moteur de recherche à l'autre. Cependant, les professionnels sont en général d'accord pour fixer cette taille optimale à une soixantaine de caractères, soit environ une dizaine de mots.

Aller plus loin

L'application de ces conseils devrait déjà avoir une incidence notable sur votre référencement. Il vous est néanmoins possible d'aller plus loin.

En tête de fichier

On tâchera de placer ce petit bout de code HTML le plus près possible du début du fichier. Il faudra donc le placer tout de suite après l'ouverture de la balise `<HEAD>` et après le META qui déclare le CHARSET. De la sorte, vous pourrez utiliser les caractères accentués dans vos balises TITLE.

L'exemple suivant spécifie l'encodage de caractères du document comme étant ISO-8859-5 (qui vous permettra l'usage de caractères accentués).

`<META http-equiv="Content-Type" content="text/html; charset=ISO-8859-5">`

Ci-dessous un exemple complet d'élément TITLE correctement formé.

Code HTML :

```
<HTML>
<HEAD>
<META http-equiv=content-type content="text/html; charset= ISO-8859-5">
<TITLE>Baladeurs mp3 - des prix en fêtes !</TITLE>
... autres éléments d'en-tête ...
</HEAD>
<BODY>
... corps du document ...
</BODY>
</HTML>
```

La forme du titre

On privilégiera un TITLE ayant la forme « phrase avec mot-clé + nom du site web » plutôt que « nom du site web + phrase avec mot-clé ».

*Par exemple, pour le mot-clé **Baladeur mp3**, le titre*

`<TITLE>Baladeur mp3 - Achat baladeurs et lecteurs mp3 – HighTechDiscount</TITLE>`

est préférable au titre,

`<TITLE>HighTechDiscount - Baladeur mp3 - Achat baladeurs et lecteurs mp3</TITLE>.`

Mise en garde

La modification à répétition de la balise TITLE d'une page web peut entraîner une pénalité pour les sites web jeune. On modifie la balise TITLE sur des pages qui reçoivent déjà du trafic des moteurs de recherche, pas sur des pages non encore affirmées. Les webmasters de sites jeunes devront donc s'armer de patience.

Le tag Meta/Description

Nous abordons maintenant l'optimisation de notre première balises Meta : la balise Meta/Description.

Le tag Meta/Description, Kezako ?

Lorsqu'on examine le code html de la page web suivante (clic droit + Afficher la source dans Internet Explorer) http://www.topachat.com/pages/g_cat_est_imageson_puis_page_est-lecteur-mp3---baladeur-mp3.html , on peut y voir en début de fichier le petit bout de code suivant,

```
<html>
<head>
<title>Lecteur MP3 - Baladeur MP3 - Achat au meilleur prix !</title>
<META NAME="Description" CONTENT="Top Achat vous rembourse la différence si vous trouvez moins cher ailleurs ! Vente
Lecteur MP3 - Baladeur MP3 au meilleur prix ! Top Achat, le high-tech moins cher...">
<META NAME="Keywords" CONTENT="Lecteur, MP3, Baladeur, achat, acheter, vente">
```

Le tag Meta/Description définit une propriété nommée (name en anglais) Description à laquelle est affecté un contenu (content en anglais).

Dans l'exemple suivant,

```
<META NAME="Description" CONTENT="Top Achat vous rembourse la différence si vous trouvez moins cher ailleurs ! Vente Lecteur MP3 - Baladeur MP3 au meilleur prix ! Top Achat, le high-tech moins cher...">
```

La propriété Description a pour valeur « Top Achat vous rembourse la différence si vous trouvez moins cher ailleurs ! Vente Lecteur MP3 - Baladeur MP3 au meilleur prix ! Top Achat, le high-tech moins cher... ».

La propriété Description sert à décrire le contenu d'une page web. Elle n'est pas destinée aux internautes, mais plutôt aux moteurs de recherche.

Le tag Meta/Description, selon la norme HTML 4.01

La spécification HTML 4.01 autorise à spécifier des métadonnées (des informations sur le document plutôt que sur le contenu du document). Cependant, cette spécification ne définit aucun ensemble de propriétés de métadonnées légales. Ces métadonnées font l'objet de convention d'usage. Le Meta/Description fait partie de cet ensemble voué à la définition des métadonnées.

Ces métadonnées obéissent toutes aux mêmes règles. La norme HTML 4.01 dit que :

- chaque document HTML peut (non obligatoire) contenir un ou plusieurs éléments META,
- ces éléments META doivent se trouver dans la section HEAD (l'entête) du document,
- l'élément META définit une propriété auquel lui est assignée une valeur,
- l'élément META peut contenir des caractères accentués ou spéciaux,
- enfin, l'élément META nécessite obligatoirement une balise ouvrante et ne doit pas avoir de balise fermante.

Ci-dessous un exemple d'éléments META correctement formés.

Code HTML :

```
<HTML>
<HEAD>
... éléments d'en-tête ...
```

```
<META name="author" content="HTML4SEO Team">
<META name="copyright" content=" HTML4SEO Copyright © 2009">
... autres éléments d'en-tête ...
</HEAD>
<BODY>
... corps du document ...
</BODY>
</HTML>
```

name définit le nom de la propriété et content sa valeur.

*On a donc ici, **author = HTML4SEO Team** et **copyright = HTML4SEO Copyright © 2009**. Attention, la norme HTML 4.01 ne dit rien des propriétés auteur et copyright présentement définies.*

La norme HTML 4.01 permet un paramétrage plus poussé des tags META. Pour aller plus loin dans ce paramétrage, consultez <http://www.la-grange.net/w3c/html4.01/struct/global.html#h-7.4.4.2>.

Il est vivement conseillé de se conformer au standard HTML (peu importe la version) car les moteurs de recherche apprécient les documents correctement formés, c'est-à-dire formés selon le standard HTML.

Le Meta/Description est-il vraiment important ?

Cette balise HTML est, en principe, unique. Il n'y en a qu'un seul exemplaire par page web, en général en début de fichier. Celle-ci ne fait pas corps avec la page web, elle ne se voit pas. Elle n'est pas destinée aux internautes, mais plutôt aux moteurs de recherche.

Depuis le début des années 2000, l'influence du tag Meta/Description sur le positionnement a sérieusement diminué du fait d'abus (blindage en mots-clés). On peut donc se demander aujourd'hui si elle est vraiment importante pour le référencement et donc si il est vraiment nécessaire d'y travailler.

Ne la négligez pas !

Elle n'intervient certes plus dans l'algorithme de classement des moteurs de recherche. Cependant, les moteurs de recherche y accordent encore de l'importance. En effet, ils y ont souvent recours pour décrire les pages web présentées dans leurs résultats de recherche.

En renseignant les tags Meta/Description de vos pages web, vous proposez aux moteurs de recherche une description de votre page web, qu'ils afficheront très probablement dans leurs résultats de recherche. De cette manière, vous maîtrisez mieux la communication faite sur votre page web dans les SERP. Vous pouvez ainsi inciter davantage l'internaute à cliquer sur votre lien plutôt que sur ceux de vos concurrents. Dans le cas contraire, vous laissez l'initiative aux moteurs de recherche de choisir n'importe quel morceau de texte de votre page web. Et il est fort probable qu'il choisisse un bout de texte sans intérêts.



On peut constater dans cet exemple que certains liens sont plus vendeurs que d'autres.

Une balise Meta/Description différente pour chaque page web

A chaque page, sa description. Cette description doit absolument refléter le contenu réel de la page. Dans le cas contraire, l'internaute floué ne tardera pas à quitter votre page. Rappelez-vous, ce n'est pas la quantité de trafic qui compte mais la quantité transformée, autrement dit la qualité (dans le sens, public que vous visez). Il est donc impératif de ne pas répéter la même description sur toutes les pages de votre site web. Car cette description serait alors trop générique. Dans l'idéal, vous ne devriez pas avoir deux pages de votre site web ayant la même balise Meta/Description.

Une description intelligible plutôt qu'un bourrage de mots-clés

Il est impératif de ne pas faire de vos balises Meta/Description des successions inintelligibles de mots-clés. La mode n'est plus au bourrage de mots-clés, mais plutôt au vrai contenu des pages web. Une ou deux phrases correctement construites ont plus de valeur qu'une série de mots-clés.

En effet, comme nous l'avons précisé plus haut, la balise Meta/Description n'entre pas directement en ligne de compte dans le référencement. Par contre, elle permet de vous dissocier des autres résultats que renvoient les moteurs de recherche (SERP) et ainsi de rendre votre page web plus attractive que celles de vos concurrents. Vous devez donner envie à l'internaute de cliquer sur votre page web.

Lorsqu'on examine les résultats Google pour le mot-clé **Baladeur mp3**, on constate que c'est la balise Meta/Description des pages web (cerclée de rouge – voir plus haut) qui est rapportée par Google dans sa liste de résultats.

Vous devez donc vous attacher à faire des balises Meta/Description en forme de présentation attractives, sans pour autant être racleurs.

Par exemple,

A conseiller, `<META NAME="Description" CONTENT="Top Achat vous rembourse la différence si vous trouvez moins cher ailleurs ! Vente Lecteur MP3 - Baladeur MP3 au meilleur prix ! Top Achat, le high-tech moins cher...">`.

A éviter, `<META NAME="Description" CONTENT="XXXXXXX : catégorie Baladeur MP3 / MP4 mp3 & hifi">`

Des mots-clés pas si importants

La balise Meta/Description ne doit pas se soucier de contenir vos mots-clés importants. Au risque d'être pénible, la balise Meta/Description n'entre pas directement en ligne de compte dans le positionnement de vos pages web dans les SERP. Elle se contente « d'assister » les moteurs de recherche dans la présentation qu'ils font de votre page web dans les SERP.

Par exemple,

A conseiller, `<META NAME="Description" CONTENT="Top Achat vous rembourse la différence si vous trouvez moins cher ailleurs ! Vente Lecteur MP3 - Baladeur MP3 au meilleur prix ! Top Achat, le high-tech moins cher...">`.

A éviter, `<META NAME="Description" CONTENT="Forum-mp3.com: télécharger mp3, Actualité, prix baladeur mp3, guide de baladeur mp3, conseil pour acheter un baladeur mp3, karaoké mp3 et sonnerie mp3...">`

La taille

La taille optimale de la balise Meta/Description varie d'un moteur de recherche à l'autre. Cependant, les professionnels sont en général d'accord pour fixer cette taille optimale à 250 caractères, soit environ une soixantaine de mots.

Les caractères accentués

Si vous utilisez des caractères accentués dans vos balises Meta/Description, il faudra alors impérativement placer vos descriptions après le META qui déclare le CHARSET.

L'exemple suivant spécifie l'encodage de caractères du document comme étant ISO-8859-5 (qui permet l'usage de caractères accentués)

`<META http-equiv="Content-Type" content="text/html; charset=ISO-8859-5">`

Ci-dessous un exemple complet d'élément Meta/Description avec charset correctement formé.

Code HTML :

```
<HEAD>
  <META http-equiv=content-type content="text/html; charset= ISO-
8859-1">
  <TITLE>...</TITLE>
  <META name="description" content="... rembourse la différence
...">
  ... autres éléments d'en-tête ...
</HEAD>
<BODY>
  ... corps du document ...
</BODY>
</HTML>
```

Enfin, la solution de l'encodage HTML est une autre approche sur laquelle le charset est sans incidence.

Ci-dessous un exemple complet d'élément Meta/Description sans charset et avec encodage html correctement formé.

Code HTML :

```
<HEAD>
  <TITLE>...</TITLE>
  <META name="description" content="... rembourse la
différence ...">
  ... autres éléments d'en-tête ...
</HEAD>
<BODY>
  ... corps du document ...
</BODY>
</HTML>
```

Le « é » de différence est « html encodé » « é ».

L'application de ces conseils devrait améliorer votre taux de clics dans les SERP.

Le tag Meta/Keywords

Le tag Meta/Keywords, Kezako ?

Lorsqu'on examine le code html de la page web suivante (clic droit + Afficher la source dans Internet Explorer) http://www.topachat.com/pages/g_cat_est_imageson_puis_page_est-lecteur-mp3---baladeur-mp3.html , on peut y voir en début de fichier le petit bout de code suivant,

```
<html>
<head>
<title>Lecteur MP3 - Baladeur MP3 - Achat au meilleur prix !</title>
<META NAME="Description" CONTENT="Top Achat vous rembourse la différence si vous trouvez moins cher ailleurs ! Vente
Lecteur MP3 - Baladeur MP3 au meilleur prix ! Top Achat, le high-tech moins cher...">
<META NAME="Keywords" CONTENT="Lecteur, MP3, Baladeur, achat, acheter, vente">
```

Le tag Meta/Keywords définit une propriété nommée (name en anglais) Keywords à laquelle est affecté un contenu (content en anglais).

Dans l'exemple suivant,

```
<META NAME="Keywords" CONTENT="Lecteur, MP3, Baladeur, achat, acheter, vente">
```

La propriété Keywords a pour valeur « Lecteur, MP3, Baladeur, achat, acheter, vente ».

La propriété Keywords n'est pas destinée aux internautes, mais plutôt aux moteurs de recherche. Elle sert à décrire le contenu d'une page web à l'aide d'une liste de mots-clés représentatif de son contenu.

Le tag Meta/Keywords, selon la norme HTML 4.01

Voir plus haut, la norme HTML 4.01 pour le Meta/Description.

Le tag Meta/ Keywords est-il vraiment important ?

Le tag Meta/Keywords ne sert strictement à rien !

Si, peut-être à vous faire perdre du temps.

Les moteurs de recherche ont beaucoup évolué depuis les débuts de l'Internet. A l'origine cette balise les aidait à classer les documents HTML. Il était alors facile aux webmasters d'abuser les moteurs de recherche en trichant sur le contenu de ces listes de mots-clés, pas toujours conforme au contenu des pages web qu'elles étaient censées décrire. Ils y

mettaient des listes interminables de mots-clés, ciblaient plusieurs marchés à la fois, tous reconnus à fort trafic et espéraient drainer ainsi un maximum de trafic souvent peu qualifié.

Les moteurs de recherche ont affiné leurs algorithmes qui maintenant se passent de cette balise. Elle ne sert donc plus à rien aujourd'hui. Les professionnelles du SEO sont tous d'accord sur ce point. Inutile donc de vous y attarder.

Les tags H1 à H6

Nous continuons notre passage en revue des balises html avec l'optimisation des balises H1 à H6.

Les tags H1 à H6, Kezako ?

Lorsqu'on examine le code HTML de la Faq HTML4SEO (clic droit + Afficher la source dans Internet Explorer) <http://www.html4seo.com/html4seo-faq.htm>, on y trouve le bout de code suivant,

```
<h1>FAQ HTML4SEO</h1>

<p>
La FAQ (Foire Aux Questions) HTML4SEO a pour objectif de répondre aux questions les plus courantes.
Nous la compléterons progressivement dans le temps, en fonction de vos attentes.
</p>

<div id="faq_toc">
  <a href="#tips"><h2 style="margin-top: 0;">Les conseils</h2></a>
  <h3><a href="#tips_signification">Signification</a></h3>
  <h4>&raquo; <a href="#tips_signification_htmltag">Tag HTML</a></h4>
  <h4>&raquo; <a href="#tips_signification_tips">Conseils</a></h4>
  <h4>&raquo; <a href="#tips_signification_url_nb">Nombre d'URLs</a></h4>
  <h4>&raquo; <a href="#tips_signification_eff">Efficacité relative</a></h4>
  <h4>&raquo; <a href="#tips_signification_diff">Difficulté</a></h4>
  <h4>&raquo; <a href="#tips_signification_roi">ROI</a></h4>
</div>
```

Ce code permet d'afficher plusieurs titres dans la page. Ci-dessous le résultat.

FAQ HTML4SEO

La FAQ (Foire Aux Questions) HTML4SEO a pour objectif de répondre aux questions les plus courantes. Nous la compléterons progressivement dans le temps, en fonction de vos attentes.

Les conseils

Signification

- » Tag HTML
- » Conseils
- » Nombre d'URLs
- » Efficacité relative
- » Difficulté
- » ROI

Le chiffre de la balise (de 1 à 6) va croissant avec le niveau de détail du titre. Les titres en H1 sont les plus importants, donc les plus visibles (en général avec la taille de police la plus grande et éventuellement une couleur différente de celle du reste du texte), alors que les tags H6 sont les moins importants.

Les tags H1 à H6 selon la norme HTML 4.01

La norme HTML 4.01 nous dit que :

- chaque document HTML peut (non obligatoire) contenir un ou plusieurs titres H1 à H6,

- chaque titre H1 à H6 nécessite obligatoirement une balise ouvrante <Hn> et une balise fermante </Hn> ,
- il existe 6 niveaux de titrages, l'élément H1 étant le plus important et H6 le moins important.

Voici un exemple d'éléments H1, H2 et H3 correctement formés.

Code HTML :

```
<HTML>
<HEAD>
... éléments d'en-tête ...
</HEAD>
<BODY>
... introduction du document ...
<H1>Les poissons d'eau douce</H1>
... ici une courte introduction sur les poissons d'eau douce
...
<H2>En Europe</H2>
... ici un texte présentant les spécificités de ces poissons en
Europe ...
<H3>En Europe du nord</H3>
... détails sur les poissons d'Europe du nord ...
<H3>En Europe de l'ouest</H3>
... détails sur les poissons d'Europe de l'est ...
...
<H2>En Asie</H2>
... ici un texte présentant les spécificités de ces poissons en
Asie ...
...
<H1>Les poissons d'eau de mer</H1>
...
</BODY>
</HTML>
```

Le respect de la hiérarchie des balises de titre H1 à H6 n'est pas obligatoire d'après la norme mais il est fortement recommandé. Dans le précédent exemple, les titres sont correctement hiérarchisés.

A éviter,

Code HTML :

```
<H1>Les éléphants d'Afrique</H1>
...
<H3>En Afrique du sud</H3>
...
<H2> Les éléphants d'Asie</H2>
...
<H3>En Asie du sud-est</H3>
...
```

Notez que le niveau d'importance de la balise ouvrante doit être strictement le même que celui de la balise fermante.

Voici un exemple de titre valide :

<H1>Les castors lapons</H1>

Le titre suivant n'est pas valide et risque même de déstructurer toute la mise en page du document :

<H1>Les castors lapons</H2>

Quelle importance ont les balises H1 à H6 ?

Les balises H1 à H6 dans le corps d'une page web

Les balises H1 à H6 sont probablement les plus importantes de l'ensemble du texte d'une page web. Après la balise TITLE, la balise H1 semble la plus importante pour le référencement. Lorsqu'ils sont bien choisis, les titres introduisent le contenu qui les suit (et d'une certaine manière, en sont représentatifs). Ils fournissent donc les termes les plus pertinents pour identifier le document.

Les balises H1 à H6 entre elles

Pour les moteurs de recherche, les balises H1 à H6 n'ont pas toutes la même importance. Les balises H1, les plus importantes parmi les titres, le sont aussi pour le référencement. Plus on augmente de niveau de titre (H2, H3, H4, H5 et H6), moins son contenu sera considéré par les moteurs de recherche. Le contenu d'une balise H1 sera donc plus important que celui d'une balise H2, lui-même plus important qu'une balise H3 et ainsi de suite jusqu'à H6. Les efforts d'optimisation devront donc se concentrer en priorité sur les titres en H1, puis en H2, etc.

Des balises stratégiques pour le référencement

Les tags H1 à H6 et les autres tags HTML

Après la balise TITLE d'une page web (titre de la fenêtre du navigateur), les balises H1 à H6 (titres et sous-titres du texte de la page) sont probablement les balises HTML les plus considérées par les moteurs de recherche. Elles sont donc presque incontournables dans l'optimisation d'une page web pour un meilleur référencement.

Les tags H1 à H6 entre eux

Toutes les balises H1 à H6 ne se valent pas. Les plus importantes parmi elles sont les balises H1, qui correspondent aux titres de premier niveau. Viennent ensuite les balises H2, puis les H3, jusqu'à H6. Les titres doivent notamment être hiérarchisés pour une bonne optimisation.

Les spécialistes SEO sont généralement d'accord pour dire que les balises H1, H2 et éventuellement H3 ont une véritable influence sur le référencement. Selon ces mêmes experts SEO, les suivantes (H4, H5 et H6) sont apparemment beaucoup plus sujettes à caution. Il est donc vivement recommandé d'insérer des mots-clés stratégiques surtout dans les balises H1 (en priorité) et H2. On veillera tout particulièrement à éviter les fautes d'orthographe.

Dans l'exemple suivant, les titres de la page ne sont pas optimisés.

Code HTML :

```
<H1>Bouger à Paris</H1>
...
<H2>Que visiter ?</H2>
...
<H3>La tour Eiffel</H3>
...
<H2>Où dormir ?</H2>
...
<H1>Partir pour Dunkerque</H1>
...
```

Il faut choisir avec soin les mots-clés à partir desquels les utilisateurs peuvent trouver la page web.

Code HTML :

```
<H1>Voyage à Paris</H1>
...
<H2>La visite des monuments historiques</H2>
...
<H3>Visiter la tour Eiffel</H3>
...
<H2>Les hôtels à Paris</H2>
...
<H1>Le tourisme à Dunkerque</H1>
...
```

Comment bien rédiger mes balises H1 à H6 ?

SEO & internautes : le bon compromis !

En bon français, un titre doit refléter au mieux le contenu qu'il annonce. Or, un titre pertinent du point de vue de la langue n'est pas nécessairement optimal pour un bon référencement. Si un titre est correctement rédigé pour des lecteurs humains, il ne l'est pas forcément pour les moteurs de recherche. Les critères d'appréciation sont différents.

Pour un bon référencement, un titre doit contenir des mots-clés stratégiques. Il faut donc trouver le meilleur compromis entre une rédaction à destination des internautes et une rédaction à destination des moteurs de recherche. Dans la pratique, la tâche peut être ardue. Il faut savoir être fin rédacteur pour glisser opportunément les bons mots-clés dans les titres.

Par exemple, le titre,

« Découvrez comment réussir votre mayonnaise »

peut convenir à des utilisateurs humains comme aux moteurs de recherche mais pour ces derniers, le titre suivant sera préférable :

« Recette d'une mayonnaise réussie »

Quelle taille pour les balises H1 à H6 ?

Il n'y a pas de taille canonique pour les titres. Ils peuvent être indifféremment courts ou longs. Il faut néanmoins surveiller la densité des mots-clés dans les titres. C'est un critère important. Cela peut justifier l'emploi plus récurrent de titres courts (ou éventuellement de longueur moyenne).

Dans l'exemple suivant,

« De ce qu'il faut retenir quant aux principes du marketing et les définitions associées »

la densité du mot-clé « marketing » est diluée par le nombre total de mots. On préférera donc le titre suivant,

« Le marketing, principes et définitions »

Les images (IMG)

Passons maintenant aux illustrations.

Les images selon la norme HTML 4.01

La norme HTML 4.01 dit que :

- l'élément IMG incorpore une image dans le document courant, à l'emplacement de la définition de l'élément,
- l'élément IMG n'a pas de contenu, il est généralement remplacé dans la ligne par l'image que désigne l'**attribut src**,
- l'**attribut alt** spécifie le texte de remplacement qui sera restitué si l'image ne peut s'afficher,
- enfin, l'élément IMG nécessite obligatoirement une balise ouvrante et ne doit pas avoir de balise fermante.

Ci-dessous un exemple d'élément IMG correctement formé.

Code HTML :

```
<IMG src="/img/logo-gnoztik.gif" alt="logo GnoZtiK">
```

On a donc,

- le fichier image se situe à l'emplacement `/img/logo-gnoztik.gif`,
- l'image a pour description alternative « logo GnoZtiK ».

Consultez <http://www.la-grange.net/w3c/html4.01/struct/objects.html#h-13.2>, pour en savoir davantage.

Quelle importance ont les images ?

En général, une image a valeur d'illustration. Elle est donc un élément important d'une page web mais également en soi. Elle peut améliorer le référencement de la page web qui l'accueille et peut aussi être référencée dans Google Image, source supplémentaire de trafic.

Cependant, les moteurs de recherche n'en comprennent pas le contenu. L'**attribut alt** permet toutefois de compenser cette défaillance en qualifiant l'image à laquelle il se rapporte.

Comment dois-je construire les images de mes pages web ?

L'attribut ALT

Chaque image porteuse de sens se doit d'avoir une description alternative. L'**attribut alt** signale aux moteurs de recherche le contenu de l'image auquel il est rattaché. Faites donc des descriptions courtes (quelques mots) et, si possible, glissez-y des mots-clés (sur le principe densité/décompte énoncé plus haut).

Par exemple,

A conseiller, .

A éviter, .

Texte et Légendes environnantes

D'après la FAQ de Google, le texte autour de l'image est capital pour contextualiser celle-ci. Faites donc en sorte que le texte environnant soit en rapport direct avec le contenu de vos images. Si en plus ce texte contient des mots-clés, c'est encore mieux !

L'attribut SRC

Tout comme les urls, les fichiers images ont un nom complet composé de répertoires (optionnels) et d'un fichier. Faites en sorte que cet ensemble soit descriptif (mots-clés) pour vos images en portant une attention particulière au nom du fichier.

Par exemple,

A conseiller, .

A éviter, .

Des images sans texte !

Placer du texte dans une image, c'est perdre du contenu. Rappelez-vous, les moteurs de recherche ne lisent pas le contenu des images. Gardez donc ce texte pour votre contenu !

Les liens (A)

Les liens, Kezako ?

Un lien hypertexte, c'est un texte présent dans une page web sur lequel on peut cliquer pour se rendre sur une autre page web. Le lien représente une connexion entre deux documents web. C'est une caractéristique fondamentale du web.

Les liens selon la norme HTML 4.01

La norme HTML 4.01 dit que :

- un lien part d'une ancre « source » et pointe vers une ancre « destination »,
- chaque élément A définit une ancre,
- le texte (cliquable) de l'élément A définit la position de l'ancre,
- l'attribut href (optionnel) fait de cette ancre, l'ancre source d'exactly un lien,
- l'attribut name (optionnel) nomme l'ancre, de sorte que celle-ci puisse être la destination d'autres liens,
- enfin, l'élément A nécessite obligatoirement une balise ouvrante et une balise fermante.

Ci-dessous un exemple d'élément A correctement formé.

Code HTML :

```
<A href="http://www.gnoztik.com">logiciel de référencement</A>
```

De façon simplifiée,

- la page web qui contient ce lien est l'ancre source (l'ancre est positionnée sur la séquence de texte « logiciel de référencement » de cette page web),
- la page web <http://www.gnoztik.com> est l'ancre destination,
- la direction va de la page web courante vers la page web <http://www.gnoztik.com>.

Consultez <http://www.la-grange.net/w3c/html4.01/struct/links.html#edef-A>, pour en savoir davantage.

Quelle importance ont les liens ?

L'ancrage des liens apparaît, en général, comme un texte différencié (bleu et souligné). En toute logique, si un webmaster prend la peine de mettre un lien sur une séquence de texte, c'est que cette séquence a probablement plus d'importance (dans une certaine mesure MOYENNE, car la fraude est possible) que le reste du texte.

Il n'y a cependant aucune certitude en la matière, si ce n'est que vos liens sont importants pour les pages web que vous pointez. Si ces pages sont les vôtres, alors mieux vaut soigner les ancres de vos liens. En plus de participer au maillage de votre site web, vous accordez quelques backlinks internes (moins valorisant que les externes) à vos pages web.

Enfin, sachez qu'ils n'occasionnent aucune perte de PageRank comme on peut parfois le lire ici ou là.

Donc dans le doute, soignez-les !

Comment dois-je construire les liens de mes pages web ?

Les liens qui ont de l'importance sont surtout ceux qui se trouvent dans le corps du texte (votre contenu). Vous veillerez donc à y glisser quelques mots-clés sans excès (sur le principe densité/décompte énoncé plus haut) et à pointer sur des pages web en rapport avec le thème de vos pages web.

Par exemple,

A conseiller, logiciel de référencement

A éviter, cliquez ici

Le texte

Le texte de vos pages web aussi est important !

Un contenu ciblé

Vos pages web doivent impérativement avoir un contenu homogène. Mieux vaut faire plusieurs pages web très ciblées qu'une seule page web abordant des thèmes multiples. Vous aurez de ce fait davantage de pages à référencer et donc autant d'opportunités supplémentaires pour être bien positionné dans les SERP.

Densité et décompte avec modération

Tout comme pour les balises majeures à optimiser vues plus haut, le décompte et la densité nécessitent une attention particulière. Il ne faut cependant pas en faire une obsession. Les répétitions excessives alourdissent votre propos et rendent la lecture de vos pages web désagréable à l'internaute. N'hésitez donc pas à user de synonymes, les algorithmes des moteurs de recherche sont suffisamment perspicaces pour ne pas vous pénaliser, bien au contraire.

Une surface d'exposition démultipliée

Dans la droite ligne de ce qui précède, le texte se doit d'être lexicalement riche et ciblé. Veillez à varier les formulations, usez de synonymes, etc. De cette manière, votre référencement ne perdra pas en qualité puisque votre contenu restera ciblé et homogène. Bien au contraire, avec un contenu de même volume, simplement plus varié votre surface d'exposition n'en sera que plus grande.

Le texte du haut !

De manière générale pour les moteurs de recherche, le début d'une page web a plus de poids que sa fin. Vous y glisserez donc, dans la mesure du possible, les mots-clés sur lesquels vous ambitionnez de vous positionner.

Mise en forme mineure

Enfin, vous utiliserez la mise en gras (balise B ou STRONG) avec parcimonie afin de faire ressortir les mots à fort contenu sémantique.

Le duplicate content

Tout contenu web en double est susceptible de subir la pénalité dite du duplicate content. Une page web qualifiée duplicate content voit alors sa capacité de positionnement dans les SERP fortement diminuée. Cette pénalité s'applique uniquement à la page web qualifiée duplicate content et non à l'ensemble du site web auquel elle appartient.

Appliquez les conseils qui suivent et vous devriez éviter à vos pages web des problèmes de duplicate content.

Pourquoi les moteurs de recherche luttent contre le duplicate content ?

Les moteurs de recherche veulent diversifier autant que possible les pages web proposées dans leurs résultats de recherche. Pour ce faire, ils tâchent d'éliminer les contenus en double.

Le duplicate content, c'est quoi ?

Vous risquez la pénalité duplicate content lorsque :

- une ou plusieurs de vos pages web sont identiques ou très proches,
- une de vos pages web est identique ou très proche d'une page d'un autre site web.

Attention, il n'est pas nécessaire d'avoir deux contenus identiques à 100% pour se voir appliquer la pénalité du duplicate content.

Comment les moteurs de recherche élisent la page web qui figurera dans les SERP ?

Lorsque un moteur de recherche trouve une page web, il la compare aux pages web déjà présentes dans sa base de données (son index). Il détermine ainsi si il existe plusieurs exemplaires (pas nécessairement identiques à 100%).

Les pages web issues des sites reconnus spammers sont alors retirées de la liste.

Les autres pages web sont classées en fonction de la quantité et de la qualité de leurs backlinks (liens entrants).

La page web avec les meilleurs backlinks (liens entrants) est retenue pour paraître dans les SERP.

Comment éviter un duplicate content à mes pages web ?

Masquez vos contenus « en double » aux moteurs de recherche

Si vous possédez différentes versions d'un même document (web classique, web mobile, impression, etc.), les moteurs de recherche peuvent retenir la mauvaise version (la moins avantageuse) pour leurs SERP.

Ajoutez (dans les headers HTML) une balise **META NOINDEX** aux versions en double de votre contenu que vous ne voulez pas voir indexer.

Code HTML :

```
<META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">
```

Signalez à Google, Yahoo et Microsoft Bing vos urls canoniques !

Plus spécifiquement, Google, Yahoo et Microsoft Bing (successeur de Microsoft Live Search) proposent de résoudre ce problème en utilisant la balise spéciale (conseillée) **link rel=canonical** qui décrit l'url canonique d'une page web (son url officielle). A l'aide de cette balise, vous indiquez pour chaque page de votre site web si elle est ou n'est pas l'url canonique du contenu qu'elle dispense. Pour ce faire, vous placez dans l'entête HTML de vos pages web canoniques la balise spéciale **link rel=canonical**, comme suit.

Code HTML :

```
<head>
...
...<link rel="canonical" href="http://www.gnoztik.com/" />
...
</head>
<body>
...
</body>
```

Exigez un backlink des sites web qui diffusent votre contenu

Si des sites web tiers reprennent une partie de votre contenu, les moteurs de recherche peuvent favoriser leur version du document plutôt que la vôtre.

Exigez de ces sites web en retour qu'ils créent un lien (backlink) vers le document d'origine (le vôtre) sur leur propre version de votre contenu.

Le duplicate content peut gravement pénaliser votre positionnement dans les SERP. En mettant en pratique les conseils ci-dessus, les crawlers des moteurs de recherche devraient non seulement indexer correctement vos pages web mais également ne pas pénaliser leur positionnement dans les SERP.

Le NetLinking

Nous abordons la quatrième et dernière partie de notre document sur le référencement naturel : le NetLinking. Certains voient en lui l'arme fatale du référencement. C'est à notre avis excessif.

Il est très imprudent de ne faire reposer le référencement de votre site web que sur le NetLinking. Ne faire reposer le référencement de votre site web que sur le NetLinking, c'est finalement mettre entre les mains d'autres que vous la visibilité de votre site web sur Internet. N'oubliez jamais que vous n'avez jamais une totale maîtrise de votre NetLinking.

Pareillement, de nombreux mots-clés très ciblés apporteurs de trafic très qualifié ne nécessitent parfois aucun travail de NetLinking. Le marché est tellement ciblé ou la concurrence tellement faible, qu'une simple optimisation de contenu suffit. Il faut donc mettre en œuvre des moyens proportionnés à vos objectifs.

Ceci est d'autant plus vrai que le NetLinking est une tâche très laborieuse et forte consommatrice de temps.

Dans le premier chapitre de cette partie, nous traiterons en profondeur la question des backlinks : ce qu'ils sont, leur poids dans votre référencement, où les trouver, comment les construire, quelle stratégie adopter pour les accumuler, les écueils à éviter, etc.

Dans le second chapitre, nous reviendrons sur deux notions qui ont fait couler beaucoup d'encre numérique : le PageRank et le TrustRank. Nous y verrons notamment en quoi ils pèsent sur votre référencement, comment ils fonctionnent, si il est possible de les utiliser comme levier pour améliorer le positionnement de vos pages web, etc.

ATTENTION, sans contenu original, le NetLinking est vain.

Le NetLinking est une arme redoutable **SOUVENT NECESSAIRE**.

Le NetLinking n'est pas toujours indispensable.

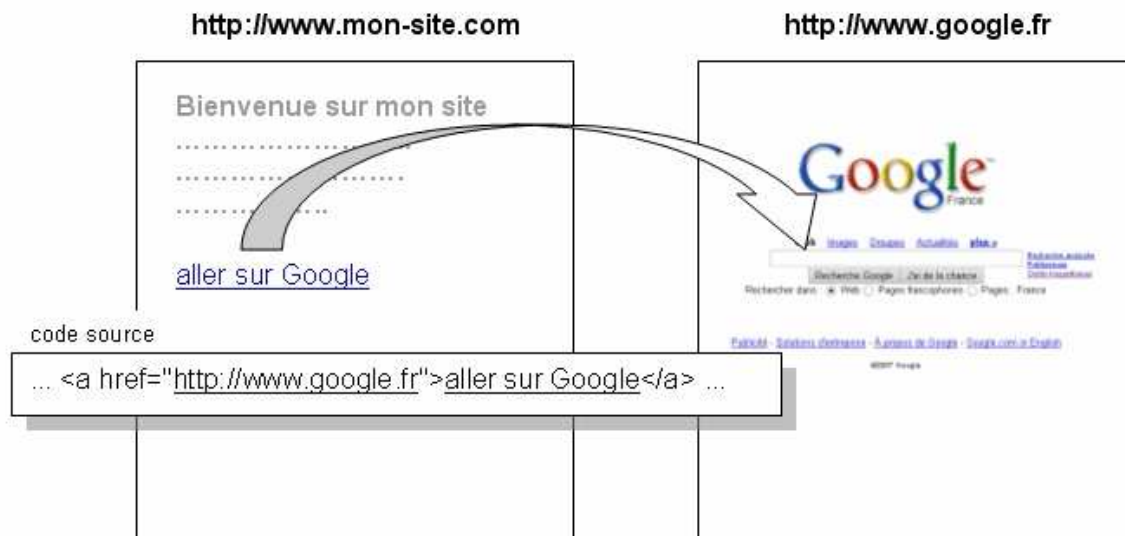
Les backlinks

Les backlinks, Kezako ?

Définition d'un lien hypertexte

Un lien hypertexte, c'est un texte présent dans une page web sur lequel on peut cliquer pour se rendre sur une autre page web.

Par exemple, sur la page ci-dessous, <http://www.mon-site.com>, quand nous cliquons sur le lien **aller sur Google**, nous arrivons sur la page <http://www.google.fr>.



Le code source de la première page (<http://www.mon-site.com>) se présente comme suit,

```
... <a href="http://www.google.fr">aller sur Google</a> ...
```

- href contient l'url (adresse) de la page web pointée, ici **<http://www.google.fr>**
- le texte affiché est contenu entre les balises <a ...> et , ici **aller sur Google**.

Pour obtenir le code source d'une page web sous Internet Explorer, cliquez sur **Affichage/Source**. Une fois le code source affiché, pour obtenir les liens hypertextes sous Internet Explorer, cliquez sur **Edition/Rechercher...** Une boîte de dialogue s'ouvre. Renseignez-la avec la séquence de mots **a href** puis validez.

Un backlink, c'est quoi ?

C'est un lien hypertexte !!!

En référencement, le terme backlink désigne la même chose (à peu près) que le terme lien hypertexte en HTML.

Lien hypertexte (en HTML) = backlink (en référencement)

Lorsqu'une page web (source) fait un lien vers une seconde page (destination) d'un autre site web, la première est un backlink de la seconde.

Dans l'exemple précédent, la page web <http://www.mon-site.com> est un backlink de la page web <http://www.google.fr>.

Si dix pages web pointent sur une page web, celle-ci a dix backlinks. En ajoutant le nombre de backlinks de chaque page web d'un site web, on obtient le nombre total de backlinks du site web.

Comment connaître mes backlinks ?

La commande « link: » de Google

Avec Google, pour connaître le nombre de backlinks de la page web <http://www.siteduzero.com/tuto-3-6-0-apprenez-a-creez-votre-site-web.html>, il suffit de taper **link: <http://www.siteduzero.com/tuto-3-6-0-apprenez-a-creez-votre-site-web.html>**.

<http://www.google.fr/search?hl=fr&q=link%3Ahttp%3A%2F%2Fwww.siteduzero.com%2Ftuto-3-6-0-apprenez-a-creez-votre-site-web.html&btnG=Recherche+Google>



A la date du 15 mai 2009, plus de 100 pages web ont un lien vers cette page web.

Attention, Google comptabilise aussi les liens internes (provenant du même site web, ici de <http://www.siteduzero.com>).

La commande « link: » de Yahoo

Avec Yahoo, pour connaître le nombre de backlinks de la page web <http://www.siteduzero.com> (uniquement la page d'accueil), il suffit de taper **link: <http://www.siteduzero.com>**. Vous serez alors redirigé vers Yahoo Site Explorer, outil Yahoo d'analyse de sites web.

<http://siteexplorer.search.yahoo.com/fr/siteexplorer/search?p=http%3A%2F%2Fwww.siteduzero.com&bwm=i&bwmf=u&bwmms=p&fr=yfp-t-501&fr2=seo-rd-se>



YAHOO! FRANCE Site Explorer

http://www.siteduzero.com Explorer l'URL

Site Explorer

- Ajouter à Mes Sites
- Préférences
- Blog (US)
- Module
- API (site US)
- Commentaires (US)

Résultats

Astuce : Votre requête "link:" a été redirigée vers Site Explorer. Vous pouvez continuer à utiliser le moteur de recherche pour d'autres requêtes.

Pages (2 299 491) Liens entrants (128 559)

Montrer les liens entrants : De n'importe quelle page pointant vers : Cette URL uniquement

Détails des résultats : Soumettre une page web ou un flux de site | Exporter les 1000 premiers résultats vers TSV

- GuppyY
text/html http://www.freeguppy.org/ - 66k - en cache Explorer
- Le Site du Zéro, site communautaire de tutoriels gratuits pour ...
text/html http://www.siteduzero.com/ - 31k - en cache Explorer
- Apprenez à créer votre site web ! - Site Web - Cours
text/html http://www.siteduzero.com/tuto-3-6-0-apprenez-a-creeer-votre-site-web.htm - 41k - en cache Explorer
- Nicolas Renous
text/html http://www.ndreams.org/ - 3k - en cache Explorer

A la date du 15 mai 2009, plus de 100 000 pages web ont un lien vers cette page web (uniquement la page d'accueil).

Attention, Yahoo comptabilise aussi les liens internes (provenant du même site web, ici de <http://www.siteduzero.com>). Cette commande est bridée chez Google. Yahoo annonce des chiffres plus élevés, ce qui ne veut pas dire que l'index de Yahoo est plus important.

La commande « linkdomain: » de Yahoo

Yahoo permet en plus de connaître l'ensemble des backlinks d'un site web en tapant, **linkdomain:http://www.siteduzero.com**.

<http://siteexplorer.search.yahoo.com/fr/siteexplorer/search?p=http%3A%2F%2Fwww.siteduzero.com&bwm=i&bwmf=s&bwm0=&fr=yfp-t-501&fr2=seo-rd-se>



YAHOO! FRANCE Site Explorer

http://www.siteduzero.com Explorer l'URL

Site Explorer

- Ajouter à Mes Sites
- Préférences
- Blog (US)
- Module
- API (site US)
- Commentaires (US)

Résultats

Astuce : Votre requête "linkdomain:" a été redirigée vers Site Explorer. Vous pouvez continuer à utiliser le moteur de recherche pour d'autres requêtes.

Pages (2 298 851) Liens entrants (215 611)

Montrer les liens entrants : De n'importe quelle page pointant vers : Tout le site

Détails des résultats : Soumettre une page web ou un flux de site | Exporter les 1000 premiers résultats vers TSV

- GuppyY
text/html http://www.freeguppy.org/ - 66k - en cache Explorer
- Le Site du Zéro, site communautaire de tutoriels gratuits pour ...
text/html http://www.siteduzero.com/ - 31k - en cache Explorer
- Apprenez à créer votre site web ! - Site Web - Cours
text/html http://www.siteduzero.com/tuto-3-6-0-apprenez-a-creeer-votre-site-web.htm - 41k - en cache Explorer
- Ubuntu - Wikipédia
text/html http://fr.wikipedia.org/wiki/Ubuntu - 132k - en cache Explorer

A la date du 15 mai 2009, plus de 200 000 pages web (externes et internes confondues) pointent vers une des pages du site web <http://www.siteduzero.com>.

Remarquez que le nombre de résultats retournés pour les deux requêtes **link:** et **linkdomain:** n'est pas le même.

Les backlinks, à quoi ça sert ?

Les backlinks vous rendent plus visible sur Internet. A quoi bon avoir un site web, si personne ne le visite ?

Plus de backlinks, c'est plus de visiteurs

Plus vous aurez de liens pointant vers vos pages web, plus vous ouvrirez de portes aux internautes pour accéder à votre site web. A chaque fois qu'un site qui fait un lien vers vous reçoit un visiteur, celui-ci peut arriver à votre site web. Vous profitez donc indirectement de la visibilité et de la qualité du référencement de ces sites web.

Autrement dit, plus il y a de backlinks, plus il y a de visiteurs.

Plus de backlinks, c'est être mieux placé dans les moteurs de recherche

Les trois principaux moteurs de recherche (Google, Yahoo et Microsoft Bing) se partagent 90% du marché de la recherche sur Internet.

Pour déterminer le classement d'une page web par rapport à un mot-clé, ils sont influencés par le nombre de backlinks qu'elle possède. Chaque backlink est considéré comme un vote.

un backlink = un vote

Plus elle aura de backlinks, plus elle sera populaire sur la toile, meilleur sera son positionnement. Mais attention, tous les backlinks ne se valent pas. Certains ont plus de poids que d'autres.

A qui demander un backlink ?

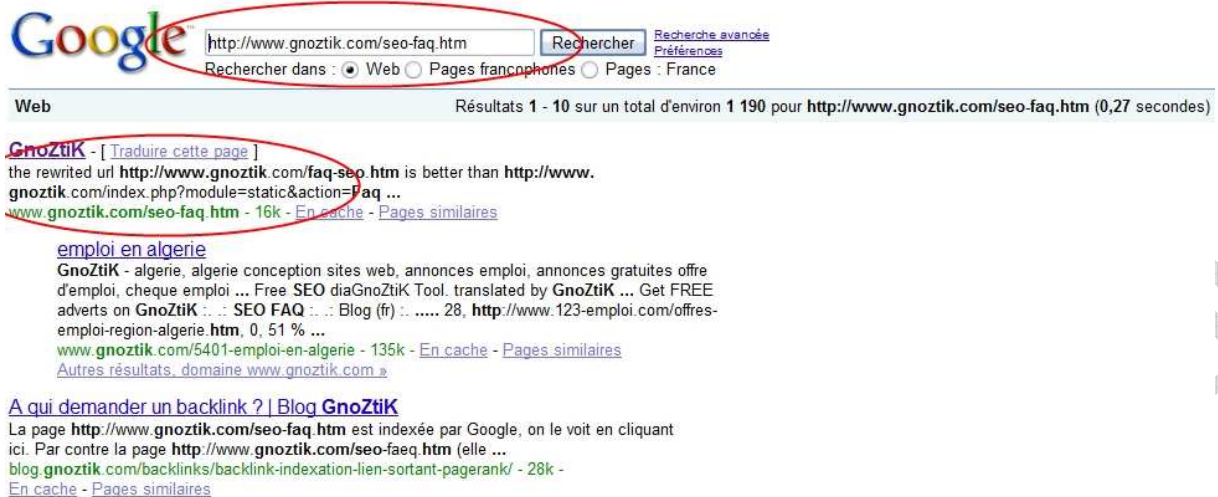
Comme nous le disions précédemment, il est important, pour avoir un bon référencement, d'avoir des backlinks de qualité. En effet, tous les backlinks ne se valent pas.

Le backlink doit être indexé

La page qui accueille le backlink doit être indexée dans les principaux moteurs de recherche (Google, Yahoo, MSN, ...). Dans le cas contraire, elle est sans intérêt et ne sera pas comptabilisée comme un backlink car elle n'existe pas au sens des moteurs de recherche. Pour le vérifier, la procédure est la même pour tous les moteurs de recherche, il suffit d'y saisir l'url de la page recherchée.

La page <http://www.gnoztik.com/seo-faq.htm> est indexée par Google, on le voit ici

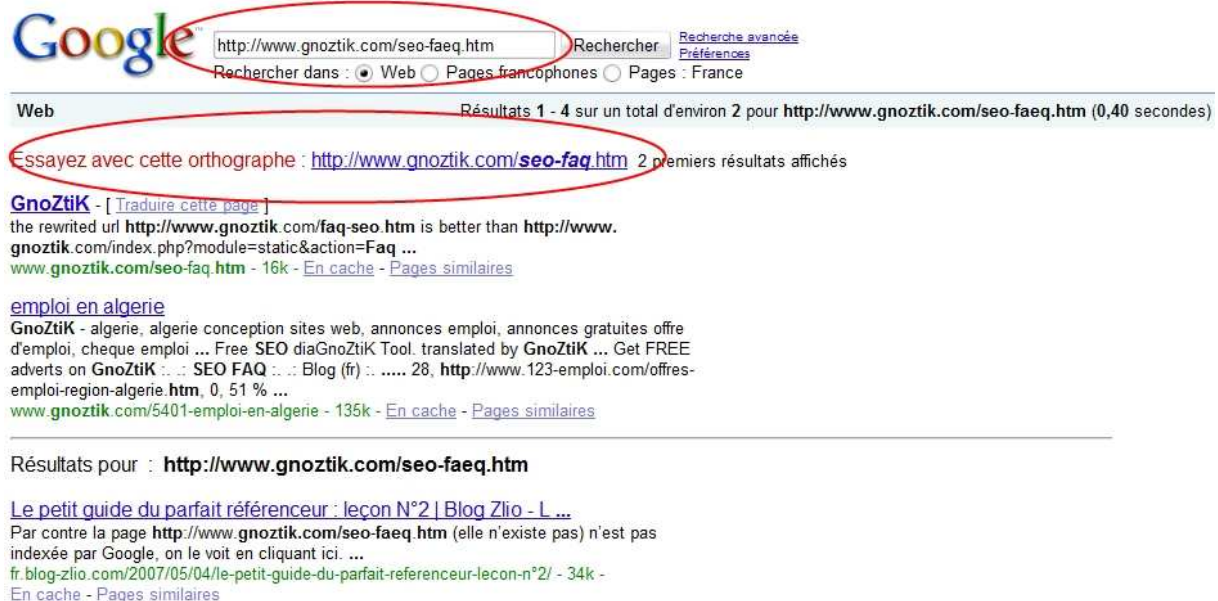
<http://www.google.fr/search?hl=fr&q=http%3A%2F%2Fwww.gnoztik.com%2Fseo-faq.htm&btnG=Recherche+Google&meta=&aq=f&oq=>



On peut constater que le premier résultat est la page web pour laquelle nous avons fait la recherche.

Par contre la page <http://www.gnoztik.com/seo-faeq.htm> (elle n'existe pas) n'est pas indexée par Google, on le voit ici.

<http://www.google.fr/search?hl=fr&q=http%3A%2F%2Fwww.gnoztik.com%2Fseo-faeq.htm&btnG=Recherche+Google&meta=&q=f&oq=>



Google propose une autre recherche indiquant par là l'inexistence de la page web (et pour cause, elle n'existe pas) dans son index.

Pas trop de liens sortants

La page qui accueille le backlink doit avoir un maximum de cent liens sortants (une vingtaine est préférable). Autrement dit, elle ne doit être un backlink que pour cent pages au maximum. Au-delà le lien est dilué et perd toute valeur aux yeux des moteurs de recherche. Pour faire simple, la page web qui accueille le backlink a une popularité propre auprès des moteurs de recherche. Elle redistribue cette popularité aux pages web qu'elle pointe. Mais cette redistribution est inversement proportionnelle au nombre de liens qu'elle fait. Si elle fait vingt liens, chacun de ces liens s'en partage un vingtième. Si elle en fait cent, c'est un

centième. Cette représentation est schématique, elle a valeur pédagogique car la réalité est plus complexe.

Le thème doit être proche

La page web qui pointe (le backlink) et celle qui est pointée (la vôtre) doivent avoir des thématiques proches.

Inutile d'avoir un backlink en provenance d'une page web qui parle de cuisine chinoise pour une page web qui traite de moto. Par contre une page web sur la sécurité routière ou sur les sports mécaniques est pertinente.

Un PageRank proportionné

La page qui pointe (le backlink) doit avoir un PageRank légèrement supérieur (1 à 2, maximum 3) à celui de votre page, au-delà ça devient suspect pour Google.

De ce qui suit ressort une première conclusion, les backlinks que proposent les annuaires sont la plupart du temps sans intérêts (à cause du nombre de liens sortants trop importants et de la thématique souvent mal ciblée).

Quel lien pour un bon backlink ?

Après un pénible travail de collecte, vous avez enfin trouvé une page web intéressante prête à faire un backlink vers la vôtre. Ce n'est pas le moment de perdre votre vigilance, prenez garde au type de lien qu'on vous propose.

Un backlink « en dur » sinon rien !!!

Ce doit être un lien « en dur », exit les redirections et le Javascript.

A éviter, logiciel référencement gratuit

Conseillé, logiciel référencement gratuit

Les annuaires offrent souvent des liens Javascript ou des redirections. Ceci est une raison supplémentaire pour ne pas y passer trop de temps.

Méfiez-vous des liens avec attribut « rel=nofollow »

Ces liens servent à indiquer aux moteurs de recherche que le lien ne doit pas être considéré comme un vrai backlink. De nombreux blogs utilisent cette technique pour éviter d'être spammé par des commentaires qui n'ont pour unique vocation que d'être des backlinks artificiels.

Nombreux sont les blogs avec des commentaires sans intérêts du genre « vraiment très intéressant comme article », « encore un super article ». La plupart du temps, ils servent à leur auteur à récupérer un backlink vers leur site web, via la signature du commentaire.

Il est positionné à l'initiative des webmasters, il faudra donc le préciser lors de votre demande de backlink.

Faux backlink, logiciel référencement gratuit

Vrai backlink, logiciel référencement gratuit

Pas de backlink en bas de page

Ce doit être un lien situé dans le contenu de la page. Privilégiez le haut de la page. Le pied de page est très déconseillé.

Variez les pages web ciblées par vos backlinks

Evitez de demander des backlinks qui pointent tous sur une même page de votre site web (en général votre page d'accueil).

Multipliez les liens vers différentes pages de votre site web, ceci apparaît plus naturel aux moteurs de recherche et augmente vos chances d'être trouvé.

Quel texte pour mes backlinks ?

Nous savons présentement comment choisir la page qui va accueillir notre backlink, le type du lien à y inclure, nous poursuivons maintenant nos pérégrinations au pays des backlinks. En matière de backlinks, rien n'est à négliger et nous arrivons enfin à l'ultime étape, celle relatif au choix du texte (ancrage) à faire figurer dans nos backlinks.

Des mots-clés stratégiques

Le texte doit nécessairement inclure des mots-clés importants. Il s'agit des mots-clés sur lesquels vous voulez positionner votre page web.

Si votre page web parle de football,

*A éviter, cuisine chinoise traditionnelle, vous serez peut-être bien positionné sur **cuisine chinoise traditionnelle**, mais ceux qui arrivent sur votre page web en tapant cette séquence dans Google risquent fort de quitter rapidement votre site web.*

*A éviter, clicquer ici, vous serez peut-être bien positionné sur **clicquer ici**, mais c'est sans intérêts.*

*Conseillé, maillot de foot marseille, vous serez peut-être bien positionné sur **maillot de foot marseille**, et ceux qui arrivent sur votre page web en tapant cette séquence dans un moteur de recherche risquent fort de rester sur votre site web.*

Des mots-clés variés

Evitez d'avoir le même texte dans tous vos backlinks.

En principe, vous recevez un backlink quand un webmaster apprécie votre page web. Le texte qu'il y met est de sa propre initiative. Il y a donc peu de chances qu'ils se passent tous le mot pour indiquer la même séquence texte dans le lien qu'ils font vers votre page web. Un même texte dans tous vos backlinks apparaît donc suspect aux moteurs de recherche et peut vous pénaliser.

Si votre page web parle d'immobilier,

*A éviter, toujours la même séquence de mots-clés **location appartement paris**.*

*Conseillé, des séquences de mots-clés variées autour de votre thématique **location appartement paris, vente appartement paris, location appartement marseille, vente appartement marseille, location maison paris, vente maison marseille, etc.***

2 à 3 mots au minimum

Le texte doit comporter deux à trois mots au minimum. Une séquence de trois mots permet de cibler plusieurs mots-clés.

Si votre page web parle d'immobilier,

*Conseillé, `location appartement paris`, vous serez peut-être bien positionné sur **location appartement paris**, mais également sur **location appartement** et **appartement paris**, ce qui augmente vos chances d'être trouvé sur les moteurs de recherche.*

Attention aux fautes d'orthographe

Le texte doit être sans fautes d'orthographe. Cette erreur est plus fréquente qu'on ne croit. Il serait dommage après tant d'efforts de réduire ceux-ci à néant par une maladresse de dernière minute.

Si votre page web parle d'immobilier,

*A éviter, `loction apprtement pars`, vous serez peut-être bien positionné sur **loction apprtement pars**, mais comme personne ne tape cette séquence dans les moteurs de recherche, c'est sans intérêt.*

Quelle stratégie adopter pour accroître mes backlinks ?

Vous en savez maintenant suffisamment pour partir à la recherche de nouveaux backlinks.

Nous recensons ici, trois stratégies classiques :

- les annuaires,
- la méthode dite du link ninja,
- et la méthode dite du link baiting.

Il est, bien entendu, possible et même souhaitable, de mixer ces trois techniques. On prendra néanmoins garde à éviter autant que possible l'effet SandBox.

Les annuaires

Les débutants en référencement passent souvent beaucoup de temps pour inscrire leur site web dans des tas d'annuaires. Si cette approche pouvait être payante il y a quelques années, aujourd'hui elle est contre-productive. Elle n'occasionne que perte de temps, éventuellement d'argent, de crédibilité et peut vous pénaliser.

La plupart des annuaires (pas tous) ont mauvaise presse auprès des moteurs de recherche. Ce sont des agglomérations de liens les plus divers sans aucun autre contenu que des publicités et des interminables descriptions de sites web qui vantent tous leurs mérites. Leurs pages web ont donc souvent des contenus très hétérogènes et des centaines de liens (souvent en Javascript, donc de faux backlinks). A la lumière de ce que vous savez maintenant sur le NetLinking, dans le meilleur des cas, ces backlinks sont de piètre qualité. A part quelques annuaires majeurs ou très spécialisés, ils ne valent pas la peine que vous vous y arrêtiez.

Certaines sociétés peu scrupuleuses vantent et vendent l'inscription automatique dans des milliers d'annuaires, alors qu'elles savent pertinemment l'inefficacité de la démarche et même son caractère nocif. Souvent, elles ont elles-mêmes mis en place ces annuaires avec des scripts automatisés (qu'il est facile de trouver sur Internet). Passez votre chemin !

Le link ninja

Que se cache t-il derrière cette expression, pour le moins, saugrenue ?

Les adeptes du link ninja s'efforcent de construire un réseau de backlinks complètement sur mesure. Ils partent (parfois à l'aide de logiciels) à la recherche des pages web candidates susceptibles d'accueillir un backlink à forte valeur ajoutée (référencement) vers l'une de leurs pages web. Pour ce faire, ils sélectionnent des pages web présentes dans l'index des moteurs de recherche majeurs et ayant,

- un thème apparenté à la page web qu'ils veulent voir pointée,
- une notoriété suffisante (nombre de backlinks),
- un nombre raisonnable (pas trop important) de liens sortants (dilution faible de la notoriété),
- et un PageRank légèrement supérieur à la page web qu'ils veulent valoriser.

Une fois la sélection faite, ils contactent les webmasters de ces sites web pour négocier un échange de liens. Les intérêts de cette approche sont multiples. Elle permet notamment :

- de choisir exactement les pages web qui seront pointées,
- de quantifier exactement le nombre des backlinks pour chacune de ces pages web,
- de choisir le niveau de proximité thématique entre backlinks et pages pointées,
- de choisir les textes (ancres) des backlinks,
- d'obtenir des backlinks de pages « anciennes » qui ont souvent davantage d'autorité.

C'est un travail de précision qui demande beaucoup de temps et est souvent laborieux (beaucoup d'échanges d'emails, vérification des backlinks, etc.). Le résultat est quant à lui très efficace lorsque la campagne est bien menée.

Le link baiting

Le link baiting (to bait = appâter) vise aussi à l'accumulation de backlinks, mais par d'autres moyens.

Dans le link baiting, les backlinks se font sur l'initiative des webmasters qui trouvent un intérêt à pointer vers l'une des pages de votre site web. On laisse ici le backlink se faire naturellement.

Mais qu'est ce qui peut pousser un webmaster à vous offrir un backlink ?

Le concept est simple, il faut lui donner envie de pointer vers votre site web.

Comment donner envie aux webmasters de pointer vers votre site web ?

Il vous faut proposer un contenu de qualité et différencié par rapport aux autres sites web. Vous pouvez, par exemple :

- publier un scoop,
- traduire un article de référence,
- concevoir un événement original,
- publier un article polémique qui fera réagir,
- proposer un service gratuit,
- rédiger un livre blanc gratuit,

- proposer un service innovant,
- vous démarquer par une compétence très spécifique,
- mettre à disposition un feed RSS (souvent repris par des sites agrégateurs de contenu),
- etc.

En bref, faire du buzz autour de votre site web.

Link Ninja contre Link Baiting

Les deux techniques qui précèdent ont chacune leurs avantages et leurs inconvénients.

Le link ninja permet de mettre en place le réseau de backlinks désirés : choix des ancres, des pages pointées comme des pages qui pointent, etc. Le principe est simple, on n'est jamais aussi bien servi que par soi-même. Toutefois, la mise en place de ce réseau est laborieuse. Elle se fait au prix de très lourds efforts et prend beaucoup de temps. De plus, souvent les backlinks sont en fait des échanges de liens qui ont moins de poids dans les moteurs de recherche car ils sont identifiés comme une tentative de gonflement artificiel de la popularité d'un site web. Les échanges de liens triangulaires sont aussi aisés à identifier par les moteurs de recherche.

Le link baiting est plus aléatoire quant au résultat final. Il est difficile de maîtriser la croissance d'un réseau de backlinks avec cette méthode. Il n'y a pas de choix sur les ancres, les pages pointées, etc. Tout ceci provoque souvent une distribution très inégale des backlinks sur l'ensemble des pages d'un site web, avec souvent une forte dominance de la page d'accueil. Par contre, l'effort est nettement moindre. Les backlinks ne sont pas des échanges de liens, ce sont de vrais backlinks (plus de valeur). Parfois, il y a la bonne surprise de récolter des backlinks de sites web de référence, qui font autorité dans leur secteur d'activité, et qui donne un vrai coup de pouce au référencement.

Ces deux techniques sont complémentaires et il faut savoir user de l'une et de l'autre au gré de la nécessité. Dans un premier temps, le link baiting semble plus approprié. Il renvoie finalement à la nécessité de produire du contenu original et de qualité, sans lequel tout travail de référencement est vain. Dans une seconde phase, le link ninja pourra venir compenser les lacunes du link baiting qui produit rarement un résultat équilibré. Par des actions plus précises ciblant des pages web et des mots-clés précis, vous étendrez la surface d'exposition de votre site web en mettant en valeur les pages web sous-exploitées.

PageRank & Trustrank

Il y a quelques années encore, le PageRank faisait couler beaucoup d'encre et était l'objet de nombreux buzz et polémiques. Le flou qu'entretenait Google autour de cette notion n'y était pas pour rien. Il concentrait toutes les attentions, à tel point qu'en matière de référencement, il était devenu le principal objet de convoitise de nombreux webmasters. Leur vision du référencement avait été pervertie par la chasse aux points de PageRank. Ils mirent en place des stratégies alambiquées en vue de faire croître le PageRank de leurs pages web et ils ne jugeaient un site web que sur la base du PageRank de sa page d'accueil.

Aujourd'hui, la situation a notablement changé. Les effets de bord produits par le PageRank ont été contenus en partie par l'avènement du TrustRank. Le TrustRank supplante désormais le PageRank, car il prend en considération davantage de paramètres. Il ne faut néanmoins pas négliger le PageRank qui reste un indice intéressant et, à la différence du TrustRank, est directement accessible (quantifiable). Nous tâcherons donc d'y porter attention sans toutefois en faire le pivot de notre stratégie de référencement.

Le PageRank

Le PageRank en clair

Le PageRank, c'est quoi ?

Le PageRank, c'est une note qu'attribue Google à chaque page web de son index. Cette note varie de 0 à 10. Plus elle est élevée, mieux c'est. Cette note varie dans le temps. Elle peut baisser ou augmenter.

Il y a quelques temps, elle était mise à jour plus ou moins régulièrement lors des Google Dance qui ont aujourd'hui disparu. Désormais, cette mise à jour se fait quasiment en temps réel.

Le PageRank, ça sert à quoi ?

Le PageRank, c'est une combine (en fait, un algorithme) de Google pour évaluer la popularité d'une page web.

Comment est évalué le PageRank ?

L'évaluation du PageRank d'une page web repose sur deux critères :

- la quantité de ses backlinks,
- et la qualité de ses backlinks.

Ce qu'on entend par qualité ici, c'est en fait le PageRank du backlink considéré, autrement dit sa popularité.

A chaque fois qu'une page web reçoit un backlink, c'est un vote en sa faveur qu'elle reçoit. Ce vote a d'autant plus de poids que ce backlink a un PageRank élevé (une popularité élevée).

Par exemple, un page web sans backlink aura son PageRank nul. Aucune page web ne lui a encore accordé son attention. En revanche, une page web qui dispose de plus d'un millier de backlinks devrait avoir un PageRank de 4 ou 5 suivant la qualité des backlinks.

Il arrive souvent qu'on parle de PageRank pour un site web. C'est en fait un abus de langage. Il s'agit en réalité du PageRank de la page d'accueil du site web.

Comment connaître mon PageRank ?

Attention, chaque page web a son propre PageRank. Il existe plusieurs manières de connaître le PageRank d'une page web.

Avec la Google Toolbar

La plus commune est probablement l'utilisation de la barre d'outil de Google (Google Toolbar). Vous pouvez la télécharger à l'adresse suivante (pour les navigateurs Internet Explorer et Firefox) http://toolbar.google.com/T4/intl/fr/index_pack.html.

Par exemple, si vous consultez la page d'accueil d'un site web avec Internet Explorer, vous devriez obtenir ceci :



Ici, la valeur du PageRank affichée est de 6 (/10).

A noter que la mise à jour du PageRank dans la Google Toolbar ne se fait pas en temps réel. Elle n'a lieu que plusieurs fois dans l'année. C'est pourquoi, on l'appelle PageRank affiché (qui ne renvoie qu'à la valeur qu'avait ce PageRank lors de sa dernière mise à jour dans la Google Toolbar). Par opposition, on nomme PageRank réel, la véritable valeur (mise à jour en temps réel) prise en compte par Google pour élaborer ses pages de résultats.

Par certains sites web

Si vous ne voulez pas télécharger la Google Toolbar (ou que vous la trouvez trop intrusive), vous pouvez aussi utiliser des outils en ligne. Ils sont nombreux, citons par exemple :

- <http://www.seoachat.com/seo-tools/pagerank-lookup/>
- <http://www.dnseo.net/seo-tools/pagerank-lookup/>
- <http://www.gopagerank.com/>

Tous ces sites vous donneront le PageRank d'une page web de votre choix.

Par exemple, avec l'outil proposé par <http://www.gopagerank.com/> il vous suffit de renseigner la page web dont vous voulez connaître le PageRank, puis de valider. Le résultat se fait sans attendre.



On constate que le PageRank de la page d'accueil de <http://www.fnac.com> est de 7, ce qui est un PageRank assez élevé.

Attention, vous pouvez saisir n'importe quelle url, pas uniquement la page d'accueil, comme montré ici dans l'exemple.

Comment augmenter mon PageRank ?

On augmente le PageRank d'une page web en obtenant de nouveaux backlinks. Leur quantité est importante mais aussi leur qualité. Le PageRank n'évolue pas linéairement avec le nombre de backlinks mais de manière logarithmique.

Logarithmique, Kesako ?

Pour que votre page web gagne un point de PageRank, il lui faut EN GROS dix fois plus de suffrage (mélange quantité, qualité des backlinks) qu'elle en a actuellement.

Autrement dit, il faut dix fois plus de suffrage pour être à PageRank 3 que pour être à PageRank 2. Il faut dix fois plus de suffrage pour être à PageRank 4 que pour être à PageRank 3, et ainsi de suite...

Si les premiers points de PageRank sont faciles à gagner, pour les suivants, c'est une autre affaire.

C'est quoi un bon PageRank ?

Le PageRank varie sur une échelle allant de 0 à 10 :

- de 0 à 2, le PageRank est faible,
- de 3 à 5, il est moyen,
- de 6 à 7, c'est du sérieux,
- enfin de 8 à 10, réservé uniquement à l'élite.

Le TrustRank

Le TrustRank en clair

Confusion autour du TrustRank

TrustRank est le nom d'un algorithme, d'un indice et d'une marque déposée.

L'algorithme du TrustRank a été mis au point (en 2004) par deux chercheurs de l'université de Stanford et un chercheur de l'entreprise Yahoo. Cet algorithme a pour objectif d'évaluer la crédibilité des pages web.

L'algorithme produit un indice évaluant le degré de confiance (trust) accordé à chaque page web. Cet indice peut ensuite être pris en compte par les moteurs de recherche dans l'élaboration de leurs résultats de recherche. Il doit, en principe, permettre d'identifier les pages de spam.

Le TrustRank est aussi une marque déposée (en 2005) par la société Google.

Yahoo et Google utilisent chacun leur propre version du TrustRank pour mettre au point leurs SERP. Dans la suite, nous nous soucierons principalement de Google.

Le TrustRank, c'est quoi ?

Le TrustRank, tout comme le PageRank, est une note qu'attribue Google à chaque page web de son index. Cette note varie de 0 à 1. Plus elle est élevée, mieux c'est. Une page ayant un TrustRank proche de 0 est peu digne de confiance (assimilée spam), alors qu'une page ayant un TrustRank proche de 1 est très crédible (assimilée page de référence).

Le TrustRank, ça sert à quoi ?

Le TrustRank est un algorithme qu'utilise Google pour évaluer la crédibilité d'une page web. Alors que le PageRank évalue davantage la popularité. On peut être populaire sans pour autant être crédible ! Sa vocation est la lutte contre les tricheries (spam) en tout genre. Il a été mis en place quelques temps après la frénésie déclenchée par le PageRank et tous les effets de bord (bourse d'échange de liens, vente de backlinks sur page à fort PageRank, etc.) qui s'en sont suivis.

Comment est évalué le TrustRank ?

Le principe est assez simple.

A la base, il y a un ensemble de sites web évalués « à la main » (par des humains), comme étant des sites dignes de confiance. Ce sont les sites de référence auquel un TrustRank de 1 est attribué. Ces sites web sont souvent ceux des organismes officiels, des grands journaux, etc.

Ces sites font des liens vers d'autres sites web qui à leur tour se voit affecter d'une crédibilité moindre mais importante quand même, puisque des sites de référence leur font confiance. A leur tour ces sites web font des liens vers d'autres sites web, et ainsi de suite. D'une certaine manière, il y a redistribution de proche en proche de la crédibilité (TrustRank).

L'algorithme du TrustRank est, bien entendu, un peu plus compliqué que cela. Il tient notamment compte :

- de l'ancienneté du nom de domaine,
- de l'existence d'une adresse postale facilement accessible,
- des backlinks de sites à fort TrustRank (.gov, .edu, .mil, grands journaux) du site web,
- de l'existence de certificat SSL, d'une politique de confidentialité, etc.

Comment connaître mon TrustRank ?

Il n'existe aucun moyen d'évaluer exactement ou même approximativement le TrustRank d'une page web.

Comment augmenter mon TrustRank ?

Les leviers pour augmenter le TrustRank sont faibles. Le mieux reste encore d'avoir un contenu différencié et à forte valeur ajoutée susceptible d'être pointé par un site à fort TrustRank. Le problème, c'est qu'il n'y a aucun moyen de connaître ce TrustRank. Le TrustRank est presque une fatalité et il est donc difficile à exploiter concrètement en matière de référencement, à la différence du PageRank.

Le logiciel GRATUIT GnoZtik

<http://www.gnozтик.com>

Présentation

Pour ceux qui ne connaissent pas encore GnoZtiK, GnoZtiK est un logiciel **GRATUIT** de référencement. Après avoir téléchargé et installé le logiciel, vous saisissez un mot-clé **UNIQUE**, vous sélectionnez une version de Google et vous saisissez l'url d'une page web que vous voulez auditer.

*Par exemple, le mot-clé **isolant mince** sur la version de Google **fr - Google - Pages francophones** :*



Vous lancez l'audit et trois minutes plus tard, vous obtenez un diagnostic complet qui vous indique :

- si votre page web peut être bien positionnée sur le mot-clé visé,
- l'effort à fournir en optimisation HTML pour y arriver,
- des pistes pour effectuer cette optimisation HTML,
- l'effort à produire en NetLinking pour y arriver.

Si votre page web n'a aucune chance de se positionner, GnoZtiK vous propose alors des mots-clés alternatifs plus ciblés pour lesquels un espoir est permis.

Analyse détaillée

L'analyse se scinde en quatre parties, chacune représentée par un onglet :

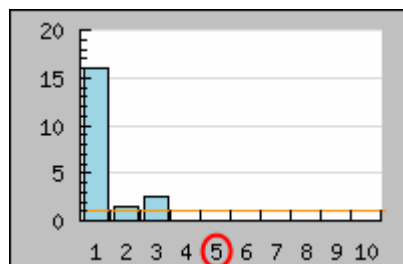
- un onglet **Classement** qui dispense des informations générales sur le mot-clé audité, sans tenir compte de l'url audité,
- un onglet **Url** qui se concentre exclusivement sur l'url audité,
- un onglet **HTML** qui met en évidence le différentiel de contenu entre la page auditée et les dix meilleures sur le mot-clé analysé (il fournit des pistes pour résorber ce différentiel, évalue l'effort à produire et le gain qui devrait en résulter),
- et un onglet **Backlinks** qui évalue l'effort à produire et le gain qui devrait en résulter, ainsi que quelques possibilités d'échanges de liens.

L'onglet Classement

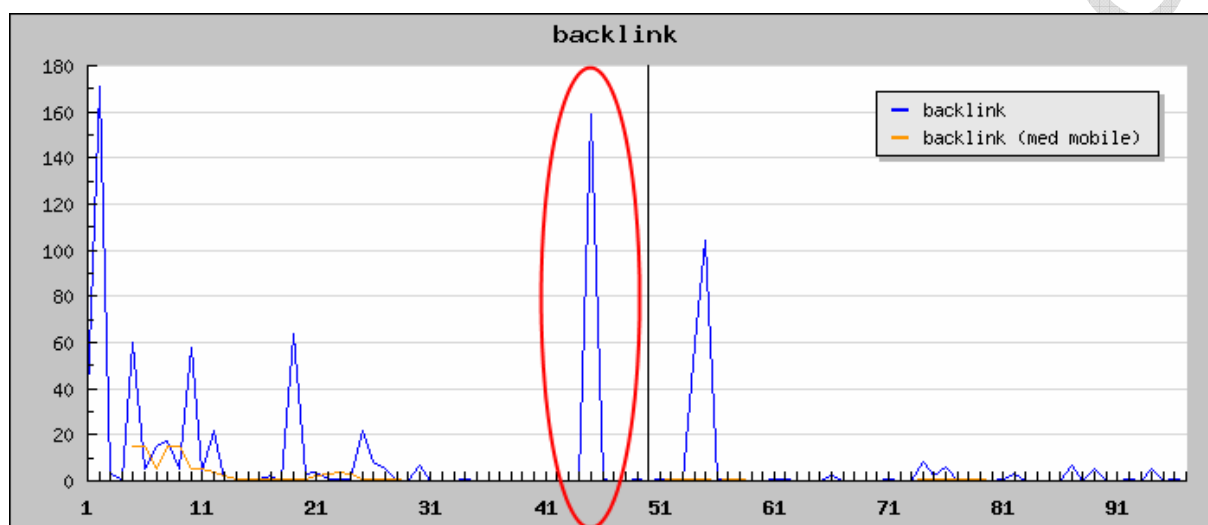
Les données brutes

Pour chaque mot-clé, GnoZtiK propose 3 graphiques (PageRank, Backlinks, contenu HTML) qui peuvent être riches d'enseignement. Ces graphiques présentent les urls groupées par page. A chaque page du classement Google correspond une barre dans l'histogramme. Cette barre représente la valeur médiane, au sens statistique du terme du critère étudié. La médiane, à la différence de la moyenne, gomme les valeurs aberrantes (les extrêmes) qui souvent altèrent la réflexion.

Par exemple, pour le mot-clé **vêtements coton bio**, on relève le graphique Backlinks suivant :



La médiane du nombre de Backlinks de la cinquième page (position 41 à 51) de résultats de Google vaut 0 (barre absente car hauteur nulle).



Par contre, la moyenne vaut 16 ($=160/10$). En effet, en cliquant sur ce premier graphique apparaît le graphique détaillé de toutes ces valeurs. Toutes les valeurs sont nulles de la position 41 à 51, excepté la position 45 qui a pour valeur 160, d'où une moyenne à 16. De ce qui précède, on constate que la médiane est plus représentative que la moyenne.

Enfin dans le premier graphique, la barre horizontale orange indique la valeur du critère (PageRank, Backlinks, contenu HTML) pour votre page. Vous avez donc, au travers de ces trois graphiques (PageRank, Backlinks, contenu HTML), un rapide aperçu de ses forces, faiblesses et potentiels.

Les données résumées

Pour chaque mot-clé, une analyse approfondie des cent premières pages web permet de définir quelle part a chaque facteur (PageRank, Backlinks, Contenu HTML) dans le classement final. Plus cette part est importante, plus le facteur correspondant est déterminant pour bien se positionner. C'est cette part que GnoZtiK appelle Poids SEO.

Par exemple, pour le mot-clé **passer annonce immobilière**, on relève les poids suivants :

| | Moyenne | Poids SEO |
|-------------|-------------|-------------|
| Pagerank | 3,8 | 31 % |
| Backlinks | 322 | 16 % |
| HTML | 67 % | 53 % |

Le contenu HTML à lui seul compte pour plus de la moitié (53%) dans la décision finale. L'optimisation du code aura donc probablement une forte incidence. Par contre les backlinks

apparaissent négligeables. Un travail sur ces derniers aura donc probablement une faible incidence.

Pour chaque mot-clé, GnoZtiK calcule la Moyenne de chaque facteur (PageRank, Backlinks, Contenu HTML) pour les cent premières pages web. Plus la moyenne est élevée pour un facteur, plus la barre est élevée pour ce facteur. Si, en plus, ce facteur a un Poids SEO élevé, il risque fort d'être très difficile de bien se positionner.

Par exemple, pour le mot-clé **world of warcraft**, on relève les moyennes suivantes :

| | Moyenne | Poids SEO |
|-----------|---------|-----------|
| Pagerank | 4,5 | 27 % |
| Backlinks | 963 | 26 % |
| HTML | 55 % | 47 % |

Le PageRank moyen des cent premières pages est de 4,5 : c'est élevé. Pareillement, le nombre moyen de Backlinks des cent premières pages est élevé : 963. Leurs Poids SEO n'étant pas négligeable, la lutte sera âpre.

Le contexte général

Tous les mots-clés n'ont pas la même valeur. Certains sont très convoités, tandis que d'autres sont délaissés par les webmasters. GnoZtiK propose une typologie des mots-clés qui doit, en principe, vous aider à trancher : maintenir votre effort ou au contraire changer votre cible. Ce n'est pas l'unique critère à retenir avant la décision finale, il est cependant très important.

GnoZtiK qualifie la requête sur laquelle vous voulez vous positionner de concurrentielle, parasitée, délaissée, niche (voir 2^{ème} partie).

Les trois premières pages de résultats de Google sont les seules intéressantes du point de vue du référencement (nombreux sont ceux qui diraient uniquement la première).

GnoZtiK vous indique les taux d'urls entrantes et sortantes de ces trois premières pages. Plus le Taux URLs sortantes est élevé, plus il y a de pages web parmi le top 30 susceptibles d'en sortir ... donc de places libérables. Plus le Taux d'urls entrantes est élevé, plus il y a de pages web candidates à l'entrée dans le top 30.

La différence entre les deux (Taux URLs sortantes - Taux URLs entrantes) donne un **contexte général**. Il est favorable si le nombre de places susceptibles d'être cédées est plus élevé que le nombre de candidats à ces places ... il y aura moins de lutte.

Par exemple, pour le mot-clé **isolant mince**, on relève le contexte général suivant :

| | |
|---------------------------|----------------------|
| Typologie de la requête ? | une requête de niche |
| Taux URLs sortantes ? | 33 % |
| Taux URLs entrantes ? | 17 % |
| Contexte général | très favorable |

Ce mot-clé se situe sur une niche, il est donc en soi intéressant : trafic très qualifié.

Le contexte général est très favorable car le nombre de places susceptibles d'être libérées est bien plus élevé que le nombre de candidats à ces places (33% est très supérieur à 17%).

L'ambiance HTML

Le Tagcloud prend les meilleurs, soit les dix premiers du classement Google comme référence. De manière générale, plus un mot est présent parmi ces pages web, plus il apparaît gros.

Par exemple, pour le mot-clé **ecommerce**, on relève le Tagcloud suivant :



Les mots-clés les plus présents parmi les dix premiers du classement Google sont donc : commerce, e-commerce, solution, boutique, internet, marketing, ... Il faudra donc probablement faire un effort particulier d'optimisation sur ces mots.

L'onglet Url

Les principaux indicateurs

GnoZtiK propose d'abord un récapitulatif des indicateurs majeurs pour votre référencement :

- la place (en position et en page) dans les résultats de recherche de Google,
- le PageRank et entre parenthèses la moyenne des cent premiers,
- les backlinks et entre parenthèses la moyenne des cent premiers,
- la pertinence HTML et entre parenthèses la moyenne des cent premiers.

Ces informations permettent de voir rapidement, où est votre page web et comment se situe son optimisation par rapport aux cent premiers.

Par exemple, pour le mot-clé **graveur**, on relève dans l'onglet URL pour la page <http://www...> (position 36 au 31 avril 2009 sur Google - Pages francophones), le récapitulatif suivant :

Votre URL à la loupe

Position : 36 ?

Page : 4 ?

Pagerank : 2 (2,6) ?

Backlinks : 0 (129) ?

Contenu HTML : 33 % (45 %) ?

http://www... est presque (inférieur) au niveau moyen des cent premiers en PageRank.

http://www... souffre d'un gros déficit en backlinks. La page est très en dessous du niveau moyen des cent premiers.

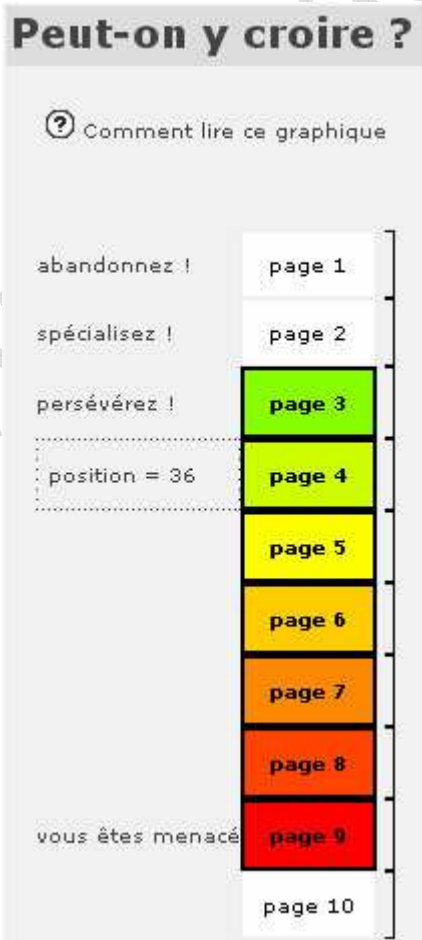
Enfin, http://www... est presque (inférieur) au niveau moyen des cent premiers en pertinence HTML.

Le pronostic ne doit pas être trop difficile à faire.

Peut-on y croire ?

Une image vaut mieux que mille mots disent les chinois. La réglette permet d'appréhender rapidement la situation de votre page web : position, potentiel, menace, conseil, etc.

*Par exemple, pour le mot-clé **graveur**, on relève dans l'onglet URL pour la page http://www... (position 36 au 31 avril 2009 sur Google - Pages francophones), la réglette suivante :*



http://www... est en position 36. Cependant la page est menacée, si rien n'est fait, elle risque fort de se retrouver aux alentours de la 9^{ème} page des résultats de Google, qui est

sans intérêts pour le référencement. En faisant un effort, <http://www...> doit pouvoir gagner quelques places de sorte à se retrouver sur la 3^{ème} page des résultats de Google.

Par contre, GnoZtiK indique qu'il est vain de tenter de se positionner sur la 1^{ère} page des résultats de Google. Quant à la 2^{ème} des résultats de Google, elle ne semble accessible qu'en spécialisant le mot-clé.

Conseils et pronostics

En référencement, seules les trois premières pages des résultats de Google sont intéressantes (uniquement la première selon certains). Le tableau des conseils permet de voir rapidement ce que peut faire votre page web pour se placer sur chacune des pages 1, 2 et 3 des résultats de Google. Pour chacune de ces pages, il y a une ligne dans laquelle sont indiqués l'effort objectif à fournir en backlinks, l'effort objectif à fournir en HTML, la difficulté objective et enfin le conseil subjectif de GnoZtiK.

Les conseils GnoZtiK sont de 3 types : « Persévérez », « Spécialisez » ou « Abandonnez ». « Persévérez » indique que le but est à votre portée. « Abandonnez », indique que vos efforts seront probablement vains. Enfin, « Spécialisez » vous invite pour plus d'efficacité à cibler un mot-clé proche et plus ciblé. Le conseil étant subjectif, vous êtes libre de ne pas le suivre.

Par exemple, pour le mot-clé **graveur**, on relève dans l'onglet URL pour la page <http://www....> (position 36 au 31 avril 2009 sur Google - Pages francophones), le tableau des conseils suivants :

| Page ? | Effort en backlinks ? | Effort HTML ? | Difficulté ? | Conseil ? |
|---|---------------------------------|---------------|---------------|----------------|
| 1 | très important (plus de 100) | important | insurmontable | Abandonnez ! |
| 2 | important (50 à 99) | important | majeure | Spécialisez !* |
| 3 | léger (1 à 24) | important | abordable | Persévérez ! |
| * Vous pouvez spécialiser la requête avec les mots mis en couleur dans le tag cloud ci-dessous. | | | | |

<http://www....> doit pouvoir gagner quelques places de sorte à se retrouver sur la 3^{ème} page des résultats de Google.

Par contre, GnoZtiK indique qu'il est vain de tenter de se positionner sur la 1^{ère} page des résultats de Google.

Quant à la 2^{ème} des résultats de Google, elle ne semble accessible qu'en spécialisant le mot-clé.

L'ambiance HTML

Le Tagcloud renvoie au contenu HTML de votre page web. Il donne une représentation graphique mettant en évidence la thématique de votre page web. Il est impératif que ce contenu soit en rapport avec le mot-clé convoité. De manière générale, plus un mot a d'importance (au sens des moteurs de recherche) dans votre page web, plus il apparaît gros. Attention, les tags HTML n'ont pas tous la même valeur en référencement. Ce Tagcloud en tient compte.

Par exemple, pour le mot-clé **graveur**, on relève dans l'onglet URL pour la page <http://www....> (position 49 au 31 avril 2009 sur Google - Pages francophones), le Tagcloud suivant :



Les mots-clés les plus présents sont donc : **graveur**, **externe**, **memorex**, **droppix**, ... ce qui colle avec le mot-clé **graveur**. Ce sont les mots dominants côté contenu HTML.

Les mots colorés en vert sont utiles lorsque vous peinez à vous positionner sur le mot-clé en cours d'analyse (GnoZtiK vous a conseillé « Spécialisez »). Ils vous permettent de construire un nouveau mot-clé connexe au mot-clé initial mais davantage ciblé de sorte à limiter la concurrence sur ce nouveau mot-clé et augmenter vos chances de drainer un trafic encore plus ciblé. Ces mots colorés en vert donnent donc des pistes pour des mots-clés plus ciblés. La plupart du temps, vous découvrirez des mots-clés inattendus à fort potentiel pour votre site web.

Dans le même exemple, si **http://www....** peine à mieux se positionner sur le mot-clé **graveur**, il est possible de construire un nouveau mot-clé à partir de **graveur** et de l'un des mots colorés en vert dans le Tagcloud. Pour le mot-clé **graveur memorex**, on constate que **http://www....** est en 18^{ème} position le 30 avril 2009 sur Google - Pages francophones. Pour le mot-clé **graveur externe**, on constate que **http://www....** est en 13^{ème} position le 30 avril 2009 sur Google - Pages francophones. Un petit effort devrait permettre de se placer en première page de Google - Pages francophones pour ces mots-clés.

L'onglet HTML

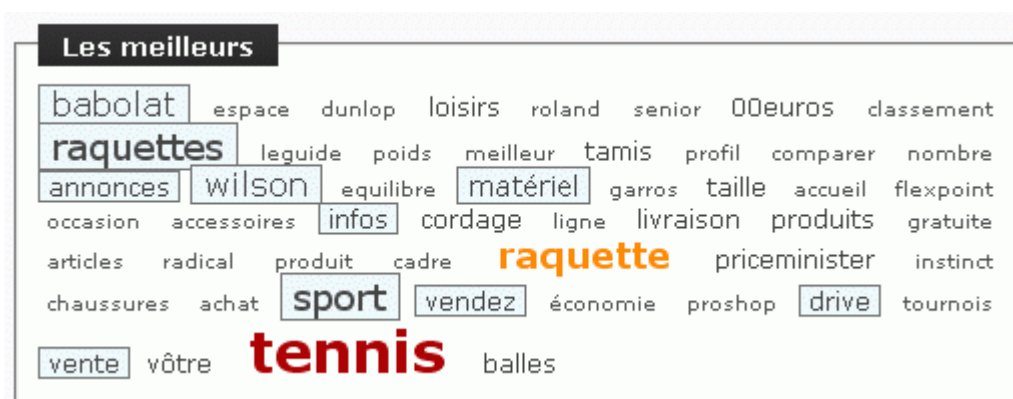
Le différentiel HTML

GnoZtiK donne les mots-clés les plus présents à travers les deux Tagclouds de l'onglet HTML. Le premier Tagcloud regroupe les mots-clés des dix premiers du classement dans Google. Il sert de référence.

Les mots-clés absents

Ce sont les mots-clés encadrés et surlignés en bleu pastel. Ils sont présents dans les pages des dix premiers du classement mais absents de votre page web.

Par exemple, pour la requête **raquette tennis**, on obtient le Tagcloud suivant :



Il apparaît que les mots-clés qui manquent à la page auditée (nous avons pris la 49^{ème}) sont : **raquettes**, **sport**, **matériel**, **wilson**, etc. Il est donc conseillé de les intégrer dans la page auditée dans la mesure du possible. Les plus gros parmi ces mots-clés sont à prendre en priorité car ils sont les plus présents parmi les pages web des dix premiers du classement.

Les mots-clés sous-représentés

Ces mots-clés sont présents dans les pages des dix premiers du classement mais moins présents dans votre page web. Ils sont représentés de trois manières :

- en rouge foncé, les mots-clés très peu présents,
- en rouge, les mots-clés assez peu présents,
- en orange, les mots-clés peu présents.

En reprenant l'exemple de la requête **raquette tennis**, les mots-clés sous-représentés dans la page auditée (toujours le 49^{ème}) sont : **raquette** et surtout **tennis**.

Cela permet de rapidement connaître les mots-clés que vous devez renforcer pour améliorer votre positionnement. Comme pour les mots-clés absents, les mots-clés les plus gros sont les plus prioritaires.

La taille des mots des Tagclouds

De manière générale, plus un mot a d'importance (au sens des moteurs de recherche), plus il apparaît gros dans le Tagcloud. Attention, les tags HTML n'ont pas tous la même valeur en référencement. Ce Tagcloud en tient compte.

GnoZtiK présente une échelle relative permettant d'évaluer la présence de chaque mot-clé d'un Tagcloud. Les pourcentages indiquent le nombre d'occurrences d'un mot relativement au mot le plus représenté.

Par exemple, le mot-clé le plus gros est à 100% car il est le plus présent. Un mot deux fois plus petit aura un pourcentage de 50% car il est deux fois moins présent dans la page.

Les effort et gain estimés

GnoZtiK évalue aussi l'effort nécessaire à fournir pour que votre contenu HTML soit au niveau des meilleurs. Cet effort peut être qualifié de léger, important ou très important en fonction de l'ampleur de la tâche qui vous attend.

Le gain estimé vous donne une évaluation du nombre de places que vous pouvez espérer gagner en appliquant l'effort demandé. En général, plus l'effort est important, plus vous aurez de chances de gagner des places dans le classement. Selon le contexte du classement, il arrive parfois qu'un léger effort s'avère très rentable et vous propulse parmi les dix premiers du classement. Surveillez donc le gain estimé avec attention.

Effort très
important



gain estimé *
de 20 à 30 places

L'onglet Backlinks

Les effort et gain estimés

Tout comme pour le contenu HTML, GnoZtiK évalue l'effort en backlinks nécessaire à fournir pour améliorer le positionnement de votre page web. Cet effort peut être qualifié de léger, important ou très important en fonction de l'ampleur de la tâche qui vous attend.

GnoZtiK vous précise entre parenthèses le nombre de backlinks correspondants à l'effort indiqué. Le gain estimé donne une fourchette du nombre de places que vous pouvez espérer gagner en appliquant l'effort demandé. En général, plus l'effort est important, plus vous aurez de chances de gagner des places dans le classement. Selon le contexte du classement, il arrive parfois qu'un léger effort s'avère très rentable et vous propulse parmi les dix premiers du classement. Surveillez donc le gain estimé avec attention.

Effort très
important
(plus de 100)



gain estimé *
de 15 à 25 places

Les backlinks possibles

Enfin, GnoZtiK vous propose une liste de pages web avec lesquelles faire des échanges de liens pour améliorer votre positionnement. Cette liste est triée par pertinence décroissante. Elle tient notamment compte de la pertinence thématique et du PageRank.

Plus la pertinence thématique est élevée en pourcentage, mieux c'est pour votre positionnement sur le mot-clé que vous convoitez.

Plus le PageRank est élevé, mieux c'est. On se gardera néanmoins de faire un échange de liens avec une page dont le PageRank est marqué du signe ⚠ qui signale une page à trop fort PageRank par rapport à celui de votre page (souvent considéré comme suspect par Google).

| Présentation Classement URL HTML Backlinks Votre avis Aide | | | | |
|---|---|----------|-------|---------|
| Pos | Url | Pagerank | HTML | Contact |
| 1 | http://www.bricoman.fr | 5 | 71 % | |
| 2 | http://discount-bricolage.zlio.com | 3 | 76 % | |
| - | http://bricolagediscount.zlio.com | 2 | 102 % | |
| 3 | http://discount-bricolage.zlio.com/c1115896--Discount-Bricol... | 0 | 85 % | |
| 13 | http://www.aidaubricolage.com | 5 | 41 % | |
| 29 | http://www.weborama.fr/top/nature-evasion/jardins-decoration... | 5 | 41 % | |
| 89 | http://www.teleachat.fr | 5 | 37 % | |
| 54 | http://www.portail-express.com | 5 | 29 % | |
| 88 | http://www.aqrojob.com/emploi/hard-discount.asp | 5 | 29 % | |
| 65 | http://www.prixdiscounts.com | 5 | 22 % | |
| 62 | http://www.ebuyclub.com/Partenaires.jsp | 5 | 17 % | |
| 64 | http://www.planeteachat.com | 5 | 15 % | |
| 46 | http://gral.gralon.net/maison-et-jardin/bricolage.htm | 4 | 51 % | |
| 14 | http://www.aidaubricolage.com/partenaire.html | 1 | 49 % | |

Enfin dans la colonne contact, en cliquant sur la petite enveloppe vous trouverez l'email argumenté à envoyer au webmaster de la page correspondante pour le convaincre de faire un échange de liens avec vous. Ensuite, il ne reste plus qu'à trouver le formulaire de contact ou l'adresse email de contact de ce site web, faire un copier/coller et envoyer.

L'audit a duré ... moins de 3 minutes pour faire le point, GnoZtiK économise votre temps.

Conclusion

Ce manuel de référencement touche à sa fin. Nous souhaitons qu'il vous sera d'un grand profit. Débuter en référencement n'est pas chose facile. Les sources sont éparses et nécessitent de nombreux recoupements avant d'avoir une vue correcte de la discipline et de pouvoir en aborder le versant opérationnel. Nous espérons au travers de cet ouvrage avoir largement contribué à cette clarification.

Ce manuel n'en est qu'à sa version 1.0 (première). Le référencement évolue. Année après année, les algorithmes des moteurs de recherche s'affinent et nécessitent une revue des techniques efficaces et sans risque. D'autres versions sont donc à prévoir qui seront l'occasion de corrections, mises à jour, compléments et divers remaniements. Pour ce faire, nous avons besoin de votre retour d'expérience. Apportez votre témoignage, signalez nous les fautes d'orthographe, les erreurs, les précisions à apporter, les aspects du référencement que vous voudriez voir développer, etc. Pour ce faire, contactez nous via le formulaire <http://www.gnoztik.com/ebook/v1/feedback.php>. Par avance, nous vous sommes reconnaissants pour votre contribution.

Si vous avez apprécié ce manuel pratique, faites en la propagation ! Rédigez des billets sur votre blog et faites le connaître à vos lecteurs, faites le connaître à vos amis et contacts en prise avec le web. Qui aujourd'hui sur la toile peut faire l'impasse sur le référencement ? Il vous est possible d'aller encore plus loin : un don Paypal (https://www.paypal.com/cgi-bin/webscr?cmd=s-xclick&hosted_button_id=5356201) du montant de votre choix nous encouragera.

Enfin, nous cherchons une ou plusieurs personnes pour traduire cet ebook en anglais (éventuellement en d'autres langues aussi). Nous voulons, bien entendu, une traduction complète et de qualité. Une fois celle-ci réalisée, chaque contributeur aura droit à une mention, probablement en pied de page (à discuter), de son site web sur les pages qu'il aura traduites. Si vous êtes intéressé, contactez nous via le formulaire <http://www.gnoztik.com/ebook/v1/translation.php>.

Glossaire

A/HREF

En langage HTML, attribut de la balise A (lien hypertexte) indiquant l'url visée par le lien.

Ancre

En langage HTML, texte lisible par l'internaute (cliquable) d'un lien hypertexte (balise A).

Backlink

Lorsqu'une page web A (source) fait un lien hypertexte vers une seconde page web B (destination), la page web A est un backlink de la page web B.

Backlink à sens unique

Lorsqu'une page web A est un backlink d'une page web B et que la page web B n'est pas un backlink de la page web A, la page web A est un backlink à sens unique de la page web B.

Backlink artificiel

Backlink obtenu à partir d'un annuaire, d'un commentaire de blog, d'un post dans un forum.

Backlink externe

Backlink d'une page web A vers une page web B appartenant à un site web différent.

Backlink interne

Backlink d'une page web A vers une page web B appartenant à un même site web.

Balise (Tag en anglais)

En langage HTML, chaîne de caractères pour identifier le type, le format et l'apparence d'un élément d'une page web.

Blacklist

Un site web est dit « blacklisté » ou « en blacklist » par tel moteur de recherche lorsqu'il n'apparaît plus dans son index, souvent après une manœuvre frauduleuse au sens des moteurs de recherche.

Black hat seo

Ensemble de techniques de référencement proscrites par les moteurs de recherche. Souvent efficaces à court terme, elles sont très lourdement sanctionnées dans la durée par les moteurs de recherche. Pas de pérennité pour le black hat seo !

Bold

En langage HTML, balise permettant d'affecter une mise en forme gras à un texte.

Bot

Voir Crawler.

Commande link:

Commande proposée par la plupart des moteurs de recherche permettant d'afficher la liste des backlinks d'une page web.

Commande linkdomain:

Commande proposée par Yahoo permettant d'afficher la liste de tous les backlinks d'un site web.

Commande site:

Commande proposée par la plupart des moteurs de recherche permettant d'afficher la liste des pages d'un site web indexées dans leur base de données.

Contenu caché

Technique dite « black hat seo », fortement déconseillée pour le référencement, consistant à afficher un texte aux internautes et un autre texte très optimisé aux moteurs de recherche.

Crawler

Un crawler est un robot (aussi appelé bot ou spider) qui parcourt le web et récupère automatiquement les pages web visitées pour ensuite les indexer.

Crédibilité (= TrustRank)

Indice d'une page web qui repose en grande partie sur les backlinks obtenus à partir de sites considérés comme référence.

Déclassement

Un site web subit un déclassement par tel moteur de recherche lorsqu'il apparaît toujours dans son index mais qu'il est en perte sévère de positionnement sur un nombre important de mots-clés. Cet événement apparaît souvent suite à un gonflement artificiel de popularité (PageRank). Si le moteur de recherche en question est Google, un effet SandBox peut être présumé. Attention, cette situation est à différencier de la désindexation.

Décompte

Le décompte correspond au nombre d'occurrences d'un mot-clé dans un texte donné. Ce texte peut être une url, un titre de document, une page web complète.

Densité

La densité est le rapport entre le décompte et le nombre total de mots. Elle mesure l'importance de la présence d'un mot-clé par rapport à l'ensemble d'un texte. Si un mot-clé apparaît 2 fois sur un total de 400 mots ($2/400 = 0,5\%$), il est moins présent que s'il apparaît 1 fois pour un total de 30 mots ($1/30 = 3,3\%$).

Désindexation

Un site web subit une désindexation par tel moteur de recherche lorsqu'il n'apparaît plus dans son index, souvent après une manœuvre frauduleuse au sens des moteurs de recherche. Il est entré en blacklist.

Duplicate content

Tout contenu web en double est susceptible de subir la pénalité dite du duplicate content. Une page web qualifiée duplicate content voit alors sa capacité de positionnement dans les SERP fortement diminuée.

Echange de liens

Lorsqu'une page web A est un backlink d'une page web B et que la page web B est un backlink de la page web A, on dit qu'il y a échange de liens (en général, résultat d'une négociation). Cette situation n'apparaît pas naturelle aux moteurs de recherche.

Flash

Technologie, fortement déconseillée pour le référencement, qui permet de créer et d'insérer des animations sur un site web.

Formulaire de soumission

Page web dédiée des moteurs de recherche permettant d'y soumettre un site web.

Frame

Technique de programmation HTML (obsolète) permettant de combiner deux documents ou plus dans des cadres séparés au sein d'une unique fenêtre de navigateur web.

GoogleBot

Crawler de Google.

H1, ..., H6

En langage HTML, balises qui permettent d'afficher les titres d'une page web. Le chiffre de la balise (de 1 à 6) va croissant avec le niveau de détail du titre. Les titres en H1 sont les plus importants, donc les plus visibles alors que les tags H6 sont les moins importants.

HTML

Acronyme signifiant HyperText Markup Language.

Langage standard utilisé pour la description des pages web.

IDM

Acronyme signifiant Indice de Densité d'un Mot-clé.

Voir Indice de densité d'un mot-clé.

IMG/ALT

En langage HTML, attribut de la balise IMG (image) informant les crawlers sur le contenu d'une image.

IMG/SRC

En langage HTML, attribut de la balise IMG (image) indiquant l'emplacement du fichier image correspondant.

Indexation

Recensement des pages d'un site web effectué par un crawler, qui les consigne dans sa base de données.

Indexation naturelle

Indexation d'un site web suite à une recommandation (backlink) faite par un autre site web.

Indice de densité d'un mot-clé

Idem Densité.

Voir Densité.

Javascript

Langage de script permettant d'étendre les possibilités du langage HTML. Le code Javascript est inclus dans la page web avec le code HTML.

Lien en dur

Lien hypertexte codé en HTML, sans l'aide du Javascript ou de tout autre technologie. Ce sont les seuls ayant un véritable intérêt pour le référencement.

Lien entrant

Idem Backlink.

Voir Backlink.

Lien hypertexte

Dans une page web, pointeur à partir d'un texte, d'une image ou d'un graphique vers une autre ressource (page ou fichier) du web. Sur le web, les liens hypertexte constituent le principal moyen pour naviguer entre les sites Web.

Lien sortant

Si une page web A est un backlink d'une page web B, alors la page web B est un lien sortant de la page web A.

Link ninja

Technique visant à la mise en place d'un réseau de backlinks complètement sur mesure reposant en grande partie sur des échanges de liens très ciblés.

Link baiting

Technique visant à la mise en place d'un réseau de backlinks à l'initiative des autres webmasters en leur donnant envie de faire des liens vers un contenu original (buzz).

Long Tail

Idem Longue Traîne.

Voir Longue Traîne.

Longue traîne

Principe qui dit que le trafic total engendré par les mots-clés à trafic faible (très nombreux) est supérieur au trafic total engendré par les mots-clés à fort trafic (peu nombreux). En pratique, ce trafic issu des mots-clés à trafic faible peut constituer plus de 80% du trafic total.

Maillage

Réseau de liens entre les différentes pages d'un site web.

Meta/description

En langage HTML, balise (importante) fournissant aux moteurs de recherche une description du contenu de la page. Son contenu est reproduit dans les résultats de recherche.

Meta/keywords

En langage HTML, balise (inutile) fournissant aux moteurs de recherche une liste de mots-clés visant à signaler le thème d'une page web.

Mot-clé

Séquence de mots renseignée par un internaute pour interroger un moteur de recherche.

Mot-clé concurrentiel

Mot-clé convoité de nombreux webmasters et engendrant un trafic très qualifié.

Mot-clé délaissé

Mot-clé non convoité des webmasters et engendrant un trafic variablement qualifié.

Mot-clé de niche

Mot-clé convoité de quelques webmasters et engendrant un trafic très qualifié.

Mot-clé parasité

Mot-clé non convoité des webmasters et engendrant un trafic peu qualifié.

Navigateur

Logiciel qui recherche et affiche des pages web pour les présenter à l'internaute. Microsoft Internet Explorer, Firefox, Opera, Safari sont des navigateurs web.

NetLinking

Partie du référencement qui renvoie à la gestion (tissage) d'un réseau de liens.

Page d'accueil

Page principale d'un site web.

PageRank

Le PageRank est une note qu'attribue Google à chaque page web de son index. Cette note varie de 0 à 10. Elle évalue la popularité de la page web.

Plan de site

Page web accueillant un plan hiérarchisé du site web auquel elle appartient. C'est une technique facilitant l'indexation d'un site web.

Popularité (= PageRank)

Indice d'une page web qui repose sur la quantité et la qualité de ses backlinks.

Positionnement

Ensemble des positions occupées dans les SERP par un site web sur une liste de mots-clés donnés.

Résultats de recherche

Page de résultats renvoyés par un moteur de recherche lorsqu'il est interrogé sur un mot-clé.

rel=nofollow

Attribut affecté à un lien hypertexte signalant aux moteurs de recherche de ne pas suivre (indexer) la page web pointée par le lien (faux backlink).

Robot

Voir Crawler.

Robots.txt

Le fichier robots.txt est le premier fichier que visitent les crawlers lorsqu'ils commencent à indexer un site web. Y sont consignées des directives de crawl de la part du webmaster pour les robots. Le fichier robots.txt (ou fichier d'exclusion des robots) est le moyen côté webmaster d'implémenter le protocole d'exclusion des robots (Robots Exclusion Protocol).

SandBox

Mécanisme, mis en place par Google, qui affecte la plupart des nouveaux sites web, mais également quelques anciens. Ils apparaissent rapidement dans l'index de Google, mais sont souvent très mal classés dans les SERP, surtout (mais pas uniquement) sur les mots-clés compétitifs.

SEO

Acronyme signifiant Search Engine Optimisation.

Optimisation pour les moteurs de recherches, souvent traduit par référencement naturel ou organique.

SERP

Acronyme signifiant Search Engine Result Page (Page de Résultats des Moteurs de Recherche).

Voir Résultats de recherche.

Sitemap

Technique, approuvée officiellement par les moteurs de recherche majeurs, qui facilite le travail d'indexation des moteurs de recherche. Elle se met en place à l'aide d'un fichier sitemap.

Soumission manuelle

Action visant à signaler l'existence d'un site web via un formulaire de soumission à un moteur de recherche mis à disposition par ce dernier.

Spamdexing

Technique dite « black hat seo » consistant à saturer une page web d'un ou de plusieurs mots-clés.

Spider

Voir Crawler.

Strong

En langage XHTML (HTML amélioré), balise permettant la mise en valeur d'un texte. Les navigateurs la rendent en général par une mise en forme gras.

Tag

Voir balise.

Taux de transformation

Rapport entre le nombre de visiteurs d'un site web et le nombre de passage à l'acte (achat, inscription newsletter, demande de documentation, etc.).

Tiret « - »

Séparateur de mots conseillé dans une URL.

Title

En langage HTML, balise (la plus importante) indiquant le titre d'une page web. Son contenu s'affiche dans la barre des titres du navigateur.

TLD

Acronyme signifiant Top Level Domain.

Voir Top Level Domain

Top Level Domain

2 à 4 dernières lettres de l'url d'un nom de domaine (ex. .com, .org, .edu, .fr, .info).

Trafic

Ensemble des visites drainées par un site web.

Trafic qualifié

Se dit d'un trafic à fort taux de transformation.

TrustRank

Le TrustRank, tout comme le PageRank, est une note qu'attribue Google à chaque page web de son index. Cette note varie de 0 à 1. Plus elle est élevée, mieux c'est. Elle évalue la crédibilité d'une page web.

Underscore « _ »

Séparateur de mots déconseillé dans une URL.

URL

Acronyme signifiant Uniform Resource Locator.

Chaîne de caractères indiquant l'adresse Internet d'une page web. Par exemple, <http://www.gnoztik.com/seofaq.htm>.

URL rewriting

Technique permettant d'optimiser pour le référencement les urls des pages d'un site web.

User Agent

Chaque crawler ou navigateur est identifiable par son User Agent (chaîne de caractères déclinant son identité qu'il s'est lui-même attribué). A chaque fois qu'il effectue une requête auprès d'un serveur web, il se présente avec son User Agent.

Cette définition est très restrictive. Elle est adaptée au besoin en référencement.

White hat seo

Ensemble de techniques de référencement appréciées par les moteurs de recherche. Seules ces techniques de référencement sont véritablement pérennes dans leurs effets.

Yahoo Site Explorer

Outil Yahoo d'analyse de site web pour le référencement.

Yahoo Slurp

Crawler de Yahoo.